Genomics in Reintroduction Biology: a case study with New Zealand hihi (*Notiomystis cincta*)

Kate D. Lee

A thesis submitted in fulfilment of the requirements for the degree of PhD in Science, the University of Auckland, 2019.

Abstract

Reintroduction is the translocation of animals to their historic range. Reintroduced populations are usually small and commonly have low population genetic diversity. They are particularly susceptible to the effects of inbreeding and genetic drift. This can lead to inbreeding depression and the loss of genetic variation, limiting population growth and adaptation respectively. Genetic diversity can be further impacted during a reintroduction both in a source population as it loses individuals and in the translocated population as it has lower genetic diversity and population size than the source. I look at the reintroduction of an endangered New Zealand passerine, hihi (*Notiomystis cincta,* 'stitchbird').

I develop a single nucleotide polymorphism array (50K SNP array) to measure genetic diversity and inbreeding in reintroduced hihi populations. SNPs can be found in large numbers throughout the whole genome and are becoming more common in the fields of ecological and conservation genetics due to their decreasing costs and reproducibility across labs. In conservation studies, samples from endangered species are often limited in the quantity and quality of DNA available. I found that a SNP array of 50K markers could genotype samples of low quantity or quality and was not affected by tissue type (feather or blood). Earlier studies on hihi relied on a panel of microsatellites from non-coding locations of the genome. I show that SNPs measure genetic diversity and inbreeding as well as microsatellites. Furthermore, SNPs can reveal the locations of long runs of homozygosity (ROHs) in a genome, i.e. parts of the genome that are more inbred. These regions mark areas of potential autozygosity, where sequences are identical due to having a common ancestor. ROH patterns in a population can result from their demographic history, with shorter ROHs from more distant ancestors and long ROHs illustrating recent inbreeding events. This is one of the first studies of ROHs in a reintroduced population of endangered animals.

Looking at a hihi population over 11 years, I show that removing individuals has no effect on the inbreeding measure F_{IS}. Small but significant changes in F_{IS} could be seen across cohorts in SNP data only (0.003, Adjusted R^2 = 0.5486, *p-value* = 0.005511). Small but significant changes could be seen in the number of SNPs in ROHs (increased by 0.02859 SNPs per year, adjusted R^2 : 0.000574, *p-value*: 0.007625) and number of ROHs per

individual per year (increased by 0.14728 per year, adjusted R^2 : 0.00231, *p-value*: 0.05345) but there were no significant changes in ROH length.

I show previously unseen increases in inbreeding across translocation bottlenecks. I found that across populations, hihi have short ROHs with a median length of 528 kb indicating no recent strong bottlenecks. Some individuals have very long ROHs, perhaps as a result of a recent consanguineous mating. A genetic marker of bottlenecks, the summed length of ROHs per individual, increases across each translocation bottleneck, indicating the genetic impact that recent repeated translocations are having on this species.

Acknowledgements

Firstly, I would like to thank my supervisor Anna Santure, who has guided me through this work with incredible patience for my crazy ways. My co-supervisor Craig Millar went above-and-beyond to help me out in the lab and focus me back on the science when I got distracted by shiny ideas. I would like to thank my co-supervisors Patricia Brekke and John Ewen who provided me with the samples they have been collecting over the years, and who have been very kind to give me invaluable advice and guidance for my chapters as my project progressed, it has greatly improved the work.

I have had an army of helpers over the few years. Melanie Hingston, who slogged through 3000 sample DNA extractions with me, Selina Patel who helped optimise my laboratory practices and Amy Tan who stepped into the trenches to help when 800 extractions needed re-doing. I am really grateful.

I would like to thank everyone who read my thesis in the last dash to the finish line (all mistakes are my own). Kathleen Collier and Leilani Walker, my writing group for the last three years – I hope we keep it up! James Brock whose authoritative voice is very soothing when you're in a panic and Carol Connolly, who has been proof- reading my science reports for as long as I can remember.

There are many people who helped me stay sane over the course of this work. The usual suspects Henny, Beccy, Ivan, Livia, James, Ulla and Irene; the Tiri gang Mhairi, Viks, Kata and Catlin and my peeps at home Mat and Larissa. Thanks for all the great times in NZ and getting me away from my computer ;)

A special thanks needs to go to Ulla Von Amon, who let me stay with her in Nelson for a month and to Sandra Anderson and Catherine Scott who have hosted me in these last few weeks while I extended my stay in Auckland to finish my thesis. I owe you all dinner, lots of them!

Finally, I would like to thank my parents, who somehow knew I could do this, and my sister, who is cheerleading for me here at the end. I love you guys ;)

Contents

Chapter 1: Introduction

Very few studies have had the opportunity to use very large numbers of genetic markers to look at populations in a conservation context. In this thesis, I describe the identification of large numbers of polymorphisms from a small amount of low coverage sequencing data and test the effects of pooling the data on downstream variant identification. I test the genotyping of samples of variable quality and different tissue type. I investigate the effects of repeatedly removing individuals from a small population of endangered birds with low genetic diversity using large numbers of genetic markers. Large numbers of markers have enabled the identification of large regions of homozygosity and I use these to infer demographic history. Finally, I observe the effects of translocation bottlenecks on genetic diversity and inbreeding of the population. Using regions of homozygosity as an indicator, I take a close look at how hihi population genomes change across translocations.

Reintroduction Biology

What is reintroduction biology?

Reintroduction of a species is the assisted colonisation of fauna to their historic ranges for conservation purposes (IUCN/SSC, 2013) and typically involves movement of small numbers of individuals to establish a new population (Ewen, 2012). The purpose of reintroductions is population restoration with the goal that the population will ultimately be self-sustaining (Seddon et al., 2012). Armstrong and Seddon (2008) proposed ten key questions that need to be addressed in reintroduction biology programs, six of which can be addressed using genetics or genomics. These include:

1 How does the composition of the release group affect establishment probability?
2 How can the genetic makeup affect the persistence of the reintroduced population?
3 How heavily should the source populations be harvested?
4 What is the optimal allocation of translocated individuals between sites?
5 Should translocations be used to compensate for isolation?
6 How does the order of reintroductions affect the ultimate species composition?

In order to address these questions, there is a need to understand the genetics of small populations, and then start querying how removing individuals affects them and how

translocation bottlenecks and founder effects impact the genetics of the population. By addressing these questions, the effectiveness of a reintroduction can be assessed, and recommendations made to improve the fate of a reintroduced population.

Genetics of small reintroduced populations

In newly established populations, founder effects caused by the sampling of alleles in founders compared to the source population can lead to reduced allele and genetic diversity. This genetic bottleneck may result in in the frequency of deleterious alleles increasing relative to the source population (Fitzsimmons et al., 1997; Laugier et al., 2016; Whitehorn et al., 2011), and can lead to the loss of adaptive alleles. In addition, in small reintroduced populations, mating between close kin can lead to an increase in the number of individuals homozygous for deleterious recessive alleles, resulting in inbreeding depression i.e. reduction in fitness (Bérénos et al., 2016; Hoffman et al., 2014; Jamieson, 2015). The combination of genetic bottlenecks and inbreeding in small isolated populations reduces genetic variation, reducing population persistence, productivity and adaptability (Laugier et al., 2016; Schmitt and Hewitt, 2004). However, if there are few alleles with deleterious effects, populations can persist with low genetic diversity without fitness consequences (Jamieson, 2015), e.g. wandering albatross (*Diomedea exulans*) (Milot et al., 2007), Tuco tuco (*Ctenomys sociabilis*) (Hadly et al., 2003), and beavers (*Castor fiber*) (Ellegren et al., 1993). In fact their low diversity, small population size and increased probability of inbreeding can increase the efficacy of purifying selection, purging deleterious alleles (Garcia-Dorado, 2012; Hedrick and Garcia-Dorado, 2016).

Small populations are more susceptible to the effects of genetic drift, as genetic diversity is lost per generation in inverse proportion to the effective population size (Kimura, 1955). The effect is greater when a population size is small for long periods of time (Ellstrand and Elam, 1993), for example in small island populations. Mating between related individuals is also more common in small populations, leading to reduced genetic diversity and increased inbreeding depression (Keller, 2002).

Studying the effects of the number and genetic makeup of individuals in a reintroduced population, and their success in terms of population establishment and growth, is therefore imperative to improving reintroduction management strategies (Armstrong and Seddon, 2008; Seddon et al., 2014).

Genetics of removing individuals from wild populations

Capturing the maximum genetic diversity of the source population in successfully translocated individuals is one of the main aims of translocating endangered animals (Weeks et al., 2011). Removing individuals from wild populations for translocation can affect the population genetics of the source population, including changing the genetic structure of the population if one particular subgroup is harvested more than others, reducing genetic diversity or acting as a selective force on remaining individuals if a particular trait is favoured among harvested individuals (Allendorf et al., 2008).

Changes in population subdivision caused by removing individuals can affect gene flow (Harris et al., 2002) and can result in the loss of local adaptations. For example, Atlantic salmon (*Salmo salar*) populations form metapopulations in the ocean where they can experience gene flow but separate into subpopulations in freshwater as groups of fish return to rivers at different times. A change in fishing regulations, whereby fish are harvested from the ocean metapopulation, or are harvested from key river populations, can affect the genetic diversity of the metapopulation (Hindar et al., 2004). Iberian lynx (*Lynx pardinus*) in Spain used to exist in 15 subpopulations in the 1940s but were reduced to two isolated subpopulations by 2000 by hunting and indiscriminate predator control (Gil-Sánchez and McCain, 2011).

Genetic diversity can be reduced by removing individuals, limiting the adaptive potential of the population (Laugier et al., 2016; Schmitt and Hewitt, 2004). Allelic diversity is a good predictor of long-term adaptive potential (Caballero and Garcia-Dorado, 2013). Removing individuals has caused reduction in the genetic diversity of many wild populations including lake trout (*Salvelinus namaycush*) (Baillie et al., 2016) and wild American ginseng (*Panax quinquefolius*) (Cruse-Sanders and Hamrick, 2004).

Removing individuals can also cause artificial selection within a wild population on the bases of sex, age and behavioural or morphological traits (Coltman, 2008; Kuparinen and Festa-Bianchet, 2017; Leclerc et al., 2017), but the effects can be reduced by regulating harvesting intensity (Kuparinen and Festa-Bianchet, 2017). For example, a comparison of tusk size in African elephants (*Loxodonta africana*) during a period of harvesting in the 1960s showed that female tusks were reduced by 37% and males by 22% compared with protected populations in the 1990s (Chiyo et al., 2015). Further, a quantitative genetics model analysing adaptive and heritable traits including boldness of behaviour, size at

maturation and size-specific reproduction predicted that a population of fish would evolve to be either more timid or more bold depending on the fishing method used (Andersen et al., 2018).

Potential genetic effects of translocation bottlenecks

Populations established by translocating animals can suffer from reduced allelic diversity caused by the bottleneck. The size of the bottleneck determines the amount of allele loss with smaller bottlenecks causing more severe loss (Allendorf, 1986; Nei et al., 1975). These founder effects are also influenced by how much genetic diversity of the source population is captured by the translocated individuals (Weeks et al., 2011). Loss of alleles and especially loss of rare alleles reduces the species' evolutionary potential (Caballero and Garcia-Dorado, 2013; Greenbaum et al., 2014; Vilas et al., 2015).

Heterozygosity is a measure of the number and spread of alleles across a population. It is also predicted to be affected by the size of the bottleneck, but only if it is over a long period of time (Nei et al., 1975). The correlation between heterozygosity and fitness in wild populations is well established (Chapman et al., 2009; Forstmeier et al., 2012; Grueber et al., 2008; Reed and Frankham, 2003; Sovic et al., 2018). How loss of heterozygosity will affect a particular species fitness will vary greatly depending on its population history and the genetic basis of the heterozygosity-fitness correlation within that population (Hedrick and Garcia-Dorado, 2016) and on its current environment (Szulkin and Sheldon, 2007).

Genetic management of a metapopulation

Measures of genetic variation can include the number of alleles at each locus across a population or allelic richness (A_R), which gives an indication of the adaptive potential of a population, i.e. if it has the genetic capacity to change when exposed to new environmental conditions (Höglund, 2009). Other measures include observed heterozygosity (H_O), which indicates the number and spread of alleles across the population (i.e. how many individuals have two different alleles at a given genetic locus) (Nei, 1973). Expected heterozygosity (H_E) measures the fraction of individuals in a population that are expected to be heterozygous at a given locus, assuming the population is at Hardy-Weinberg equilibrium (Nei, 1973).

Minimising inbreeding and maintaining genetic variability (A_R, H_E and H_O) are of primary concern when planning translocations and managing small populations (Reed and

Frankham, 2003; Weeks et al., 2011). In the case that inbreeding is accumulating rapidly and the population shows signs of inbreeding depression, one strategy to reinforce genetic variation is to translocate further individuals from the original source or another population (IUCN/SSC, 2013). Jamieson (2011) cautions that to avoid long-term accumulation of inbreeding, populations should not be reintroduced to small islands, or genetic rescue measures should be taken. Genetic rescue is the introduction of new genetic material into a population. It should reduce inbreeding and increase the genetic diversity of a population (Frankham, 2015; Whiteley et al., 2015). Ideally this would mitigate the effects of loss of fitness in a population with reduced heterozygosity and high inbreeding load (Hedrick and Garcia-Dorado, 2016). It has been seen to occur naturally in wild populations e.g. where immigration has counteracted the effect of genetic drift in snow voles (*Chionomys nivalis*) (García-Navas et al., 2015).

However, for management purposes, the negative effects of inbreeding within a small population may be outweighed by other conservation considerations such as the availability of more individuals to translocate and top-up the population and the demographics of the potential source and current population (Harding et al., 2016). Further, the cost of such top-ups and the cost of ongoing genetic monitoring to be able to detect inbreeding accumulation is significant, and with limited resources the advantages of genetic top-ups must be weighed against other effective strategies for ensuring population growth and persistence, such as habitat restoration and provision of supplementary feeding.

Methods and Technology

Measuring genetic diversity

Genetic diversity can be measured in a number of ways including assessing A_R or H_E and H_O. Greater diversity means the population has more evolutionary potential and is more likely to be able to adapt to new environments (Höglund, 2009). A higher level of heterozygosity indicates that the population is less inbred. Inbreeding can be measured more directly using the inbreeding coefficient (F_{IS}). This measures inbreeding levels of an individual relative to the subpopulation and can show differences between subpopulations (Wright, 1951).

Microsatellites are repeated motifs of 2-5 nucleotides that vary in length depending on the number of repeats. They were discovered in 1984 (Weller et al., 1984) and have been widely used for estimating genetic diversity and inbreeding because each locus will typically have multiple alleles (Tautz, 1989) resulting in high heterozygosity values. However, they are prone to genotyping errors, especially at more variable marker sites, and it can be difficult to standardise microsatellite allele lengths across laboratories (Gill et al., 1998; LaHood et al., 2002). In particular, null alleles, which are alleles that consistently fail to amplify to detectable levels via the polymerase chain reaction, have been reported in many microsatellite studies and have to be carefully managed in the data analysis to avoid errors (Dakin and Avise, 2004). In ecological studies, microsatellites from non-coding regions of the genome are most frequently used. Microsatellite data is multiallelic and can therefore show large variance in allelic diversity compared with a panel of biallelic SNPs (Morin et al., 2004).

Single nucleotide polymorphisms (SNPs) are single DNA nucleotide differences in genomic sequences in a population. They are distributed genome-wide and generally at higher density in the genome than microsatellite markers (Hernández et al., 2015), and may be found in coding regions, non-coding regions, and regions under selection in the genome (Syvänen, 2001).SNP data is usually bi-allelic and can typically be genotyped in much larger marker numbers than microsatellites, resulting in more precise measurements of genetic diversity (Morin et al., 2004). SNPs are used with increasing frequency and in larger numbers in population genetics in parallel with or replacing microsatellite markers (Shafer et al., 2015). SNPs have low genotyping error rates and are reproducible across laboratories. Furthermore, genotyping SNPs is becoming more accessible due to the falling costs of DNA sequencing (Morin et al., 2004). When genotyped on a SNP array (see section 2.3), an added advantage of this marker type is their utility on poor quality or degraded DNA (Smith et al., 2011).

Measuring inbreeding

Inbreeding is the mating of related animals and results in individuals carrying alleles that are identical by descent (IBD). The measurement of relatedness within an individual, the coefficient of inbreeding (*F*), is the probability that two alleles will be IBD (Falconer and Mackay, 1996). *F* can be measured using pedigree information (F_{ped}), and is based on the

assumption that the pedigree founders are unrelated (Wright, 1922). Consequently, the accuracy of *F* in estimating IBD varies with the quality (i.e., accuracy of assigning family relationships) and depth (i.e., the number of generations) of the pedigree (Nietlisbach et al., 2017; Santure et al., 2010).

Molecular-based methods of calculating *F* have been developed for study species for which pedigrees are not available and they have the added advantage that they measure the realised IBD directly from gene sequences or marker alleles, rather than the expected IBD estimated from the relatedness values of the pedigree (Forstmeier et al., 2012; Keller et al., 2011). For example, an offspring of a mating between two full siblings has an expected pedigree inbreeding coefficient of 0.25, but at any position in the genome will share 0, 1 or 2 alleles identical by descent and so the expected and realised relatedness will differ (Allendorf and Luikart, 2007). Further, pedigrees are often inaccurate due to missing data (Bérénos et al., 2016) and pedigrees derived from small panels of molecular markers, which are frequently used for wild populations, can suffer from a lack of discriminatory power (Sardell et al., 2010; Taylor, 2015), particularly in species with low genetic diversity (Taylor, 2015). Molecular-based *F* measurements are therefore particularly useful in a conservation context in which pedigrees may be unavailable or incomplete, and genetic diversity is typically low. Molecular markers used to measure inbreeding include microsatellites and SNPs.

The distribution of SNPs across the genome has enabled the identification of runs of homozygosity (ROHs), contiguous regions of the genome that are dominated by homozygous markers (Broman and Weber, 1999a). These regions are identical by state (IBS) and may also be identical by descent (IBD) (Howrigan et al., 2011). They mark shared ancestry of genetic haplotypes, and longer ROHs represent IBD regions inherited from a common ancestor in more recent generations, while shorter ROHs have been inherited from more distant ancestors (Keller et al., 2011; McQuillan et al., 2008). ROH are increasingly used in studies of humans and livestock (Ceballos et al., 2018). They are ubiquitous in humans (Gibson et al., 2006) in which long ROHs are common (Broman and Weber, 1999a). ROHs have a defined pattern in some populations, for example in Europeans, there are 77 regions with a higher occurrence of ROHs when compared to other human populations (Nothnagel et al., 2010). They have been shown to be enriched

for homozygous deleterious variants compared with non-deleterious variants in human populations (Alsalem et al., 2013). ROHs have also been linked to disease such as schizophrenia (Lencz et al., 2007), and they can affect complex traits such as height and cognition (Joshi et al., 2015). Higher levels of homozygosity in humans does not affect survival to old age (Kuningas et al., 2011) or increase risk of breast or prostate cancer (Enciso-Mora et al., 2010). However, in red deer (*Cervus elaphus*), homozygosity is associated with annual breeding success and maternal inbreeding coefficient is tightly linked to offspring survival (Bérénos et al., 2016).

Patterns of ROHs have been used to investigate genetic diversity (Gibson et al., 2006) and population demography, e.g. short ROHs dominate outbred populations (Bosse et al., 2012; Pemberton et al., 2012). The sum of total length of ROH (sROH) compared with the number of ROH in individuals in the population changes depending on demography (Ceballos et al., 2018). Larger populations have shorter sROH and fewer ROH than smaller populations, admixed populations have the smallest. Consanguineous populations have a larger range of sROH length and bottlenecks result in both longer sROH and numbers of ROH (Figure 1 in Ceballos et al., (2008) illustrates this beautifully). Effective population size (N_e) can be measured by binning homozygous and heterozygous regions of the genome (Li and Durbin, 2011; MacLeod et al., 2013). ROHs can be used to assess inbreeding across the genome (reviewed in Curik et al., 2017, 2014). It has been of particular interest to those studying inbreeding in cattle, where inbreeding is measured using different length categories of ROH (Bjelland et al., 2013; Ferenčaković et al., 2013; Marras et al., 2015; Peripolli et al., 2018). ROHs can be used to identify regions that might be under selection (Metzger et al., 2015; Purfield et al., 2012).

SNPs enable calculation of an inbreeding *F*-statistic based on these runs of homozygosity for individuals within a population (F_{ROH}). F_{ROH} is highly correlated with pedigree-based inbreeding (F_{ped}) for long runs of homozygosity and has been proposed as an alternative when pedigree data is not readily available (McQuillan et al., 2008; Zhang et al., 2015). ROH measurements from high density SNP data offer some new methods to conservation biologists looking for more accurate inbreeding measures, methods to uncover population structure and demography and identify regions under selection.

Large SNP arrays in conservation biology

A SNP array is an oligo array, designed to genotype SNP polymorphisms within a population. SNP arrays are generally designed by looking at genome sequence data of a small selection of individuals and identifying polymorphic sites. Probes developed for a subset of these sites are used to manufacture the array (Kothiyal et al., 2009). SNP arrays differ from other SNP methods as they genotype a particular set of loci, ensuring comparable results across samples and they also work well with poor quality samples (Johnston et al., 2013; Mead et al., 2008). While they are used widely for animal and plant breeding purposes, only a few large SNP arrays have been designed for wild populations that are not of commercial interest. These include a 10K and 500K SNP array for great tit (*Parus major*) (Kim et al., 2018; van Bers et al., 2012), which was used to create a linkage map (van Oers et al., 2014) and dissect the genetic architecture of phenotypic traits (Kim et al., 2018; Santure et al., 2013, 2015). A 9K BeadChip for polar bear (*Ursus maritimus*) showed that linkage disequilibrium was rapidly decaying in the populations and SNPs showed genetic structure of the population much more clearly than earlier studies with microsatellites (Malenfant et al., 2015). The polar bear SNP array data was applied to improve a microsatellite pedigree and resulted in the identification of adopted cubs, and monozygotic twinning (Malenfant et al., 2016). It was also used to gain a great understanding of the genetic architecture of body size in these animals (Malenfant et al., 2018). A SNP50 BeadChip array designed for commercial sheep (*Ovis aries*) breeds (Kijas et al., 2009) was used to dissect recombination rate among individuals in a population of Soay sheep (Johnston et al., 2016) and to understand the genomic basis of morphology in bighorn sheep (*Ovis canadensis*; Miller et al. 2018). A 10K SNP array for house sparrow (*Passer domesticus*) (Hagen et al., 2013a) was used to dissect the genetic architecture of bill morphology (Lundregan et al., 2018; Silva et al., 2017). A 50K SNP array for the flycatcher genus (*Ficedula*) showed that hybridisation between collared flycatchers and pied flycatchers did not pass the F1 generation suggesting that the populations may have become reproductively incompatible (Kawakami et al., 2014a). The flycatcher array was also used to create a high-density linkage map, improve the flycatcher genome assembly and investigate recombination rates and chromosomal evolution (Kawakami et al., 2014b).

Hihi (*Notiomystis cincta*, 'stitchbird') Populations

Species biology

Hihi or stitchbird *(Notiomystis cincta*) is a threatened endemic New Zealand passerine and one of the few sexually dimorphic birds in New Zealand (see Figure 1). The males have a black head with white ear tufts and bright yellow patches on their shoulders. The females have an olive-brown colour and a white bar on their wing. The vocalisations of the males include a piercing single, double or triple note call that sounds much like two stones hit together. The females have more subtle vocalisations (Robertson et al., 2015).

The hihi mating system varies from monogamy to polygynandry (Castro et al., 1996) with social monogamy and extra-pair copulation. Unusually, hihi engage in forced copulation using face-to-face mating (Castro et al., 1996; Low, 2004; Low et al., 2005), which results in high levels of extra pair paternity (Brekke et al., 2013, 2012; Ewen et al., 2004). Heritability of traits in hihi is low, for example heritability of fledgling mass is h^2 = 0.0329; tarsus length is h^2 = 0.123 and head-bill length is h^2 = 0.0581(de Villemereuil et al., 2019a; de Villemereuil et al., 2019b).

Figure 1: Hihi in Tiritiri Mātangi. From left to right a female bird, male bird and two 21-day old chicks (pictures taken by Kate D Lee).

Hihi are phylogenetically distinct, representing the sole species in their family *Notiomystidae* (Driskell et al., 2007). From a conservation standpoint, they are important pollinators in the forest ecosystem (Castro and Robertson, 1997), and they are of cultural

importance to New Zealand Māori, with their presence historically seen as an indicator of a healthy, mature forest system.

Historically, hihi were widespread across the North Island of New Zealand but following European settlement hihi were extirpated by the 1880s due to the introduction of mammalian predators and forest clearing. They were reduced to a single remnant population, on the island of Te Hauturu-o-Toi (Little Barrier Island) in the Hauraki Gulf (36°12'S, 175°05'E). Since the 1980s hihi have been translocated from the remnant population to various predator-free locations around the North Island, with currently six successfully established reintroduced populations. As a result hihi is an excellent model for reintroduction biology (Ewen et al., 2011; Thorogood et al., 2013).

Translocation history of hihi on the North Island of New Zealand

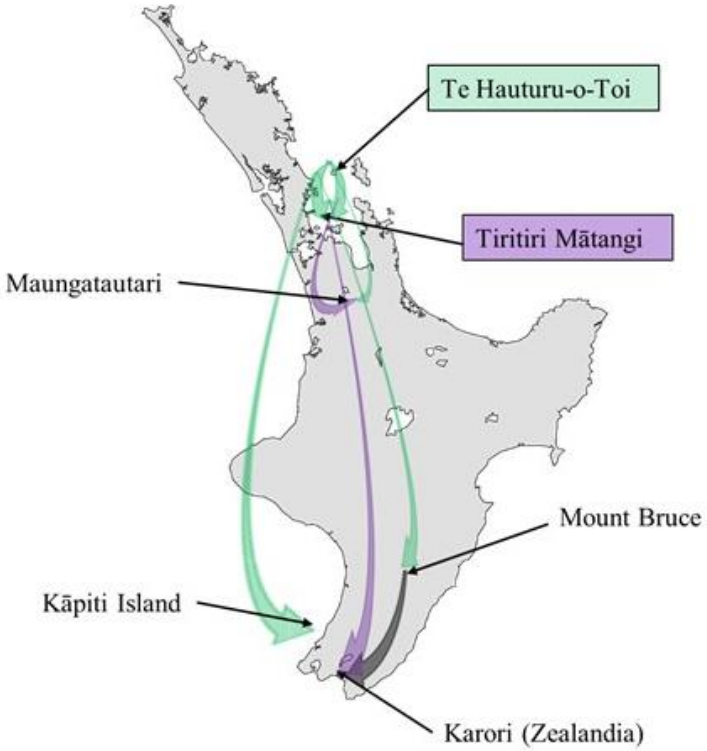


Figure 2: Translocation history of hihi samples used in this study. Green arrows indicate birds translocated from Te Hauturu-o-Toi, purple arrows from Tiritiri Mātangi and the black arrow from Mt. Bruce.

Figure 2 shows the translocation history of birds used in this study. Hihi have been translocated to various predator free locations across the New Zealand North Island since the 1980s to varying degrees of success (Brekke et al., 2011). All the birds originate from the remnant wild population in Te Hauturu-o-Toi in the Hauraki Gulf. The translocation history up until 2011 is shown in Table 1; a number of other populations have since been established. The hihi population on the island of Tiritiri Mātangi, also in the Hauraki Gulf, has been intensively studied, with all individuals in the population banded and tracked through their lifetimes, and a long term pedigree reconstructed with the help of a panel of 19 microsatellite markers (Brekke et al., 2013, 2009). Supported by demographic modelling, Tiritiri Mātangi has frequently been used as a source for establishing other populations (reviewed in Thorogood et al., 2013). As Tiritiri Mātangi is directly established from the remnant population in Te Hauturu-o-Toi (a first-order translocation) and is used to establish further populations (second order translocations), it also provides an opportunity to investigate the effects of multiple translocation bottlenecks.

Table 1: Translocations to establish and replenish populations.

Population	Translocated from	Year	Birds
Mount Bruce	Te Hauturu-o-Toi	1979	6
		1980	4
		1985	6
		1990	4
		1994	2
		1998	3
Kāpiti	Te Hauturu-o-Toi	1983	23
		1984	30
		1990	14
		1991	50
		1992	49
	Mokoia	1999	1
		2000	1
		2002	12
	Mount Bruce	2002	7
	Tiritiri Mātangi	2000	2
		2001	5
		2004	11
		2010	30
Tiritiri Mātangi	Te Hauturu-o-Toi	1995	37
		1996	14
		2010	20
Karori	Mount Bruce	2005	11
		2006	14
		2007	9
		2008	11
		2009	2
		2011	2
	Tiritiri Mātangi	2005	60
		2007	5
		2010	5
Maungatautari	Tiritiri Mātangi	2009	59
		2010	37
		2011	39
		2009	20

Management history

Hihi have undergone numerous translocations to new predator-free sites since the 1980s. A great deal is understood about hihi biology (Thorogood et al., 2013) and extensive research has been done to optimise hihi management. This includes improving the supplementary feeding systems to improve reproductive success (Doerr et al., 2017), survival (Armstrong and Perrott, 2000) and post-translocation survival (Chauvenet et al., 2012), identifying diseases affecting mortality (Alley et al., 1999; Cork et al., 1999), observing the direct effects of translocating birds on mortality and behaviour (Ewen et al., 2011), and how they adapt to their new environments (Richardson et al., 2010). Supplementary feeding is now carried out all year round on Tiritiri Mātangi and chicks are monitored intensively until they are fledged and treated for mites should the need arise.

Thesis Outline

Aims

The overarching aim of this thesis is to investigate to what degree genomic data can be of service to reintroduction biology programs. I address two of the reintroduction biology questions posed by Armstrong and Seddon (2008) including looking at the effects of removing individuals from the source population and how the order of reintroductions (i.e. first and second order translocation bottlenecks and genetic top-ups) affects the genetic variation and inbreeding levels in reintroduced populations. I also address (but don't directly test) a third question of whether translocations should be used to compensate for the isolation of the reintroduced population and the hihi metapopulation managed as a whole. I make novel use of ROHs to gain better understanding of inbreeding across the hihi genome and how the genome is affected by translocations during the management of a metapopulation of an endangered bird.

Chapter 2: Pooling sample data for assembly improves SNP discovery when using low coverage whole genome sequencing.

Single nucleotide polymorphisms (SNPs) are used frequently in ecology and conservation due to decreasing costs and ease of reproducing results across laboratories. However, the cost of high-depth whole genome sequencing (WGS) coverage can still be prohibitively high. Here I look at how to identify polymorphisms in a species with no annotated reference genome, with only low coverage WGS data available. Additionally, for

conservation studies which may not have access to state-of-the-art facilities, samples stored over long periods can provide DNA of variable quality and might not all be from the same tissue type. SNP arrays in particular are useful as they can genotype poor quality samples (Johnston et al., 2013; Mead et al., 2008). I design a SNP array for manufacture and test the effects of DNA quality and sample type on genotyping success.

Low coverage WGS data from ten birds and restriction site associated DNA sequencing (RAD-seq) data for 31 birds was used to identify polymorphic sites in the hihi genome and to select a subset of these SNPs for genotyping on a custom SNP array. I took the opportunity to investigate the effects of pooling data for assembly on downstream variant calling success as measured by conversion rates on the SNP array. WGS data was assembled for each sample separately and for two pooled sample data sets (i) a subset of three samples and (ii) all ten samples together. Variants were called from all 12 assemblies, merged based on homologous position on the zebra finch genome and SNPs chosen based on quality filters and putative positions on the genome. 1,536 hihi samples were genotyped on the SNP array and the conversion rates of SNPs determined.

Samples were collected from multiple locations including Te Hauturu-o-Toi, Tiritiri Mātangi, Karori (Zealandia Wildlife Sanctuary), Kāpiti Island and Sanctuary Mountain Maungatautari. DNA extracted from the samples was of variable quality and quality depending on the age of sample and how it was stored. Samples were either of blood or feathers. I analysed the effects of DNA quantity and quality and sample type on conversion rate on the SNP array. I generate a large dataset of the genotypes of 1,536 samples at ~50K genomic loci.

Chapter 3: Evaluating the impact of removing individuals and new immigrants on inbreeding in a reintroduced population using microsatellite and single nucleotide polymorphism markers

Translocation is a tool with many uses including increasing the number of populations of endangered species. Reintroduction is the translocation of a population back into its historic range. Typically, a small number of individuals are moved. These small populations can be particularly susceptible to the effects of genetic drift and inbreeding, resulting in loss of genetic variation and limiting population growth and adaptation. Usually genetic studies of translocation events are concerned with capturing the maximum genetic diversity of the source population in the translocated birds. Here I investigate instead the effects of removing individuals from a small source population

(Tiritiri Mātangi) over an 11-year period where birds have been removed for translocation seven times. I look for changes in genetic diversity and inbreeding over time and model the effects of removing individuals on inbreeding. I use both microsatellite markers and SNP array data in the analysis and note the differences in sensitivity of each marker to different metrics. To better understand changes in inbreeding across the genome, I look at ROHs in the population and note how they change over time. These shared haplotypes mark regions of homozygosity that may be autozygous (IBD) and can illuminate patterns of demographic history, selection, measure inbreeding and detect changes in inbreeding across the genome in the population. I look to see how length and number of ROHs in individuals are changing across time and if particular regions of the genome are increasing or decreasing in ROHs over time. I also identify the genes present in these regions to see whether they are over-representative of any particular functions.

Chapter 4: Effects of translocation bottlenecks on population genomics of an endangered species

Translocating small populations to new habitats can act as a bottleneck event reducing genetic diversity and heterozygosity of a population. This reduction affects the ability of the population to increase and adapt to new environments. Hihi have been reintroduced into predator-free locations around New Zealand's North Island to enable the number of populations to increase. There are currently six of these managed populations all derived from the remnant wild population on Te Hauturu-o-Toi. I use samples from the remnant wild hihi population (Te Hauturu-o-Toi) and samples from two of the managed populations (Tiritiri Mātangi and Karori) taken nine years after establishment to investigate both the genomic effects of the initial translocation and the effect of being maintained at a relatively small population size over a period of time. I use SNP data from a 50K SNP array to look at the erosion of genetic diversity across bottlenecks using measures of allelic richness and heterozygosity. In order to see how quickly the populations are diverging, I identified regions of the genome that were particularly susceptible to translocation bottlenecks by identifying SNPs that had increased frequency in ROHs after both translocations.

Chapter 2: Pooling sample data for assembly improves SNP discovery when using low coverage whole genome sequencing

Abstract

In the fields of ecological and conservation genetics, the detection and assessment of genetic variation allows inference of evolutionary processes and estimation of the adaptive potential of populations. The use of single nucleotide polymorphisms (SNPs) as a measure of genetic diversity is becoming increasingly common in the fields of ecological and conservation genetics due to the decreasing costs of genomic sequencing. In species where a reference genome is not available, the process of identifying large numbers of polymorphisms in a population may require the assembly of genomic sequence data to create a draft genome, and subsequent mapping of sequence reads from a representative subset of the population to detect SNPs. The cost of the initial high-depth whole genome sequencing coverage and the technical skills required to assemble and annotate a draft genome can be a barrier to SNP discovery in endangered species. In addition, samples from endangered species are often limited in the quantity and quality of DNA available. One approach, which offers the advantage of being able to genotype samples of variable quality and of different types, is to genotype previously identified polymorphisms on a SNP array. Here I use low-coverage whole genome sequencing from ten bird samples, and restriction digest sequencing (RAD-seq) of 31 bird samples of the threatened hihi or stitchbird (*Notiomystis cincta*) to identify variants for use on a 50K SNP array. I overcome the limitations of having a low quantity of sequence data available by assembling the RAD-seq data *de novo* using Stacks, by assembling the ten samples separately using SOAP-de-novo, and by creating two further draft assemblies by pooling reads from three samples or from ten samples. SNPs were identified in each assembly and their relative locations were mapped to the zebra finch (*Taeniopygia guttata*) genome using BLAST. The SNPs were filtered and pooled together and a subset spanning the zebra finch genome was selected for manufacture of the hihi SNP microarray. I found that pooling reads from different samples in the assembly resulted in fewer failed probes and that each of the assemblies resulted in a large portion of unique SNPs. I suggest that in instances in which small amounts of sequencing data is available per individual it may be appropriate to pool reads from samples in different combinations for assembly and SNP discovery in addition to using assemblies from single individuals to maximise the number of polymorphisms obtained and to increase the conversion rate. I also found that sample type (blood or feather) had no impact on genotyping success and that DNA of lower

quality than recommended for an Affymetrix SNP array can sometimes be successfully genotyped.

Introduction

Single nucleotide polymorphism (SNP) data is becoming increasingly common in ecological studies as falling sequencing costs have enabled the discovery and genotyping of hundreds to millions of variants across the genome. Recent reviews in conservation genetics have shifted focus from 'genetic' approaches (i.e., information from a single locus, or small panel of markers such as microsatellites) to 'genomic' approaches (i.e., large numbers of genome-wide polymorphisms) and reflects the excitement in the field for applying these technologies (Benestan et al., 2016; Corlett, 2017; Galla et al., 2016; Garner et al., 2016; Harrisson et al., 2014; Ouborg et al., 2010; Shafer et al., 2015). While SNPs have been employed in the study of wild populations for applications such as parentage assignment and relatedness measures (Kleinman-Ruiz et al., 2017), ancestry (Foote and Morin, 2016; Kawakami et al., 2014; Kleinman-Ruiz et al., 2017) population structure, genetic divergence and identifying loci under selection (Zhen et al., 2017), research in 'conservation genomics' has only recently begun to develop.

Genomics has been used to inform the conservation of a small number of species such as the Florida scrub jay (*Aphelocoma coerulescens*; Chen et al., 2016) and Bell's vireo (*Vireo bellii*) in North America (Klicka et al., 2016), reef fish (*Scarus niger*) in the South East Asia (Stockwell et al., 2016) and Irish grouse (*Lagopus lagopus scotica*) in Europe (Meyer-Lucht et al., 2016); but its full potential has yet to be realised in conservation management, due to issues around access to funding, expertise or communication between researchers and conservation practitioners (Taylor et al., 2017).

Two main methods are currently employed for SNP genotyping. 1) Array-based methods in which flanking probe sequences interrogate pre-identified SNPs (termed 'SNP arrays' or 'SNP chips'), and 2) 'genotyping-by-sequencing' (GBS) methods such as RAD-seq which sequence, assemble and call SNPs directly from genomic sequence data (Davey and Blaxter, 2010; Elshire et al., 2011). SNP chips offer a robust and easily replicable way of genotyping samples, with the added advantage that they can be used successfully on degraded DNA (Johnston et al., 2013; Mead et al., 2008), potentially allowing for museum and other historic samples to be included in wild species studies (Decker et al., 2009).

SNP arrays have been routinely used to study species of commercial value such as cows *(Bos taurus*; Matukumalli et al., 2009), sheep (*Ovis aries*; Kijas et al., 2009), goats (*Capra aegagrus hircus*; Tosser-Klopp et al., 2014), red deer (*Cervus elaphus;* Bixley et al., 2009), salmon (genus *Salmo*; Houston et al., 2014; Karlsson et al., 2011), rice (genus *Oryza*; Thomson et al., 2017) and soy bean (*Glycine max;* Wang et al., 2016). In contrast, only a handful of SNP arrays have been designed for non-commercial wild species such as house sparrows (*Passer domesticus;* Hagen et al., 2013b), great tit (*Parus major*; Kim et al., 2018; van Bers et al., 2012), polar bear (*Ursus maritimus*; Malenfant et al., 2015) and flycatcher (*Ficedula;* Kawakami et al., 2014). These SNP chips for wild species generate a quantity of genomic data that can be used to analyse the genomic architecture of traits (Husby et al., 2015; Johnston et al., 2016, 2011; Kardos et al., 2016; Kim et al., 2018; Lundregan et al., 2018; Miller et al., 2018; Robinson et al., 2013; Santure et al., 2013; Silva et al., 2017), assess linkage disequilibrium (Kawakami et al., 2014a), and characterise copy number variants in the genome (da Silva et al., 2018; Kim et al., 2018). SNP arrays can also be used for identifying adaptation, for example in cattle (Gautier et al., 2009) and to estimate evolutionary potential (Harrisson et al., 2014).

Hihi or stitchbird *(Notiomystis cincta*) is a threatened endemic New Zealand passerine that has undergone a program of translocations to predator-free sites across the North Island of New Zealand since the 1980s. Hihi are characteristic of species undergoing conservation as they were reduced to only a small remnant population, and current translocated populations require intervention to support robust population growth and persistence (Chauvenet et al., 2012; Doerr et al., 2017). They are a good model system to test how to translate genomic tools to wild threatened populations. Limitations to studying this species include variable and small sample sizes across time, unbalanced sample sizes across populations and inconsistencies in the collection and preservation of samples e.g. different types of samples (feathers or blood) were collected at different sites. Hihi are a phylogenetically distinct species, the only representative of the family *Notiomystidae* (Driskell et al., 2007), and therefore there is no closely related reference genome available. In this study, I describe how I overcame the limitations of low coverage genome sequencing, resulting in highly fragmented genomes, in order to identify polymorphisms. I outline the sequencing, assembly and SNP detection from second generation sequencing reads from two library types – RAD-seq of 31 individuals, and low coverage whole genome sequencing (WGS) from ten individuals. A subset of

detected SNPs was selected for inclusion on a custom SNP array and this was used to genotype 1,536 samples of varying quality and type from across different hihi populations. I test the conversion rates of different approaches for assembly and SNP detection, and consider the effects of pooling data for assembly on downstream variant calling. Lastly, I also discuss the effects of DNA quality and type on genotyping success rates. This SNP array will provide data to assess linkage disequilibrium, genomic architecture of traits and genomic differences between populations of hihi in future studies and help pave the way for more genomic studies in conservation biology.

Methods

Restriction-site associated DNA sequencing (RAD-seq) and assembly

Hihi individuals from two islands, the remnant population on Te Hauturu-o-Toi (24 individuals) and a translocated population on Tiritiri Mātangi (7 individuals) were selected for RAD-seq. DNA was isolated from the 31 hihi samples using an ammonium acetate precipitation method at the NERC Biomolecular Analysis Facility, University of Sheffield. Samples were inspected visually on an agarose gel for degradation, quantified with a DNA fluorometer (Hoefer DynaQuant 200), and normalised to approximately 50ng/µl. Samples were submitted to Floragenex (Inc.), Portland, Oregon, for RAD-seq, with one duplicate sample (to assess genotyping reproducibility), to give 32 samples in total for sequencing. Samples were digested with the restriction enzyme SbfI, sample libraries prepared and pooled, and single-end 90-bp fragments were sequenced across two lanes of an Illumina HiSeq™. A total of 257,833,998 reads was generated, with a median of 6,709,382 reads per sample.

The quality of demultiplexed raw reads received from Floragenex (Inc.) was evaluated using FastQC (Andrews, 2014). Except for in the first <10 bases (representing the 6-base fixed restriction site sequence for all reads), the inter-quartile range of the per-base quality scores always exceeded 28 for all 32 samples. As a result, reads were not trimmed but instead the software Stacks version 1.32 (Catchen et al., 2013, 2011) was used to remove reads with low quality scores, using the *process_radtags* program. Raw reads from the replicated individual were merged into one file. *Process_radtags* was then run on each of the 31 samples to clean the data and remove any read with an uncalled base (-c option), and discard reads (-q) where the average score within a sliding window of 15% (-

w 0.15) of the read length dropped below 15 (-s 15). Reads were further filtered using the *kmer_filter* program to remove reads that contained very rare (--rare) or very abundant (--abundant) kmers.

SNP detection in RAD-seq data

Filtered reads were then assembled *de novo* and SNPs detected using Stacks. Loci were assembled per individual using the *ustacks* program, with a minimum depth of coverage of six reads required to create a stack (-m 6) and up to two nucleotide mismatches (-M 2) allowed between stacks. The deleveraging algorithm was enabled (-d) to help resolve over-merged tags. A catalogue of loci across individuals was assembled using cstacks, with one mismatch allowed when merging loci in the catalogue (-n 1), resulting in a total of 131,412 loci. Individual reads were matched back to this catalog using sstacks. The *populations* program was then used to create an output vcf file of SNPs, with all individuals assigned to the same population, and SNPs filtered so that SNPs were present in at least 5 of the 31 individuals (-r 0.16), individuals had to have at least eight reads mapping to the locus (-m 8), and heterozygosity at the locus did not exceed 75% (--max_obs_het 0.75). A total of 30,835 SNPs were detected.

Whole genome sequencing (WGS) and assembly

Low coverage whole genome sequencing of ten individuals (a subset of the samples used in RAD-seq) were used to identify further polymorphisms. To maximise the variation captured, seven of the samples were from the remnant population on Te Hauturu-o-Toi and three were from the reintroduced population on Tiritiri Mātangi. Samples were multiplexed and two PCR-free DNA libraries were prepared by New Zealand Genomics Limited and used to generate 100bp paired-end Illumina reads over two lanes of Illumina HiSeq™ sequencing. This resulted in a total of 879,894,554 reads with a median of 44,782,143 per sample.

Sequence quality was assessed using FastQC (Andrews, 2014). Adapters and poor quality reads were removed with Trimmomatic-0.33 (Bolger et al., 2014) under strict conditions (ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:20, MINLEN:70, CROP:110), over-represented reads identified in FastQC were also removed by appending them to the TruSeq3-PE-2.fa file. The sample with the largest number of reads was used to run SOAPdenovo2 version 1.5.14 (Luo et al., 2012) at kmer sizes ranging from 25 - 95. The optimum kmer length of 36 bases was determined by N50 value and length of assembly and was used in subsequent assemblies.

Each of the samples was assembled using SOAPdenovo2 with kmer 36 and insert size 210. The samples with the three largest sets of reads - samples 6, 9 and 10 - were also assembled together ('3 in 1' assembly), as well as an assembly of all ten samples together ('10 in 1' assembly). Each assembly was mapped back to zebra finch with bwa-mem version 0.7.12 (Li and Durbin, 2010) and the zebra finch coverage was calculated using bedtools genomecov (Quinlan and Hall, 2010). Genome completeness was also assessed using CEGMA v3 (Parra et al., 2007).

Mapping and variant calling

For each of the twelve WGS assemblies, reads from all individuals were combined and mapped back to the assemblies using bwa-mem version 0.7.12 (Li and Durbin, 2010) adding read group headers (-R) and marking shorter reads as secondary (-M). As hihi is not a model species with known variable sites, GATK targetrealigner requires months to run and cannot be parallelised. To overcome this, assembly contigs over 200bp from the hihi draft genomes were grouped into 50 lists of approximately similar summed length, and these lists were used to split the bam files using an in-house perl script (divide_into_equal_regions.pl).

Local realignment was carried out with GATK version 1.3 (DePristo et al., 2011; McKenna et al., 2010) RealignerTargetCreator and IndelRealigner commands. Quality scores were then adjusted using the BaseRecalibrator and Printreads commands. The list of split bam files for each assembly was then used to merge these files back to a single re-alignment file with the SAMtools version 1.3 (Li, 2011; Li et al., 2009) merge, sort and index commands. Variants were called with SAMtools mpileup version 1.3 (Li, 2011; Li et al., 2009).

SNP selection from WGS

SAMtools output was parsed and annotated with an in-house perl script (vcf_annotate.pl), and the genotypes and alleles represented in each location (Tiritiri Mātangi and Te Hauturu-o-Toi) were recorded. BLAST 2.3.0 (AltschuP et al., 1990; Camacho et al., 2009) was used to map assembly contigs back to the Ensemble 86 zebra finch genome with an evalue cut off of 5. The BLAST output was parsed with an in-house perl script (blastparse.pl) which calculated the proportion of the query that aligned, retrieved the hit with the lowest evalue for each contig, checked the number of hits, skipped matches of less than 80% of the total query length, and checked if the SNP position on the hihi contig aligned to the zebra finch genome was in a gene using Ensemble biomart 86 gene

positions. All this information was then added to the SAMtools vcf output file using an in-house perl script (vcf_pos_ann.pl). An initial filtering of SNPs with read depth <10 was carried out using an in-house perl script (filter_read_depth.pl). For ease of downstream processing, files were split by zebra finch chromosome hits using an in-house bash script (setup_split_vcfs.sh), and vcf files from all hihi assemblies that corresponded with each zebra finch chromosome were merged and filtered using an in-house perl script (merge_filter.pl). The filtering resulted in SNPs being discarded if their hihi assembly contig hit more than one zebra finch chromosome; if the number of BLAST hits was greater than ten (in order to skip repetitive regions); if the variants were indels, monoallelic or multiallelic; if the distance to the nearest identified SNP was less than 40bp; if the SNP was not polymorphic in the Tiritiri Mātangi population (expected to have a reduced genetic diversity as a result of the reintroduction bottleneck); if the read depth < 19, minimum flanking contig sequence < 6 or maximum flanking contig sequence < 35; and if SNP types weren't A/G, C/T, G/T, or A/C (as these only require one probe on the SNP chip). Once merged, the best version of the SNP across all assemblies was identified with the required minimum flanking sequence and highest quality score and the distances to the next SNP or end of chromosome on the zebra finch genome was calculated (including the potential differences due to gap size present in the BLAST output). A total of 9,403,082 SNPs remained after the above quality filters and merging.

Selection of SNPs for the hihi SNP chip

A total target number of ~200,000 SNPs were selected (see below) and submitted to Affymetrix, Thermo Fisher Scientific, Santa Clara, California for assessment of the suitability of the SNPs for inclusion on a custom AXIOM 384HT SNP chip (which includes a check for duplicate flanking information suggesting repetitive elements, and an assessment of the complexity of the flanking sequence).

From the 30,835 RAD-seq SNPs, 9,484 SNPs were selected for consideration on the custom SNP chip by setting the minimum number of individuals genotyped for a SNP to 10. RAD-seq contigs containing SNPs were aligned to the zebra finch genome as described above for the WGS SNPs.

From the WGS, three steps were used to select SNPs for consideration. First, a target number of SNPs for each zebra finch chromosome were determined, proportional to the

length of the chromosome. Chromosomes were categorised into three SNP-density classes: high (chromosomes 10-28, LG1, LG5, LGE22, Z and the mitochondria), medium (chromosomes 1-9, 1A, 1B and 4A) and low (chromosome Un, which indicates zebra finch sequence of unidentified chromosome location), to reflect higher gene densities and recombination rates on the micro- compared to macro-chromosomes (Axelsson, 2005; Backström et al., 2013). Densities were adjusted so that high density chromosomes had approximately nine times more selected SNPs per megabase than chromosome Un, and medium density chromosomes had approximately 5.5 times more selected SNPs per megabase than chromosome Un. SNPs were further filtered to be at least 80 base pairs from the next known SNP. They were ranked based on SAMtools quality score and the appropriate number of SNPs taken for each chromosome, with a total of 185,647 selected in this step. Second, a total of 4,000 SNPs which mapped to random zebra finch chromosomes (e.g. 1_random) were also included, proportional to the total number of SNPs detected on each chromosome. Third, from a list of 1,185 high quality SNPs which did not map to the zebra finch genome, 560 were selected to represent each of the contigs with only one high quality SNP. Fasta sequences of flanking sequence each side of the SNP were extracted for all SNPs and formatted for Affymetrix according to their specifications using an in-house perl script (get_snps_fasta.pl), with 35 bases both upstream and downstream of the SNP extracted where possible; if not possible, 35 bases on one side and a minimum of one base on the other side.

Of the 199,691 SNPs submitted to Affymetrix, a total of 79,451 were deemed by Affymetrix to not be ‘designable’ in either the forward or reverse flanking sequence. From the remaining 120,240 designable SNPs, 59,928 SNPs were selected for inclusion on the SNP chip. All 654 RAD-seq SNPs designable in both forward and reverse directions were included, along with a further 73 SNPs designable in one direction and neutral in the other, selected by ranking SNPs based on their combined Affymetrix pconvert score. From the WGS SNPS, 48,220 were selected using a similar approach to the previous density selection. Densities were again adjusted such that that high and medium density chromosomes had approximately nine times and 5.5 times more SNPs per megabase than chromosome Un respectively. Two thousand one hundred and twelve (2,112) SNPs which were the only SNP within an annotated gene but failed to be selected among the best SNPs on a chromosome were also added, along with 559 high quality SNPs that failed to map to zebra finch. A further 4,155 SNPs were tiled in both directions

as both their forward and reverse flanking sequence was assessed to be neutral. The 59,928 SNPs were submitted to Affymetrix, and 58,466 of these were manufactured on the custom Hihi50K AXIOM 384HT array, which included 1,286 RAD-seq SNPs, 9,056 SNPs within annotated zebra finch genes (including 528 duplicates tiled in both directions), 727 unmapped SNPs and an overall total of 4,112 SNPs tiled in both directions. SNP probe flanking sequences were re-aligned to the zebra finch genome using BLAST. These results were then compared with the earlier full-contig mappings to see whether the probes aligned to the expected chromosome, on the random part of the chromosome, or on an alternative chromosome or on a combination of the above, where the flanking sequence mapped to a number of locations in the zebra finch genome.

Samples for genotyping

Genotyping was carried out on samples from five hihi populations. This study used blood samples collected from the main hihi study site of Tiritiri Mātangi between the 1996/1997 and 2014/15 austral breeding seasons, blood samples of birds translocated from the remnant population in Te Hauturu-o-Toi Island in the 2003/04, 2006/07, 2008/09 and 2010/11 breeding seasons, blood samples from Kāpiti Island in the 2003/04 breeding season, blood samples from Sanctuary Mountain Maungatautari in the 2011/12 breeding season and feather samples from Zealandia Wildlife Sanctuary from the 2013/14 and 2014/15 breeding seasons.

Blood samples were collected by brachial venipuncture (approximately 70 µl) and stored in 95% ethanol as described previously (Brekke et al., 2010). For Zealandia, two or three feathers were plucked and stored in 95% ethanol. DNA for ~2,500 individuals was extracted from the blood and feather samples using Qiagen DNeasy Blood and Tissue kits as recommended by the manufacturer. DNA was quantified on a NanoDrop 8000. One thousand five hundred and thirty-six (1,536) samples were chosen for genotyping on the hihi SNP chip based on their DNA quality (260/280 ratio of ~1.8-1.9 where possible), concentration (≥30 ng/µl where possible) and ensuring representation across cohorts. In total, 1,290 Tiritiri Mātangi, 55 Te Hauturu-o-Toi, 14 Kāpiti, 12 Maungatautari and 163 Zealandia samples were genotyped, plus 3 samples of unknown origin (see Supplementary Table S2). Samples were quantified before genotyping by Affymetrix using PicoGreen.

Results

Sequencing and assembly

SNP calling in Stacks returned 30,835 SNPs, 9,484 SNPs of which passed the quality filter and 2,388 of these SNPs mapped to zebra finch and were unique to the RAD data set and not found among the WGS SNPs.

Filtering and assembly of whole genome sequencing draft genomes from reads of single samples and pooled reads from three (3 in 1) or ten samples (10 in 1) resulted in genomes slightly smaller than the median bird genome length, but well within their known range of ~0.96–2.2 Gb (Kapusta et al., 2017). Draft genomes of pooled reads were marginally larger than those from single samples (see Table 1).

Assembled genomes were mapped to zebra finch to ascertain their contig positions and coverage of the genome. Neighbouring contigs for all assemblies overlapped 54% of the when mapped to the zebra finch genome, the median overlap between contigs in all single assemblies was -30 bases with a range from -110 to -1, and the median overlap between contigs in all pooled assemblies was also -30, with a range from -109 to -1 (Supplementary Figure 2a). Neighbouring contigs for all assemblies had gaps between them 46% of the time. In single assemblies, the median gaps between neighbouring contigs was 17,878 with a range between 0 and 1,083,088 bases, while the median gaps in pooled assemblies were smaller at 6431 bases, with a range of 0 to 610,970 bases (Supplementary Figure 2b).

SNP identification and characterisation

Pooled sample draft genome assemblies resulted in a smaller number of SNPs before and after SNP array design filtering (see Table 1). Across all assemblies, a total of 9,403,082 SNPs remained after quality filtering and merging of SNPs with homologous zebra finch positions (note that some of these SNPs are represented across multiple assemblies).

To select SNPs for inclusion on the SNP chip, hihi SNPs were classified by their mapping position on zebra finch chromosomes, with the aim to include a slightly higher density of markers on the microchromosomes, given expected higher recombination rates and gene densities on these chromosomes. SNPs from each chromosome were then chosen predominantly based on their quality (see Methods). Chromosomes with high density had 74.5 SNPs per Mb, those with medium density had 43.4 SNPs per Mb, and those with low density had 13.4 SNPs per Mb. The distribution of gap length between adjacent SNPs

in each of these groups is illustrated in Supplementary Figure 3 and the SNPs per chromosome are listed in Supplementary Table 2.

Table 1: SNPs per assembly before and after filtering. For each assembly, the size of the assembled draft genome (excluding N), the N50 value and the coverage of the zebra finch genome is shown along with gene completeness as assessed with CEGMA. The number of SNPs identified in SAMtools after filtering and the number of SNPs that are only found in that assembly is shown.

Assembly	size without N (Gb)	N50	% zebra finch	CEGMA % completeness	Filtered SNPs	Unique SNPs
1	979,320,412	1,175	63.5	7.7	1,050,028	734,870
2	929,250,757	676	59.0	1.2	963,210	690,776
3	836,706,703	379	51.3	0.0	720,939	527,184
4	967,240,548	989	62.3	6.1	1,041,331	735,443
5	971,406,841	1,086	62.8	5.7	1,058,210	745,124
6	991,036,624	1,559	65.0	10.9	1,046,922	719,408
7	988,256,031	1,371	65.2	7.7	1,093,037	767,461
8	951,613,123	834	60.9	2.8	1,014,469	719,369
9	991,536,846	1,482	65.6	6.1	1,088,067	755,084
10	1,002,019,675	1,928	66.0	16.1	1,019,426	685,923
3 in 1	1,048,884,582	3,137	68.3	23.8	806,025	420,806
10 in 1	1,046,305,858	1,960	67.6	15.3	764,792	416,123
RAD	11,827,080	90			2,887	2,388

SNP probes re-aligned to the zebra finch genome showed that most SNPs exhibited the same mapping as the hihi contigs from which they were identified (93.2 %). Only a small number aligned to a different chromosome than expected (4.9%) and many of these were on the ‘Un’ chromosome. A small number were found on the ‘random’ section of the same chromosome only (0.2%) which indicates that they are known to be on that chromosome, but the exact location is uncertain; on both the expected chromosome and either the ‘random’ section of the same chromosome (0.3%) or a second chromosome (1.4%); or on the ‘random’ section of the expected chromosome as well as an alternative chromosome (0.0%) (see Figure 1).

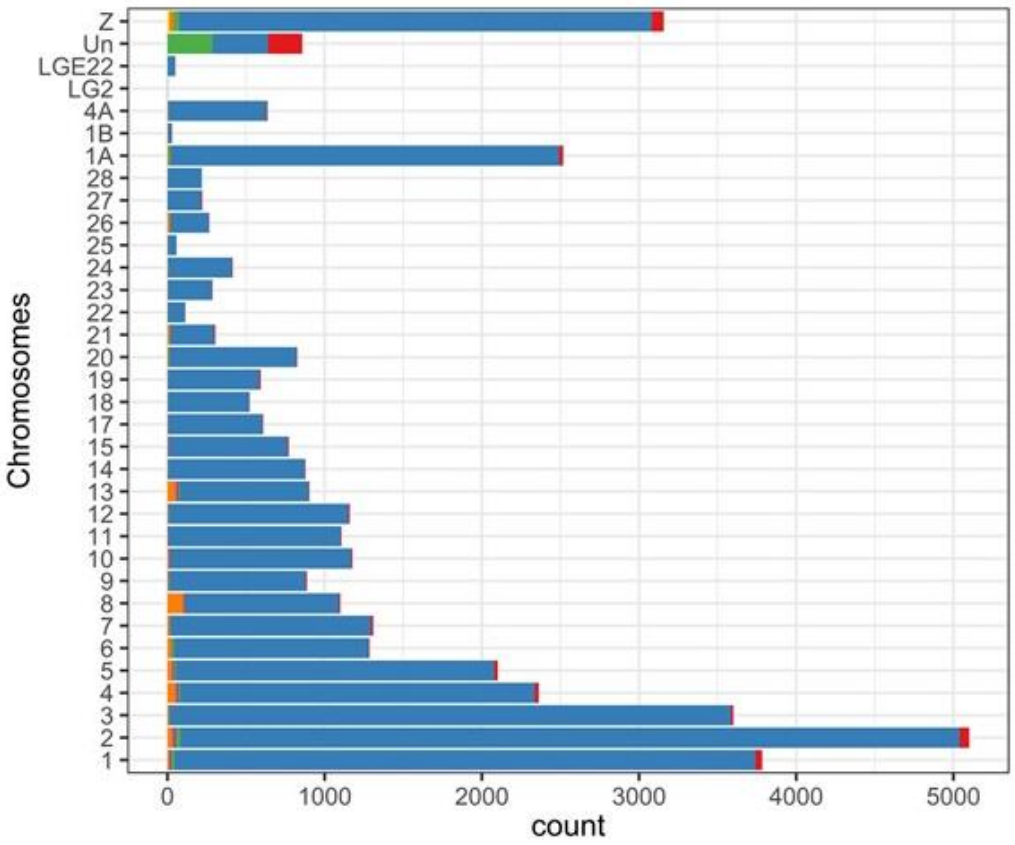


Figure 1: Rechecking SNP probe positions on the zebra finch genome. Homologous SNP positions were initially estimated from their position within their hihi contig's best-hit to the zebra finch genome. To ensure they were placed correctly, once the probes were designed, they were re-aligned to the zebra finch using only the SNP and its flanking sequences (37–71 bases). The majority aligned where expected (blue), a small number were found on the 'random' part of the same chromosome (orange), on an alternative chromosome (red), or on both the expected and random chromosomes (purple) or on both the expected and an alternate chromosome (green).

Of the 58,466 SNPs manufactured on the custom Hihi50K AXIOM 384HT array 42,212 markers (77.66% of the total) were polymorphic and passed Affymetrix filtering metrics in the Axiom Analysis Suite software, 7,898 (14.53%) passed filtering metrics but were monomorphic, and the remainder (7.81%) failed due to low call rates or other quality filters.

Sample type, quantity and quality

Of the total 1,536 samples, 96.03% were successfully genotyped according to Axiom Analysis Suite filtering metrics. Although no duplicate samples were intentionally included on the array, Axiom Analysis Suite identified six replicated samples, likely due

to plating errors during sample extractions. From these samples, genotyping reproducibility could be calculated, and was very high at 99.98%. There was no significant difference in success rate between blood and feather samples processed on the array (Pearson's Chi-squared test: Chi-squared = 0.040316, df = 1, *p-value* = 0.8409; there were 1,318 blood samples that passed and 55 that failed, feathers had 157 pass and 6 fail).

DNA sample quantity as measured by Affymetrix using PicoGreen had an impact on SNP failure rates as shown in Figure 2(a). A Mann-Whitney-Wilcoxon Test demonstrates that the mean DNA quantity of the fail group (30.5 ng/µl) is significantly smaller than that of the pass group (60.4 ng/µl), (W = 22557, *p-value* = 3.915e-11).

NanoDrop measurements of DNA quality showed no significant impact of being outside the recommended 260/280 ratio (1.8 – 1.9) on the pass rate of the samples (Pearson's Chi-squared test: Chi-squared = 3.6412, df = 1, *p-value* = 0.05637; of the DNA that fell in the recommended 260/280 ratio, 558 passed and 16 failed, of the DNA outside the recommended 260/280 ratio, 898 passed and 45 failed).

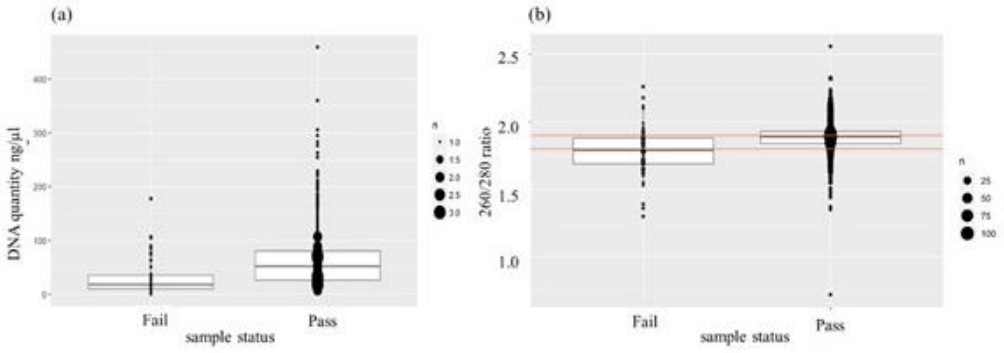


Figure 2: Impact of DNA quantity and quality on failure of SNPs. SNP failure was determined by Axiom filtering (Recommend.ps file) (a) DNA quantity as measured by Affymetrix using PicoGreen. The DNA quantity of samples that failed to pass Affymetrix filters were significantly lower than those that passed. (b) DNA quality as measured by NanoDrop 260/280 ratio. There was no significiant difference in failure rates for samples within and outside the recommended 260/280 ratio. The recommended ratio of 1.8-1.9 is between the two horizontal red lines.

RAD vs WGS failure rates

With 524 of 1,286 SNPs generated from RAD data failing genotyping compared to 15,730 of the 57,180 SNPs from WGS data, there was a much higher failure rate among the RAD SNPs (Pearson's Chi-squared test with Yates' continuity correction: Chi-squared = 109.79, df = 1, *p-value* < 2.2e-16).

SNP failure rate per assembly

Seven of the twelve assemblies showed similar levels of failure rate when their total SNP numbers were normalised, including the two assemblies from pooled data and five of the ten assemblies from single samples (Figure 3). On average, the two pooled assemblies showed lower failure rates than the ten single assemblies (Pearson's Chi-squared test with Yates' continuity correction: Chi-squared = 156.54, df = 1, *p-value* < 2.2e-16; see Figure 4).

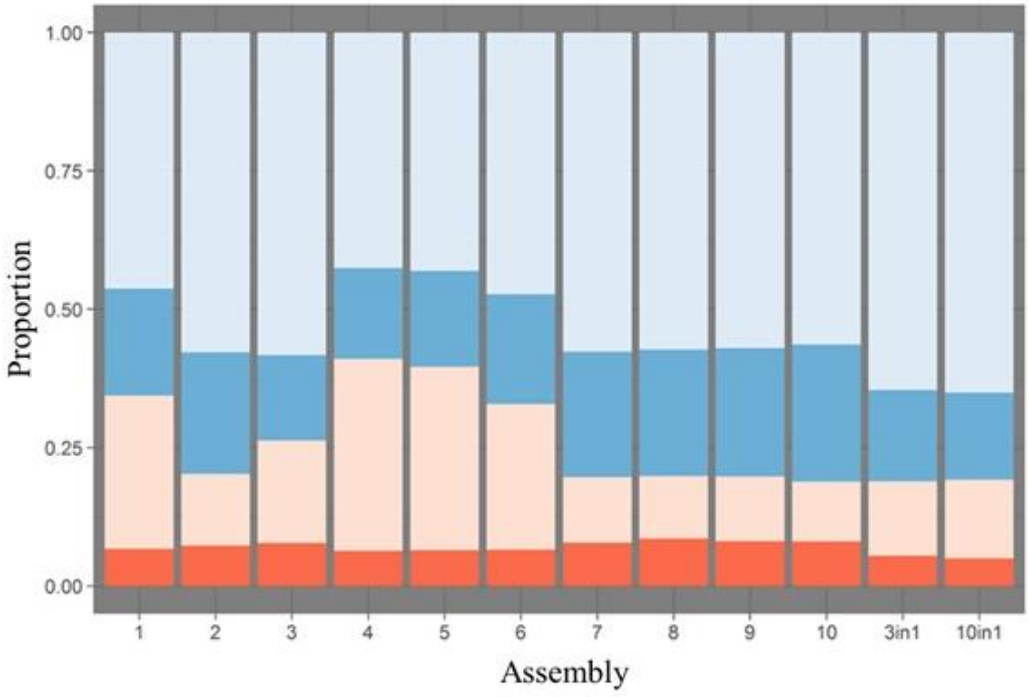


Figure 3: Failure of SNPs from each assembly. Pass and failure of SNPs were determined by the Recommended.ps file from Axiom filtering. Normalised proportions of the total number of SNPs on the array from each assembly were compared. For each assembly, there were SNPs that were unique to that assembly and passed the Axiom Analysis Suite filters (light blue) or failed the filtering step (light orange). There were also SNPs that are found in other (single) assemblies that passed filtering (dark blue) and failed filtering (dark orange).

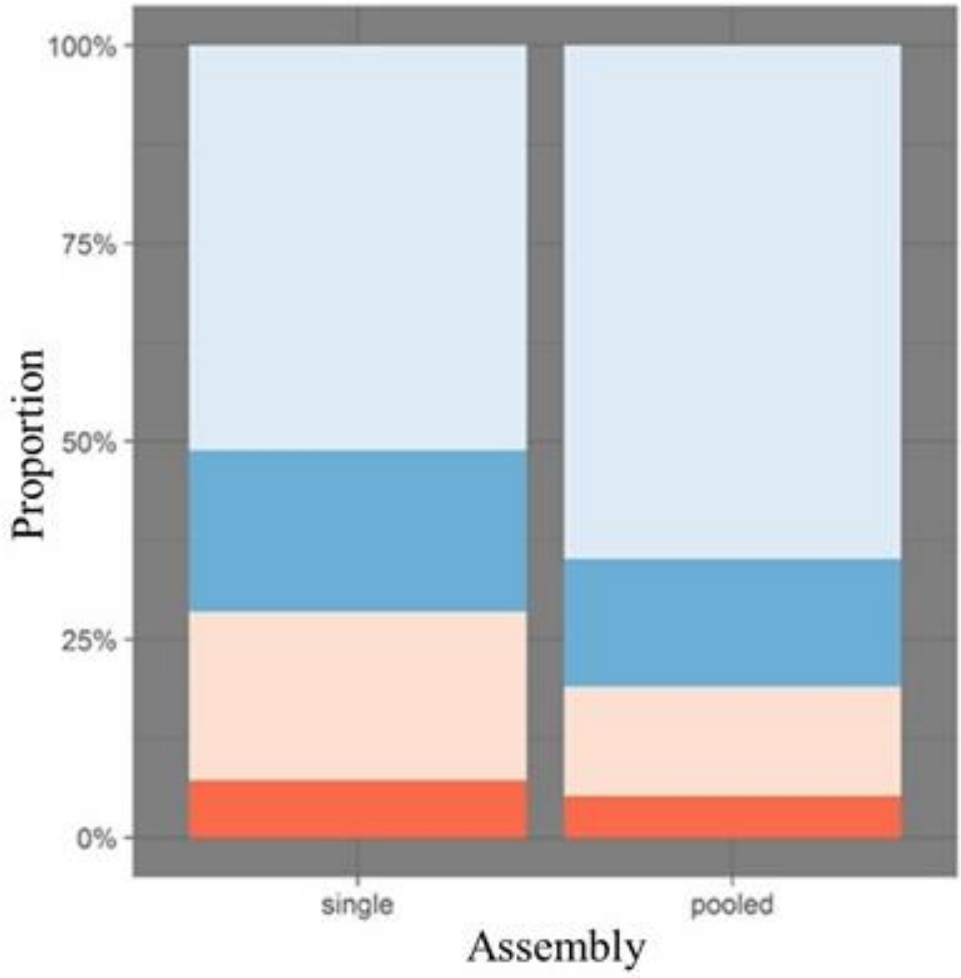


Figure 4: Averaged SNP performance for probes from single and pooled assemblies. The single column represents SNPs found on all draft assemblies for each of the ten samples assembled separately. The Pooled column represents SNPs from the assembly of pooled reads from samples 6, 9 and 10 (3 in 1) and pooled reads from all ten samples (10 in 1). The graph shows SNPs that were unique to one assembly and passed the Axiom Analysis Suite filters (light blue) and SNPs that are found in other (single) assemblies that passed filtering (dark blue). It also shows those that were unique to one assembly and failed Axiom Analysis Suite filters (light orange) and SNPs that were also found in other (single) assemblies that failed filtering (dark orange).

A very large proportion of SNPs were unique to only one assembly as shown by the 'Unique Passed' and 'Unique Failed' segments of Figures 3 and 4 and also in the 'Unique SNPs' in Table 3. In general, there were marginally more SNPs shared with pooled assemblies than with single assemblies (Table 3).

Pearson's Chi-squared tests with Yates' continuity correction were carried out on unique SNPs from either single or pooled assemblies, compared with those shared only with

single assemblies, those shared only with pooled assemblies and those shared with both single and pooled assemblies (see Table 2). SNPs in single assemblies had a significantly greater chance of success if they were also found in other single assemblies or in pooled assemblies (see Table 3 for test statistics). Pooled assembly SNPs had a significantly greater chance of success if they were also found in one or more single assemblies, but if they were also found in the second pooled assembly this had no significant impact on their success rate (Table 3).

Table 2: SNP probe performance on Affymetrix SNP-array. Absolute numbers for the performance of the SNPs from each assembly on the Affymetrix SNP-array. Each assembly had some overlapped SNPs shared with single assemblies and also some that were shared with pooled assemblies. A large number of the total SNPs were only found in one assembly, either single or pooled.

	total			unique			shared with an assembly from a single individual's data			shared with an assembly from pooled data		
assembly	**pass**	**fail**	**conversion**	**pass**	**fail**	**conversion**	**pass**	**fail**	**conversion**	**pass**	**fail**	**conversion**
1	4,753	2,492	66	3,187	1,960	62	1,399	486	74	714	241	75
2	3,562	903	80	2,516	556	82	982	325	75	442	114	79
3	2,276	811	74	1,773	564	76	476	240	66	191	62	75
4	4,447	3,092	59	3,096	2,576	55	1,238	480	72	605	213	74
5	4,629	3,036	60	3,187	2,502	56	1,325	495	73	671	230	74
6	5,125	2,507	67	3,323	1,940	63	1,514	503	75	936	285	77
7	4,894	1,198	80	3,313	679	83	1,380	478	74	735	226	76
8	3,943	979	80	2,755	536	84	1,123	421	73	527	175	75
9	4,910	1,209	80	3,285	656	83	1,419	494	74	811	281	74
10	5,049	1,174	81	3,149	596	84	1,540	499	76	1,042	334	76
3in1	5,447	1,274	81	2,493	487	84	2,142	625	77	5,447	1,274	81
10in1	5,145	1,222	81	2,433	488	83	1,900	572	77	5,145	1,222	81

Table 3: Chi-squared tests of SNPs that are unique or shared between assemblies from data of single individuals or pooled samples. Unique SNPs are those that are only found on one assembly (either from a single individual or pooled samples), shared SNPs are those that are shared across assemblies – either SNPs from an assembly of a single individual or SNPs from an assembly from pooled individuals, shared with assemblies from either single individuals, pooled data or both.

Unique SNPs in single/pooled assembly	Passed	Failed		Shared SNPs or Unique SNPs	Passed	Failed	X-squared	df	*p-value*
single	29584	12565	**vs**	SNPs from single assembly shared with one or more single assemblies	2893	1068	**13.97**	1	1.86E-04
single	29584	12565	**vs**	SNPs from single assembly shared with one or more pooled assemblies	1608	415	**79.96**	1	< 2.2e-16
single	29584	12565	**vs**	SNPs from single assembly shared with single and pooled assemblies	1627	545	**21.86**	1	2.94E-06
pooled	4926	975	**vs**	SNPs from one pooled assembly shared with one or more single assemblies	2428	723	**55.18**	1	1.10E-13
pooled	4926	975	**vs**	SNPs from pooled assembly shared with the other pooled assembly	812	162	**0.00**	1	9.69E-01
pooled	4926	975	**vs**	SNPs from pooled assembly shared with single and pooled assemblies	807	237	**75.73**	1	< 2.2e-16

Discussion

These results show that while pooling samples gives larger draft assemblies, it does not yield more SNPs after downstream variant calling and filtering (Table 1), but it may mitigate some of the effects causing failure in some assemblies from single samples. Each of the draft hihi genome assemblies from low coverage data resulted in a large proportion of SNP discoveries that were unique to that assembly regardless of whether they were from one single sample or pooled samples. Furthermore, samples with DNA quantity lower than the recommended yield can be successfully genotyped, and DNA quality and the type of sample has no significant effect on the efficacy of the SNP array genotyping. This study has created a large dataset of polymorphic data for 1,536 samples at 50K genomic loci across five populations of an endangered birds which can be used for future genomic studies of the genetic effects of management practices on the populations, assessment of linkage disequilibrium, genomic architecture of traits and estimation of evolutionary potential.

In our hihi 50K SNP array design, based mainly on low coverage whole genome sequencing from ten birds, 77.66% of the SNPs on the array passed Affymetrix filtering and were found to be polymorphic. By comparison, SNP discovery for the flycatcher 50K

SNP array was based on mapping to a 85X coverage reference genome and had a 90% conversion rate (Kawakami et al., 2014); SNP discovery for the house sparrow 200K SNP array was based on mapping to a ~130X coverage reference genome (Elgvin et al., 2017) and had a 92.8% conversion rate (Lundregan et al., 2018); SNP discovery for a low density great tit 10K SNP array was based on a transcriptome assembly from pooled sample data, and four genome assemblies from three libraries of reduced representation genomic DNA and one library of whole genome DNA, resulting in an 86% conversion rate (van Bers et al., 2012); and those for the high density great tit 500K SNP array, based on mapping to a high quality reference genome (Laine et al., 2016) had an 87% conversion rate for SNPs previously typed on the 10K SNP array, and an 82% conversion rate for unvalidated SNPs (Kim et al., 2018). In this dataset, the low coverage data is unlikely to have been able to adequately resolve repetitive or duplicated regions, and may have also resulted in high error rates in assembled contiguous sequences. In future, it would be valuable to re-map SNP flanking sequences from failed SNPs to a more contiguous genome sequence to determine whether these are more likely to appear in duplicated regions that failed to be identified from the low-coverage genome assemblies.

No significant difference in the genotyping success rate of samples extracted from feather or blood was found. Taking feathers is considered a non-invasive sampling method (the tip of the feather shaft is plucked, not cut), but it still requires bird handling. It is a useful method where bird handling is difficult and drawing blood might present a danger to the bird. One drawback is that the DNA extraction tends to use the whole sample (two-three plucked feathers), so there is no opportunity for reanalysis of the sample. Nucleated erythrocytes make bird blood an effective source of DNA, but here, in agreement with previous studies (Harvey et al., 2006; Maurer et al., 2010), I demonstrate that feathers are sufficient to successfully genotype an individual in cases where obtaining blood is not possible.

Overall individual genotyping success rates were very high (96.03%), but DNA quantity had a significant effect on sample success. Affymetrix recommends a minimum of 25µl at a minimum concentration of 23 ng/µl, with a recommended concentration of 30 ng/µl, but this threshold was relaxed in order to accommodate representation of cohorts and populations with fewer available DNA samples, such that 368 samples fell below the recommended concentration. Given that 333 of these samples genotyped successfully (albeit with a lower success rate than those above the recommended minimum [1,142 of

the 1,168 samples passed]), for important samples it may be worth attempting to genotype them even if DNA quantity is low. Failure rate here is much lower than has been reported elsewhere for samples of low DNA quantity (Kim et al., 2018).

To maximise cohort and population representation, of the 1,517 samples with 260/280 NanoDrop measurements taken, 798 were included on the SNP array even though they had DNA quality measures outside the recommended 260/280 ratio of 1.8-1.9 for DNA (Figure 2). I found that DNA quality had no significant effect on the overall genotyping success of the sample on the microarray, as has been shown elsewhere with human saliva samples genotyped on an Illumina OmniExpress array (Gudiseva et al., 2016) and fish scale samples genotyped on an Illumina iSelect array (Johnston et al., 2013).

WGS showed significantly greater success than RAD-seq data in SNP pass rates on the chip. It was expected that RAD-seq data would be more robust as it was generated from samples of 31 birds, however as there were many fewer SNPs, the filtering process was less stringent (i.e., the SNP needed to be present in at least 10 of the 31 individuals, individuals had to have at least eight reads mapping to the locus, and heterozygosity at the locus could not exceed 75%) versus a much more stringent filtering process for whole genome sequencing SNP discovery (i.e., SNPs had a read depth of at least 19, and in addition had been filtered to be polymorphic in the Tiritiri Mātangi population, at least 80 base pairs from the next known SNP or end of hihi contig, were detected in a hihi contig that only mapped to the zebra finch genome in one place, with a minimum flanking contig greater than 6 and a maximum flanking contig greater than 35). The higher success rates of the WGS SNPs is likely therefore a consequence of including a higher proportion of stringently filtered SNPs on the array compared to the RADseq SNPs. While other SNP arrays have been designed with SNPs discovered using a mixture of sequencing technologies and genome data (for example, SNPs discovered from transcriptome sequencing had a much higher failure rate than SNPs discovered from reduced representation sequencing in great tit; van Bers et al. 2012), to my knowledge no other SNP array has combined WGS and RAD-seq SNPs.

Pooling the reads for assembly may have had an effect on SNP pass rates on the chip. While on average, the assemblies from pooled data showed a greater pass rate than the assemblies from single individuals (Figure 4), there were only two assemblies from pooled data, compared with 10 assemblies from single sample data. Looking across all of

the single assemblies in the normalised data in Figure 3, it can be seen that half of the single assemblies had a similar pass rate to the pooled data, and that the lower average pass rate from the single assemblies came from only five of these single assemblies. This data suggests, however, that pooled assemblies may be able to attenuate the effects of variation in the quality of individual assemblies, as the 3 in 1 assembly had data from sample 6 (with a high failure rate) as well as samples 9 and 10 (with lower failure rates), and the 10 in 1 assembly had five samples with higher failure rates and five samples with lower failure rates.

The lower failure rates of the assemblies of pooled data may be as a result of how sequencing error and population polymorphisms introduce challenges for assembly algorithms (Bradbury *et al.*, 2011; *Hahn et al.*, 2014). In the case of a single individual, low levels of sequencing error and heterozygosity are more easily identified by de Bruijn graph genome assemblers such as SOAPdenovo2. For assemblies from pooled data, a much higher overall level of polymorphism is likely to lead to regions being duplicated in the assembly. This may translate into lower downstream mapping scores and lower numbers of variants called, as reads mapped back to the assembly can match more than one location. Lower mapping scores may then result in lower quality scores for these SNPs, and fewer SNPs from these regions reaching the quality threshold. The inflated size of the assemblies from pooled samples and their lower SNP yield (Table 1) indicates that this may have occurred here. Further, the filtered SNPs that are called from assemblies from pooled data may be less likely to be from regions of high polymorphism, as these regions may have been duplicated in the assembly, lowering SNP quality scores. This might lead to higher rates of SNP conversion overall, as polymorphisms in flanking regions that interfere with the SNP probe binding will be minimised.

SNPs that were shared across more than one assembly had a higher conversion rate on average than those that were unique to one assembly. Van Bers et al. (2012) similarly found that SNPs identified in both the United Kingdom and Netherlands great tit populations had higher conversion rates.

One interesting outcome of this SNP chip design was the observation that a very large proportion of the SNPs discovered on each of the assemblies was unique to that assembly. The assemblies each covered between 51-68% of the zebra finch genome, and CEGMA estimates that very few gene sequences were present in-full in any of the assemblies, and

so each could be representing a large proportion of the genome not assembled in the others. As bird genomes are highly conserved in gene synteny and chromosomal structure (Zhang et al., 2014), it is expected that the zebra finch genome coverage will represent a good estimate of how much of the whole hihi genome each of these draft assemblies cover.

The pooled assemblies shared only a small proportion of the SNPs discovered in the single assemblies. Furthermore pooling reads from 10 birds together identified different SNPs than pooling just reads from 3 birds. Crucially, these results indicate that assembling and remapping data in different ways can enhance the utility of the dataset, and led to the discovery of high-quality SNPs that would not otherwise be detected from a one-off assembly using a single or pooled sample.

Conclusion

Pooling data is considered an effective strategy to reduce overall costs (Wang et al., 2013), and is commonly used to build draft reference transcriptomes with RNA seq data (Perina et al., 2017; van Bers et al., 2012). Here I illustrate some of the limitations of data pooling, and suggest that in the absence of data with high genomic coverage, sequencing individuals separately and pooling data in different combinations for assembly may increase downstream variant discovery.

I also show that DNA quantity recommendations can be relaxed, but with lower pass rates and that DNA quality and sample type have no significant effect on genotyping by Affymetrix SNP array, making it an ideal method for lower quality DNA e.g. DNA from tissues that are poorly preserved or historic samples.

Acknowledgements

Many thanks to Rachel Tucker and Jon Slate at the NERC Biomolecular Analysis Facility, University of Sheffield, for completing the DNA extractions for the WGS and RAD-seq individuals. Thank you to Jason Boone and the Floragenex Team for the RAD-seq, and to the New Zealand Genomics Limited Team for WGS sequencing. I extend many thanks to Alayna Burrett, Millennium Science (NZ) Pty Ltd, for brokering the contract for SNP chip development and genotyping with Affymetrix. I also thank everyone in the Affymetrix Scientific Services and Bioinformatics Services teams who helped with the design, manufacture, genotyping and data delivery of the SNP chip, in

particular Christofer Bertani who designed the array. I am very grateful to the volunteers, past students and Department of Conservation staff who have contributed to monitoring the hihi populations and to the Hihi Recovery Group, Department of Conservation, and Supporters of Tiritiri Mātangi for maintaining such a long term vision in monitoring and management of the Tiritiri Mātangi population. I acknowledge Ngati Manuhiri as Mana Whenua and Kaitiaki of Te Hauturu-o-Toi and its taonga, including hihi. AWS, KDL, PB and JGE were supported by a Marsden Grant (UOA1408) awarded to AWS from the New Zealand Royal Society Te Aparangi. AWS was also supported by a New Zealand National Science Challenge Biological Heritage Project Grant, Project 1.4, and PB was also supported by an AXA Fellowship.

Ethical statement

Permissions to conduct the research and collect hihi blood samples were granted by the New Zealand Department of Conservation, permit numbers 15073-RES, 13939-RES, 246-RES, 36186-FAU, 24128-FAU, 32213-FAU and 44300-FAU.

Chapter 3: Evaluating the impact of removal of individuals and new immigrants on inbreeding in a reintroduced population using microsatellite and single nucleotide polymorphism markers

Abstract

Removing individuals from wild populations for translocation to new habitat can help increase populations of endangered species. Reintroduction is the translocation of animals to their historic range. It typically involves movement of a small number of individuals and these reintroduced populations can be particularly susceptible to the effects of inbreeding and genetic drift. This can lead to inbreeding depression and the loss of genetic variation, limiting population growth and adaptation respectively. When considering the optimal number of individuals to be removed from a source population, studies usually focus on capturing the maximum genetic diversity for translocation, and the effect of removing individuals on the source population is seldom quantified. Here I use a 50K single nucleotide polymorphism (SNP) array and microsatellite data to assess genetic diversity and test the effects of removing individuals from a small population of endemic New Zealand birds (hihi, *Notiomystis cincta*) on the inbreeding coefficient (F_{IS}) in the source population. I also observe the effect of a genetic top-up from the larger remnant wild population. I find that changes in genetic diversity from 2004-2014 were minimal and for the most part not significant for both marker types (metrics included allelic richness, expected and observed heterozygosity). Inbreeding increased over time when measured both by microsatellite and SNP markers, although the changes were small, and significant only for the SNP markers. Removing individuals had no significant effects on inbreeding measures, but I detected a decrease in inbreeding following the genetic top-up as measured by microsatellite markers. Using the SNP data, I detect regions of the genome that show consistent increases in inbreeding over time and found that genes in these regions were highly varied in function. Overall, these results suggest that removing individuals from this small population of hihi has not negatively impacted them genetically and that microsatellite and SNP markers provide complementary insight into inbreeding accumulation in small populations.

Introduction

Humans have always relied on wild plants and animals to survive. In modern times, studies of removing individuals from wild populations are primarily concerned with industries such as fishing (Swain et al., 2007) and forestry (Schaberg et al., 2008; Shackleton, 2015). In this context, the main concern is often the effect of artificial population decline on a particular phenotype or trait e.g. size of fish, (Swain et al., 2007) or antlers (Pozo et al., 2016), and sustainability for future harvests (Kreziou et al., 2016; Natusch et al., 2016; Walsh et al., 2006). A major aspect of sustainably removing a portion of the population is maintaining the genetic diversity of the current population (Baillie et al., 2016; Cruse-Sanders and Hamrick, 2004; Schaberg et al., 2008) and this is one of the main considerations within a conservation context (Allendorf et al., 2008).

Removing individuals can affect population genetics by changing the structure of the population (i.e. changing or removing population subdivision), decreasing genetic variation and causing selective genetic changes where individuals with particular traits are preferentially taken (Allendorf et al., 2008). Removing individuals has been shown to affect genetic variation – for example, changes to allele frequencies in zebrafish (*Danio rerio)* (Uusi-Heikkilä et al., 2017), eastern hemlock (*Tsuga canadensis*) (Schaberg et al., 2008) and white pine (*Pinus strobus* L.) (Rajora et al., 2000). Effects on population structure were shown in a culled red deer population (Nussey et al., 2005), harvested lake trout (*Salvelinus namaycush*) (Baillie et al., 2016) and wild American ginseng (*Panax quinquefolius*) (Cruse-Sanders and Hamrick, 2004). Selection pressures caused by harvesting wild populations have been reviewed by Kuparinen and Festa-Bianchet (2017). Genetic changes caused by the removal of individuals can be ameliorated by gene flow (Allendorf et al., 2008) or artificially by genetic rescue (Hedrick and Garcia-Dorado, 2016). However, excessive genetic changes in a population can reduce its capacity to recover to historic population sizes (Kuparinen and Festa-Bianchet, 2017; Walsh et al., 2006).

There are several metrics that can be used to assess the genetic diversity in a population and correlate these with changes in diversity over time. Allelic richness (A_R) is the number of alleles per locus (Höglund, 2009), it takes sample size into account and measuring it over time will convey whether alleles are being lost from the population. The most widely used metric for genetic diversity is observed heterozygosity (H_O) (Nei, 1973), which indicates the number and spread of alleles across a population. Expected

heterozygosity (H_E) measures the fraction of individuals in a population that are expected to be heterozygous at a given locus, assuming the population is at Hardy-Weinberg equilibrium (Nei, 1973). Any change in inbreeding levels between subpopulations over time can be measured with the inbreeding coefficient (F_{IS}). This measures the inbreeding level of an individual relative to the subpopulation (Wright, 1951).

Changes in genetic diversity in a population (such as measured by A_R, H_E, H_O and F_{IS} mentioned above) have traditionally been measured by microsatellites and more recently by SNPs (Morin et al., 2004). However, large amounts of SNP data also allows investigation of genomic changes (i.e. across the whole genome), and can give a more sensitive analysis of genetic diversity which may be required in a species that has undergone drastic decline e.g. Tasmanian devil (*Sarcophilul harrisii*) (Wright et al., 2015). For example, the distribution of SNPs across the genome has enabled the identification of runs of homozygosity (ROHs), contiguous regions of the genome that are dominated by homozygous markers, these are identical by state (IBS) and may also be identical by descent (IBD) (Howrigan et al., 2011; Rebelato and Caetano, 2018). These regions mark shared ancestry of genetic haplotypes, and longer ROHs represent IBD regions inherited from a common ancestor in more recent generations, while shorter ones have been inherited from more distant ancestors (Keller et al., 2011; McQuillan et al., 2008). SNPs enable calculation of an inbreeding *F*-statistic based on these runs of homozygosity for individuals within a population (F_{ROH}). F_{ROH} is highly correlated with pedigree-based inbreeding (F_{ped}) for long runs of homozygosity and has been proposed as an alternative when pedigree data are not readily available (McQuillan et al., 2008; Zhang et al., 2015). In addition to facilitating the measure of inbreeding (Curik et al., 2014; Forutan et al., 2018), the prevalence and location of ROHs have been used in population studies to elucidate population structure and demography (Bosse et al., 2012a; MacLeod et al., 2013) and identify regions under selection (Purfield et al., 2017).

Notiomystis cincta (hihi) is an endemic New Zealand passerine which has been managed as part of a reintroduction program since the 1980s. Tiritiri Mātangi Island is the source of the main population used for new translocations. It was founded from the progeny of 16 birds taken from the remnant wild population, on the island of Te Hauturu-o-Toi, a population estimated to be between 2,500-3,400 birds (Toy et al., 2018). This small number of founders is unfortunately common in reintroduction programs e.g. (Laugier et al., 2016; Miller et al., 2009; Seddon et al., 2005) and makes this a representative focal

population for reintroduction genetics. The Tiritiri Mātangi hihi population was shown to have a moderate to low genetic load, exhibiting male-specific inbreeding depression (Brekke et al., 2010) and is known to have low genetic diversity (Brekke et al., 2010; de Villemereuil et al., 2019a) and a lack of heritability in its adaptive traits (de Villemereuil et al., 2019a, 2019b). A genetic top-up of a further 20 birds from Te Hauturu-o-Toi to Tiritiri Mātangi was carried out in 2010 to address the expected accumulation of inbreeding in this small closed population since founding.

Removing hihi from a source population for translocation to new sites has thus far relied on demographic information. The demographics of the source population of hihi on Tiritiri Mātangi is well understood along with requirements for hihi survival and optimal management strategies for population growth (Armstrong et al., 2007; Armstrong and Perrott, 2000; John G. Ewen et al., 2011). Prior to each translocation, Armstrong and Ewen (2013) used demographic models to assess whether Tiritiri Mātangi was at carrying capacity (limited due to the island size; 220 ha), and predict whether removing juveniles for translocation would impact the population. The model estimates (i) the mean number of chicks fledged per female and the (ii) adult and (iii) juvenile survival probability per annum. It assumes that the probability of juvenile survival decreases with density until above a critical number of females (~24), and then it remains constant (at an equilibrium level determined by estimates (i) and (ii)). Ultimately, it was decided that retaining ~70 adult females in the source population was optimal (Armstrong and Ewen, 2013).

Here I investigate the genetic effects of repeatedly removing individuals from a small population of endangered birds, using microsatellite markers and a 50K SNP array as a complementary study to the demographic models. I look for any changes in genetic diversity measures (allelic richness (A_R), expected and observed heterozygosity (H_E and H_O) and inbreeding measures (F_{IS} and F_{ROH})) in the population over time. I test the effects of removing individuals on F_{IS} and observe the effect of the genetic top-up in 2010. I observe whether there are differences in inbreeding across the genome using ROHs and investigate if they change over time. I expected to see a decrease in genetic diversity over time as alleles are lost to drift, and also expect an increase in inbreeding values. I expect that the large number of 39,445 SNP markers will provide more precision in these genetic measurements than the 19 microsatellite markers. I hope that this study will offer insight into any potential pitfalls of removing individuals from a small population with low

genetic diversity, and whether genetic top-ups may be required to maintain its genetic equilibrium.

Methods

Study area and species

Tiritiri Mātangi Island sanctuary (36°36'S, 174°53'E) is a 220-ha wildlife sanctuary in the Hauraki Gulf 23 km northeast of Auckland City in New Zealand. Stripped of most of its forest by historic human activities, it was replanted between 1984 and 1994. Pacific rats were eradicated from the island in 1993 and 15 species of endangered fauna, including hihi, were translocated there between 1973-2011 (Galbraith and Cooper, 2013).

Hihi were eradicated from the mainland in the late 1800s by mammalian predators, but a remnant population survived on Te Hauturu-o-Toi. A hihi population was established on Tiritiri Mātangi from an initial translocation of 51 birds from Te Hauturu-o-Toi between 1995 and 1996. Of the translocated birds, twelve males and four females survived and bred, and the population steadily grew to a carrying capacity of ~150 adult birds. From 2004, juvenile hihi were removed from the population to establish populations elsewhere when predictive models indicated the loss would not greatly impact the Tiritiri Mātangi source population (Armstrong and Ewen, 2013). A second translocation of 20 birds from the Te Hauturu-o-Toi population was carried out in 2010 with the aim of increasing the genetic diversity of the Tiritiri Mātangi population. Population size on Tiritiri Mātangi was estimated from pre-breeding survey data taken in September before the breeding season and post-breeding survey data taken in February after the breeding season each year.

Samples

All fledglings on Tiritiri Mātangi are banded and a ~70 µl blood sample taken via brachial venepuncture at 21 days. Hihi blood samples were stored in 95% ethanol as described previously (Brekke et al., 2010). Here I use 1,189 individual samples representing a subset of the fledglings between the 2004/05 and 2014/15 austral breeding seasons. The number of samples and the total number of fledglings in the cohort is provided in Table 1.

Microsatellite data

Since 2006, as part of ongoing research all fledglings on Tiritiri Mātangi have been genotyped for a panel of 20 microsatellite markers, including two Z-chromosome-linked loci, as described previously (Brekke et al., 2009; Dawson et al., 2013). Each microsatellite locus has between two and ten alleles, and the observed heterozygosity ranges between 0.29 and 0.91 (Brekke et al., 2011). For the subset of individuals in this study, I excluded the Nci014 marker due to missing data, leaving 19 microsatellite markers (Brekke et al., 2009).

SNP data

Single nucleotide polymorphism (SNP) data was used from a hihi 50K SNP array manufactured by Affymetrix (described in Chapter 2). Briefly, SNPs were identified from low coverage whole genome sequencing and restriction-associated DNA (RAD) sequencing. To select SNP markers across the genome, SNP positions were approximated from homology with the zebra finch chromosome (Ensembl 86) and were chosen for inclusion on the SNP chip such that their density was designed to be high for micro chromosomes 10-15, 17-28, LG1, LG5, LGE22 and the Z chromosome and medium for macro chromosomes 1-9 as well as chromosomes 1A, 1B and 4A. DNA was extracted from all available blood samples using Qiagen DNeasy Blood and Tissue kits and was quantified on a NanoDrop 8000. Samples were chosen for genotyping on the hihi SNP chip based on their DNA quality (260/280 ratio of ~1.8-1.9 where possible), concentration (≥30 ng/µl where possible) and ensuring representation across cohorts (Table 1). Samples with less than 97% call rate on the first round of Axiom clustering were excluded from the analysis. 42,212 SNP markers were polymorphic on the array. The average reproducibility of calls was 99.98%. Polymorphic SNP density averaged 27.3 and 23.7 SNPs/kb for high and medium density chromosomes respectively.

Genetic Diversity

Allelic richness (A_R) was calculated to asses if alleles were being lost over time in the population. There are multiple possible alleles per locus for microsatellites, but only two per SNP locus due to design constraints for the SNP array. Allelic richness was calculated with diveRsity version 1.9.90 (Keenan et al., 2013) for both microsatellite and SNP data.

To test for changes in the distribution of alleles across the population, observed and expected heterozygosity measures (H_O and H_E) were calculated with diveRsity version 1.9.90 (Keenan et al., 2013) for both microsatellite and SNP data.

Genetic diversity measures (A_R, H_O and H_E) were graphed per cohort using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). Changes in genetic diversity over time were evaluated with linear models (A_R ~ cohort; H_O ~ cohort; H_E ~ cohort) for each marker type using R version 3.4.3 (R Core Team, 2017). Year was treated as a continuous variable as in this case it is like an index ratio. Models were tested as follows: test that the mean of residuals is zero, or close to zero; test that the x variables and residuals are uncorrelated using Pearson's product-moment correlation (scores between -1 and 1, a score of 0 means there is no correlation) and test for homoscedasticity, or equal variance of residuals (check visually by plotting the lm model) with R version 3.4.3 (R Core Team, 2017). Test for autocorrelation, the correlation of a time series with lags of itself, using the Durbin-Watson test (scores between 0 and 4, a score of 2 means there is no autocorrelation) using R package lmtest (Zeileis and Hothorn, 2002). Assume no perfect correlation between variables (check this visually) using R package corrplot (Wei and Simko, 2017). See Supplementary Table 4 and Supplementary Figures 4, 5 and 6.

Inbreeding coefficient

The changes in inbreeding across cohorts were assessed by calculating Weir and Cockerham's (1984) estimate of F_{IS} with GENEPOP 4.1 (Raymond and Rousset, 1995; Rousset, 2008) for both microsatellite and SNP data for each cohort. Linear models (F_{IS} ~ cohort) were used for each marker type to assess any changes in inbreeding levels in the population over time using R version 3.4.3 (R Core Team, 2017). To test the impacts of removing individuals on inbreeding, microsatellite and SNP F_{IS} were investigated using linear models in R (R Core Team, 2017). For each marker type, two models were tested: (i) F_{IS} ~ removing individuals /not removing individuals + pre-breeding population size + year, and (ii) F_{IS} ~ removing individuals /not removing individuals + post-breeding population size + year. Year is treated as a continuous variable here as in this case it is a ratio/interval. Models were tested as above, see Supplementary Table 4 and Supplementary Figures 4, 5, 6, 7 and 8.

R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016) was used to graph (i) the number of birds removed or added to the population per annum, (ii) F_{IS} for both microsatellite and SNP data and (iii) pre-breeding and post-breeding population size, all into one figure for comparison.

Runs of homozygosity (ROH)

In addition to a global estimate of inbreeding from ROH data (i.e., F_{ROH}), the local distribution of ROHs can be identified to detect changes in inbreeding levels across regions of the genome. The local distribution of ROHs was determined with PLINK v 1.07 (stable version; Purcell et al., 2007), perl v5.22.0 (Christiansen et al., 2012) and R version 3.4.3 (R Core Team, 2017) scripts. ROHs across the genome were detected using the --homozyg function in PLINK v 1.07 (stable version; Purcell et al., 2007). Using a perl script, the SNP data was split into macro chromosomes that were represented at medium density on the SNP chip and micro chromosomes that were represented at high density, as PLINK v 1.07 is limited to a maximum of 25 chromosomes for this function. Chromosome names were replaced with numbers and hyphens were removed from the SNP names for PLINK.

A sliding window of 20 SNPs was used instead of base pair window length, as SNP density was variable across the genome. The thresholds for scanning window hits were set at 0 heterozygotes, and the missing SNP call threshold was left at the default of 5. The default window threshold for each SNP of 0.05 homozygosity was used; this is the minimum proportion of sliding windows that qualify as an ROH for the SNP to be considered within an ROH. A recent study on data simulated to represent the equivalent of human diversity showed that ROHs with 45 SNPs over 949 kilobases could detect true autozygosity and did not have (ungenotyped) heterozygous sites between markers (Howrigan et al., 2011). Due to the lower genetic diversity of this wild population compared with humans, I took a lower threshold of 20 SNPs. In this data, although overall minor allele frequencies are low (Supplementary Figure 3), homozygotes made up 70% of the total calls in the dataset, suggesting that the probability of getting 20 homozygotes in a row by chance (i.e., identical by state but not identical by descent) is < 0.001. To further increase the likelihood that identified regions are identical by descent, the minimum length of the ROH was set at 10 kb and no heterozygous SNPs were allowed. The --homozyg-group parameter was used to identify groups of ROHs that were common across individuals.

Original chromosome names were restored to the output files using awk version 4.2.1 (Aho et al., 1987) scripts (available upon request). Data from all chromosomes were then merged, excluding data from the zebra finch random and unmapped chromosomes, resulting in information for 39,445 SNP markers for 1,189 individual hihi. Minor allele

frequency (MAF) for all SNP markers was plotted using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016).

The PLINK hom file output was used to graph the length of ROHs in each cohort, the number of SNPs per ROH in each cohort and the number of ROHs per individual for each cohort using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). Data distributions were assessed with histograms and linear models (ROH length ~ cohort; number of SNPs ~ cohort; ROHs per individual ~ cohort) were used to evaluate their trends over time using R version 3.4.3 (R Core Team, 2017). Models were tested as above, see Supplementary Table 4 and Supplementary Figures 9 and 10. Cohort was treated as a continuous variable as in this case it is a ratio/interval.

The PLINK hom file output was used to graph the percentage of SNPs in ROHs per cohort. The number of SNPs in ROHs was summed per individual. The percentage of SNPs in ROHs was calculated by dividing the number of SNPs in ROHs by the total number of SNPs (39,445) to identify more and less inbred individuals using R version 3.4.3 (R Core Team, 2017). Data was summed per cohort in a violin plot using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). A linear model (percentage SNPs in ROH ~ cohort) was used to identify any trends across time R version 3.4.3 (R Core Team, 2017). Percentage data is bounded which breaks one of the assumptions of linear model. There were only a small number of 0 percentages (at the bounds) and so the effect of bounding was minimal. Models were tested as above, see Supplementary Table 4 and Supplementary figures 11 and 12.

An emulator of the SNPSinRUNs function from the detectRUNS R package version 0.9.5 (Biscarini et al., 2018) was created in Perl v.5.26.2 to facilitate graphing of the data. The aim of the graph is to identify regions of the genome with high and low inbreeding levels. A custom script (Count_snps_in_ROH_per_group.pl, available upon request) was used to summarise ROH data. PLINK map files were used to identify the locations of SNPs in each ROH and PLINK hom files to identify the start and end of ROHs. Cohorts were grouped into early, middle and late cohort groups, which represent 2004-2007, 2008-2010 and 2011-2014 cohorts respectively. This allowed us to identify SNPs in the genome that increased or decreased in inbreeding levels over time. The number of times a SNP was found in an ROH was tallied for both an overall count and a per cohort group count. The overall percentage of SNPs appearing in ROHs was calculated for all the data (i.e. the

number of times the SNP was found in an ROH divided by the total number of samples) and also per group (i.e. the number of times a SNP was found in an ROH in samples in the cohort group divided by the total number of samples in that cohort group).

The overall percentage of SNPs in ROHs was plotted per chromosome using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). A linear model (percentage of SNPs in ROHs ~ cohort group) was calculated with R 3.4.3 (R Core Team, 2017).

Changes in the percentage of SNPs in ROHs between each group of cohorts (early to middle and middle to late) were graphed in a histogram for both macro chromosomes and micro and Z chromosomes using the R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). A one-sample Student's *t*-test was used to compare the mean to zero using R 3.4.3 (R Core Team, 2017).

SNPs were identified that had a consistent increase of at least 1% presence in ROHs between the early and middle cohort groups and between the middle and late cohort groups using R 3.4.3 (R Core Team, 2017). That is, if they increased between groups by 1% consistently, by the late cohort they appeared at least 2% more often within an ROH than in the early group. Similarly, consistent 1% decreases in percentage in ROH were detected over time. Accumulated changes were graphed with R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016).

Inbreeding measure from runs of homozygosity (F_{ROH})

Inbreeding measures can also be inferred from estimating levels of genome-wide autozygosity within individuals (i.e. when two genetic strands may have originated as identical by descent from the same individual) based on runs of homozygosity above a specified length (F_{ROH}) (Keller et al., 2011; McQuillan et al., 2008). The genome size was estimated directly from the Fasta format Ensembl 86 zebra finch genome (Zerbino et al., 2018). ROH data was taken from the PLINK hom file output and F_{ROH} was calculated for each individual for the SNP data set using the F_{ROH} _inbreeding function in the detectRUNS R package version 0.9.5 (Biscarini et al., 2018). Mean F_{ROH} per cohort was calculated and the trend in F_{ROH} across translocation events was tested with a linear model (F_{ROH} ~ cohort) calculated in R (R Core Team, 2017). Models were tested as above, see Supplementary Table 4 and Supplementary figures 11 and 12. A violin plot of the F_{ROH} per cohort was graphed using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016) and compared with the percentage of SNPs in ROHs per cohort.

Gene annotation

Genes in regions of accumulating ROH in SNPs across time were identified. Regions were defined as half the mean ROH length on either side of a SNP, and overlapping regions were joined. Gene annotations were obtained for these regions from the zebra finch (*Taeniopygia guttata*) Ensembl 86 genome gtf3 file (Zerbino et al., 2018) using a custom perl script (grab_gene_annotations_for_regions_from_gtf3.pl, available on request). GO term values were obtained using the Ensembl stable gene IDs from biomart Ensembl Genes 95 (Zerbino et al., 2018).

Results

Genetic diversity measures

Overall, the microsatellite data returned higher values for allelic richness, and expected and observed heterozygosity. Mean allelic richness (A_R), expected heterozygosity (H_E) and observed heterozygosity (H_O) were [4.76, 0.66, 0.65] and [1.82, 0.31, 0.30] for microsatellites and SNPs respectively. Changes in A_R, H_E and H_O over time were generally not significant for either marker type. Of these, only microsatellite H_E and SNP H_O showed significant trends across cohorts (shown in Supplementary Figure 1). Microsatellite H_E showed an increase of 0.0036 per year and SNP H_O showed a 0.0011 decrease per year.

Effect of removing individuals on inbreeding coefficient

F_{IS} measurements from microsatellite and SNP data showed a similar trend of about a 0.003 increase per year (Figure 1, data in Table 1). However, this was only significant in the SNPs but not in the microsatellite markers. SNP measurements were higher and less variable than the microsatellite data. Juvenile hihi were regularly removed from the Tiritiri Mātangi population to found or top-up other populations of hihi in predator-free locations around New Zealand (open circles in Figure 1a), with the number of birds removed from the population ranging from 11 birds to 72. The years when no individuals were removed from the population did not reduce the overall accumulation of inbreeding, as F_{IS} measurements from either microsatellite or SNP data were not impacted. Linear models of removing individuals showed no significant effects when including (a) whether the cohort was removed, the pre-breeding population and year, or (b) whether the cohort was removed, the post-breeding population and year, for either SNP or microsatellite values of F_{IS} (Supplementary Table 1). The addition of 20 birds from Te Hauturu-o-Toi in 2010 (black diamond in Figure 1a) appeared to lead to a decrease in the inbreeding

coefficient which was stable in the three subsequent cohorts for microsatellite data before increasing again. However, a similar trend was not evident from the SNP data (Figure 1b).

Table 1: Comparison of microsatellite and SNP Inbreeding coefficient (F_{IS}). F_{IS} was calculated in genepop for both microsatellite and SNP data.

Cohort	Total fledged	Total samples	Microsatellite F_{IS}	SNPs F_{IS}	SNPs F_{ROH} mean	SNPs F_{ROH} median
2004	179	109	-0.02	0.01	0.24	0.23
2005	161	39	-0.04	-0.01	0.23	0.22
2006	158	39	0.00	0.01	0.24	0.24
2007	200	150	0.00	0.03	0.25	0.24
2008	196	144	0.02	0.02	0.24	0.24
2009	194	148	0.03	0.02	0.24	0.24
2010	251	103	-0.01	0.03	0.25	0.24
2011	152	90	-0.01	0.02	0.24	0.24
2012	184	155	-0.01	0.03	0.25	0.24
2013	154	129	0.01	0.03	0.24	0.24
2014	91	83	0.03	0.03	0.25	0.25

Runs of homozygosity

Across time, there were no significant changes in the overall length of ROHs or in the total number of ROHs per individual, but small and significant increases in the number of SNPs contained within an ROH (Figure 2). The median length of ROH was 525.8 kb and increases in ROHs length were not significant, with an increase of 652 bp per year (length ROH ~ cohort; p-value: 0.2473, 10,658 df, adjusted R^2 : 3.175e-05, *t-value*: 1.157). While median ROH lengths did not change significantly, the length of the longest ROHs spiked from just under 1,391 kb in 2004 to just over 1,721 kb in 2007, 1,634 kb in 2009 and 1,679 kb in 2012. The number of SNPs within ROHs also changed very slightly, but significantly, with a mean increase of 0.02859 SNPs per year (SNPs in ROH ~ cohort; *p-value*: 0.007625, adjusted R^2 : 0.000574, *t-value:* 2.669). There was a slight trend of increasing number of ROHs per individual over time with a mean increase of 0.14728 per year (ROHs per individual ~ cohort; *p-value*: 0.05345, 1,180 df, adjusted R^2 : 0.002312, *t-value*: 1.933).

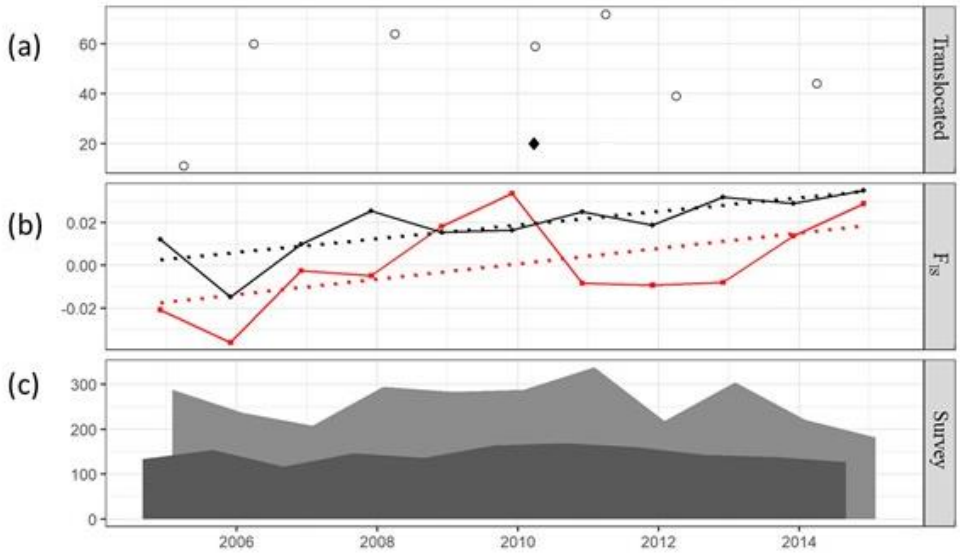


Figure 1: Microsatellite and SNP-based F_{IS} calculated for each cohort on Tiritiri Mātangi from 2004 to 2014. (a) Number of individuals removed or added to the population at the end of the breeding season in April. Empty circles represent the number of juveniles that have been removed from the population for translocation purposes. The filled diamond represents the addition of adult and juvenile birds from the remnant population Te Hauturu-o-Toi. (b) F_{IS} values for SNP and microsatellite data across all birds within that austral breeding season, plotted in the middle of the breeding season. SNP data is shown in black (y = -6.395644 + 0.003193x, adjusted R^2 = 0.5486, *p-value* = 0.005511) and microsatellite data is shown in red (y = -7.210038 + 0.003589x, adjusted R^2 = 0.2431, *p-value* = 0.07036). (c) Survey data of the number of birds observed in the population. Pre-breeding survey data (individuals counted in September) is shown in dark grey, post-breeding survey data (counted in February) is shown in light grey.

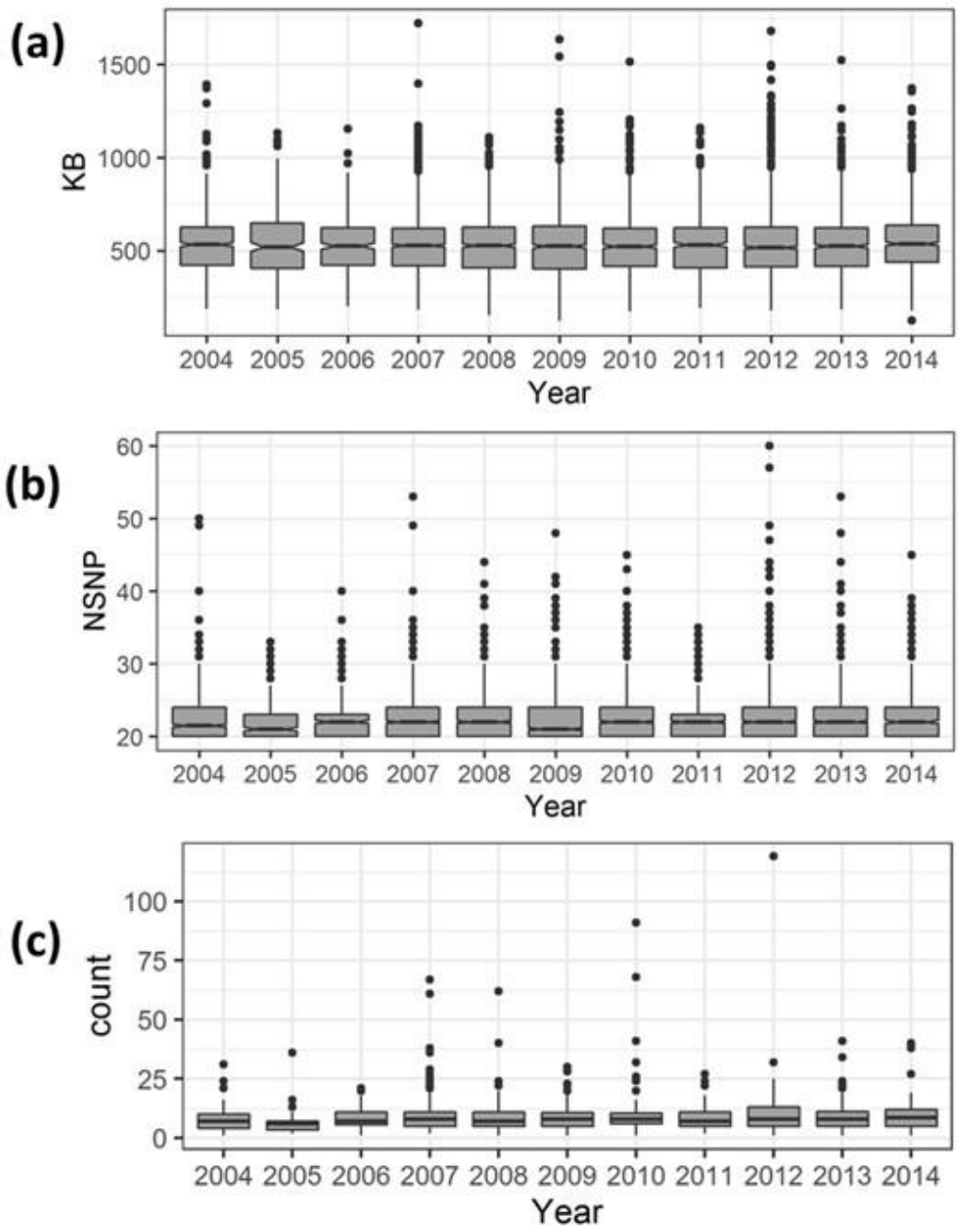


Figure 2: Average changes in runs of homozygosity per cohort. (a) length in kilobases of ROHs per cohort (linear model: length ROH ~ cohort; y = 530736.9 + 652.7x, adjusted R^2: 3.175e-05, *p-value*: 0.2473). **(b)** number of SNPs within ROHs per cohort (linear model: number of SNPs in ROH ~ cohort; y = 22.39601 + 0.02859x, adjusted R^2: 0.000574, *p-value*: 0.007625). **(c)** number of ROHs per individual per cohort. (linear model: ROHs per individual ~ cohort; y = 8.08335 + 0.14728x, adjusted R^2: 0.002312, *p-value*: 0.05345)

The distribution of the change in the percentage of SNPs contained within ROHs between early and middle and middle and late cohort groups was calculated separately for macro chromosomes that were represented at medium density and micro and Z chromosomes that were represented at high SNP density on the SNP array (shown in Supplementary Figure 2). A one-way Student's *t*-test of the change in percentage of SNPs in ROH between cohort groups (i.e. 4-5-year period) showed very small but significant changes over time. The percentage change for macro chromosomes had a mean of 0.028 (*t-value* = 10.89, df = 51771, *p-value* < 2.2e-16) and for micro and Z chromosomes had a mean of 0.0199 (*t-value* = 6.0708, df = 27117, *p-value* < 1.289e-09).

F_{ROH} change across individual cohorts was significant but small, increasing by 1.101e-04 annually (linear model: F_{ROH} ~ cohort; Adjusted R^2: 0.00310, *p-value*: 0.03048). Similar to the increase in F_{ROH}, the percentage of SNPs in ROHs increased significantly across all cohorts but was also small at 0.011997 per annum (linear model: percentage SNPs in ROH ~ cohort; *p-value*: 0.02614, Adjusted R^2 = 0.003327, see Figure 3). The rate of increase in average F_{ROH} per year did not appear to be influenced by the top-up in 2010 (see Supplementary Table 2). This same trend of an increase in the total percentage of SNPs within ROHs was seen across all chromosomes (data not shown).

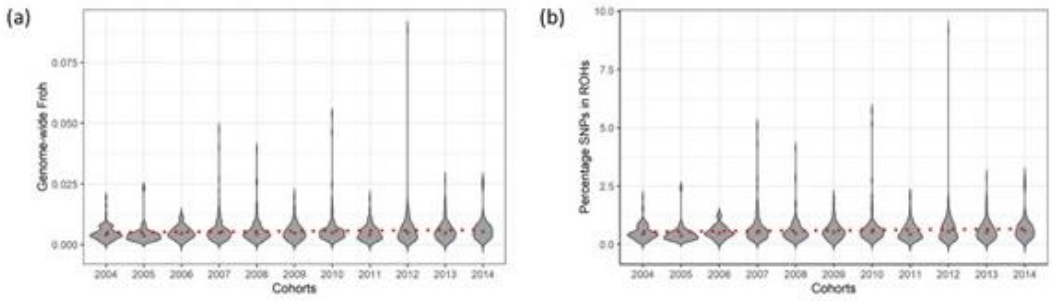


Figure 3: Measurements of average runs of homozygosity for each individual or SNP per cohort: The median for each cohort is shown as a black circle, the best-fit linear regression line is shown in red. (a) Genome-wide F_{ROH} is the sum of the ROH lengths divided by the genome length as calculated by detectRUNs R package. (y = -2.156e-01 + 1.101e-04x, Adjusted R^2 = 0.003104, *p-value* = 0.03048). (b) Percentage of SNPs located with a run of homozygosity per SNP per cohort. (y = 0.528163 + 0.011997x, Adjusted R^2 = 0.003327, *p-value* = 0.02614)

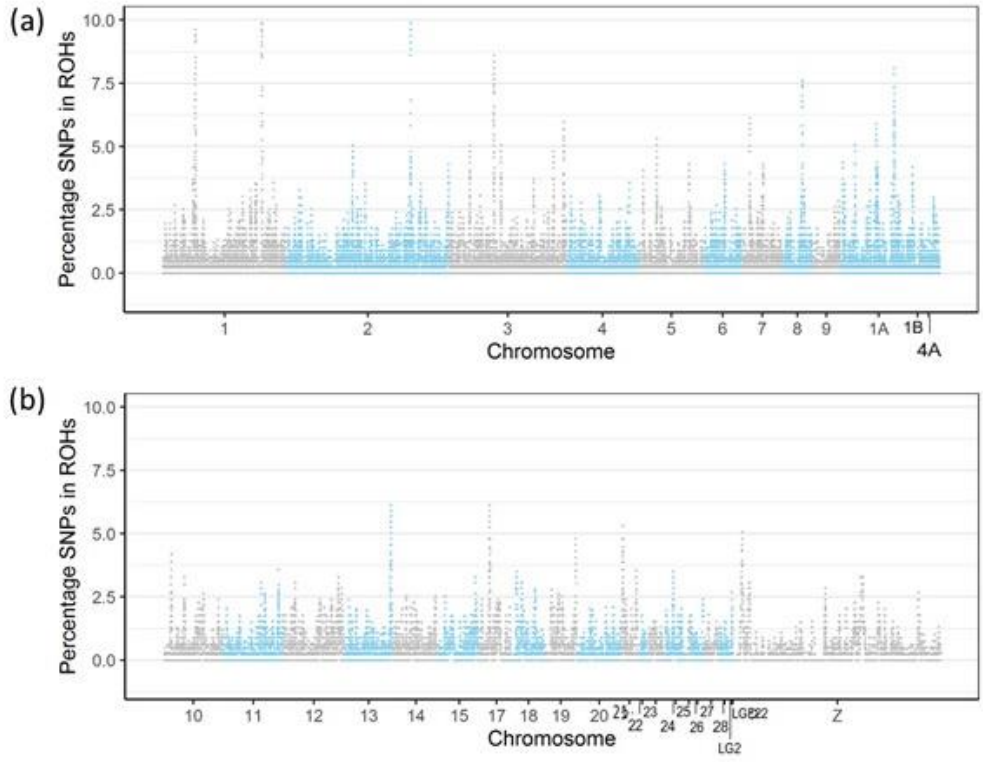


Figure 4: Percentage of SNPs in runs of homozygosity for Tiritiri Mātangi individuals in the late cohort group (2011-2014). (a) Macro chromosomes, which were represented on the SNP array at lower density show higher levels of percentage SNPs in ROHs (i.e. number of individuals in a population with a particular SNP in an ROH). (b) Micro chromosomes and the Z chromosome which were represented on the SNP array at higher density, show lower levels of percentage SNPs in ROHs.

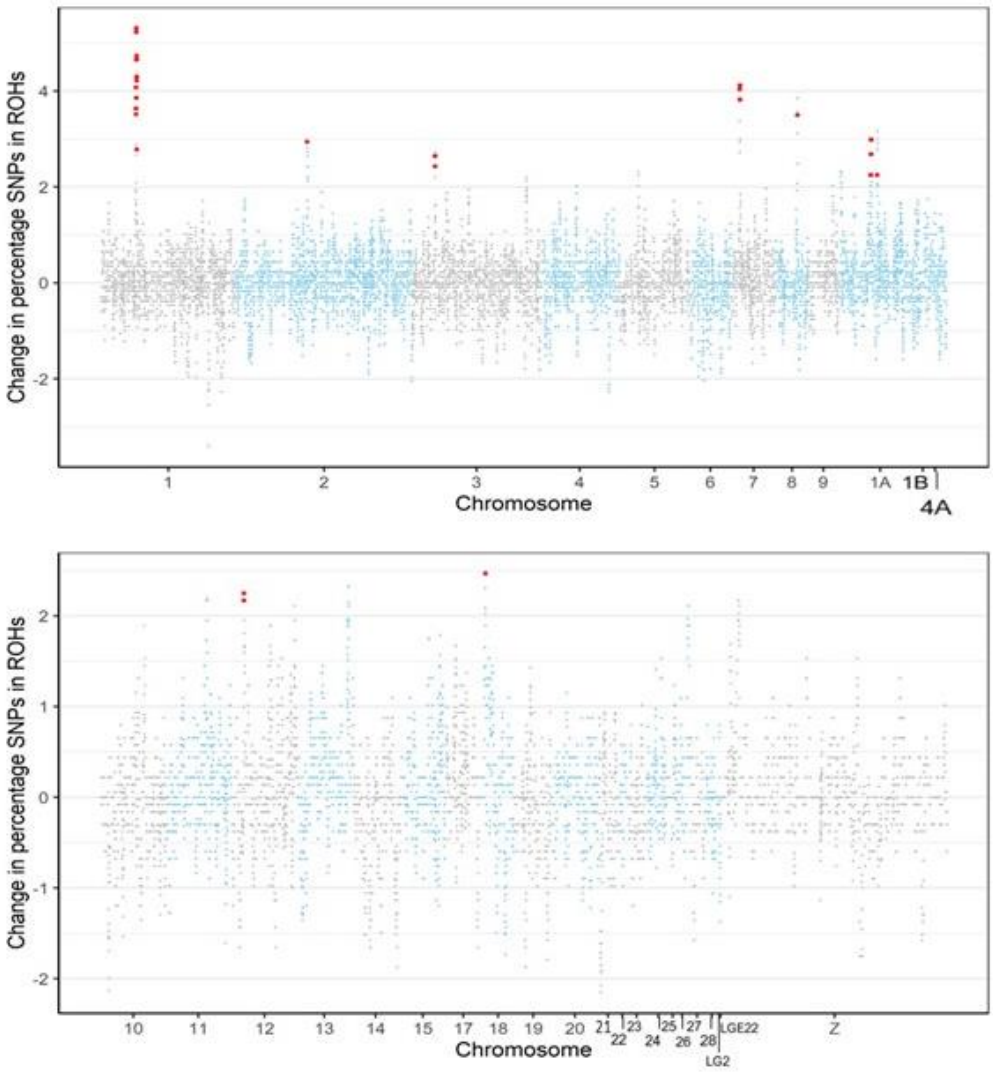


Figure 5: Accumulated change in the percentage of individuals in which each SNP appears within ROHs. The total percentage change for each SNP from the early to late cohorts is illustrated for each chromosome in alternating grey and blue colours. SNPs that have changed by at least 1% between early (2004-2007) to middle (2008-2010) cohort groups and again by at least 1% between middle (2008-2010) to late (2011-2014) cohort groups are highlighted, i.e. changed by a minimum of 2% overall. Accumulated 2% increases are shown in red, no accumulated 2% decreases observed. Regions which continuously accumulated changes in percentage SNPs in ROH are present on chromosomes 1, 2, 3, 7, 8, 1A, 12 and 18.

Figure43 illustrates the percentage SNPs in ROH for each SNP in both (a) macro and (b) micro and Z chromosomes in the late cohort group. Overall each SNP is only in ROHs for a small proportion of the population, but there are some areas, especially in the macro chromosomes that have higher proportions of SNPs in ROHs marked by large spikes on

the graph. This may be due to a slightly lower density of SNPs on macro-chromosomes compared with micro-chromosomes and the Z chromosome, or could reflect the higher recombination rates across the micro-chromosomes which may break up ROHs. There is no regular pattern to the location of these spikes along the chromosomes.

The total percentage change of SNP presence in ROHs for each SNP is illustrated in Figure 5. Highlighted SNPs were accumulating SNPs in ROHs across time. These SNPs were increasing in percentage in ROH by at least 1% across both cohort group transitions (early to middle and middle to late) and therefore had at least 2% difference between the early and late cohort groups.

Regions that were accumulating SNPs in ROHs contained 94 genes when compared against the zebra finch genome (see Supplementary Table 3). Regions were defined as half the mean length of ROHs before and after each SNP, and overlapping regions were joined. Of the 94 genes found, 61 had gene name annotations, and 80 had gene ontology records in Ensembl with 400 unique GO-term names and 104 GO-slim values spanning a wide range of functions. These included DNA and RNA binding, circulatory, nervous system, immune system and carbohydrate metabolic processes, cell cycle, embryo development, developmental maturation, aging, transmembrane transport and response to stress.

Discussion

Here I examined the effects of regularly removing individuals from a small population of endangered birds, quantified the changes in the population genetics over time, and measured the effects of a genetic top-up. I used two marker types and found that genetic diversity measures were lower for SNP markers than microsatellite markers. I found that inbreeding was low but increasing slightly over time (but was significant in the SNP data only). Removing individuals showed no significant effects on the F_{IS} inbreeding measures of the population, but the addition of 20 birds into the population decreased microsatellite F_{IS} measurements for the period over which the newly added birds were breeding. I found regions of the genome that were becoming more inbred over time and showed that the functions of the genes in these regions were highly varied. These results suggest that the ongoing removal of individuals from the Tiritiri Mātangi population has had little effect on the population.

Genetic diversity

Measures of genetic diversity for the Tiritiri Mātangi population were higher when calculated with microsatellite markers compared to SNP markers. Using microsatellites, Brekke et al. (2011) found a similar level of allelic richness, observed and expected heterozygosity in the population - 4.71, 0.66 and 0.64 respectively compared with 4.76 , 0.65 and 0.66 in this study. The small differences between the two studies can be attributed to the fact that here I used only samples that also had SNP data available, and to the inclusion of additional data up to the 2013/14 breeding season. The SNP data shows comparatively reduced measures of diversity, likely due to the small number of alleles per locus when compared to the microsatellite loci. Microsatellites are not a representative measure of genome-wide diversity as there is only a weak relationship between microsatellite diversity and nucleotide diversity, and high microsatellite mutation rates and genotyping errors can also bias overall measurements of genome-wide diversity (Väli et al., 2008). SNP markers, particularly those genotyped on SNP arrays, also suffer from bias such as the loss of rare polymorphisms, ascertainment bias during SNP discovery and selection for genotyping arrays (Albrechtsen et al., 2010) and the inclusion of both coding and non-coding loci (Helyar et al., 2011). However, SNP based approaches to measure diversity are preferable (Morin et al., 2004; Väli et al., 2008) and here it is likely that the SNP measurement is a more accurate reflection of actual genome-wide diversity. While the results on changes in genetic diversity are mostly not significant here, SNP data is expected to give a much more precise measurement of diversity than microsatellites.

Inbreeding measures (F_{IS})

Inbreeding measures of F_{IS} were very close to what is expected under Hardy-Weinberg equilibrium with microsatellite measures indicating the population has slightly lower heterozygosity than expected (ranging from -0.02 to 0.03) and SNP measures indicating slightly higher levels of heterozygosity than expected (ranging from 0.01 to 0.03). There was much less variance shown in the SNP markers than microsatellite inbreeding measures. Santure et al. (2010) has shown that a few hundred SNPs can give accurate inbreeding estimates. The 39,445 SNP markers used in this study are therefore likely to be many times what is needed to accurately measure inbreeding. In this study, SNP data is giving much more precise measurements than microsatellite data (for example across all groups, the average SNP F_{IS} is 0.018 with confidence intervals 0.011 and 0.023, average microsatellite F_{IS} is 0.005 with confidence intervals -0.014 and 0.019). This level

of sensitivity is very useful in a conservation context where genetic diversity of endangered animals is already compromised.

The Tiritiri Mātangi hihi population has experienced some increase in inbreeding over time, which is significant only for the SNP measures (F_{IS} from microsatellite data, F_{IS} from SNP data and SNP F_{ROH} data increases by 0.003589 [*p-value* = 0.07036], 0.003193 [*p-value* = 0.005511], and 0.0001 [*p-value* = 0.03048] respectively per year). The level of inbreeding estimated from hihi SNP data is much less than, for example, wild and captive Hawaiian tree snails (*Achatinella lila*). Six founder snails captured in 1997 had F_{IS} of 0.395 which reduced to 0.244 in 45 juvenile captive snails (<1 year old) and 0.224 in 92 wild snails measured 16 years later in 2013 (F_{IS} *p-values* all < 0.001) (Price and Hadfield, 2014). Further, compared to this study, in a study of human SNP data, Keller et al., (2011) used much larger numbers of samples to measure F_{ROH}, and suggested that due to the large variance in ROHs in populations of humans, 12,000 or more individuals may be required for accurate measurements of inbreeding. Humans have more genetic variation than hihi, thus a closer investigation of ROH variation across the hihi population might show if this study is under-powered.

Genetic rescue

Although not evident from the SNP data, the effects of the genetic rescue can be seen in the microsatellite F_{IS} measures. This suggests that the highly variable microsatellite markers may be better suited to detect increases in genetic diversity, particularly increases in the frequency of rare alleles, in this instance derived from birds that successfully bred following translocation from Te Hauturu-o-Toi. Of the birds released, four females and two males successfully reproduced in the 2010-2011 breeding season, and five females survived to the following breeding season, with their last recorded breeding in 2012 (Brekke et al., unpublished work). Microsatellite F_{IS} shows no significant change overall, but the F_{IS} decreased in the 2010/11 cohort, stayed down for two years at the end of the 2011/12 and 2012/13 breeding years and then returned to previous higher measurements in 2013, the year after the last immigrant had offspring. SNP data was not sensitive to any changes in inbreeding either through F_{IS} or F_{ROH} measurements for those same cohorts. However, although the allelic richness (A_R) of SNP markers has no significant trend, these measures also matched the expected pattern after genetic top-up. A_R showed a marked increase following the input of birds from Te Hauturu-o-Toi in early 2010. The upward trend continued until the last immigrant had offspring in 2012, levelled off and dropped

again in 2014. Microsatellite A_R was much more varied from year to year, but is not expected to reflect genome-wide nucleotide heterozygosity well given the small number of markers and biases in using microsatellite loci to measure nucleotide heterozygosity (Väli et al., 2008). The impact of the marked decrease on inbreeding measures in microsatellite data and spike in allelic richness in SNP data both decline over time, which may suggest that the introduction of individuals to Tiritiri Mātangi had only a temporary effect on the accumulation of inbreeding in the population.

Just one migrant recruiting into a small population can increase genetic diversity (Gustafson et al., 2017), and between one and ten migrants per generation is suggested to be enough to sustain the level of genetic diversity in a small population (Mills and Allendorf, 1996; Steinbach et al., 2018). Genetic top-ups can have long-term fitness effects in the population (Bijlsma et al., 2010; Frankham, 2015; Frankham et al., 2010) even when the top-up is from inbred lines (Bijlsma et al., 2010). Loss of specific alleles was shown in the translocation of birds from Te Hauturu-o-Toi to Tiritiri Mātangi (Brekke et al., 2011). Therefore, further genetic top-ups may still introduce new alleles and serve to mitigate the effects of low diversity and slowly increasing inbreeding measures in the Tiritiri Mātangi population.

Effects of removing individuals

My analysis confirms that removing individuals did not affect inbreeding measures of the population. This may be due to the fact that many juveniles naturally do not recruit to the next breeding season and the loss of juveniles to translocation had no more impact than general recruitment losses. Removing individuals is also unlikely to have affected changes in population subdivision and loss of individuals with favourable phenotypes from the population due to their selection for translocation (Allendorf et al., 2008). Subdivision is not likely because Tiritiri Mātangi is 220 ha and juveniles disperse randomly across the island in their first year (Rutschmann et al., in prep), with females traveling slightly further than males (Richardson et al., 2010). Further, phenotype changes in the population as a result of selection of individuals for translocation is unlikely as there is very little heritable genetic diversity to select upon (de Villemereuil et al., 2019a).

Runs of homozygosity have mainly been used in human population studies and animal breeding (Curik et al., 2014; Rebelato and Caetano, 2018). Their formation and location across the genome is affected by recombination and recent selection (Pemberton et al.,

2012). For example captively-bred animals have longer ROHs than are found in human populations (Curik et al., 2014). Here I used a much smaller number of markers than in some human studies (McQuillan et al., 2008), but similar numbers to many cattle studies and a recent study of ROH in European ibex (*Capra ibex*; Grossen et al., 2018). Cattle have much larger genomes than birds (3 Gb vs 1.2 Gb) so the genotyped hihi SNPs are denser across the genome. Genome sequences that are identical, or 'identical by state' (IBS), may also share ancestry, i.e. be identical by descent (IBD). A long string of homozygous markers strongly indicates IBD rather than IBS (Broman and Weber, 1999a; Knief et al., 2015) and the chances of a contiguous string of homozygous markers being homozygous by chance becomes less the longer the string gets. I assume that ROHs of 20 SNPs or more are indicative of IBD in this population, but it isn't clear how well the markers reflect the homozygosity of sequence between them (Howrigan et al., 2011). This problem is also confounded by the lack of phase information in the genotype (Browning and Browning, 2012). This data is likely further confounded with false-positives due to low minor allele frequencies (MAF) of many SNPs. A MAF of 10% would be expected to arise in a non-IBD segment with a frequency of 1.6%, which means longer IBD segments are required for positive identification of IBD (Browning and Browning, 2012). The correlation between markers and whole genome sequence homozygosity cannot be addressed until a high-quality hihi genome has been assembled and more whole genome sequencing data is made available.

The mean of ROH length in the hihi population is small and does not change significantly from year to year. Short ROHs are indicative of an outbred population (Pemberton et al., 2012), and may reflect the effects of the polygynandrous mating system of hihi. There are some spikes in length of some ROH in 2007, 2009 and 2012. Long ROHs are caused by the mating of close relatives (Bosse et al., 2012a) and it is possible that these much longer ROHs, indicate matings of closely related birds.

Genes identified in ROHs were highly varied in function and this broad array implies a lack of selection on particular traits. By comparison, genes associated with milk production were found in ROHs in Gyr (*Bos indicus*) dairy cattle (Peripolli et al., 2018); genes associated with body size, muscle formation and skin colour in six commercial sheep breed ROHs (Purfield et al., 2017); and genes for defence mechanisms and adaptations to novel environments in ROHs of commercial and wild porcine genomes (Bosse et al., 2012a).

Conclusion

The general trend of minimal change in genetic diversity and inbreeding is evident from both data types. Microsatellites were more sensitive to changes in inbreeding in this population caused by a genetic top-up and they also show greater overall variation. SNP data can also illuminate regions of the genome that are more diverse or inbred. As has been suggested elsewhere (Funk et al., 2012; Narum et al., 2008), both marker types should be combined to make optimal management decisions. However, SNP data gives much more precise measurements in general, which is especially useful in a conservation context where genetic diversity is minimal.

It will be of interest in future studies to assess local LD at ROH sites that have increased frequency over time in the population and to measure the overall variation in ROH across the population to get a better estimate of the power of F_{ROH} measures. The survival rates of individuals with much longer ROHs than the general population and the fitness impact of genes in regions that are becoming more inbred would also be interesting avenues of inquiry.

The fact that removing individuals does not seem to impact the population inbreeding measures indicates it will be safe to continue to use this as a source population to reintroduce hihi to new predator-free locations around New Zealand's North Island. Due to the slow accumulation of inbreeding shown in SNP F_{IS} data and also the microsatellite data F_{IS} changes after genetic top-up, which lasted over three years, it also seems advisable to continue to carry out genetic top-ups over each generation (every 4-5 years) to mitigate the losses of allelic richness and reduce inbreeding measures.

Chapter 4: Effects of translocation bottlenecks on population genomics of an endangered species

Abstract

Translocations are a widely used tool for conservation of endangered species and understanding their genetic effects is important to maximise genetic diversity in conservation programs. Genetic studies of translocations are most commonly carried out with small numbers of genetic markers such as microsatellites, at a few loci in the genome. Here we can test the effects of translocations of an endangered New Zealand passerine hihi (*Notiomystis cincta*, 'stitchbird') across the whole genome using a 50K single nucleotide polymorphism (SNP) array. I look at genetic diversity measures A_R, H_O and H_E and inbreeding coefficients F_{IS} and F_{ROH} as well as the fixation index F_{ST} and compare them with earlier studies with microsatellites. I characterise runs of homozygosity, which mark regions of higher inbreeding within the hihi genome and investigate how they change across translocation bottlenecks. I found that SNP genetic diversity measures were all reduced over translocation bottlenecks, which was only evident in A_R but not H_O and H_E measures in microsatellite studies. Inbreeding coefficient (F_{IS}) for each population, and the proportion of genetic variance explained by population subdivision (F_{ST}) between populations showed similar trends for both data types. Changes in the minor allele frequency distribution in SNP data illustrated a loss of diversity over each translocation event. Runs of homozygosity in hihi were short (median = 528 kb) and increased in frequency after every translocation bottleneck. They were more prevalent in macro chromosomes which are likely to have higher linkage disequilibrium. Overall, the SNP data showed a decline in genetic diversity over translocation bottlenecks and in the small managed populations the extent of which was not previously seen in microsatellite data.

Introduction

Translocations to establish new populations have been used extensively in species recovery across many taxa (Swan et al., 2018; Weeks et al., 2011) such as coho salmon (*Oncorhynchus kisutch)* (Campbell et al., 2017), speckled dace (*Rhinichthys osculus)* (Kinziger et al., 2011), goat (*Rupicapra rupicapra*) (Crestanello et al., 2009), bighorn sheep (*Ovis canadensis)* (Jahner et al., 2019), bats (*Chalinolobus tuberculatus* and *Mystacina tuberculata*) (Sherley et al., 2010), Merriam's turkey (*Meleagris gallopavo merriami*) (Mock et al., 2004), Canada goose (*Branta canadensis occidentalis*) (Talbot et al., 2003), natterjack toad (*Bufo calamita*) (Rowe et al., 1998) and Great Capricorn beetle (*Cerambyx cerdo*) (Drag and Cizek, 2015). They are a common practice in the management of endangered populations in New Zealand (Jamieson, 2009; Miskelly and Powlesland, 2013) e.g. New Zealand saddleback (*Philesturnus carunculatus rufusater*) (Lambert et al., 2005), South Island robins (*Petroica australis australis*) (Boessenkool et al., 2007), little spotted kiwi (*Apteryx owenii*) (Ramstad et al., 2013) and takahe (*Porphyrio hochstetteri*) (Grueber and Jamieson, 2011).

A large group of individuals would, ideally, be used to capture a large proportion of the genetic diversity of the source population, convey it to a new location, and the group supported to increase quickly to an effective population size of >1,000 individuals (Weeks et al., 2011). However, a large group is rarely available for translocation and small numbers are often used. The reduced size of the new population compared with the source population creates a genetic 'bottleneck' effect in which some genetic variability is lost (Allendorf, 1986; Wright, 1931). The strength of the bottleneck is determined by the number and genetic diversity of individuals successfully translocated and recruited to breed (founder effects) (Jamieson, 2011; Kinziger et al., 2011) and how fast the population grows after the bottleneck (Allendorf, 1986; Nei et al., 1975). Smaller populations are more susceptible to loss of alleles through genetic drift (Allendorf et al., 2013). Prolonged bottlenecks of small population size reduce heterozygosity (Allendorf, 1986) and increase inbreeding measures as mating opportunities between unrelated individuals become less frequent (Keller, 2002). Reduced heterozygosity can lead to inbreeding depression and a reduction in fitness (Falconer and Mackay, 2009).

Frequently, microsatellite markers are used to assess the genetics of endangered populations (Ouborg et al., 2010). Microsatellites are multiallelic markers from non-coding regions of the genome. Their multiallelic nature gives them particular sensitivity

to changes in allelic richness (A_R) (Tautz, 1989), and they can also be used in many kinds of conservation genetic studies including assessing heterozygosity, genetic differentiation between subpopulations, inbreeding and selection (Hodel et al., 2016; Jehle and Arntzen, 2002; Moura et al., 2017). One drawback of microsatellites is that typically only a small panel of markers is used, representing a small portion of the genome. They are also not randomly distributed across the genome (Lawson and Zhang, 2006). Furthermore, low numbers of microsatellite marker numbers do not fully reflect homozygosity levels in endangered species (Wan et al., 2004).

SNP data can be used to calculate most of the same measures as microsatellites, such as genetic diversity, population differentiation and inbreeding for conservation purposes (Allendorf, 2017; Ouborg et al., 2010; Shafer et al., 2015). They can also identify regions of the genome that are identical by descent. Runs of homozygosity (ROH), are contiguous stretches of the genome that are dominated by homozygous markers that were first described in Broman and Weber, (1999). These regions often mark shared ancestry of genetic haplotypes, with longer ROHs from recent generations and shorter ones from distant ancestors (Keller et al., 2011; McQuillan et al., 2008). ROHs have been used to elucidate population demography (Bosse et al., 2012b; Gibson et al., 2006; MacLeod et al., 2013; Pemberton et al., 2012), assess inbreeding across the genome (Bjelland et al., 2013; Curik et al., 2014; Marras et al., 2015; Peripolli et al., 2018), identify regions that might be under selection (Metzger et al., 2015; Purfield et al., 2017) and identify regions of the genome that have shared ancestry (Ferenčaković et al., 2013; Howrigan et al., 2011; Rebelato and Caetano, 2018).

Hihi, like many New Zealand birds, underwent an abrupt range contraction during the 1800s due to habitat loss and lack of defences against mammalian predation (Duncan and Blackburn, 2004). They were confined to a remnant population of approximately 2,500 - 3,400 birds on Te Hauturu-o-Toi (Toy et al., 2018). Since the 1980s, hihi have been reintroduced to predator-free sites within their former range across New Zealand's North Island from the remnant wild population and also from previously established reintroduced hihi sites (Brekke et al., 2011). The translocation bottleneck effects in this study results from a combination of founder effects and population size, with reintroduced populations reaching carrying capacity at <200 adult birds (Brekke et al., 2011). An adaptive management program has improved the birds' survival and speed of increasing population size post translocation by studying the birds' needs and adjusting

the management of the birds (Armstrong et al., 2007) including their supplementary feeding regime (Castro et al., 2003; Chauvenet et al., 2012; Walker et al., 2013), care against pathogens such as *Aspergillus* (Alley et al., 1999; Cork et al., 1999), *Salmonella* (Ewen et al., 2007) and mites (Low et al., 2007a), assessment of habitat requirements and environment (Armstrong et al., 2001; Ewen et al., 2011; Richardson and Ewen, 2016), and understanding the pre and post-translocation behaviour of the birds (Franks and Thorogood, 2018; Panfylova et al., 2016; Richardson et al., 2010). The demographic history of the reintroduced hihi populations make them an ideal model to test the impact of genetic bottlenecks on the genetic viability of reintroduced or bottlenecked populations. Bottlenecks can also occur naturally due to disease or environmental changes, and this study may help us understand how these processes might translate to other non-managed populations.

An earlier study of the genetics of hihi populations by Brekke et al. (2011) used microsatellites to show that the remnant wild hihi population has a higher genetic diversity than other threatened New Zealand and global avifauna and that it showed no genetic signs of recent bottlenecks. The study also showed that the ratio between effective population size and population count (N_e/N_c) was very high in Te Hauturu-o-Toi but decreased in reintroduced populations. There was significant differentiation between populations (F_{ST}), but genetic diversity including total number of alleles (N_A), alleles per locus (H_O and H_E), allelic richness (A_R), F_{IS} and average genetic diversity per subpopulation (H_S) were not significantly reduced in most reintroduced populations except Kāpiti Island, where the population had taken a long time to increase in size after the translocation (Brekke et al., 2011).

I will repeat some of the previous microsatellite study (Brekke et al., 2011) using SNP data from a 50K SNP array, to see if genomic data can shed more light on genetic changes to an endangered bird population following translocation bottlenecks. I test genetic diversity measures such as allelic richness (A_R) and heterozygosity measures H_O and H_E, and check inbreeding coefficient F_{IS} and genetic differentiation (F_{ST}) between populations. To better understand translocation effects on inbreeding across the genome I assessed ROH length and distribution before and after translocations. I expect that the number of ROHs will increase across bottlenecks as has been seen in other species (e.g. pigs, Bosse et al., 2012) and some regions of the genome will contain more ROHs than others due to differences in the recombination landscape of the genome, different

selection pressures and stochastic sampling of regions transmitted to the next generation. I also tested changes in the percentage of ROH at each SNP location and identified regions of the genome that had increasing numbers of ROH over translocations. These regions mark areas of increased inbreeding the population. Finally, I test for changes in the sum of the total length of ROH (sROH) to infer demographic characteristics of the populations.

Methods

Study area and species

Hihi (*Notiomystis cincta*, or 'stitchbird') is an endemic and endangered New Zealand passerine. These birds nest in tree cavities and as a result were among those devastated by mammalian predators over the 18th and 19th centuries (Duncan and Blackburn, 2004). Hihi live for an average of 3-4 years, but some birds live as long as 9 years (Low and Pärt, 2009). Their optimal reproductive years are between 2-4 (Brekke et al., 2013; Low et al., 2007b). The breeding season runs from September to February and females usually have two clutches per year (Doerr et al., 2017). Hihi have one of the highest records of extra pair paternity (EPP) (Brekke et al., 2013) and females use post-copulatory inbreeding avoidance mechanisms (Brekke et al., 2012).

The remnant population of hihi is situated on an off-shore island Te Hauturu-o-Toi (36°12′ S, 175°05′ E) in the Hauraki Gulf of New Zealand. The population size on this 2800 ha site has been estimated to be between 2,500-3,400 birds (Toy et al., 2018). Tiritiri Mātangi Island sanctuary (36°36'S, 174°53'E) is 220 ha wildlife sanctuary twenty-three km northeast of Auckland city in New Zealand. Thirty eight hihi from Te Hauturu-o-Toi were initially translocated to Tiritiri Mātangi in 1995 and 13 more in 1996; of these four females and twelve males survived to the 1996/97 breeding season (Armstrong et al., 2002) and the population grew rapidly to capacity of approximately 160 adult birds (Brekke et al., 2011). Zealandia Sanctuary (41.2902° S, 174.7535° E) is a 225 ha forest wildlife sanctuary in Karori, Wellington city (hereafter Karori), surrounded by a mammalian predator-proof fence. The hihi population was established there in 2005 with birds 60 birds translocated from Tiritiri Mātangi and 32 from a captive population at Mt Bruce. The number of birds that reproduced following translocation is not known. The population now stands at approximately 112 adults (Rutschmann et al., in prep).

Samples

Blood or feather samples are available from individuals from the three populations. For Te Hauturu-o-Toi, 55 hihi blood samples were collected in sampling expeditions to the island in between 1995 and 2010 from adult birds of unknown age. Individuals in Tiritiri Mātangi and Karori are routinely monitored and all individuals are banded at fledging and a blood or feather sample taken. For Tiritiri Mātangi, blood is sampled from 21-day-old chicks via brachial venipuncture (approximately 70 µl) and stored in 95% ethanol as described previously (Brekke et al., 2010), while in Karori feathers are sampled at fledging and stored in 95% ethanol. To ensure equal sample sizes, a subset of 55 samples from the 2004 cohort in Tiritiri Mātangi were used in this study to evaluate birds fledged nine years after the initial population was established with birds from Te Hauturu-o-Toi. In addition, 55 samples from the 2014 cohort in Karori were used in this study to evaluate birds fledged nine years after this population was established with birds from Tiritiri Mātangi.

SNP data

A 50K SNP array manufactured by Affymetrix (described in chapter 2) provided the SNP data used here. Briefly, Qiagen DNeasy Blood and Tissue kits were used to extract DNA from the blood and feather samples and DNA was quantified on a NanoDrop 8000. Genotyped samples were selected from all hihi samples available based on their DNA quality (260/280 ratio of ~1.8-1.9 where possible), concentration (≥30 ng/µl where possible) and ensuring representation across cohorts. Samples were excluded from the analysis if they had less than 97% call rate on the first round of Axiom clustering. 42,212 polymorphic SNP markers from the array were retained for analysis. Reproducibility of calls averaged 99.98%. Hihi SNP genome positions were approximated from homology with the zebra finch genome (Ensembl 86) and their density was designed to be high (average 27.3 SNPs/kb) for micro chromosomes 10-15,17-28, LG1, LG5, LGE22 and the Z chromosome and medium (average 23.7 SNPs/kb) for macro chromosomes 1-9 as well as chromosomes 1A, 1B and 4A.

Population structure

The genetic structure among populations was first visualised by principle component analysis (PCA) using LEA version 2.0.0 software (Frichot and François, 2015) in the R statistical software version 3.6.1 (R Core Team, 2017). However, PCA calculates both the variation within and between groups. The within group variation should ideally be left out

of the assessment of between group variation; Jombart et al. (2010) created Discriminant analysis of Principal Components (DAPC) to address this issue. I used DAPC as outlined in Paradis et al. (Paradis et al., 2017) and implemented using the adegenet version 2.1.1 (Jombart and Ahmed, 2011) and HWxtest 1.1.7 (Engels, 2014) R packages. DAPC first reduces the data by identifying principal components as in a PCA analysis, but only retaining some of them (<100). A discriminant analysis is then performed on this reduced data set. Here we used the first 100 principal components to construct the discriminant analysis as the cumulative variance of additional PCs did not plateau earlier (Supplementary Figure 3). Sequential k-means clustering with model selection based on the Bayesian information criterion (BIC) was used to infer prior clusters and confirm the given population groups (Paradis et al., 2017). The results were plotted on 2-dimensional axes.

A STRUCTURE-like analysis was carried out using the LEA (Frichot and François, 2015) R package (R Core Team, 2017) to visualise the admixture components of each population. Sparse non-negative matrix factorisation (snmf) was used to estimate the individual admixture coefficients from the genotype matrix and provides least-squares estimates of ancestry proportions for each number K of ancestral populations. It estimates the entropy criterion which can then be used to help choose the number of ancestral populations that likely fits the data (the K value at the first upward bend of the graph in Supplementary Figure 1). The genotype matrix was then sorted by population and optimum number of admixture components and plotted on a bar plot using R (R Core Team, 2017).

Traditional measures such as F_{ST} provide good estimates of the proportion of genetic variance explained by population sub-division. Genetic drift plays a much bigger role in recently separated populations than mutation (Slatkin, 1995). F_{ST} measures for biallelic markers range from zero to one and unlike multiallelic markers, they behave as was described by Wright's classic F_{ST} (Wright, 1965, 1943) with no standardization necessary (Meirmans and Hedrick, 2011). F_{ST} was calculated from the SNP data using the Genepop R package (Raymond and Rousset, 1995; Rousset, 2008).

MAF for each SNP was calculated using PLINK v 1.07 (stable version; Purcell et al., 2007) --freq algorithm allowing for nonfounders, a chromosome set of 80 and samples with no-sex information. To prepare the SNP genotype data for use in PLINK, numbers

were used in lieu of chromosome names and hyphens were removed from the SNP names. MAF results were graphed using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016).

Genetic diversity measures

Genetic diversity measures of allelic richness (A_R), observed and expected heterozygosity (H_O and H_E) were calculated using diveRsity (Keenan et al., 2013; Rousset, 2008).

Inbreeding coefficient

The inbreeding coefficient F_{IS} measures the inbreeding of an individual compared with H_E under Hardy-Weinberg equilibrium (Wright, 1951). It is the proportion of the variance in a subpopulation that is present in an individual, with higher values indicating more inbred individuals (Falconer and Mackay, 2009). To assess any changes in inbreeding levels between the populations, Weir and Cockerham's (1984) estimate of F_{IS} was calculated using the R package diverRsity version 1.9.90 (Keenan et al., 2013; Rousset, 2008).

Runs of homozygosity (ROH)

ROH are homozygous regions of the genome that can be used to infer inbreeding levels across the genome (Keller et al., 2011). To identify ROHs in hihi populations, PLINK ped files were split into macro chromosomes and chromosomes 1A, 1B and 4A that were represented at medium SNP density on the array (27.3 SNPs/kb) and micro chromosomes and the Z chromosome that were represented at high density (23.7 SNPs/kb), as PLINK v 1.07 (stable version; Purcell et al., 2007) can only handle 25 chromosomes for the --homozyg function. SNPs that mapped to random and Un (unmapped genomic region) zebra finch chromosomes were removed, resulting in information for 39,445 SNP markers.

Runs of homozygosity (ROH) were detected in the SNP data using PLINK v 1.07 --homozyg function (Purcell et al., 2007) on data from both sets of chromosomes. Autozygous sequences are DNA segments that are identical by descent (IBD) from a common ancestor and not just identical by state (IBS). PLINK software was shown to be the most effective way to detect autozygosity (Howrigan et al., 2011). A sliding window of 20 SNPs was used instead of window length to account for the variation in SNP density across the genome. To identify an ROH, the minimum number of SNPs in the ROH was set at 20, minimum length of the ROH at 10 kb and no heterozygous SNPs were allowed. The default window threshold of 0.05 homozygosity was used, which

ensured that five percent of the sliding windows that passed over a SNP had to have an ROH for the SNP to be considered part of an ROH. Groups of ROHs common across individuals were identified using the --homozyg-group parameter. Full chromosome names were returned to the output ped, map and hom files using awk version 4.2.1 (Aho et al., 1987) scripts (available upon request).

The length of ROHs in each population, the number of SNPs per ROH in each population, the number of ROHs per individual and the sum of the total length of ROH (sROH) for each population was graphed from the PLINK hom file output using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). Linear models (ROH length ~ population; ROHs per individual ~ population) and a generalised linear model (number of SNPs per ROH ~ population) were used to evaluate their trends over sequential bottlenecks using R version 3.4.3 (R Core Team, 2017). Models were tested as follows: test that the mean of residuals is zero, or close to zero; test that the x variables and residuals are uncorrelated using Pearson's product-moment correlation (scores between -1 and 1, a score of 0 means there is no correlation) and test for homoscedasticity, or equal variance of residuals (check visually by plotting the lm model) with R version 3.4.3 (R Core Team, 2017). Test for autocorrelation, the correlation of a time series with lags of itself, using the Durbin-Watson test (scores between 0 and 4, a score of 2 means there is no autocorrelation) using R package lmtest (Zeileis and Hothorn, 2002). T Assume no perfect correlation between variables (check this visually) using R package corrplot (Wei and Simko, 2017). See Supplementary Table 4 and Supplementary Figures 5 and 6.

In addition, the percentage of SNPs in ROHs per population was graphed from the PLINK hom file output. To do so, the number of SNPs in ROHs was summed per individual. The percentage of SNPs in ROHs was calculated by dividing the number of SNPs in ROHs by the total number of SNPs (39,445) to identify more and less inbred individuals using R version 3.4.3 (R Core Team, 2017). Data was summarised per population in a violin plot using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016). A linear model (percentage SNPs in ROH ~ population) was used to identify any trends across population in R version 3.4.3 (R Core Team, 2017). While the percentage data is bounded, there were not many zero counts overall, so the linear model was used. Models were tested as above, see Supplementary Table 4 and Supplementary Figures 7 and 8.

SNP counts in ROHs were calculated using an emulator of the SNPSinRUNs function from the detectRUNS R package version 0.9.5 (Biscarini et al., 2018) created in Perl v.5.26.2. (Christiansen et al., 2012). This was due to the limitation of chromosome visualisation within the detectRUNS package for a genome with many chromosomes. The custom script Count_SNPs_in_ROH_with_groups.pl (available upon request) was used to count the percentage of SNPs that were in runs of homozygosity for each SNP within each population on each chromosome. The percentage SNPs in runs of homozygosity was calculated at each SNP position per population and was summarised on a violin plot using R version 3.4.3 (R Core Team, 2017). A linear model comparing percentage SNPs in ROHs ~ number of bottlenecks was calculated using R 3.4.3 (R Core Team, 2017). Models were tested as above, see Supplementary Table 4 and Supplementary Figures 7 and 8.

Differences between SNPs percentages in ROH across populations from the Count_SNPs_in_ROH_with_groups.pl output was visualised using R package ggplot2 version 3.1.0 (R Core Team, 2017; Wickham, 2016), highlighting SNPs that increased or decreased by at least 5% in each of the two translocations.

To easily visualise how ROHs may have proliferated or declined in the population, the accumulated change in percentage of SNPs in ROHs across populations (i.e. after each translocation bottleneck) for each SNP was calculated and illustrated with R 3.4.3 (R Core Team, 2017). SNPs were identified that had a consistent increase of at least 3% in percentage in ROHs for both translocations. That is, if they increased by 3% consistently across both translocations, in Karori they were at least 6% higher in percentage SNPs in ROHs than in Te Hauturu-o-Toi. Consistent decreases of percentage in ROH were also detected across translocations.

ROHs can also be used to estimate individual inbreeding levels using a global F_{ROH} measure (Keller et al., 2011; McQuillan et al., 2008). F_{ROH} is calculated as the sum of ROH length divided by the length of the genome (Biscarini et al., 2018). The genome was estimated from the PLINK map file and calculated as the sum of the maximum SNP position for each chromosome. F_{ROH} was calculated for each of the populations for the SNP data set using the F_{ROH} inbreeding function in detectRUNS R package version 0.9.5 (Biscarini et al., 2018; R Core Team, 2017). The mean and median F_{ROH} and a linear

model of genome-wide F_{ROH} ~ number of bottlenecks were calculated using R 3.4.3 (R Core Team, 2017). Models were tested as above, see Supplementary Table 4 and Supplementary Figures 7 and 8.

Gene annotation

Genes were identified in regions where SNPs presence in ROHs was continuously increasing or decreasing by at least three percent across translocations. Regions were defined as half the average ROH length on either side of the first and last SNP in the region. Gene annotations were obtained for these regions from the zebra finch (*Taeniopygia guttata*) Ensembl 86 genome gtf3 file (Zerbino et al., 2018) using a custom perl script (grab_gene_annotations_for_regions_from_gtf3.pl, available on request). Gene ontology (GO) term values were obtained using the Ensembl stable gene IDs from BioMart Ensembl Genes 95 (Zerbino et al., 2018).

Results

Population structure

Using the 55 Te Hauturu-o-Toi samples, 55 samples from Tiritiri Mātangi (2004) and 55 samples from Karori (2014), an initial PCA analysis of the SNP data in LEA (Frichot and François, 2015) showed that three populations were clearly separated across the first two principal components and across the first and third components. A comparison of the second and third principal components showed an overlap between Karori and Te Hauturu-o-Toi datasets, where the Te Hauturu-o-Toi samples were clustered tightly in the middle of a long spread of Karori samples (see Supplementary Figure 2). As PCA does not perform well when analysing datasets larger than 10,000 data points (Jombart et al., 2010), the data was also analysed by DAPC. The increase in variance explained by PCA is shown in Supplementary Figure 3a and three was identified as the optimum number of clusters, which corresponded with the number of populations (Supplementary Figure 3b). In agreement with the first two principal components of the PCA, each of the populations was grouped very distinctly from the others in the DAPC analysis (Figure 1).

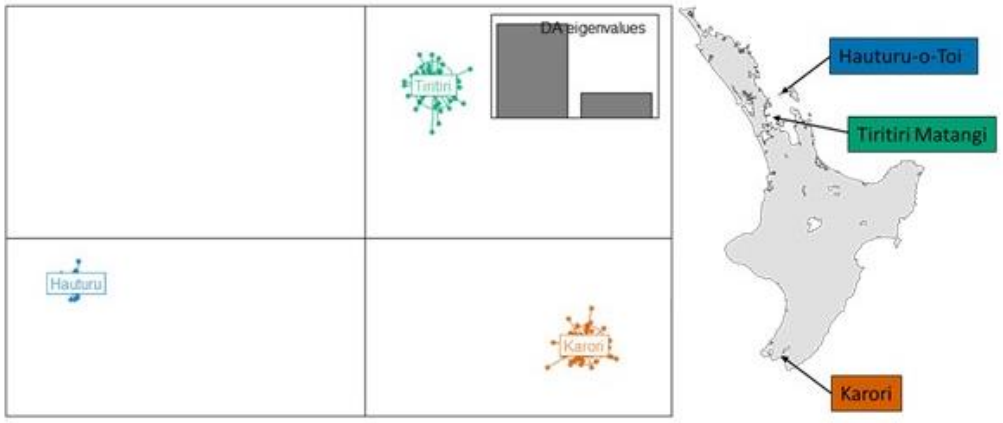


Figure 1: DAPC analysis of hihi population differences. The samples from Te Hauturu-o-Toi are shown in blue, samples from Tiritiri Mātangi are shown in green and those from Karori are shown in orange. In the left panel, the upper right corner displays the magnitude of the first and second eigenvalues. The right panel shows the location of the three populations on the North Island of New Zealand.

A STRUCTURE-like analysis illustrates the differences between the populations (Figure 2). Individuals in Te Hauturu-o-Toi, Tiritiri Mātangi and Karori are each predominantly characterised by a distinct single ancestry, reflected in the admixture coefficients. Although all three ancestries are present in all populations, these represent a minor component of individuals compared to the dominant ancestry for each population.

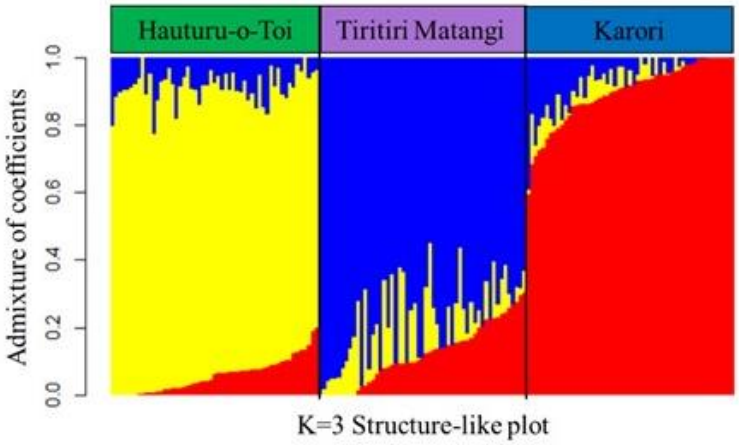


Figure 2: SNP Structure-like plot of Te Hauturu-o-Toi, Tiritiri Mātangi and Karori. The coloured bar above the plot indicates the corresponding population. Individuals within each population are sorted by their major admixture coefficient, with each admixture cluster represented by a colour.

Pairwise F_{ST} was calculated between the three populations, shown in Table 1. F_{ST} between hihi populations shows little to moderate genetic differentiation (Hartl and Clark, 1989). F_{ST} values show similar distances between population pairs that are one translocation bottleneck apart (Te Hauturu-o-Toi vs Tiritiri Mātangi, and Tiritiri Mātangi vs Karori), and a greater distance between those that are two translocation bottlenecks apart (Te Hauturu-o-Toi vs Karori), but these values are not statistically significant.

Table 1: F_{ST} values for each population pair. F_{ST} is a measure of the proportion of genetic variance explained by population sub-division. Lower and upper values reflect the 95% confidence interval on the estimate of F_{ST}.

Population pairs	F_{ST}	Range
Te Hauturu-o-Toi vs Tiritiri Matangi	0.05	0.046, 0.056
Te Hauturu-o-Toi vs Karori	0.08	0.077, 0.083
Tiritiri Matangi vs Karori	0.05	0.046, 0.056

Genetic diversity

Allelic richness (A_R) showed a less pronounced decline over the first bottleneck than the second, with a small decrease in the bottleneck between Te Hauturu-o-Toi and Tiritiri Mātangi and a large decrease during the second bottleneck between Tiritiri Mātangi and Karori, but these differences are small (Supplementary Figure 1). Both expected and observed heterozygosity had a similar pattern of greater decline over the second bottleneck, but with a more pronounced decline over the first bottleneck than was seen in the A_R data, however only microsatellite expected heterozygosity and SNP observed heterozygosity are significant (Supplementary Figure 1). In contrast, the inbreeding measure F_{IS} decreased much more in the initial bottleneck than the second bottleneck, but again was only a small change (Figure 3, Supplementary Table 1).

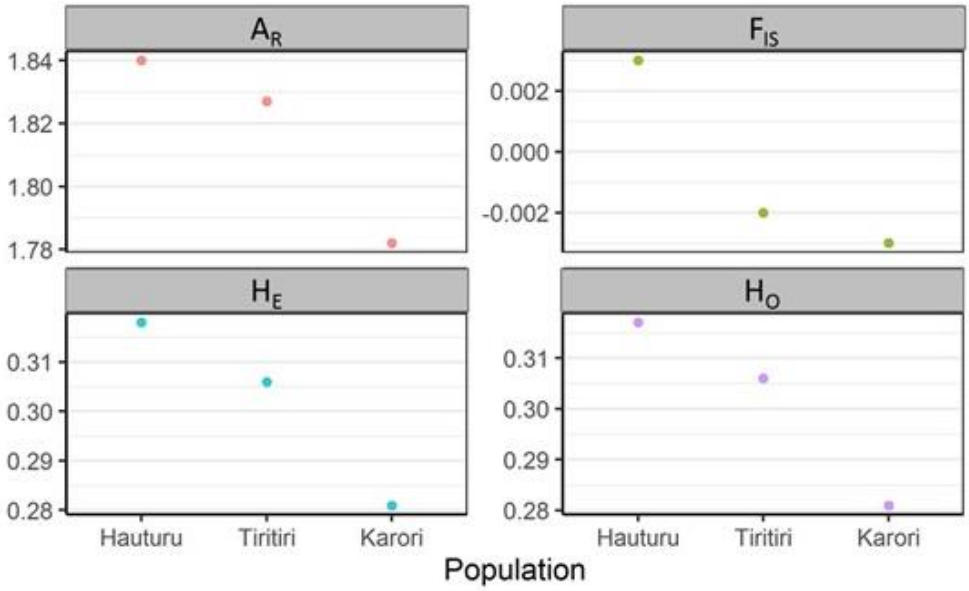


Figure 3: Genetic diversity and inbreeding measures for three hihi populations. A_R: allelic richness, F_{IS}: inbreeding measure, H_E: expected heterozygosity, H_O: observed heterozygosity.

The frequency of alleles with very low minor allele frequency increased after each translocation. The number of SNPs with rare alleles (MAF less than 0.01) increased from 188 in Hauturu-o-Toi, to 357 in Tiritiri Mātangi, and 349 in Karori (Figure 4). The number of SNPs with MAF of zero (not shown on the graph) also increased through the bottlenecks, with 6307 in Hauturu-o-Toi, 6787 in Tiritiri Mātangi (480 lost in the first order bottleneck) and 8558 in Karori (1771 lost in the second order bottleneck, 2251 lost in total). Overall, the remnant population of Te Hauturu-o-Toi shows the smallest number of rare alleles. Tiritiri Mātangi, which was populated with birds from Te Hauturu-o-Toi leading to a bottleneck, has more rare alleles. Karori, populated with birds from Tiritiri Mātangi, has slightly fewer rare alleles than Tiritiri Mātangi, but overall has lost more than four times the number of alleles as Tiritiri Mātangi compared with the remnant wild population.

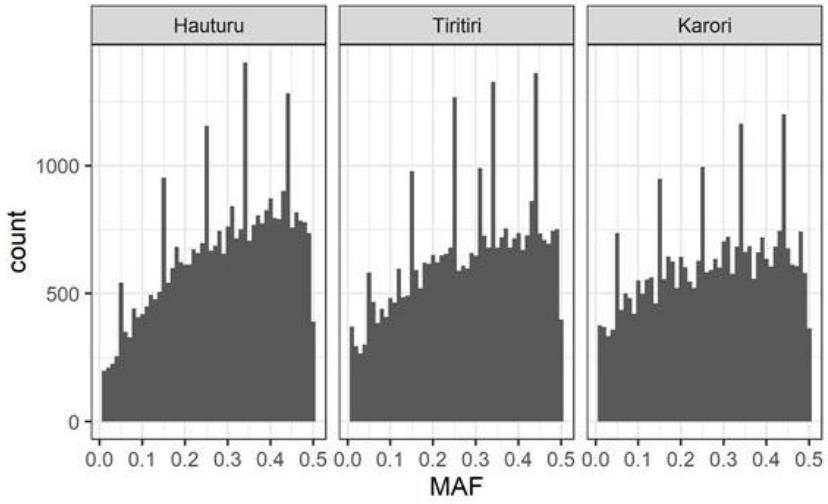


Figure 4: The distribution of SNP minor allele frequency for three hihi populations.

Inbreeding within populations

As described above, F_{IS} inbreeding measurements were 0.003 for Te Hauturu-o-Toi, -0.002 for Tiritiri Mātangi and -0.003 for Karori (Figure 3). F_{ROH} is an inbreeding coefficient calculated from runs of homozygosity (ROH). Mean F_{ROH} inbreeding measurements were 0.0035 for Te Hauturu-o-Toi, 0.0047 for Tiritiri Mātangi and 0.0091 for Karori (Figure 5a). In contrast to F_{IS}, genome-wide F_{ROH} across populations increased marginally but significantly across translocations (0.003 increase per translocation, *p-value* = 2.627e-14). F_{ROH} increases were more pronounced over the second bottleneck (Figure 5a), as was seen with the genetic diversity measures (Figure 3).

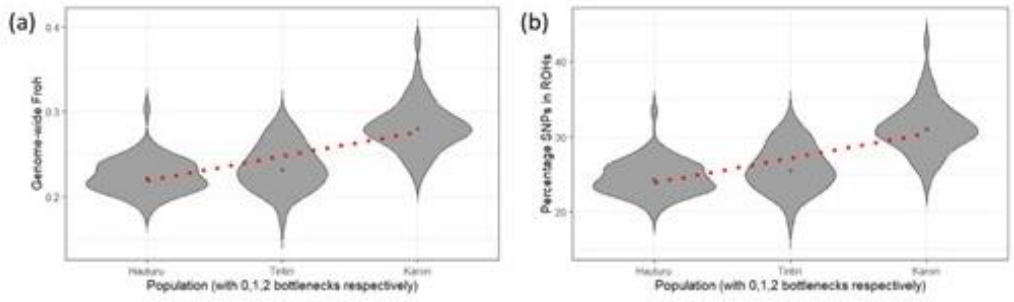


Figure 5: Genome-wide F_{ROH} and Percentage SNPs in ROHs. (a) Genome-wide F_{ROH} across hihi populations. (Genome-wide F_{ROH} ~ number of bottlenecks, y = 0.189890 + 0.0287049x, Adjusted R^2: 0.4273, p-value: 2.2e-16). **(b)** Percentage SNPs in ROHs across hihi populations. (Percentage SNPs in ROHs ~ number of bottlenecks, y = 20.6409 + 3.2665x, Adjusted R^2: 0.4317, *p-value*: 2.2e-16)

Runs of homozygosity

The length of ROHs did not change significantly between populations, the average number of SNPs per ROH between populations increased by a small, but significant amount and a generalised linear model (GLM) showed that number of ROHs per individual in the population increased significantly after each translocation (Figure 6). The mean number of ROHs per individual more than doubled between the Te Hauturu-o-Toi and Karori populations (Figure 6). A small number of outliers are seen in each metric where a few individuals have very long ROHs (either their length or the number of SNPs in the ROH), and individuals have many more ROHs than the majority of the population (one individual in Te Hauturu-o-Toi and one in Kārori). The number of individuals with higher sROH increased for each translocation, most strikingly in the second translocation from Tiritiri Mātangi to Karori (Figure 7).

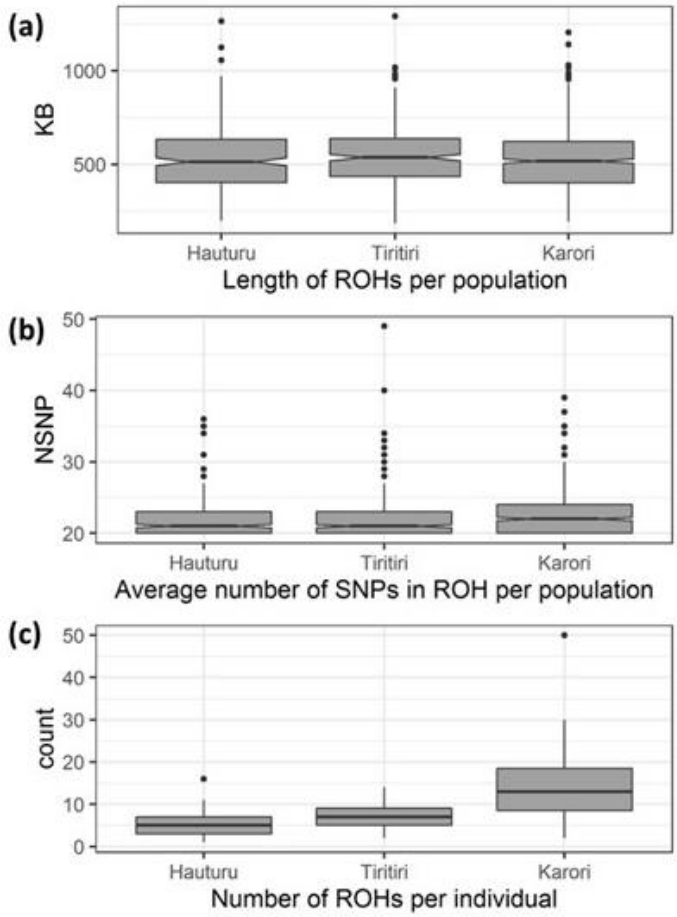


Figure 6: Characteristics of runs of homozygosity (ROH) across populations (a) Distribution of length of ROHs per individual within each population, the maximum lengths of a ROH in Te Hauturu-o-Toi, Tiritiri Mātangi and Karori were 1264, 1290 and 1204 respectively. The mean lengths in the same populations were 534 kb, 542 kb and 524 kb respectively. Linear model (length of ROHs in individuals ~ population) showed y = 547594 – 7085x, Adjusted R^2 = 0.0005317, *p-value* = 0.1807. **(b)** Distribution of number of SNPs in a ROH in each population. The maximum for each population was 36, 49 and 39 and the mean was 22.16, 22.24 and 22.61 respectively. Linear model (number of SNPs in ROH ~ cohort) shows y = 21.7105 + 0.3300x, Adjusted R^2: 0.003866, *p-value*: 0.008231. **(c)** Distribution of number of ROHs per individual in each population. The maximum number was 16, 14 and 50 and the mean was 5.56, 7.4 and 14.44 respectively. Generalised linear model (ROH per individual ~ population) shows y = 1.03790 + 0.60202 x, *z-value* = 14.622, *p-value* = <2e-16.

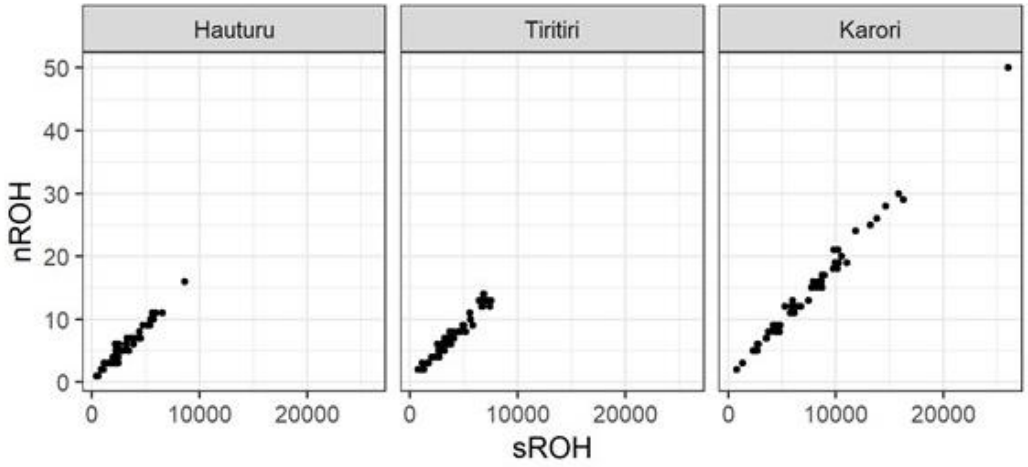


Figure 7: Sum of the total ROH (sROH) for individuals in each population, plotted against the number of ROHs for each individual. Individuals with more and longer ROHs can be found after each translocation.

Looking at the percentage of SNPs in ROHs for each SNP marker is helpful to visualize how ROHs are spread across the genome. The overall pattern of percentage SNPs in ROHs for the whole genome is similar to the pattern of genome wide F_{ROH} (Figure 5b), with an increase across the bottlenecked populations. Changes in the percentage of SNPs in ROHs are slightly greater for macro chromosomes and chromosomes 1A, 1B and 4A, than for micro chromosomes and the Z chromosome (see Supplementary Figure 4). Samples from the Tiritiri Mātangi to Karori translocation had a more pronounced mean change of 0.449 and 0.634 for macro and micro chromosomes respectively, than those of the earlier Te Hauturu-o-Toi to Tiritiri Mātangi translocation, with means of 0.135 and 0.084 respectively (Supplementary Figure 4).

The percentage of SNPs in ROHs changes over each translocation (Figure 8). Overall, there are more increases in percentage SNPs in ROHs than decreases. These are spread out across the genome. The changes after the translocation between Tiritiri Mātangi and Karori are greater than those between Te Hauturu-o-Toi and Tiritiri Mātangi. Across both translocations, fewer SNPs decreased by 5% or more on the macro chromosomes than the microchromosomes, but the opposite pattern was observed with SNP increases.

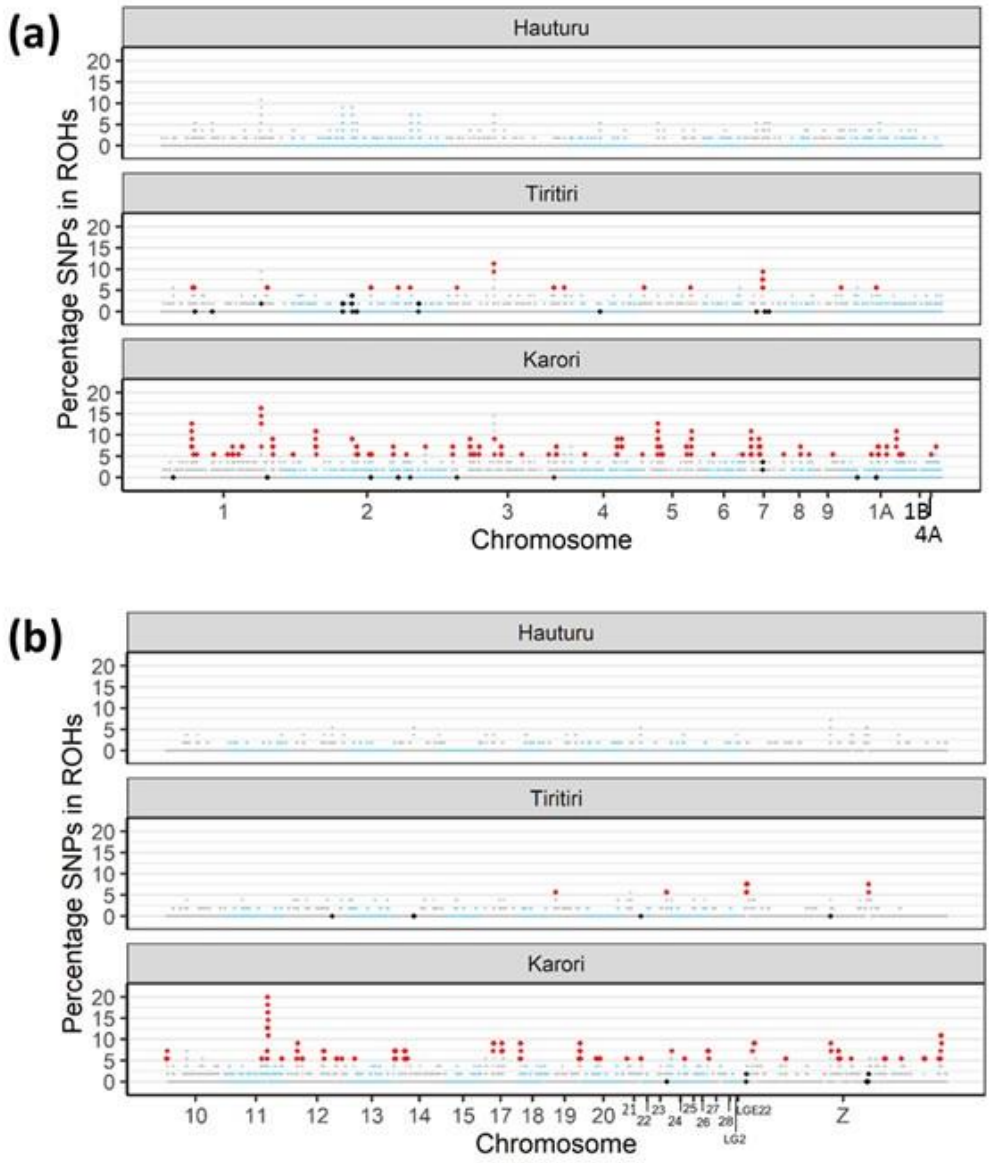


Figure 8: Percentage of SNPs in ROHs across chromosomes for each population. (a) macro chromosomes and chromosomes 1A, 1B and 4A (b) micro and Z chromosomes. SNPs are represented in grey and light blue. SNPs highlighted in red in the Tiritiri panel have increased in percentage by 5% or more when translocated between Te Hauturu-o-Toi and Tiritiri Mātangi and SNPs highlighted in red in the Karori panel have increased by 5% or more between Tiritiri Mātangi and Karori. SNPs highlighted in black in the same panels have decreased by 5% or more following translocation.

Accumulated changes were also observed to identify if there are any areas of the genome that might be particularly susceptible to translocation bottlenecks (Figure 8). Regions on chromosomes 2, 3, 4, 12 and the Z chromosome all had their percentage of SNPs in

ROHs at specific marker positions increase by at least 3% over each translocation – resulting in a 6% or greater increase of SNPs in ROHs in those locations between Te Hauturu-o-Toi and Karori. Only one region, on chromosome 2, decreased consistently across translocations. Genes identified in these regions can be found in Supplementary Tables 2 and 3. Within the region of decreasing ROHs in SNPs, ENSTGUG00000011226 had gene ontology (GO) terms associated with "roof of mouth development".

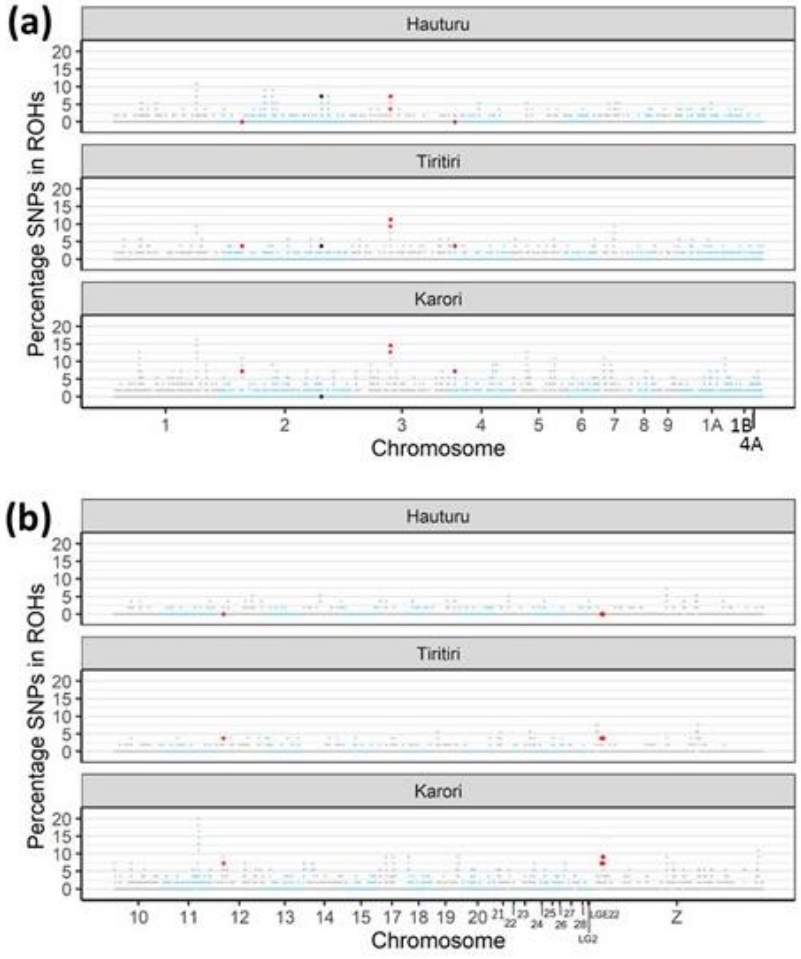


Figure 9: Accumulated change of SNPs in ROHs across multiple translocation bottlenecks. (a) SNPs highlighted in red have increased by at least 3% in the translocation between Te Hauturu-o-Toi and Tiritiri Mātangi and increased a second time by at least 3% during the translocation between Tiritiri Mātangi and Karori. The single SNP highlighted in black consistently decreased by at least 3% in each of the two translocations.

Discussion

I investigated the population genomic effects of imposing multiple translocation bottlenecks on an endangered bird population. The samples on Tiritiri Mātangi were taken 9 years after the initial bottleneck and I found that allelic richness, expected and observed heterozygosity were all reduced as was expected theoretically (Allendorf, 1986) and has been shown with microsatellites in an earlier study (Brekke et al., 2011). These metrics were further reduced over a subsequent bottleneck to Karori, as was also expected (Chakraborty and Nei, 1977; Nei et al., 1975). Inbreeding measures F_{IS} and F_{ROH} were low. In contrast to expectations, F_{IS} measures showed a slight decrease across the translocations showing the birds were becoming less related than expected under Hardy-Weinberg equilibrium. However, in accordance with expectations, both F_{ROH} and the number of ROHs increased in individuals across translocation bottlenecks, capturing a genome-wide signal of inbreeding accumulation. Measures of inbreeding at the SNP level indicated that inbreeding was affected in some regions of the genome more than others. I identified regions of the genome that were particularly susceptible to translocation bottlenecks by identifying SNPs that had increased frequency in ROHs after both translocations. Overall, I show that while total inbreeding levels are changing incrementally, some regions of the genome are becoming inbred at a much faster rate.

Population structure

PCA, DAPC and STRUCTURE-like plots show that the three populations are genetically distinct despite the fact that the F_{ST} values between them show only little to moderate genetic differentiation (Hartl and Clark, 1989). Measures of F_{ST} are higher than earlier work with microsatellites which showed F_{ST} between Te Hauturu-o-Toi and Tiritiri Mātangi to be 0.035 (Brekke et al., 2011) compared with 0.051 in this study. The F_{ST} values were larger between populations separated by two bottlenecks (Te Hauturu-o-Toi and Karori) compared with population pairs separated by only one bottleneck (Te Hauturu-o-Toi and Tiritiri Mātangi; Tiritiri Mātangi and Karori).

Genetic diversity

SNP allelic richness was depressed following each bottleneck as had been expected (Allendorf, 1986). It also corroborated the earlier study with microsatellites which showed an allele reduction of 9% following the initial translocation of hihi from Te Hauturu-o-Toi to Tiritiri Mātangi (Brekke et al., 2011). Both expected and observed heterozygosity measures were also reduced across each translocation bottleneck which

was not seen in the microsatellite study (Brekke et al., 2011). Heterozygosity is not expected to be as greatly affected as allelic richness immediately following a bottleneck, but an extended duration of a bottleneck can erode it (Allendorf, 1986). The period after translocation at a relatively small population size may have been a factor in reducing heterozygosity.

The reduction of A_R, H_E and H_O in our samples taken nine years following each bottleneck and the interaction between them is also seen in the distribution of MAF in each population. Rare alleles are more likely to be lost following a bottleneck (Luikart, 1998; Nei et al., 1975), and the reduced post-bottleneck A_R in this study supports that. It may seem surprising that there is an increase, or similar numbers of loci with very low MAF over each translocation, as the rare alleles are lost with reduced A_R. However, this reflects the decreased heterozygosity where the frequency of particular common alleles goes down and some alleles become rarer. The overall lowering of MAF across SNPs creates more alleles which might be susceptible to loss due to genetic drift over time as was shown with the very high loss of rare alleles in the second order translocation, highlighting the need for genetic management of these birds.

Bottlenecks are anticipated to reduce heterozygosity and rapidly increase genetic divergence in early generations. Over time, mutation will introduce new variation and hence increase heterozygosity, but the rate of recovery of heterozygosity is dependent on the size of the population and the rate of mutation per locus and therefore takes a very long time (Chakraborty and Nei, 1977; Nei et al., 1975). Populations of over 1,000 are recommended as a minimum threshold for maintaining enough genetic diversity to adapt to any environmental changes (Willi et al., 2006). While the effective to census population size ratios (N_e/N_c) in hihi have been seen to be high in remnant and reintroduced hihi populations (Brekke et al., 2011; Castro et al., 2004; Wang et al., 2010), the size of reintroduced populations is low e.g. the Tiritiri Mātangi carrying capacity is ~150 adults with supplementary feeding (Chauvenet et al., 2012; Graham et al., 2013). In such cases, a program of gene flow from the source population is recommended to maintain genetic diversity (Allendorf, 1983; Franklin, 1980; Mills and Allendorf, 1996).

Inbreeding measures

Inbreeding levels were expected to increase after each bottleneck (Lambert et al., 2005; Luikart, 1998), but the F_{IS} inbreeding coefficient reduced between Te Hauturu-o-Toi and

Tiritiri Mātangi and again to a lesser extent in the move from Tiritiri Mātangi to Karori as had been seen with microsatellite data (Brekke et al., 2011, see Supplementary Table 1). It could be a result of fluctuating F_{IS} as has been seen in golden mounted ground squirrels (*Spermophilus lateralis*) following a bottleneck (McEachern et al., 2011). The Te Hauturu-o-Toi population is at least an order of magnitude larger than the others studied here. There may be more structure in the Te Hauturu-o-Toi population that has not yet been investigated. For example, F_{IS} values for Te Hauturu-o-Toi ranged from -0.015 to 0.003 compared with a lower but overlapping range of -0.026 to 0.001 for Tiritiri Mātangi and of -0.026 to 0.003 for Karori. A small number of more inbred individuals may have affected the overall measure in the remnant wild population. F_{ROH} is considered a much more accurate measure of inbreeding (Howrigan et al., 2011; Keller et al., 2011), but it does not account for sample size or the distribution of allele frequencies per population. In contrast with F_{IS}, F_{ROH} measures of inbreeding showed a small but significant increase of 0.0028414 in inbreeding measures per translocation bottleneck. Despite their opposing trajectories, both F_{IS} and F_{ROH} measures of inbreeding are low in the hihi population.

Polyandry has long been known to reduce inbreeding (Stockley et al., 1993; Yasui, 1998) especially in small populations where the possibility of mating with a relative is higher (Cornell and Tregenza, 2007). Extra pair paternity reduces inbreeding in the offspring of alpine marmots (*Marmota marmota*) (Cohas et al., 2007), song sparrows *(Melospiza melodia)* (Reid et al., 2011) and noisy miner (*Manorina melanocephala*) (Barati et al., 2018). Brekke et al. has shown that hihi have one of the highest recorded levels of extra pair paternity in birds (Brekke et al., 2013), and that females engage in postcopulatory inbreeding avoidance (Brekke et al., 2012). Both of these behaviours may contribute to reducing the overall inbreeding levels for the populations.

Runs of homozygosity

Runs of homozygosity maintained similar lengths and number of SNPs in ROHs between populations, but the number of ROHs found in each individual increased greatly in bottlenecked populations, indicating that some regions of the genome are becoming more inbred across translocation bottlenecks. The length of the ROHs were short (median = 528 kb), which most likely reflect ancient haplotype patterns rather than more recent sharing of ancestry (Pemberton et al., 2012) and suggest recent genetic diversity loss due to bottleneck or founder events (Al-Mamun et al., 2015). Longer ROHs may reflect areas that have undergone recent selection (Bosse et al., 2012b; Gibson et al., 2006; Metzger et

al., 2015; Purfield et al., 2017) or recent inbreeding (Al-Mamun et al., 2015). Their absence here may reflect what appears to be an overall lack of adaptive potential in the species, i.e. that there is little adaptive genetic variation for selection to act upon (de Villemereuil et al., 2019a). Further, the small population sizes suggest that selection would likely be a weak force in contrast to genetic drift (Willi et al., 2006). Their absence also supports the inbreeding coefficient measurements discussed above which also illustrate that recent inbreeding measures are low. A small number of longer ROHs in some individuals and could be caused by mating between closely related individuals, or it could result from insufficient resolution in SNP markers where heterozygous loci are overlooked. However, the majority of ROHs over 500 kb in another cavity nesting passerine, the Baltic collared flycatcher (*Ficedula albicollis*), were between 500-1000 kb in length, and they were still well represented at lengths of 5000 kb as measured with whole genome sequencing (Kardos et al., 2017). Therefore, it is possible that longer ROHs are truly present in hihi, although the longest one in this study is only 1721 kb. The tight range of ROH lengths, number of SNPs per ROH as well as ROHs per individual, indicate that the hihi population is likely in general outbred. The trend towards more ROHs and longer sROH in translocated birds reflects the bottleneck imposed by translocation (Ceballos et al., 2018).The percentage of SNPs in ROHs follow a similar trend to the F_{ROH} inbreeding measure in this study. Visualising the percentage of SNPs in ROH is a convenient way to evaluate the location of overlapping ROHs across the genome (Biscarini et al., 2018). The location of these ROH show regions of the genome which are more inbred. There are more ROHs in macro chromosomes than in the micro chromosomes, which may be caused by lower linkage disequilibrium in the micro chromosomes breaking up ROHs more frequently. However, the fact that there are not more ROHs decreasing over translocations in the micro chromosomes compared with the macro chromosomes indicates that current changes are not being driven by recent recombination events.

There are clear increases in the number of ROHs in individuals across each of the translocations, but the increases are not equally distributed across the genome. The changes in percentage of SNPs in ROH across translocation bottlenecks indicates that the transition between Te Hauturu-o-Toi to Tiritiri Mātangi had much less impact than the one between Tiritiri Mātangi and Karori on regions that have increasing numbers of ROH present. Population bottlenecks have previously been shown to cause increases in ROHs

in pigs (*Sus scrofa*) (Bosse et al., 2012b; Groenen et al., 2012) cows (*Bos taurus*; Purfield et al., 2012) and sheep (*Ovis aries*; Purfield et al., 2017).

Five regions are increasing in their frequency of ROHs across both translocations and one region of ROHs became less frequent across translocations. These changes could be stochastic, and once a genome assembly is available (see below), simulations could test the expected distribution of ROHs in the three populations and determine whether these concordant regions of inbreeding accumulation are significant. Regions with relatively low frequencies of ROHs might also be an indication of regions that are under selection pressure, as 'cold spots' of ROHs are expected to have loci of critical function (Curik et al., 2014; Pemberton et al., 2012). All these regions have genes of a wide variety of functions (see Supplementary Tables 2 and 3 for GOSlim GOA accessions and descriptions). Among the gene annotations of the region decreasing in SNPs in ROHs across translocations, ENSTGUG00000011226 may be of interest as it is associated with "roof of mouth development". One issue among hihi in the reintroduced populations is a sublingual oral fistula (Alley and Low, 2007), it is currently unclear whether there is a genetic basis for this trait, but ENSTGUG00000011226 is a potential candidate for further investigation.

Caveats and future work

There are some caveats around the ROH measures used in this study. It is unclear how well the SNP markers represent the variation of the sequences around them (Curik et al., 2014), particularly given the moderate density of genotyped SNPs. Not all homozygous sequences are autozygous (Howrigan et al., 2011). Additionally, SNP array bias can affect ROH identification as the markers were selected for polymorphism across the genome (Lencz et al., 2007). It has been suggested from studies of ROH in sheep and cattle, that 50K chips overestimate ROHs under 4 Mb (Ferenčaković et al., 2013; Purfield et al., 2017). However, both animals have much larger genomes than hihi and so our SNP array is much denser. Higher F_{ROH} for livestock is thought to be caused by artificial selection and small effective population size (Curik et al., 2014). A study of porcine (*Sus scrofa*) genomes showed that domesticated populations had long ROHs which were absent in wild populations (Bosse et al., 2012). Fully evaluating potentially undetected heterozygosity and autozygosity in hihi, and the optimal SNP marker parameters for detecting them will require a complete genome (currently under construction) and whole genome sequencing data.

ROH length is also augmented in regions of low recombination rate due to higher power to detect them in these regions (Kardos et al., 2017). This effect can be reduced by mapping with genetic (centimorgan) rather than physical (kilobase) coordinates (Kardos et al., 2017). I have evaluated the genomic position of ROHs by comparison with the zebra finch genome physical positions. A linkage map is under construction for the hihi, but it is not currently at high enough resolution to evaluate ROH by genetic, rather than physical, length. However, we do know there are some inversions in the hihi genome compared with zebra finch (Scherer and Santure, unpublished data). The linkage map and a well annotated hihi genome will enable a more rigorous examination of inbreeding and demographic history using ROHs in hihi in future studies.

Proposed management actions

The distinct genetic diversity of each of the populations and the accumulation of ROHs after each translocation event suggests that this metapopulation of hihi may be amenable to the positive effects of artificial migration. By moving birds between populations every five years (once per generation), rarer alleles that might be lost due to genetic drift within one population, could be maintained in the metapopulation as a whole (Mills and Allendorf, 1996; Weeks et al., 2011). Further, movement of individuals is likely to introduce new combinations of alleles and could reduce the frequency of regions of inbreeding that are shared across many individuals in the populations currently. Heber et al., (2012) has shown that reciprocal translocations of highly inbred donors has increased heterozygosity in South Island robin populations. Re-establishing connectivity between populations has also been shown to rescue genetic diversity (Jangjoo et al., 2016), and populations of the Eltham copper butterfly, which like hihi have low genetic diversity and are recently separated, have also been flagged for genetic rescue via translocations (Roitman et al., 2017).

Conclusions

I showed that allelic richness, expected and observed heterozygosity were all reduced across translocation bottlenecks. The hihi population still has enough variation that genetic diversity and inbreeding are affected by translocations and there are clear differences between populations. Applying SNP array data has greatly increased the resolution at which we can investigate inbreeding measurements. The change in MAF of SNPs across the genome shows that translocations are changing the distribution of allele frequencies. I have illustrated that some regions of the genome are becoming more

homozygous (SNP markers with an increasing proportion in ROHs across the populations) through translocation bottlenecks and that at least one region is becoming less homozygous (SNP markers with a decreasing proportion in ROHs across the populations). When comparing our data to an earlier study with microsatellites (Brekke et al., 2011) we found that (i) A_R measures were much smaller for SNPs than microsatellites, but were also sensitive to decreasing numbers of alleles across translocations and heterozygosity measures H_O and H_E were more sensitive in SNP data indicating a decline that wasn't seen in microsatellite data (ii) inbreeding coefficients F_{IS} from SNP and microsatellites showed a similar decrease across translocations, but the SNP measure of F_{ROH} increased slightly over the translocations; (iii) measures of F_{ST} were larger for SNP data, but it is unclear whether this is due to marker differences or differences in the years that individuals were sampled when comparing my and the previous study. Further analysis with the SNP data showed that (iv) the hihi populations have short ROHs, corroborating Brekke et al.'s (2011) work showing lack of genetic signal for a bottleneck in the microsatellite data, (v) that the number of ROHs increased across bottlenecks, indicating increases in inbreeding in translocated populations and (vi) that some regions of the genome contained more ROHs than others, potentially due to differences in linkage disequilibrium. Genomic data has also allowed us to identify putative genes based on homology with zebra finch in areas with increasing and decreasing ROHs across translocations. Overall, the SNP data is giving us a clearer picture of changes in population genomics across the translocations. From my analysis we can see a clear reduction in diversity across the genome and increase in homozygosity in some regions of the genome. I recommend artificial migration among hihi populations to mitigate potential losses of rare alleles.

Chapter 5: General Discussion

The potential of genomic studies and SNP technologies to benefit conservation biology and add to our understanding of wild populations has been highlighted in a number of reviews on the subject (Benestan et al., 2016; Corlett, 2017; Galla et al., 2016; Garner et al., 2016; Harrisson et al., 2014; Ouborg et al., 2010). Shafer et al., (2015) notes that genomics offers higher resolution with which to study demography, adaptive genetic variation, quantitative genetic variation, taxonomic identification, inbreeding and genetic monitoring among other advantages. There is also enthusiasm for understanding the genomic architecture of adaptive traits in wild populations (Johnston et al., 2011; Kim et al., 2018; Slate et al., 2010). However, Harrisson (2014) cautions that focusing on particular adaptive traits does not translate well into management practice given our current understanding of the genomic basis of adaptation and future threats, but that a generalized measure of adaptive potential can usefully inform adaptive management programs.

SNP arrays are useful for characterising the architecture of traits under selection but can also be used to calculate generic measures of genetic diversity which indicate adaptive potential (Chapter 2). Heritability of traits in hihi is low, for example heritability of fledgling mass is $h^2 = 0.0329$; tarsus length is $h^2 = 0.123$ and head-bill length is $h^2 = 0.0581$, suggesting low evolutionary potential in the species (de Villemereuil et al., 2019a; Villemereuil et al., 2019b).
I focused on measuring genetic diversity and inbreeding to assess the effect of conservation management processes on the general population and across translocation bottlenecks. I also used the higher resolution of SNPs to confirm that removing individuals was not adversely affecting the population.

One utility of SNPs that has only been lightly explored in a conservation context is their ability to illuminate inbreeding across the genome at a fine scale. Runs of homozygosity (ROH) are contiguous regions of the genome dominated by homozygous markers (Broman and Weber, 1999a). Runs of homozygosity mark regions of inbreeding across the genome and in some cases have been associated with human diseases such as Alzheimer's (Ghani et al., 2015), and coronary artery disease (Christofidou et al., 2015). Furthermore, regions marked by ROH can be under positive selection e.g. in regions controlling reproductive traits in horse (genus *Equus*; Metzger et al., 2015) and those

controlling muscle formation and skin colour in sheep (*Ovis aries*; Purfield et al., 2017). There is a potential to explore how inbreeding is spread across the genome and identify regions that might be under particular pressure in small populations.

In hihi, it seems that while the global levels of inbreeding are increasing incrementally, there are some regions of the genome that are more heavily impacted than others (Chapters 3, 4). This pattern was observed in the Tiritiri Mātangi population across time and more markedly across translocations, where the prevalence of ROHs were increasing in some genomic regions, and a few were decreasing. Some possible causes are positive selection, regions of high linkage disequilibrium, and random chance. Although we studied the genes in these regions of changing ROH prevalence, we did not observe a clear trend in gene ontology. A study in pigs (*Sus scrofa*) also showed no significant correlation between gene content and ROH count (Bosse et al., 2012). Future research should investigate whether the ROHs that are increasing are located in regions of low recombination and those that are decreasing in regions of high recombination (when the linkage map is available).

In the remaining hihi populations, short ROHs are predominant indicating that the signature of inbreeding comes from an ancient ancestor as a sign of historic bottleneck events (Chapters 3, 4). Examination of ROHs as undertaken in this thesis uncover historic genetic events as well as current genetic diversity that are key to conservation management.

Whilst ROHs are widely used in studies of human (*Homo sapiens*; McQuillan et al., 2008; Pemberton et al., 2012) and cattle (Peripolli et al., 2017; Purfield et al., 2012), there have been few applications of ROH identification in ecology or conservation biology to date. An in-depth study of flycatcher genomes (*Ficedula* spp.) combined linkage data with ROHs to determine the expected population size (N_e) of current flycatcher populations and their ancestral populations to identify levels of hybridisation across insular populations (Kardos et al., 2017). A genome study of mountain gorillas (*Gorilla gorilla*) demonstrated a recent decline in the population resulted in ~34% homozygosity across the genomes of most individuals (Xue et al., 2015). In woolly mammoths (*Mammuthus primigenius*) a 28-fold increase in ROHs is evident at the end of the last glaciation showing a significant increase in inbreeding (Palkopoulou et al., 2015). Bosse et al. (2012) carried out an extensive study of ROHs in wild and domestic pig (*Sus scrofa)*

and found that nucleotide diversity outside ROHs was high in populations that were historically large. Reintroduction programmes of European ibex (*Capra ibex*) found that long ROHs (20Mb) were only present in the reintroduced populations when compared with the source population which is directly comparable to the changes in ROH signature in hihi across translocations (Chapters 3, 4; Grossen et al., 2018).

Many of these studies have focussed on general metrics of ROHs across a population including length of ROH and number of ROH per individual. These general metrics inform the demographic history of the population of the study species.

A recent review illustrates how different demographic histories can affect patterns of ROH based on studies of human populations, in particular how the sum of total length of ROH (sROH) compared with the number of ROH in individuals in the population changes depending on demography (Ceballos et al., 2018). The tight clustering of ROH lengths, number of SNPs per ROH as well as ROHs per individual, indicate that the hihi population is likely outbred. However, there are a small number of outliers for each metric. This could mean that there are a small number of highly related individuals mating every year, increasing the length of ROH, number of SNPs and SNPs per individual. Alternatively, it could mean that heterozygous SNPs between markers are not being picked up that would otherwise break the ROH in two. However, the majority of ROHs over 500 kb in another cavity nesting passerine, the Baltic collared flycatcher (*Ficedula albicollis*), were between 500-1000 kb in length, and they were still well represented at lengths of 5000 kb as measured with whole genome sequencing (Kardos et al., 2017). Therefore it is possible that longer ROHs are truly present in hihi, although the longest one in this study is only 1721 kb.

The work in this thesis illustrates how ROH can be used to illustrate the recent demographic history of a species of conservation concern. I demonstrated the effects of first and second order translocations in further increasing ROHs across the genome and especially increasing in the number of ROHs per individual. While the mean length of individual ROHs may not have changed significantly, the sROH in each individual gets longer as the mean number of ROHs per individual increases marking the effects of the bottleneck (and moving the ROH characteristics from those of a small population towards those of a bottlenecked population). I showed that genetic diversity is low in hihi (Chapter 3), supporting results from a recent publication (de Villemereuil et al., 2019a). It

is not currently known how long hihi have persisted at these low levels of genetic diversity as no recent bottlenecks have been detected in microsatellite data (Brekke et al., 2011). SNP data supports this as ROHs are mostly short in the population, indicating that they were generated in an ancient bottleneck (Ceballos et al., 2018).

One reason for the small number of studies of ROHs in a conservation or ecological context is the need for whole genome sequencing. Here we have used SNP array data, but it is possible that there will be some false positives in the ROHs we have detected (Purfield et al., 2012). A threshold of 20 or more consecutive homozygous SNPs was chosen in order to identify ROHs, given that if all SNPs are at the average minor allele frequency of 0.23 (Chapter 4), this would occur by chance with a probability of only 1.7e-13. A study in European ibex (genus *Capra;* Grossen et al., 2018) used 41,907 restriction associated DNA sequencing (RAD-seq) markers, a comparable number to many SNP arrays in cattle studies; however, it has been shown that 50K chips overestimate ROHs under 4 Mb in cattle (Ferenčaković et al., 2013; Purfield et al., 2017). In Iberian ibex (*Capra ibex)*, ROHs are predominantly less than 5Mb, so some may be false positives, but there are also much longer segments (Grossen et al., 2018). Without whole genome sequencing, it is not possible to quantify the limitations of ROH identification (Chapters 3, 4).

ROH are often divided into length categories to distinguish different ancestral origins or timing (e.g. in Pemberton et al., (2012), 3 categories were used (i) short ROHs, tens of kbs in length (ii) intermediate ROH measuring hundreds of kb to several Mb and (iii) long ROH that are multiple Mb long). Human (*Homo sapiens*) genomes are 2.85 Gb (Human Genome Sequencing Consortium, 2004) and cow (*Bos taurus*) is 2.86 Gb (Zimin et al., 2009) compared with 1.2 Gb in zebra finch (Taeniopygia guttata; Warren et al., 2010) and a maximum length of 1.05 Gb for the hihi (*Notiomystis cincta*) draft assembly in this study (passerine genomes are generally of similar size; (The Avian Genome Consortium et al., 2014)). However, genome length *per se* does not seem to affect potential ROH size; the collared flycatcher study on ROHs above shows a distribution of ROHs between 953 bp to 17.5 Mb, with sROH ranging from 0.07 to 0.16 of the genome (Kardos et al., 2017). Hihi, therefore, appear to have very short ROHs even by the standards of bird populations.

Management suggestions

In our study we have shown that isolated reintroduced populations are genetically distinct and that translocations have reduced the genetic diversity of new populations as has been seen with microsatellite data (Brekke et al., 2011). There was a decrease in the microsatellite coefficient of inbreeding in Tiritiri Mātangi after a genetic top-up from Te Hauturu-o-Toi that lasted as long as the migrant birds were reproducing. This supports the idea that moving individuals between populations may help mitigate erosion of genetic diversity as is recommended by the one-migrant-per-generation rule of thumb (Mills and Allendorf, 1996). Simulations show that mating individuals with an aim to avoid shared ROHs can outperform mating based on pedigree distance in populations with low genetic load (Bosse et al., 2015; de Cara et al., 2013).

I have shown clear changes in the ROH genomic landscape in each of the hihi populations. The key management recommendation as an outcome of our study is to identify individuals for translocation that are dominated by a different set of ROHs than the recipient population, which may help increase genetic diversity in the recipient population.

SNP arrays are robust enough to genotype degraded samples as has been seen in this study and elsewhere (Johnston et al., 2013; Mead et al., 2008). One interesting study might be to collect hihi museum samples and investigate if these show different levels of diversity and inbreeding, and how ROH distribution has changed across the genome. A study of takahē (*Porphyrio hochstetteri*) museum specimens from the 1900s showed that they had only slightly more genetic diversity at that time (Grueber and Jamieson, 2011). In a review of small and reintroduced populations, Jamieson (2015) noted that museum specimens showed the same tendency toward greater losses in MHC variation than neutral diversity as contemporary populations. This illustrates both the need to check historic sample diversity and the need for SNP data, which spans both coding and non-coding regions of the genome.

Linkage maps are available for other passerines such as zebra finch (Stapley et al., 2008), collared flycatcher (*Ficedula albicollis;* Kawakami et al., 2014), house sparrow (*Passer domesticus*) and blue tit (*Cyanistes caeruleus*; Hansson et al., 2010) and will soon be available for hihi. A hihi linkage map and a fully annotated hihi genome in conjunction with the SNP data will allow more in-depth analysis. ROH length is dependent on mutation and recombination rates (Bosse et al., 2012; Kardos et al., 2017) and a linkage

map will help distinguish regions more likely to have long ROHs (and where they might be false positives). Also, it will help elucidate whether the ROH hotspots across translocations and increasing in the Tiritiri Mātangi population are in regions of low recombination rate. Linkage data is required to accurately measure expected population size (N_e) based on ROHs (Browning and Browning, 2015), as the length of segments measured in centimorgans (cM) is inversely proportional to the number of generations to the common ancestor. Further study of linkage disequilibrium combined with ROH information could elucidate the timing and nature of the bottleneck that the Te Hauturu-o-Toi population must have gone through to establish on this island. This might complement the study of museum specimens suggested above. The availability of whole genome resequencing in the future may help benchmark the ROHs detected here.

Appendices A

Supplementary data for Chapter 2

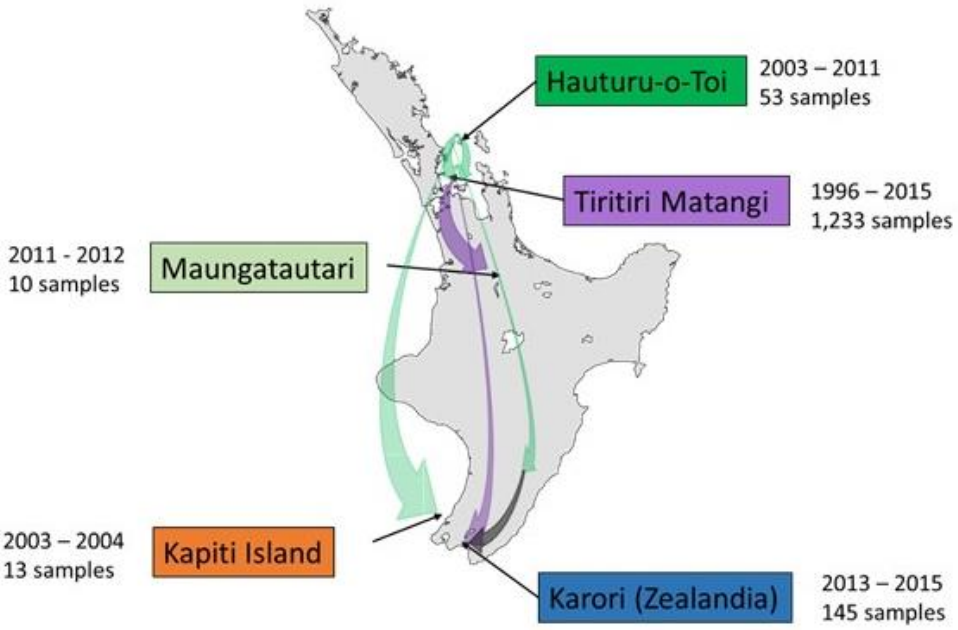


Figure S1: Map of New Zealand North Island with population locations. Each population is listed with the total number of samples and the range of years they were sampled.

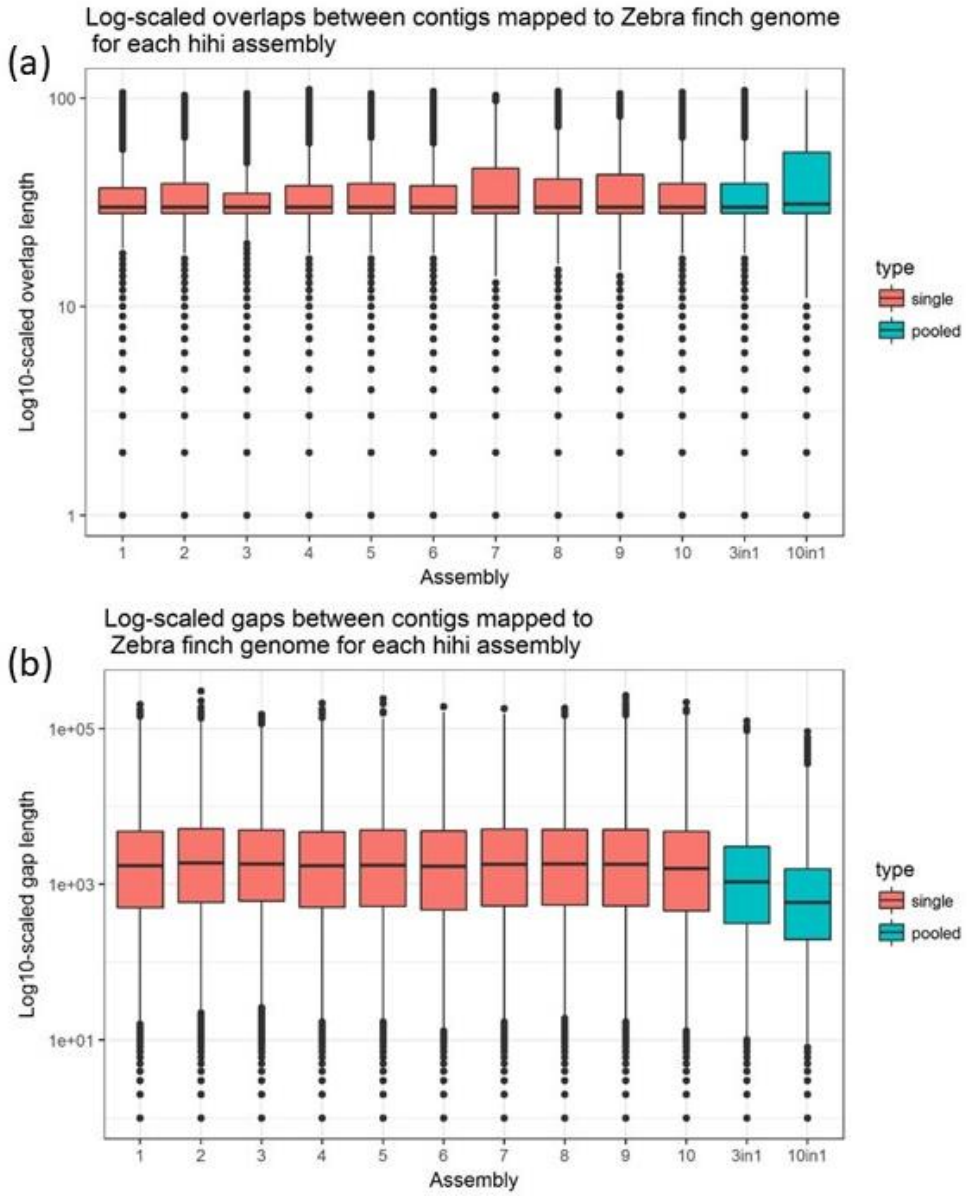


Figure S2: Overlaps and gap lengths between contigs per assembly. a) Log-scaled overlaps between adjacent contigs mapped to zebra finch genome for each hihi assembly **b)** Log-scaled gaps between contigs mapped to zebra finch genome for each hihi assembly.

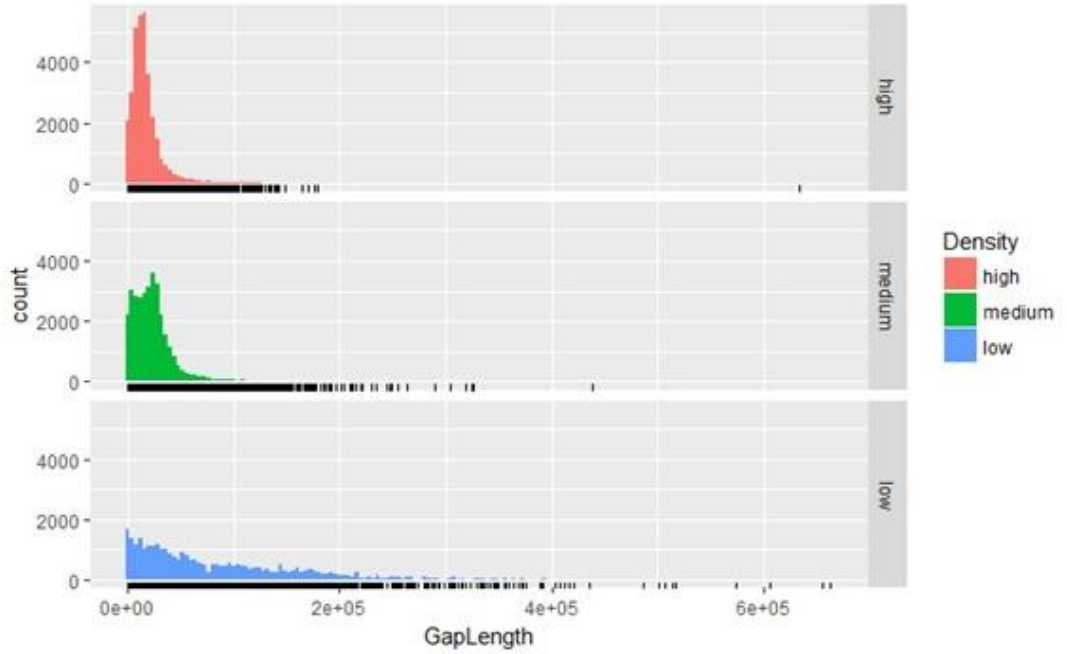


Figure S3: SNP distribution across homologous positions on the zebra finch genome. The histograms show the gap lengths between adjacent SNPs based on their homologous positions on the zebra finch genome. Chromosomes were divided into three SNP-density classes shown here: **a)** high (chromosomes 10-28, LG1, LG5, LGE22, Z and the mitochondria), **b)** medium (chromosomes 1-9, 1A, 1B and 4A) and **c)** low (chromosome Un and 'random' chromosomes).

Table S1: Sample cohort and population

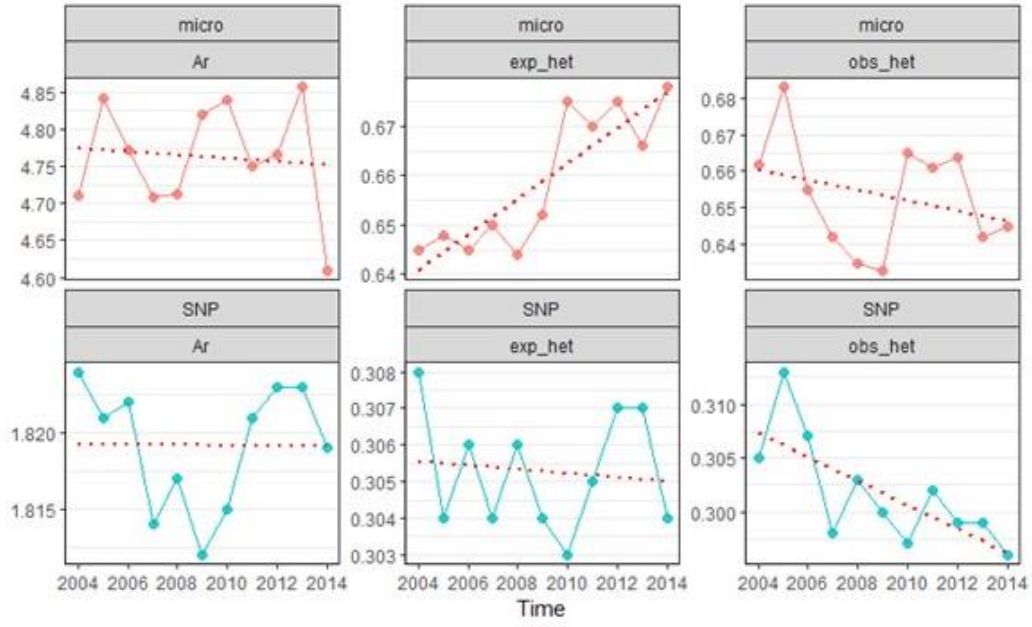
population	cohort	sum per cohort	number failed / total* (proportion)	total* per population
Hauturu-o-Toi	2003/2004	9	1/9 (0.11)	55
	2006/2007	14	1/14 (0.07)	
	2008/2009	8	0/8 (0.00)	
	2010/2011	24	0/24 (0.00)	
Kāpiti	2003/2004	14	1 /14 (0.07)	14
Karori	2013/2014	90	4/90 (0.04)	163
	2014/2015	73	2/73 (0.03)	
Maungatautari	2011/2012	10	0/10 (0.00)	
	2012/2013	2	0/2 (0.00)	12
Tiritiri Mātangi	1996/1997	1	0/1 (0.00)	1278
	1997/1998	1	0/1 (0.00)	
	1999/2000	3	1/3 (0.33)	
	2000/2001	6	1/6 (0.17)	
	2001/2002	8	1/8 (0.13)	
	2002/2003	16	4/16 (0.25)	
	2003/2004	20	3/20 (0.15)	
	2004/2005	115	4/113 (0.04)	
	2005/2006	50	11/50 (0.22)	
	2006/2007	50	11/50 (0.22)	
	2007/2008	150	0/148 (0.00)	
	2008/2009	150	6/148 (0.04)	
	2009/2010	149	1/149 (0.01)	
	2010/2011	108	5/108 (0.05)	
	2011/2012	92	1/92 (0.01)	
	2012/2013	155	1/152 (0.01)	
	2013/2014	130	0/130 (0.00)	
	2014/2015	86	2/83 (0.02)	
Unknown		2		
Total		1536		1520

*total is the total number of individuals once six replicated individuals and two unknown samples were removed

Table S2: SNP failure per chromosome

Chromosome	Passed	Failed	Chromosome	Passed	Failed
1	4045	1151	1_random	25	14
2	5388	1507	2_random	55	12
3	3841	1225	3_random	17	9
4	2526	675	4_random	107	35
5	2115	759	5_random	49	28
6	1278	436	6_random	45	14
7	1418	428	7_random	21	4
8	974	331	8_random	182	59
9	935	347	9_random	9	3
10	1176	375	10_random	15	5
11	1224	395	11_random	10	2
12	1203	431	12_random	4	2
13	900	329	13_random	107	40
14	899	358	14_random	6	1
15	794	328	15_random	11	3
16	0	1	16_random	2	0
17	617	264	17_random	6	1
18	599	232	18_random	7	4
19	577	274	19_random	4	1
20	832	368	20_random	8	2
21	327	146	21_random	28	12
22	178	72	22_random	8	8
23	334	169	23_random	9	6
24	429	191	24_random	7	2
25	58	34	25_random	10	2
26	274	137	26_random	20	23
27	219	145	27_random	1	5
28	255	136	28_random	3	3
1A	2626	848	1A_random	13	8
1B	36	13	1B_random	8	1
4A	704	255	4A_random	8	1
LGE22	65	38	LGE22_random	1	1
Z	2598	2500	Z_random	27	33
LG2	1	3			
Un	1172	484			
MT	0	1			
RAD	762	524			

Supplementary data for Chapter 3



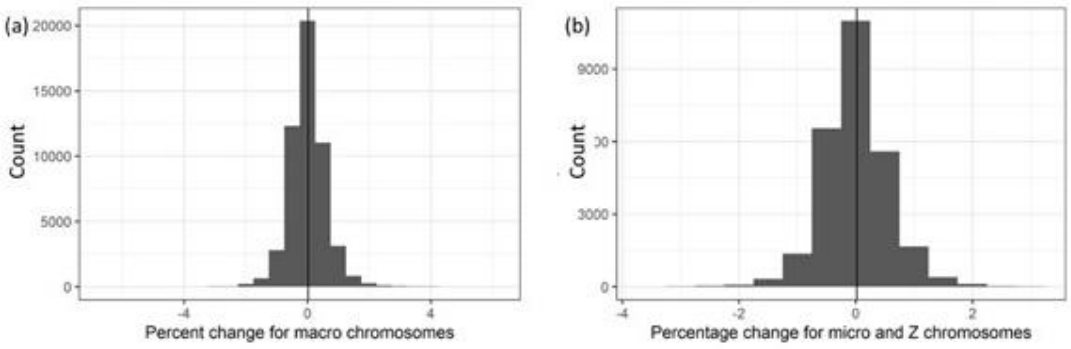
Supplementary Figure 1: Allelic Richness (A_R), Expected Heterozygosity (exp_het) and Observed Heterozygosity (obs_het) for microsatellite (micro) and SNP data. The red dotted line is a best fit regression line, but only microsatellite expected heterozygosity and SNP observed heterozygosity are significant. **(A_R)** Allelic Richness of microsatellites decreases by -0.002218x (*p-value*: 0.7738), allelic richness of SNPs decreases by -1.818e-05x (*p-value*: 0.9657). **(exp_het)** Expected heterozygosity of microsatellites increases by 0.0036182 (*p-value*: 0.0005303) and expected heterozygosity of SNPs decreases by -5.455e-05x (*p-value*: 0.7435). **(obs_het)** Observed heterozygosity of SNPs decreases by -0.0011182 (*p-value*: 0.009877) and observed heterozygosity of microsatellites decreases by -0.001400 (*p-value*: 0.3645).

Supplementary Table 1: Effects of removing individuals on F_{IS} measures

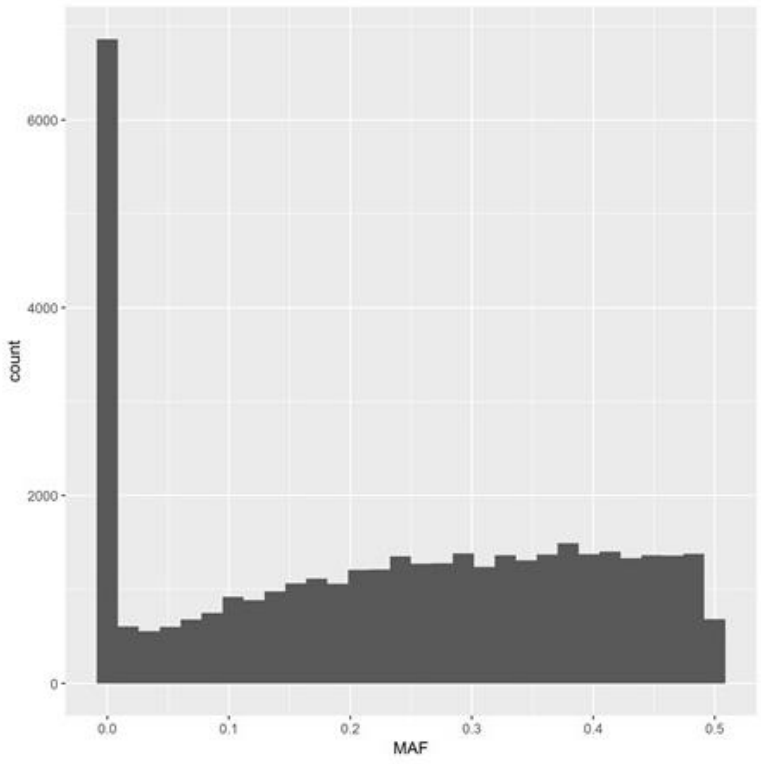
Model	Adjusted R^2	p-value	Intercept	p-value	harvested	p-value	Pre-breeding	p-value	year	p-value
Microsatellite Fis vs harvesting/no-harvesting + pre-breeding population + year	0.2103	0.22	-8.071208	0.063	-0.013096	0.31	-0.0002677	0.4901	0.004041	0.0621
Microsatellite Fis vs harvesting/no-harvesting + post-breeding population + year	1.62E-01	0.2643	-7.62E+00	0.0889	-1.32E-02	0.3269	-4.04E-05	0.7657	3.80E-03	0.0879
SNP Fis vs harvesting/no-harvesting + pre-breeding population + year	0.5665	0.0313	**-6.901823**	**0.0058**	-0.007798	0.2192	-0.0001478	0.4328	**0.003458**	**0.0057**
SNP Fis vs harvesting/no-harvesting + post-breeding population + year	6.08E-01	0.0223	**-7.28E+00**	**0.0039**	-6.31E-03	0.2939	7.15E-05	0.2583	**3.63E-03**	**0.0038**

Supplementary Table 2: Average F_{ROH} across years.

Year	mean	median
2004	0.2358	0.2317
2005	0.2260	0.2158
2006	0.2395	0.2390
2007	0.2483	0.2393
2008	0.2438	0.2402
2009	0.2440	0.2438
2010	0.2481	0.2432
2011	0.2419	0.2417
2012	0.2464	0.2441
2013	0.2447	0.2421
2014	0.2548	0.2495



Supplementary Figure 2: Percentage change between (a) early, middle and late groupings in macro chromosomes ('medium' density of SNPs). mean = 0.028, t = 10.89, df = 51771, *p-value* < 2.2e-16. (b) early, middle and late groupings in micro and Z chromosomes ('high' density of SNPs). mean = 0.0199, *t* = 6.0708, df = 27117, *p-value* < 1.289e-09. Black vertical line indicates the mean value.



Supplementary Figure 3: SNP minor allele frequency distribution for all individuals across all cohorts.

Supplementary Table 3: Genes found in regions of accumulating ROH in SNPs across early-mid and mid-late cohort groups. Some regions had increased ROH frequency across both cohort groups. Regions here are defined as half the average ROH length on either side of a SNP that has increasing frequency in ROHs. Annotations are from the zebra finch (*Taeniopygia guttata*) Ensembl 86 genome gtf3 file.

chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
1	**30668785**	**31625145**	30644456	30675156	ENSTGUG00000010057	LONRF2	LON peptidase N-terminal domain and ring finger 2
			30741102	30800866	ENSTGUG00000010070		
			30895913	31069032	ENSTGUG00000010075		
			31115697	31156550	ENSTGUG00000010082	REV1	REV1%2C DNA directed polymerase
			31161029	31195560	ENSTGUG00000010099	EIF5B	eukaryotic translation initiation factor 5B
			31199707	31206777	ENSTGUG00000010125		
			31219902	31232302	ENSTGUG00000010130		
			31234806	31236754	ENSTGUG00000010133		
			31237189	31240225	ENSTGUG00000010137	MITD1	microtubule interacting and trafficking domain containing 1
			31242795	31243913	ENSTGUG00000010145	LIPT1	lipoyltransferase 1
			31322421	31333015	ENSTGUG00000010146		
			31333748	31347850	ENSTGUG00000010153		
			31367962	31441120	ENSTGUG00000010155	MGAT4A	mannosyl (alpha-1%2C3-)-glycoprotein beta-1%2C4-N-acetylglucosaminyltransferase%2C isozyme A
			31446803	31450942	ENSTGUG00000010176	UNC50	unc-50 inner nuclear membrane RNA binding protein
			31451603	31455673	ENSTGUG00000010191	COA5	uncharacterized protein LOC100228523 Source:RefSeq peptide%3BAcc:NP_001184232]
			31528138	31570414	ENSTGUG00000010196	INPP4A	inositol polyphosphate-4-phosphatase type I A
			31584720	31666545	ENSTGUG00000010237	TMEM131	transmembrane protein 131
2	**63182824**	**63568824**	63198548	63204921	ENSTGUG00000004502		
			63211335	63212387	ENSTGUG00000004507		
			63224393	63225445	ENSTGUG00000004511		
			63244706	63245653	ENSTGUG00000018123		
			63285699	63286571	ENSTGUG00000004516	XCR1	X-C motif chemokine receptor 1
			63309226	63342652	ENSTGUG00000004520	FYCO1	FYVE and coiled-coil domain containing 1
			63349911	63351050	ENSTGUG00000004525	CCR9	C-C motif chemokine receptor 9
			63360437	63361753	ENSTGUG00000004527	TMPPE	transmembrane protein with metallophosphoesterase domain
			63367420	63403725	ENSTGUG00000004528	GLB1	galactosidase beta 1
			63398281	63421774	ENSTGUG00000018201	CCR4	C-C motif chemokine receptor 4
			63435149	63436207	ENSTGUG00000018365	CX3CR1	C-X3-C motif chemokine receptor 1
			63455854	63456921	ENSTGUG00000004574		
			63461079	63467604	ENSTGUG00000004577	SLC25A38	solute carrier family 25 member 38
			63475376	63479933	ENSTGUG00000004584	RPSA	Taeniopygia guttata laminin receptor 1 (LAMR)%2C mRNA.
			63493066	63521851	ENSTGUG00000004612		A6A2:H33A1:H33H61A15:H33A2:A1:H33

chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
3	**19718725**	**20124232**	19761500	19766212	ENSTGUG00000004700		
			19767112	19772004	ENSTGUG00000004706		
			19772354	19779565	ENSTGUG00000004725		
			19786724	19796586	ENSTGUG00000004738	SRF	serum response factor
			19803910	19815873	ENSTGUG00000004755	PTK7	protein tyrosine kinase 7 (inactive)
			19840067	19843990	ENSTGUG00000004794		
			19844259	19857695	ENSTGUG00000004797		
			19867530	19869093	ENSTGUG00000004831	MRPL2	Taeniopygia guttata mitochondrial ribosomal protein L2-like (LOC100190175)%2C mRNA.
			19870532	19873547	ENSTGUG00000004839		
			19879289	19879670	ENSTGUG00000004840		
			19880957	19934419	ENSTGUG00000004842	PACS1	phosphofurin acidic cluster sorting protein 1
			19941715	19942542	ENSTGUG00000004900		
			19958458	19971226	ENSTGUG00000004902	KLHDC3	kelch domain containing 3
			19977883	19978960	ENSTGUG00000004930	MEA1	male-enhanced antigen 1
			19981662	20006850	ENSTGUG00000004938	PPP2R5D	protein phosphatase 2 regulatory subunit B'delta
			20013862	20051363	ENSTGUG00000004985	ABHD12	abhydrolase domain containing 12
			20054171	20065675	ENSTGUG00000005015	PYGB	phosphorylase%2C glycogen%3B brain
			20071902	20082627	ENSTGUG00000005097	ENTPD6	ectonucleoside triphosphate diphosphohydrolase 6 (putative)
			20108682	20111562	ENSTGUG00000005139		
			20116765	20129188	ENSTGUG00000005146	PEX6	peroxisomal biogenesis factor 6
7	**7558298**	**8044332**	7834568	8039714	ENSTGUG00000004748	CNTNAP5	contactin associated protein like 5
8	**18644654**	**19030654**	18644011	18644697	ENSTGUG00000008231	LURAP1	leucine rich adaptor protein 1
			18728050	18745045	ENSTGUG00000008233	RAD54L	RAD54-like (S. cerevisiae)
			18746390	18753798	ENSTGUG00000008262	LRRC41	leucine rich repeat containing 41
			18754304	18755563	ENSTGUG00000008278	UQCRH	Taeniopygia guttata ubiquinol-cytochrome c reductase hinge protein (UQCRH)%2C mRNA.
			18756976	18760529	ENSTGUG00000008304		
			18761676	18771832	ENSTGUG00000008319		
			18791842	18799899	ENSTGUG00000008323	FAAH	fatty acid amide hydrolase
			18833732	18840550	ENSTGUG00000008353	DMBX1	diencephalon/mesencephalon homeobox 1
			19006195	19020882	ENSTGUG00000008358		

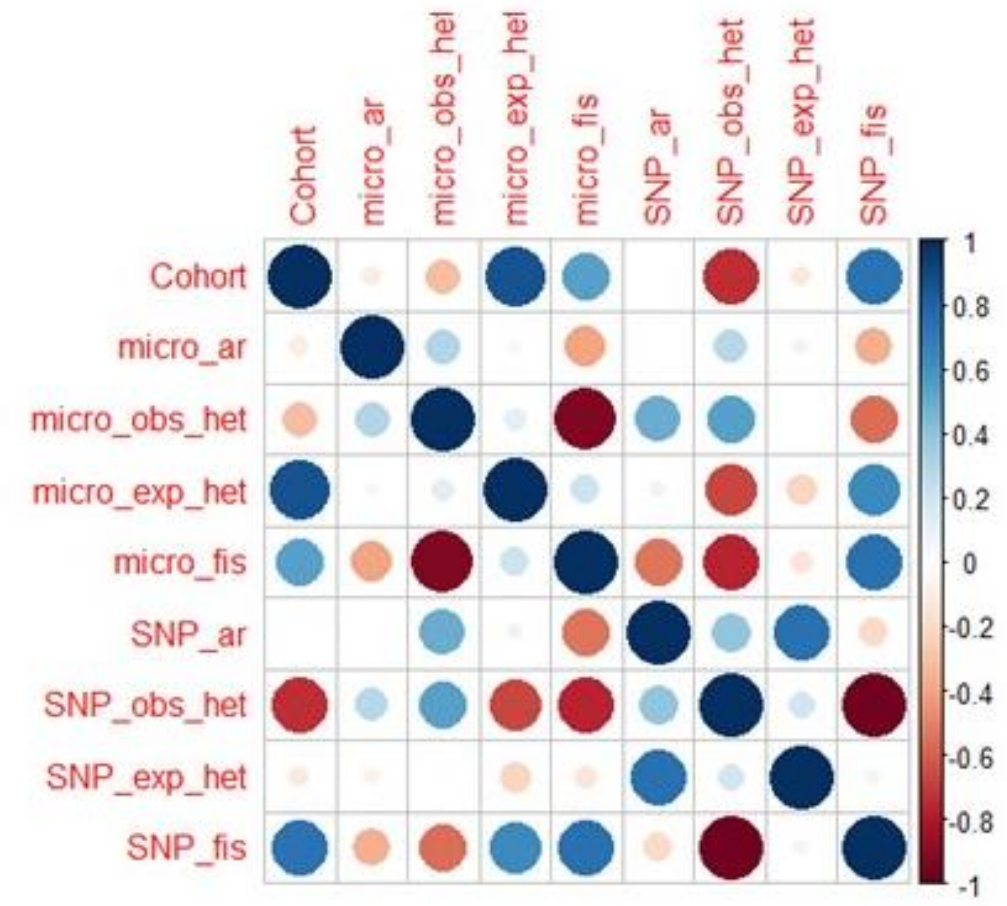
chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
12	**3931094**	**4325468**	4151320	4155816	ENSTGUG00000004715	C3orf18	chromosome 3 open reading frame 18
			4162341	4185084	ENSTGUG00000004718	HEMK1	HemK methyltransferase family member 1
			4241293	4242494	ENSTGUG00000004724	CISH	cytokine inducible SH2 containing protein
			4256355	4306559	ENSTGUG00000004729	MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3
18	**1287942**	**1673942**	1302259	1308543	ENSTGUG00000003057	ENDOV	endonuclease V
			1311661	1319551	ENSTGUG00000003062		
			1338798	1339241	ENSTGUG00000003067		
			1341444	1351954	ENSTGUG00000003069		
			1363585	1370661	ENSTGUG00000003079	SLC26A11	solute carrier family 26 member 11
			1370883	1373370	ENSTGUG00000003089	SGSH	N-sulfoglucosamine sulfohydrolase
			1374230	1379314	ENSTGUG00000003105	GAA	glucosidase alpha%2C acid
			1382237	1392666	ENSTGUG00000003135	CCDC40	coiled-coil domain containing 40
			1401181	1424254	ENSTGUG00000003143	TBC1D16	TBC1 domain family member 16
			1442104	1445251	ENSTGUG00000003149	CBX4	chromobox 4
			1463425	1465062	ENSTGUG00000003156	CBX8	chromobox 8
			1470054	1470911	ENSTGUG00000003161		
			1479287	1481601	ENSTGUG00000003166	ENPP7	ectonucleotide pyrophosphatase/phosphodiesterase 7
			1583371	1586770	ENSTGUG00000003184	ENGASE	endo-beta-N-acetylglucosaminidase
			1589531	1594019	ENSTGUG00000003204		
			1596614	1598146	ENSTGUG00000003205	CANT1	calcium activated nucleotidase 1
			1614043	1645771	ENSTGUG00000003207		
			1647569	1662393	ENSTGUG00000003319	PGS1	phosphatidylglycerophosphate synthase 1
			1664631	1665260	ENSTGUG00000003345	SOCS3	suppressor of cytokine signaling 3
1A	**28276735**	**29005362**	28304477	28536443	ENSTGUG00000005793	PDZRN4	PDZ domain containing ring finger 4
			28717760	28737586	ENSTGUG00000005804	GXYLT1	glucoside xylosyltransferase 1
			28773864	28796368	ENSTGUG00000005817	YAF2	YY1 associated factor 2
			28830702	28836447	ENSTGUG00000005829	ZCRB1	zinc finger CCHC-type and RNA binding motif containing 1
			28847595	28906754	ENSTGUG00000005837	PPHLN1	periphilin 1
			28913871	28921925	ENSTGUG00000005851	PRICKLE1	prickle planar cell polarity protein 1
1A	**33822572**	**34208572**	33789742	33865928	ENSTGUG00000006648	GRIP1	glutamate receptor interacting protein 1
			34143137	34169941	ENSTGUG00000006691		

Linear Model Testing

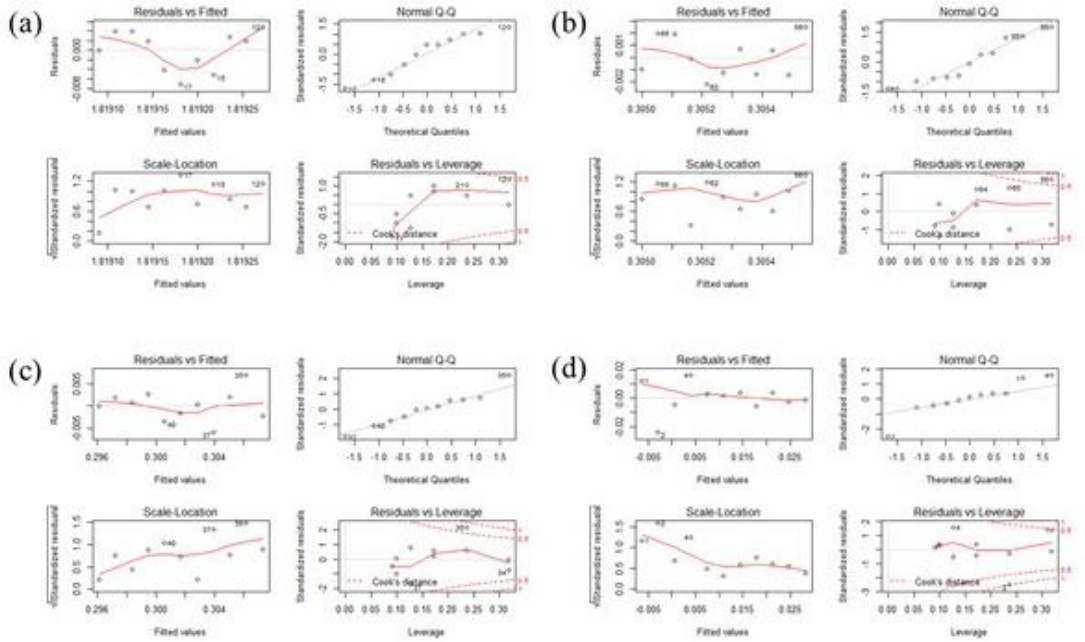
Supplementary Table 4: Results of linear model tests. Test mean of residuals, autocorrelation using the Durbin-Watson test, that the x variables and residuals are uncorrelated using Pearson's product-moment correlation, test for correlation between variables and homoscedasticity.

			Pearson's product-moment correlation								
model	**mean of residuals**	**Durbin-Watson**	***p-value***	***corr***	***t***	***df***	***p-value***	**correlated variables**	**Supp Fig**	**homoscedasticity**	**Supp Fig**
SNP A_R ~ cohort	-2.76E-19	1.0312	0.01156	0.999892	203.67	9	< 2.2e-16	no	4	near zero	5a
SNP H_E ~ cohort	-4.93E-20	1.9182	0.2963	0.99373	26.665	9	7.10E-10	no	4	no	5b
SNP H_O ~ cohort*	-3.95E-20	2.2236	0.502	0.677426	2.7628	9	2.20E-02	yes	4	no	5c
SNP F_{IS} ~ cohort*****	1.58E-19	2.3428	0.586	0.673075	0.586	9	2.32E-02	yes	4	near zero	5d
Microsatellite A_R ~ cohort	4.10E-18	2.0715	0.3959	0.995162	30.389	9	2.21E-10	no	4	near zero	6a
Microsatellite H_E ~ cohort*	1.97E-19	1.782	0.2182	0.4964	1.7155	9	1.20E-01	yes	4	no	6b
Microsatellite H_O ~ cohort	7.89E-19	1.3921	0.06606	0.952871	9.4227	9	5.86E-06	yes	4	no	6c
Microsatellite F_{IS} ~ cohort	5.91E-19	1.3398	0.05375	0.838937	4.6245	9	1.25E-03	yes	4	no	6d
SNP F_{IS} ~ harvesting/no-harvesting + pre-breeding pop size + year	3.75E-19	2.5463	0.7362	0.550877	1.8439	9	9.83E-02	yes	7	near zero	8a
SNP F_{IS} ~ harvesting/no-harvesting + post-breeding pop size + year	0	2.5773	0.8108	0.523633	1.8439	9	9.83E-02	yes	7	no	8b
microsatellite F_{IS} ~ harvesting/no-harvesting + pre-breeding pop size + year	5.12E-19	1.2071	0.02188	0.743499	3.3354	9	8.72E-03	yes	7	no	8c
microsatellite F_{IS} ~ harvesting/no-harvesting + post-breeding pop size + year	1.10E-18	1.416	0.05827	0.765912	3.5738	9	5.99E-03	yes	7	no	8d
ROHs per individual ~ cohort	2.75E-16	1.8261	0.001259	0.99842	610.38	1180	2.20E-16	no	9	no	10a
length ROHs ~ cohort	2.95E-15	1.9705	0.06242	0.999937	9212.4	10658	2.20E-16	no	9	no	10b
number of SNPs in ROHs ~ cohort*****	2.66E-16	1.95	0.004779	0.999666	3993.6	10658	2.20E-16	no	9	no	10c
F_{ROH} ~ cohort*****	-7.16E-20	1.8356	0.002094	0.998026	547.01	1185	2.20E-16	no	11a	no	12a
percentage SNPs in ROH ~ cohort*****	-3.42E-18	1.8307	0.0016	6.36E-17	2.19E-15	1185	1.00E+00	no	11b	no	12b

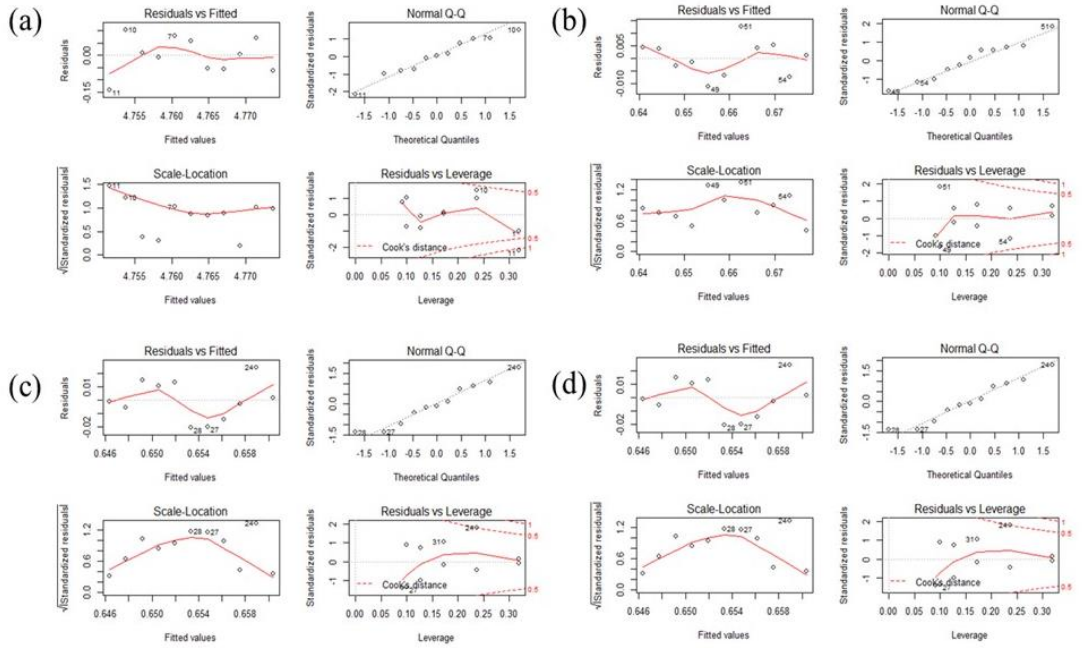
* models that were significant are highlighted with grey rows.



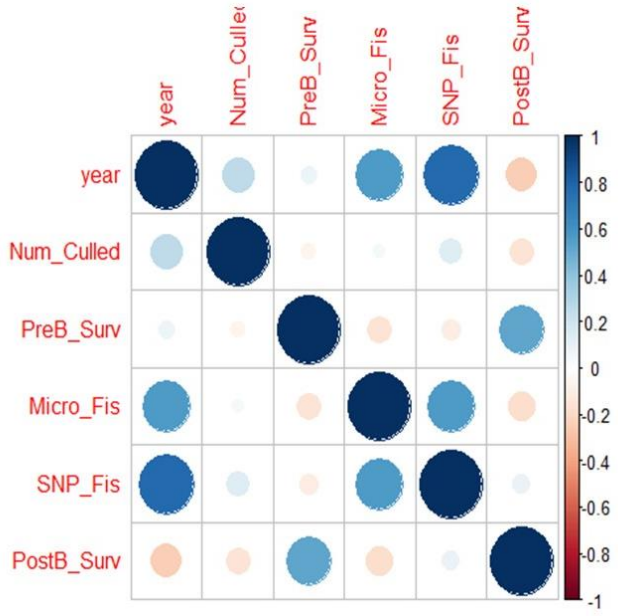
Supplementary Figure 4: Correlation between variables in genetic diversity linear models including year, allelic diversity, observed heterozygosity, expected heterozygosity and F_{IS} for both SNP and microsatellite data.



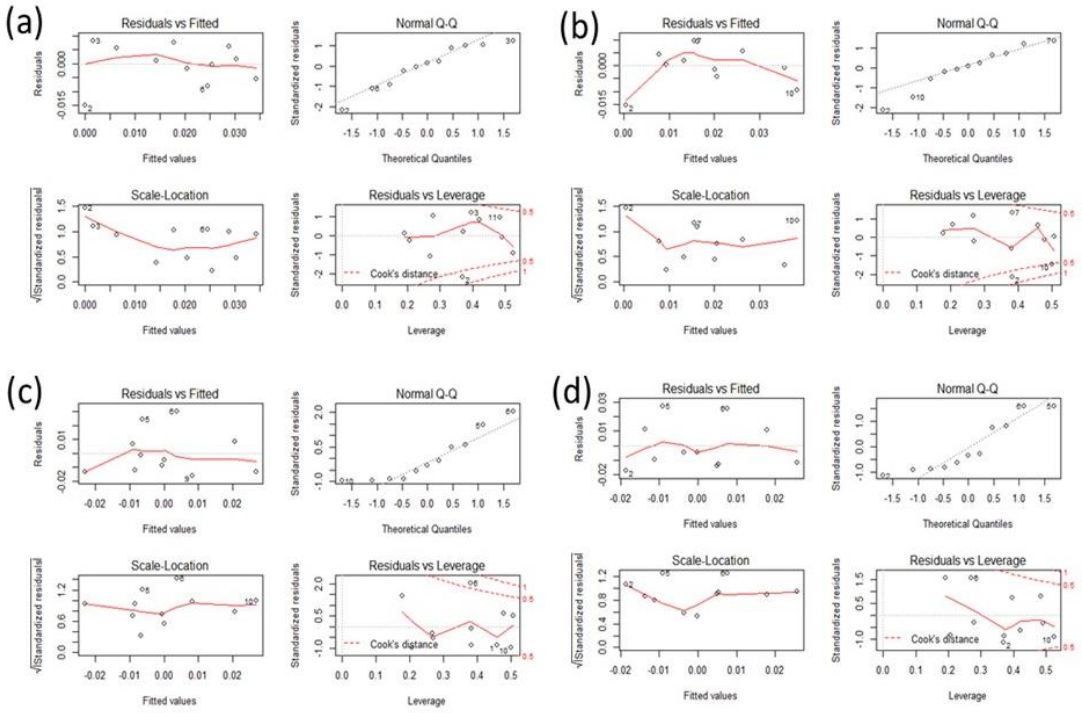
Supplementary Figure 5: Variable correlation and heteroscedasticity for SNP genetic diversity and inbreeding models (a) allelic richness, **(b)** expected heterozygosity, **(c)** observed heterozygosity, **(d)** inbreeding $F_{IS.}$



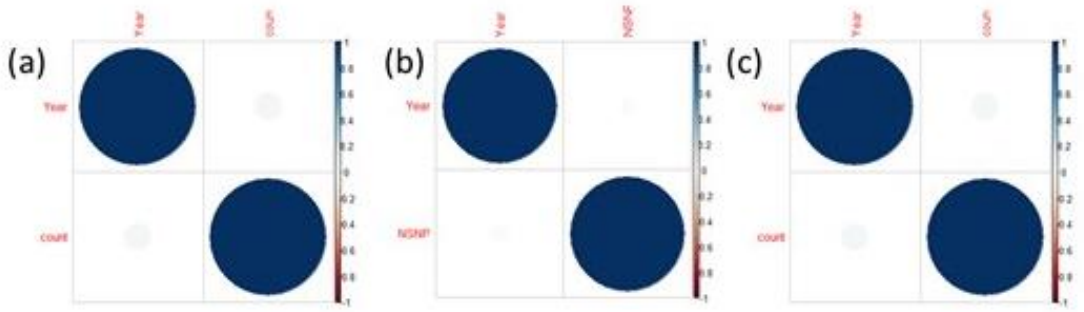
Supplementary Figure 6: Variable correlation and heteroscedasticity for microsatellite genetic diversity and inbreeding models (a) allelic richness, **(b)** expected heterozygosity, **(c)** observed heterozygosity, **(d)** inbreeding $F_{IS.}$



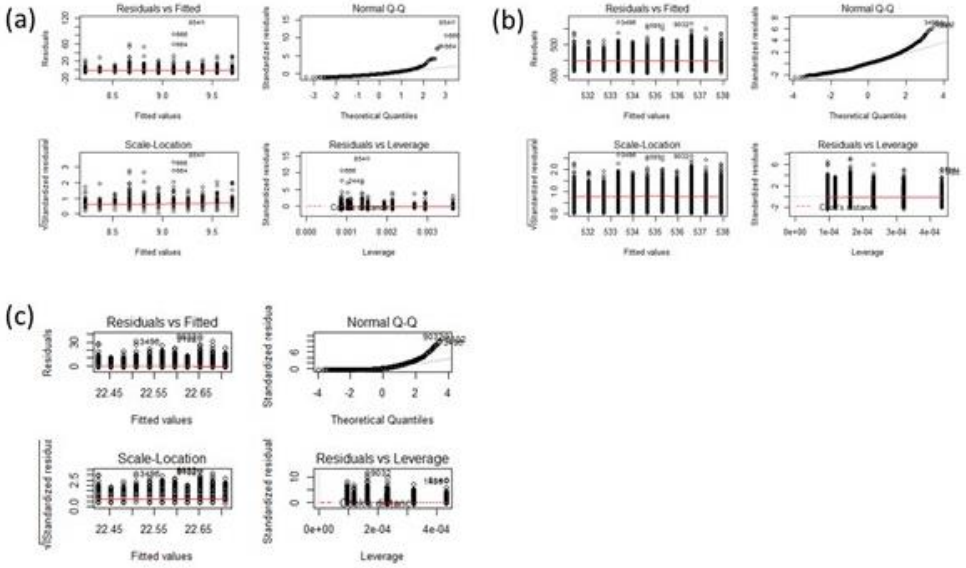
Supplementary Figure 7: Correlation between variables in harvesting models including the year, the number of birds removed, the pre-breeding survey population numbers, the microsatellite *FIS*, SNP *FIS* and post-breeding survey population numbers.



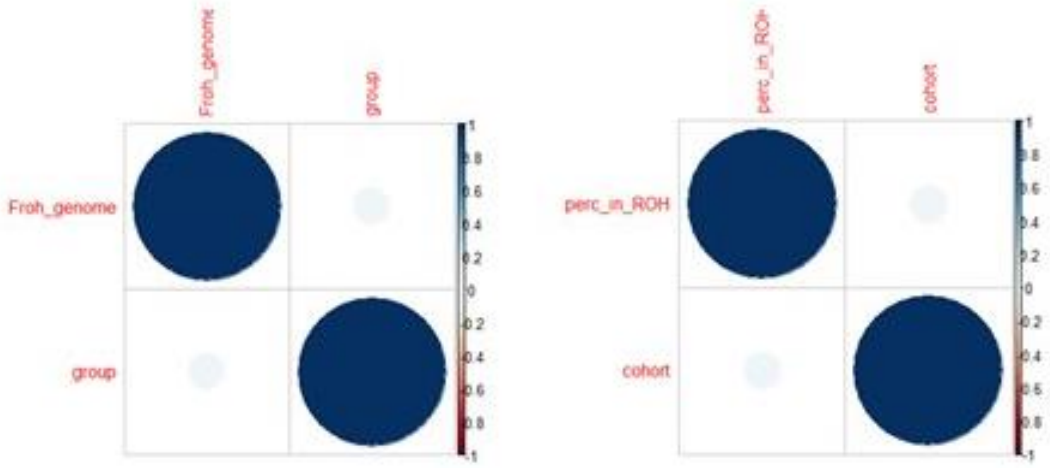
Supplementary Figure 8: Variable correlation and heteroscedasticity for SNP F_{IS} vs harvesting models (a) Heteroscedasticity tests for SNP F_{IS} pre-breeding survey linear model. **(b)** Heteroscedasticity tests for SNP F_{IS} post-breeding survey linear model. **(c)** Heteroscedasticity tests for microsatellite F_{IS} pre-breeding survey linear model. **(d)** Heteroscedasticity tests for microsatellite F_{IS} post-breeding survey linear model.



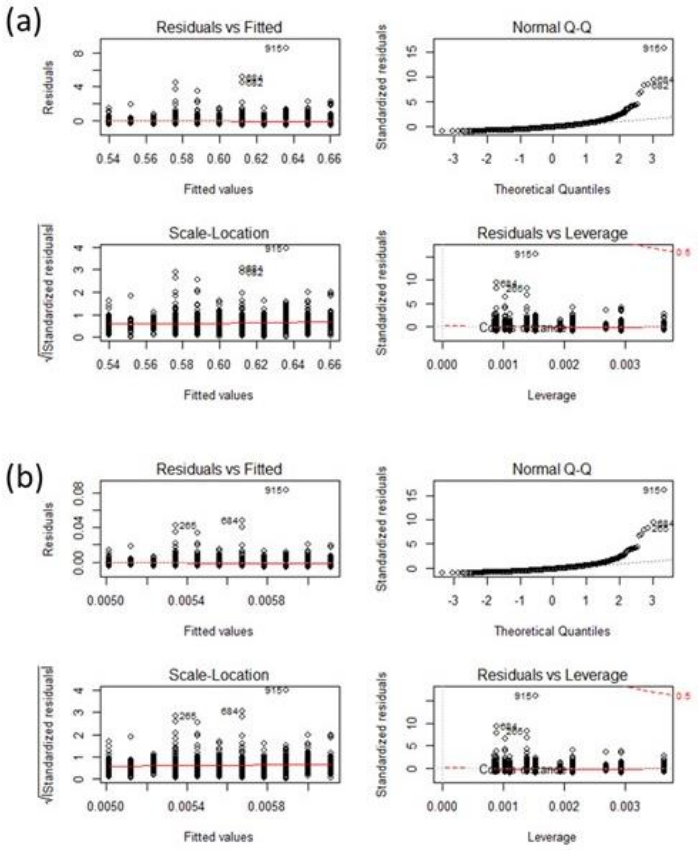
Supplementary Figure 9: Correlation between variables in models of runs of homozygosity. (a) year and count of runs of homozygosity per sample, **(b)** year, length (KB) of runs of homoyzogsity, (c) year and number of SNPs per run of homozyogsity.



Supplementary Figure 10: Variable correlation and heteroscedasticity for linear models of runs of homozygosity (ROH). (a) number of ROH per sample per cohort, **(b)** length of ROHs per cohort, **(c)** number of SNPs in ROHs per cohort.

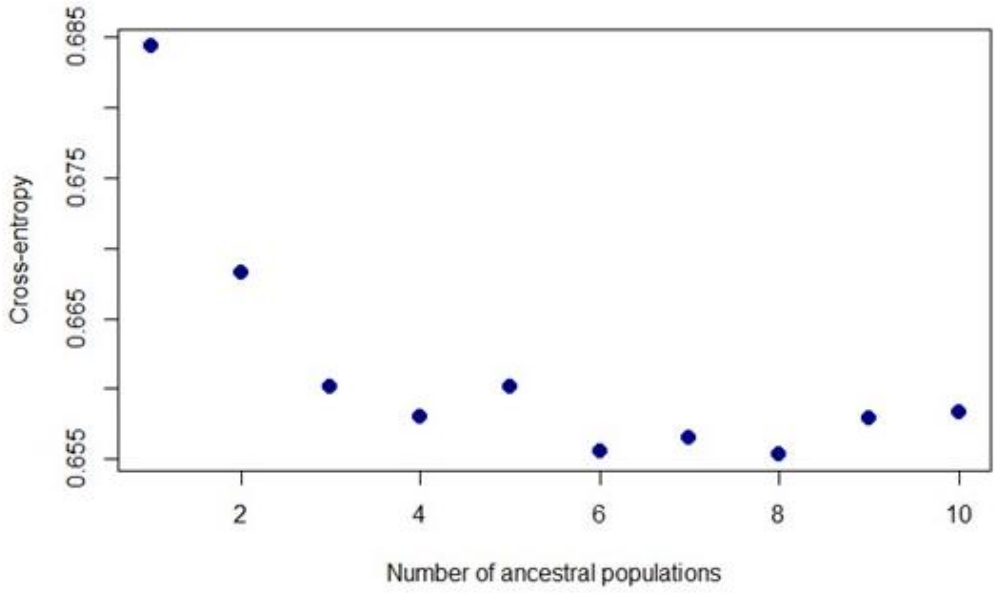


Supplementary Figure 11: Correlation between variables in models of runs of homozygosity. (a) F_{ROH} and year **(b)** percentage of SNPs in ROHs and year.

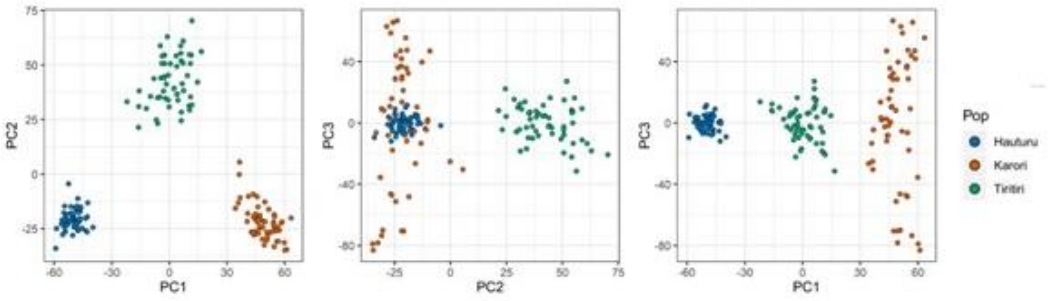


Supplementary Figure 12: Variable correlation and heteroscedasticity for linear models of runs of homozygosity (ROH). (a) F_{ROH} and year **(b)** percentage of SNPs in ROHs and year.

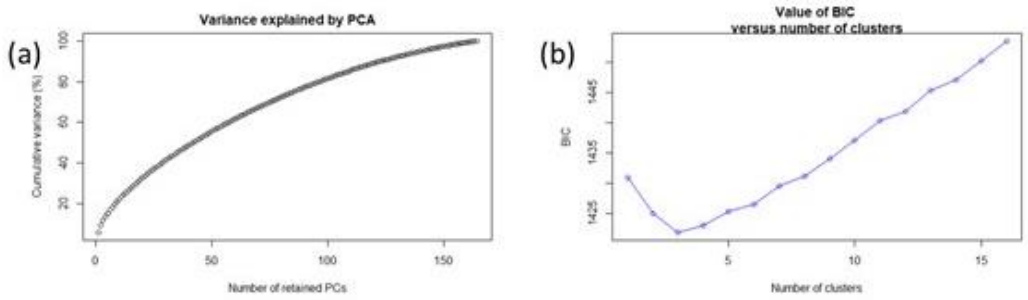
Supplementary data for Chapter 4



Supplementary Figure 1: Estimates of how well genomic data potentially describes different numbers of ancestral populations. The figure illustrates sparse nonnegative matrix factorisation (snmf) estimates of the entropy criterion of the number of ancestral populations



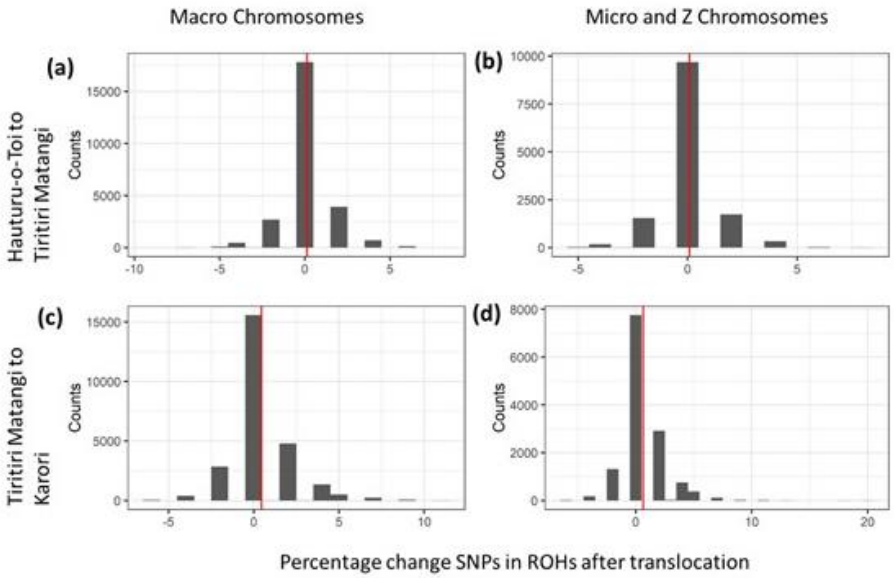
Supplementary Figure 2: PCA analysis of Te Hauturu-o-Toi, Tiritiri Mātangi and Karori populations. Figure shows that on PC1 and PC2 all three populations are distinct. In PC2 and PC3 Te Hauturu-o-Toi maps onto the Karori population and in PC1-PC3 the populations are again distinct. The Karori population shows a larger spread than either of the others.



Supplementary Figure 3: Cumulative variance explained and Bayesian information criterion (BIC) data for DAPC **analysis.** Cumulative variance explained by the PCA components shows a continual increase in variance explained by PCA given the number of retained PCAs. The Bayesian Information Criterion (BIC) for different number of clusters identifies 3 clusters as the optimum for this data set.

Supplementary Table 1: Genetic diversity measures from SNP data in this study and microsatellite data from Brekke et al. (2011)

	SNP data				**Microsatellite data**			
Population	A_R	H_O	H_E	F_{IS}	A_R	H_O	H_E	F_{IS}
Hauturu	1.84	0.317	0.318	0.003	5.19	0.66	0.68	0.042
Tiritiri	1.827	0.306	0.306	-0.002	4.71	0.66	0.64	-0.028
Karori	1.782	0.281	0.281	-0.003	4.42	0.65	0.66	NA



Supplementary Figure 4: Percentage change in SNPs in ROHs after translocation. Percentage change of SNPs in ROHs in (a) macro chromosomes and chromosomes 1A, 1B and 4A (mean = 0.135) and (b) micro and Z chromosomes (mean = 0.084) for the Te Hauturu-o-Toi translocation to Tiritiri Mātangi. Percentage change in SNPs in ROHs in (c) macro chromosomes (mean = 0.449) and (d) micro and Z chromosomes (mean = 0.634) for the translocation from Tiritiri Mātangi to Karori.

Supplementary Table 2: Genes found in regions of accumulating ROH in SNPs across translocations. Some regions had increased ROH frequency across both translocations. Regions here are defined as half the average ROH length on either side of a SNP that has increasing frequency in ROHs. Overlapping regions are joined. Annotations are from the zebra finch (*Taeniopygia guttata*) Ensembl 86 genome gtf3 file.

chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
2	28259568	28815294	28588024	28656673	ENSTGUG00000001694	STAC	SH3 and cysteine rich domain
3	42363464	43354992	42765532	42767039	ENSTGUG00000010041		
			42796437	42798723	ENSTGUG00000010059		
			43138810	43162476	ENSTGUG00000010090		
			43204849	43385568	ENSTGUG00000010096		
			42713738	42728774	ENSTGUG00000010005	ABCB10	ATP binding cassette subfamily B member 10
			43075601	43107544	ENSTGUG00000010081	EGLN1	egl-9 family hypoxia inducible factor 1
			43054563	43057353	ENSTGUG00000018403	EXOC8	exocyst complex component 8
			42477296	42515172	ENSTGUG00000009986	GALNT2	polypeptide N-acetylgalactosaminyltransferase 2
			42731549	42762548	ENSTGUG00000010013	NUP133	nucleoporin 133
			42421711	42454611	ENSTGUG00000009982	PGBD5	piggyBac transposable element derived 5
			42806929	42813597	ENSTGUG00000010060	RAB4A	RAB4A%2C member RAS oncogene family
			43022926	43027394	ENSTGUG00000010073	RHOU	ras homolog family member U
			43057675	43062085	ENSTGUG00000010080	SPRTN	SprT-like N-terminal domain
			42681362	42695745	ENSTGUG00000010002	TAF5L	TATA-box binding protein associated factor 5 like
			42660429	42676677	ENSTGUG00000009990	URB2	URB2 ribosome biogenesis 2 homolog (S. cerevisiae)
4	4015762	4617080	4251353	4292632	ENSTGUG00000001856		
			4348602	4364213	ENSTGUG00000001867		
			4614729	4661710	ENSTGUG00000001899		
			4600262	4602158	ENSTGUG00000001896	FABP2	fatty acid binding protein 2
			4531033	4545798	ENSTGUG00000001887	MYOZ2	myozenin 2
			4303091	4327893	ENSTGUG00000001861	PRSS12	protease%2C serine 12
			4382505	4419798	ENSTGUG00000001870	SEC24D	SEC24 homolog D%2C COPII coat complex component
			4493352	4499188	ENSTGUG00000001883	SYNPO2	synaptopodin 2
			4569066	4579269	ENSTGUG00000001892	USP53	ubiquitin specific peptidase 53

chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
12	3781098	4373284	3781440	3784272	ENSTGUG00000004711		
			3899654	3899773	ENSTGUG00000004714		
			4151320	4155816	ENSTGUG00000004715	C3orf18	chromosome 3 open reading frame 18
			4241293	4242494	ENSTGUG00000004724	CISH	cytokine inducible SH2 containing protein
			4162341	4185084	ENSTGUG00000004718	HEMK1	HemK methyltransferase family member 1
			4256355	4306559	ENSTGUG00000004729	MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3
Z	4328891	5714217	4608897	4688550	ENSTGUG00000000293		
			4691266	4691616	ENSTGUG00000000304		
			4744048	4750326	ENSTGUG00000000307		
			4799420	4831602	ENSTGUG00000000336		
			5387296	5388408	ENSTGUG00000018154		
			5608473	5609492	ENSTGUG00000000382		
			4756484	4784165	ENSTGUG00000000310	ABCA1	ATP binding cassette subfamily A member 1
			5246780	5372137	ENSTGUG00000000365	ADAMTS19	ADAM metallopeptidase with thrombospondin type 1 motif 19
			5413331	5567837	ENSTGUG00000000380	CHSY3	chondroitin sulfate synthase 3
			5119773	5135157	ENSTGUG00000000362	ISOC1	isochorismatase domain containing 1
			4543512	4552432	ENSTGUG00000000287	LPL	lipoprotein lipase
			5054799	5090331	ENSTGUG00000000343	SLC27A6	solute carrier family 27 member 6
			4959219	5015496	ENSTGUG00000000337	SLC44A1	solute carrier family 44 member 1

Supplementary Table 3: Genes found in regions of decreasing ROH in SNPs across translocations. Some regions had decreased ROH frequency across both translocations. Regions here are defined as in Supplementary Table 1.

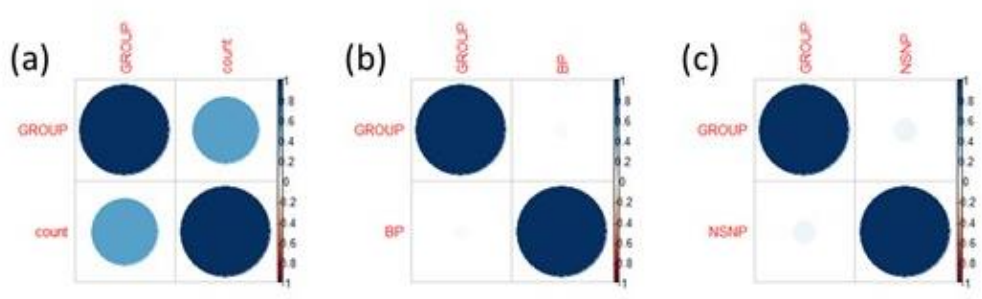
chr	region_start	region_end	gene_start	gene_end	ID	gene_name	description
2	119460939	120037840	119407264	119524999	ENSTGUG00000011226		
			119961880	119974901	ENSTGUG00000011234		
			119763882	119941542	ENSTGUG00000011233	NKAIN3	Na+/K+ transporting ATPase interacting 3
			119977440	119991331	ENSTGUG00000011241	TTPA	alpha tocopherol transfer protein
			120014103	120022247	ENSTGUG00000011243	YTHDF3	YTH N6-methyladenosine RNA binding protein 3

Linear Model Tests:

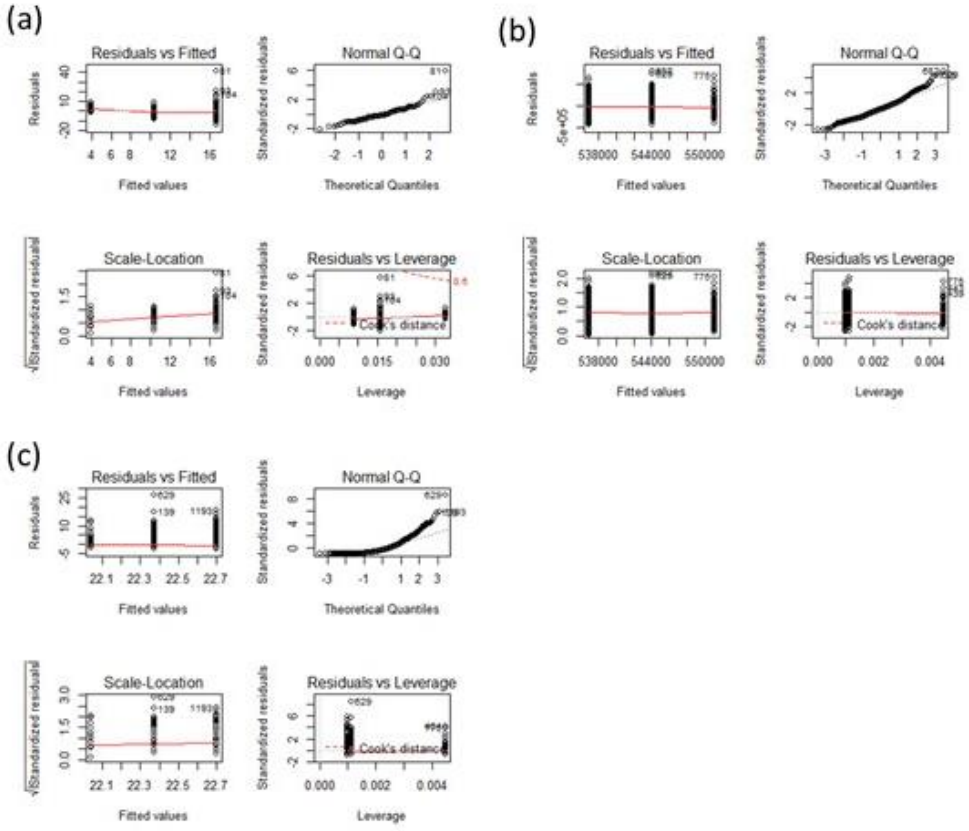
Supplementary Table 4: Results of linear model tests. Test mean of residuals, autocorrelation using the Durbin-Watson test, that the x variables and residuals are uncorrelated using Pearson's product-moment correlation, test for correlation between variables and homoscedasticity.

			Pearson's product-moment correlation								
Model	**mean of residuals**	**Durbin-Watson**	***p-value***	***corr***	***t***	***df***	***p-value***	**correlated variables**	**Supp Fig**	**homoscedasticity**	**Supp Fig**
ROH length ~ number of bottlenecks	-5.28E-12	1.7625	1.31E-06	0.9996553	1496.7	1545	2.20E-16	no	4a	no	5a
number of SNPs per ROH ~ number of bottlenecks*	-1.35E-16	2.0204	0.6464	0.9977421	583.93	1545	2.20E-16	no	4b	no	5b
ROHs per individual ~ number of bottlenecks*	-2.13E-16	2.2992	0.947	8.39E-01	1.73E+01	126	2.20E-16	yes	4c	no	5c
genome-wide FROH ~ number of bottlenecks*	1.57E-18	2.0812	0.6734	0.7544675	14.676	163	2.20E-16	yes	6a	no	7a
percentage SNPs in ROH ~ number of bottlenecks*	-1.60E-17	2.1137	0.7452	1.11E-16	1.42E-15	163	1	yes	6b	no	7b

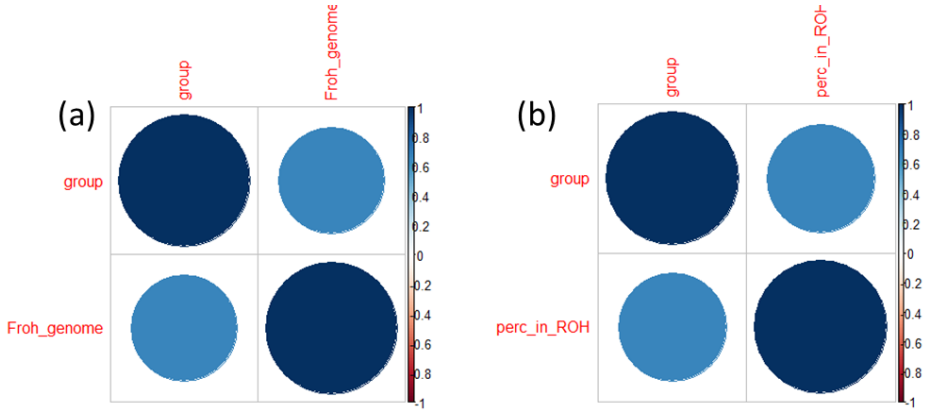
* models that were significant are highlighted with grey rows.



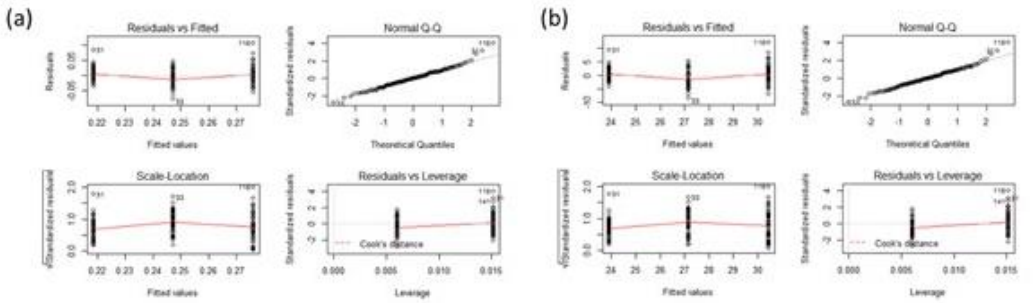
Supplementary Figure 5: Correlation between variables in linear models of runs of homozygosity. (a) group (0,1,2 translocations) and SNP count per individual, (b) group and length (BP) per run of homozygosity, (c) group and number of SNPs per run of homozygosity.



Supplementary Figure 6: Variable correlation and heteroscedasticity for linear models of runs of homozygosity (ROH). (a) number of ROH per individual per number of translocation bottlenecks, **(b)** length of ROHs per number of translocation bottlenecks, **(c)** number of SNPs in ROHs per number of translocation bottlenecks.



Supplementary Figure 7: Correlation between variables in linear models of **(a)** FROH and population (number of bottlenecks) **(b)** Percentage of individuals with a particular SNP in an ROH and population (number of bottlenecks).



Supplementary Figure 7: Variable correlation and heteroscedasticity for **(a)** FROH and population (number of bottlenecks) **(b)** Percentage of individuals with a particular SNP in an ROH and population (number of bottlenecks).

References

Aho, A.V., Kernighan, B.W., Weinberger, P.J., 1987. The AWK Programming Language. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Albrechtsen, A., Nielsen, F.C., Nielsen, R., 2010. Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. Mol. Biol. Evol. 27, 2534–2547. https://doi.org/10.1093/molbev/msq148

Allendorf, F., England, P., Luikart, G., Ritchie, P., Ryman, N., 2008. Genetic effects of harvest on wild animal populations. Trends Ecol. Evol. 23, 327–337. https://doi.org/10.1016/j.tree.2008.02.008

Allendorf, F.W., 2017. Genetics and the conservation of natural populations: allozymes to genomes. Mol. Ecol. 26, 420–430. https://doi.org/10.1111/mec.13948

Allendorf, F.W., 1986. Genetic drift and the loss of alleles versus heterozygosity. Zoo Biol. 5, 181–190. https://doi.org/10.1002/zoo.1430050212

Allendorf, F.W., 1983. Isolation, gene flow, and genetic differentiation among populations., in: C. Schonewald-Cox, S. Chambers, B. MacBryde, and L. Thomas Eds. Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations. Benjamin Cummings Publ. Co., Menlo Park, CA.

Allendorf, F.W., Luikart, G., 2007. Conservation and the genetics of populations. Blackwell Pub, Malden, MA.

Allendorf, F.W., Luikart, G., Aitken, S.N., 2013. Chapter 6: Small populations and genetic drift, in: Conservation and the Genetics of Populations. Blackwell Publishing Ltd.

Alley, M., Low, M., 2007. Oral fistulas in hihi (*Notiomystis cincta*): Are they genetic or acquired? N. Z. Vet. J. 55, 50–50. https://doi.org/10.1080/00480169.2007.36739

Alley, M.R., Castro, I., Hunter, J.E.B, 1999. Aspergillosis in hihi (*Notiomystis cincta*) on Mokoia Island. N. Z. Vet. J. 47, 88–91.

Al-Mamun, H.A., A Clark, S., Kwan, P., Gondro, C., 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 90. https://doi.org/10.1186/s12711-015-0169-6

Alsalem, A.B., Halees, A.S., Anazi, S., Alshamekh, S., Alkuraya, F.S., 2013. Autozygome Sequencing Expands the Horizon of Human Knockout Research and Provides Novel Insights into Human Phenotypic Variation. PLoS Genet. 9, e1004030. https://doi.org/10.1371/journal.pgen.1004030

AltschuP, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215, 403–410.

Andersen, K.H., Marty, L., Arlinghaus, R., 2018. Evolution of boldness and life history in response to selective harvesting. Can. J. Fish. Aquat. Sci. 75, 271–281. https://doi.org/10.1139/cjfas-2016-0350

Andrews, S., 2014. FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Armstrong, D., Seddon, P., 2008. Directions in reintroduction biology. Trends Ecol. Evol. 23, 20–25. https://doi.org/10.1016/j.tree.2007.10.003

Armstrong, D.P., Castro, I., Griffiths, R., 2007. Using adaptive management to determine requirements of re-introduced populations: the case of the New Zealand hihi. J. Appl. Ecol. 44, 953–962. https://doi.org/10.1111/j.1365-2664.2007.01320.x

Armstrong, D.P., Davidson, R.S., Dimond, W.J., Perrott, J.K., Castro, I., Ewen, J.G., Griffiths, R., Taylor, J., 2002. Population dynamics of reintroduced forest birds on

New Zealand islands. J. Biogeogr. 29, 609–621. https://doi.org/10.1046/j.1365-2699.2002.00710.x
Armstrong, D.P., Ewen, J.G., 2013. Consistency, continuity and creativity: long-term studies of population dynamics on Tiritiri Matangi Island. N. Z. J. Ecol. 37, 10.
Armstrong, D.P., Perrott, J.K., 2000. An Experiment Testing Whether Condition and Survival are Limited by Food Supply in a Reintroduced Hihi Population. Conserv. Biol. 14, 1171–1181.
Armstrong, D.P., Perrott, J.K., Castro, I., 2001. Estimating impacts of poison operations using mark-recapture analysis: hihi (*Notiomystis cincta*)on Mokoia Island. N. Z. J. Ecol. 25, 49–54.
Axelsson, E., 2005. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. 15, 120–125. https://doi.org/10.1101/gr.3021305
Backström, N., Sætre, G.-P., Ellegren, H., 2013. Inferring the demographic history of European Ficedula flycatcher populations. BMC Evol. Biol. 13, 2. https://doi.org/10.1186/1471-2148-13-2
Baillie, S.M., Muir, A.M., Scribner, K., Bentzen, P., Krueger, C.C., 2016. Loss of genetic diversity and reduction of genetic distance among lake trout *Salvelinus namaycush* ecomorphs, Lake Superior 1959 to 2013. J. Gt. Lakes Res. 42, 204–216. https://doi.org/10.1016/j.jglr.2016.02.001
Barati, A., Andrew, R.L., Gorrell, J.C., McDonald, P.G., 2018. Extra-pair paternity is not driven by inbreeding avoidance and does not affect provisioning rates in a cooperatively breeding bird, the noisy miner (*Manorina melanocephala*). Behav. Ecol. 29, 244–252. https://doi.org/10.1093/beheco/arx158
Benestan, L.M., Ferchaud, A.-L., Hohenlohe, P.A., Garner, B.A., Naylor, G.J.P., Baums, I.B., Schwartz, M.K., Kelley, J.L., Luikart, G., 2016. Conservation genomics of natural and managed populations: building a conceptual and practical framework. Mol. Ecol. 25, 2967–2977. https://doi.org/10.1111/mec.13647
Bérénos, C., Ellis, P.A., Pilkington, J.G., Pemberton, J.M., 2016. Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. Mol. Ecol. 25, 3152–3168. https://doi.org/10.1111/mec.13681
Bijlsma, R., Westerhof, M.D.D., Roekx, L.P., Pen, I., 2010. Dynamics of genetic rescue in inbred *Drosophila melanogaster* populations. Conserv. Genet. 11, 449–462. https://doi.org/10.1007/s10592-010-0058-z
Biscarini, F., Cozzi, P.A., Gaspa, G., Marras, G., 2018. detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.5.
Bixley, M.J., Ward, J.F., Brauning, R., Archer, J.A., Fisher, P.J., 2009. Building a deer SNP chip. Proc. Assoc. Adv. Anim. Breed. Genet. 18, 300–303.
Bjelland, D.W., Weigel, K.A., Vukasinovic, N., Nkrumah, J.D., 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy Sci. 96, 4697–4706. https://doi.org/10.3168/jds.2012-6435
Boessenkool, S., Taylor, S.S., Tepolt, C.K., Komdeur, J., Jamieson, I.G., 2007. Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges. Conserv. Genet. 8, 705–714. https://doi.org/10.1007/s10592-006-9219-5

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bosse, M., Megens, H.-J., Madsen, O., Crooijmans, R.P.M.A., Ryder, O.A., Austerlitz, F., Groenen, M.A.M., de Cara, M.A.R., 2015. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Res. 25, 970–981. https://doi.org/10.1101/gr.187039.114

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L.A.F., Schook, L.B., Crooijmans, R.P.M.A., Groenen, M.A.M., 2012a. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. PLoS Genet. 8, e1003100. https://doi.org/10.1371/journal.pgen.1003100

Bosse, M., Megens, H.-J., Madsen, O., Paudel, Y., Frantz, L.A.F., Schook, L.B., Crooijmans, R.P.M.A., Groenen, M.A.M., 2012b. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. PLoS Genet. 8, e1003100. https://doi.org/10.1371/journal.pgen.1003100

Brekke, P., Bennett, P.M., Santure, A.W., Ewen, J.G., 2011. High genetic diversity in the remnant island population of hihi and the genetic consequences of re-introduction. Mol. Ecol. 20, 29–45. https://doi.org/10.1111/j.1365-294X.2010.04923.x

Brekke, P., Bennett, P.M., Wang, J., Pettorelli, N., Ewen, J.G., 2010. Sensitive males: inbreeding depression in an endangered bird. Proc. R. Soc. B Biol. Sci. 277, 3677–3684. https://doi.org/10.1098/rspb.2010.1144

Brekke, P., Cassey, P., Ariani, C., Ewen, J.G., 2013. Evolution of extreme-mating behaviour: patterns of extrapair paternity in a species with forced extrapair copulation. Behav. Ecol. Sociobiol. 67, 963–972. https://doi.org/10.1007/s00265-013-1522-9

Brekke, P., Dawson, D.A., Horsburgh, G.J., Ewen, J.G., 2009. Characterization of microsatellite loci In The hihi *Notiomystis cincta* (Notiomystidae, Aves). Mol. Ecol. Resour. 9, 1255–1258. https://doi.org/10.1111/j.1755-0998.2009.02626.x

Brekke, P., Wang, J., Bennett, P.M., Cassey, P., Dawson, D.A., Horsburgh, G.J., Ewen, J.G., 2012. Postcopulatory mechanisms of inbreeding avoidance in the island endemic hihi (*Notiomystis cincta*). Behav. Ecol. 23, 278–284. https://doi.org/10.1093/beheco/arr183

Broman, K.W., Weber, J.L., 1999a. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. Am. J. Hum. Genet. 65, 1493–1500. https://doi.org/10.1086/302661

Broman, K.W., Weber, J.L., 1999b. Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. Am. J. Hum. Genet. 65, 1493–1500. https://doi.org/10.1086/302661

Browning, S.R., Browning, B.L., 2015. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. Am. J. Hum. Genet. 97, 404–418. https://doi.org/10.1016/j.ajhg.2015.07.012

Browning, S.R., Browning, B.L., 2012. Identity by Descent Between Distant Relatives: Detection and Applications. Annu. Rev. Genet. 46, 617–633. https://doi.org/10.1146/annurev-genet-110711-155534

Caballero, A., Garcia-Dorado, A., 2013. Allelic Diversity and Its Implications for the Rate of Adaptation. Genetics 195, 1373–1384. https://doi.org/10.1534/genetics.113.158410

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Campbell, N.R., Kamphaus, C., Murdoch, K., Narum, S.R., 2017. Patterns of genomic variation in Coho salmon following reintroduction to the interior Columbia River. Ecol. Evol. 7, 10350–10360. https://doi.org/10.1002/ece3.3492

Castro, I., Brunton, D.H., Mason, K.M., Ebert, B., Griffiths, R., 2003. Life history traits and food supplementation affect productivity in a translocated population of the endangered Hihi (Stitchbird, *Notiomystis cincta*). Biol. Conserv. 114, 271–280. https://doi.org/10.1016/S0006-3207(03)00046-6

Castro, I., Mason, K.M., Armstrong, D.P., Lambert, D.M., 2004. Effect of Extra-Pair Paternity on Effective Population Size in a Reintroduced Population of the Endangered Hihi, and Potential for Behavioural Management. Conserv. Genet. 5, 381–393. https://doi.org/10.1023/B:COGE.0000031146.51681.b0

Castro, I., Minot, E.O., Fordham, R.A., Birkhead, T.R., 1996. Polygynandry, face-to-face copulation and sperm competition in the Hihi *Notiomystis cincta* (Aves: Meliphagidae). Ibis 138, 765–771. https://doi.org/10.1111/j.1474-919X.1996.tb08834.x

Castro, I., Robertson, A.W., 1997. Honeyeaters and the New Zealand forest flora: the utilisation and profitability of small flowers. N. Z. J. Ecol. 21, 11.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140. https://doi.org/10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. G3amp58 GenesGenomesGenetics 1, 171–182. https://doi.org/10.1534/g3.111.000240

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F., 2018. Runs of homozygosity: windows into population history and trait architecture. Nat. Rev. Genet. 19, 220–234. https://doi.org/10.1038/nrg.2017.109

Chakraborty, R., Nei, M., 1977. Bottleneck Effects on Average Heterozygosity and Genetic Distance with the Stepwise Mutation Model. Evolution 31, 347–356.

Chapman, J.R., Nakagawa, S., Coltman, D.W., Slate, J., Sheldon, B.C., 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. Mol. Ecol. 18, 2746–2765. https://doi.org/10.1111/j.1365-294X.2009.04247.x

Chauvenet, A.L.M., Ewen, J.G., Armstrong, D.P., Coulson, T., Blackburn, T.M., Adams, L., Walker, L.K., Pettorelli, N., 2012. Does supplemental feeding affect the viability of translocated populations? The example of the hihi: Management affects translocated populations' viability. Anim. Conserv. 15, 337–350. https://doi.org/10.1111/j.1469-1795.2012.00522.x

Chen, N., Cosgrove, E.J., Bowman, R., Fitzpatrick, J.W., Clark, A.G., 2016. Genomic Consequences of Population Decline in the Endangered Florida Scrub-Jay. Curr. Biol. 26, 2974–2979. https://doi.org/10.1016/j.cub.2016.08.062

Chiyo, P.I., Obanda, V., Korir, D.K., 2015. Illegal tusk harvest and the decline of tusk size in the African elephant. Ecol. Evol. 5, 5216–5229. https://doi.org/10.1002/ece3.1769

Christiansen, T., Foy, B.D., Wall, L., 2012. Programming Perl, 4th ed. ed. O'Reilly, Beijing ; Sebastopol.

Christofidou, P., Nelson, C.P., Nikpay, M., Qu, L., Li, M., Loley, C., Debiec, R., Braund, P.S., Denniff, M., Charchar, F.J., Arjo, A.R., Trégouët, D.-A., Goodall, A.H., Cambien, F., Ouwehand, W.H., Roberts, R., Schunkert, H., Hengstenberg, C.,

Reilly, M.P., Erdmann, J., McPherson, R., König, I.R., Thompson, J.R., Samani, N.J., Tomaszewski, M., 2015. Runs of Homozygosity: Association with Coronary Artery Disease and Gene Expression in Monocytes and Macrophages. Am. J. Hum. Genet. 97, 228–237. https://doi.org/10.1016/j.ajhg.2015.06.001

Cohas, A., Yoccoz, N.G., Allainé, D., 2007. Extra-pair paternity in alpine marmots, *Marmota marmota*: genetic quality and genetic diversity effects. Behav. Ecol. Sociobiol. 61, 1081–1092. https://doi.org/10.1007/s00265-006-0341-7

Coltman, D.W., 2008. Molecular ecological approaches to studying the evolutionary impact of selective harvesting in wildlife. Mol. Ecol. 17, 221–235. https://doi.org/10.1111/j.1365-294X.2007.03414.x

Cork, S.C., Alley, M.R., Johnstone, A.C., Stockdale, P.H.G., 1999. Aspergillosis and other causes of mortality in the stitchbird in New Zealand. J. Wildl. Dis. 35, 481–486. https://doi.org/10.7589/0090-3558-35.3.481

Corlett, R.T., 2017. A Bigger Toolbox: Biotechnology in Biodiversity Conservation. Trends Biotechnol. 35, 55–65. https://doi.org/10.1016/j.tibtech.2016.06.009

Cornell, S.J., Tregenza, T., 2007. A new theory for the evolution of polyandry as a means of inbreeding avoidance. Proc. R. Soc. B Biol. Sci. 274, 2873–2879. https://doi.org/10.1098/rspb.2007.0926

Crestanello, B., Pecchioli, E., Vernesi, C., Mona, S., Martínková, N., Janiga, M., Hauffe, H.C., Bertorelle, G., 2009. The Genetic Impact of Translocations and Habitat Fragmentation in Chamois (Rupicapra) spp. J. Hered. 100, 691–708. https://doi.org/10.1093/jhered/esp053

Cruse-Sanders, J.M., Hamrick, J.L., 2004. Genetic diversity in harvested and protected populations of wild American ginseng, *Panax quinquefolius* L. (Araliaceae). Am. J. Bot. 91, 540–548. https://doi.org/10.3732/ajb.91.4.540

Curik, I., Ferenčaković, M., Sölkner, J., 2017. Genomic dissection of inbreeding depression: a gate to new opportunities. Rev. Bras. Zootec. 46, 773–782. https://doi.org/10.1590/s1806-92902017000900010

Curik, I., Ferenčaković, M., Sölkner, J., 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. Livest. Sci. 166, 26–34. https://doi.org/10.1016/j.livsci.2014.05.034

da Silva, V.H., Laine, V.N., Bosse, M., Oers, K. van, Dibbits, B., Visser, M.E., M. A. Crooijmans, R.P., Groenen, M.A.M., 2018. CNVs are associated with genomic architecture in a songbird. BMC Genomics 19, 195. https://doi.org/10.1186/s12864-018-4577-1

Dakin, E.E., Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. Heredity 93, 504–509. https://doi.org/10.1038/sj.hdy.6800545

Davey, J.W., Blaxter, M.L., 2010. RADSeq: next-generation population genetics. Brief. Funct. Genomics 9, 416–423. https://doi.org/10.1093/bfgp/elq031

Dawson, D.A., Ball, A.D., Spurgin, L.G., Martín-Gálvez, D., Stewart, I.R.K., Horsburgh, G.J., Potter, J., Molina-Morales, M., Bicknell, A.W.J., Preston, S.A.J., Ekblom, R., Slate, J., Burke, T., 2013. High-utility conserved avian microsatellite markers enable parentage and population studies across a wide range of species. BMC Genomics 14, 176. https://doi.org/10.1186/1471-2164-14-176

de Cara, M.Á.R., Villanueva, B., Toro, M.Á., Fernández, J., 2013. Using genomic tools to maintain diversity and fitness in conservation programmes. Mol. Ecol. 22, 6091–6099. https://doi.org/10.1111/mec.12560

de Villemereuil, P., Rutschmann, A., Lee, K.D., Ewen, J.G., Brekke, P., Santure, A.W., 2019a. Little adaptive potential of a threatened passerine. Curr. Biol.

de Villemereuil, P., Rutschmann, A., Ewen, J.G., Santure, A.W., Brekke, P., 2019b. Can threatened species adapt in a restored habitat? No expected evolutionary response in lay date for the New Zealand hihi. Evol. Appl. https://doi.org/10.1111/eva.12727

Decker, J.E., Pires, J.C., Conant, G.C., McKay, S.D., Heaton, M.P., Chen, K., Cooper, A., Vilkki, J., Seabury, C.M., Caetano, A.R., Johnson, G.S., Brenneman, R.A., Hanotte, O., Eggert, L.S., Wiener, P., Kim, J.-J., Kim, K.S., Sonstegard, T.S., Van Tassell, C.P., Neibergs, H.L., McEwan, J.C., Brauning, R., Coutinho, L.L., Babar, M.E., Wilson, G.A., McClure, M.C., Rolf, M.M., Kim, J., Schnabel, R.D., Taylor, J.F., 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. Proc. Natl. Acad. Sci. 106, 18644–18649. https://doi.org/10.1073/pnas.0904691106

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. https://doi.org/10.1038/ng.806

Doerr, L., Richardson, K., Ewen, J., Armstrong, D., Armstrong, D., 2017. Effect of supplementary feeding on reproductive success of hihi (stitchbird, *Notiomystis cincta*) at a mature forest reintroduction site. N. Z. J. Ecol. 41, 34-40. https://doi.org/10.20417/nzjecol.41.15

Drag, L., Cizek, L., 2015. Successful reintroduction of an endangered veteran tree specialist: conservation and genetics of the Great Capricorn beetle (*Cerambyx cerdo*). Conserv. Genet. 16, 267–276. https://doi.org/10.1007/s10592-014-0656-2

Driskell, A., Christidis, L., Gill, B.J., Boles, W.E., Barker, F.K., Longmore, N.W., 2007. A new endemic family of New Zealand passerine birds: adding heat to a biodiversity hotspot. Aust. J. Zool. 55, 73. https://doi.org/10.1071/ZO07007

Duncan, R.P., Blackburn, T.M., 2004. Extinction and endemism in the New Zealand avifauna. Glob. Ecol. Biogeogr. 13, 509–517. https://doi.org/10.1111/j.1466-822X.2004.00132.x

Elgvin, T.O., Trier, C.N., Tørresen, O.K., Hagen, I.J., Lien, S., Nederbragt, A.J., Ravinet, M., Jensen, H., Sætre, G.-P., 2017. The genomic mosaicism of hybrid speciation. Sci. Adv. 3, e1602996. https://doi.org/10.1126/sciadv.1602996

Ellegren, H., Hartman, G., Johansson, M., Andersson, L., 1993. Major histocompatibility complex monomorphism and low levels of DNA fingerprinting variability in a reintroduced and rapidly expanding population of beavers. Proc. Natl. Acad. Sci. 90, 8150–8153. https://doi.org/10.1073/pnas.90.17.8150

Ellstrand, N.C., Elam, D.R., 1993. Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. Annu. Rev. Ecol. Syst. 24, 217–242. https://doi.org/10.1146/annurev.es.24.110193.001245

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6, e19379. https://doi.org/10.1371/journal.pone.0019379

Enciso-Mora, V., Hosking, F.J., Houlston, R.S., 2010. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. Eur. J. Hum. Genet. 18, 909–914. https://doi.org/10.1038/ejhg.2010.53

Engels, B., 2014. HWxtest: Exact Tests for Hardy-Weinberg Proportions.

Ewen, J.G. (Ed.), 2012. Reintroduction biology: integrating science and management, Conservation science and practice series. Wiley-Blackwell, Chichester.
Ewen, J.G., Armstrong, D.P., Ebert, B., Hansen, L.H., 2004. Extra-pair copulation and paternity defense in the hihi (or stitchbird). N. Z. J. Ecol. 28, 8.
Ewen, John G, Parker, K.A., Richardson, K., Armstrong, D., Smuts-Kennedy, C., 2011. Translocation of hihi *Notiomystis cincta* to Maungatautari, a mainland reserve protected by a predator-exclusion fence, Waikato, New Zealand 8, 58-65.
Ewen, John G., Thorogood, R., Armstrong, D.P., 2011. Demographic consequences of adult sex ratio in a reintroduced hihi population: Demography and adult sex ratio. J. Anim. Ecol. 80, 448–455. https://doi.org/10.1111/j.1365-2656.2010.01774.x
Ewen, J.G., Thorogood, R., Nicol, C., Armstrong, D.P., Alley, M., 2007. *Salmonella* Typhimurium in Hihi, New Zealand. Emerg. Infect. Dis. 13, 788–790. https://doi.org/10.3201/eid1305.060824
Falconer, D.S., Mackay, T.F.C., 2009. Introduction to quantitative genetics, 4. ed., [16. print.]. ed. Pearson, Prentice Hall, Harlow.
Falconer, D.S., Mackay, T.F.C., 1996. Introduction to Quantitative Genetics (4th Edition), 4th ed. Pearson.
Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genet. Sel. Evol. 45, 42. https://doi.org/10.1186/1297-9686-45-42
Fitzsimmons, N.N., Buskirk, S.W., Smith, M.H., 1997. Genetic Changes in Reintroduced Rocky Mountain Bighorn Sheep Populations. J. Wildl. Manag. 61, 863. https://doi.org/10.2307/3802195
Foote, A.D., Morin, P.A., 2016. Genome-wide SNP data suggest complex ancestry of sympatric North Pacific killer whale ecotypes. Heredity 117, 316–325. https://doi.org/10.1038/hdy.2016.54
Forstmeier, W., Schielzeth, H., Mueller, J.C., Ellegren, H., Kempenaers, B., 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. Mol. Ecol. 21, 3237–3249. https://doi.org/10.1111/j.1365-294X.2012.05593.x
Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F.S., Sargolzaei, M., 2018. Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics 19, 98. https://doi.org/10.1186/s12864-018-4453-z
Frankham, R., 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. Mol. Ecol. 24, 2610–2618. https://doi.org/10.1111/mec.13139
Frankham, R., Ballou, J.D., Briscoe, D.A., 2010. Introduction to conservation genetics, 2nd ed. ed. Cambridge University Press, Cambridge, UK ; New York.
Franklin, I.R., 1980. Evolutionary change in small populations, in: Eds. M. Soule, and B. Wilcox Conservation Biology: An Evolutionary- Ecological Perspective. Sinauer Associates Sunderland, MA, pp. 135–149.
Franks, V.R., Thorogood, R., 2018. Older and wiser? Age differences in foraging and learning by an endangered passerine. Behav. Processes 148, 1–9. https://doi.org/10.1016/j.beproc.2017.12.009
Frichot, E., François, O., 2015. LEA : An R package for landscape and ecological association studies. Methods Ecol. Evol. 6, 925–929. https://doi.org/10.1111/2041-210X.12382

Funk, W.C., McKay, J.K., Hohenlohe, P.A., Allendorf, F.W., 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27, 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Galbraith, M., Cooper, H., 2013. Tiritiri Matangi – an overview of 25 years of ecological restoration. N. Z. J. Ecol. 37, 258–260.

Galla, S.J., Buckley, T.R., Elshire, R., Hale, M.L., Knapp, M., McCallum, J., Moraga, R., Santure, A.W., Wilcox, P., Steeves, T.E., 2016. Building strong relationships between conservation genetics and primary industry leads to mutually beneficial genomic advances. Mol. Ecol. 25, 5267–5281. https://doi.org/10.1111/mec.13837

Garcia-Dorado, A., 2012. Understanding and Predicting the Fitness Decline of Shrunk Populations: Inbreeding, Purging, Mutation, and Standard Selection. Genetics 190, 1461–1476. https://doi.org/10.1534/genetics.111.135541

García-Navas, V., Bonnet, T., Waldvogel, D., Wandeler, P., Camenisch, G., Postma, E., 2015. Gene flow counteracts the effect of drift in a Swiss population of snow voles fluctuating in size. Biol. Conserv. 191, 168–177. https://doi.org/10.1016/j.biocon.2015.06.021

Garner, B.A., Hand, B.K., Amish, S.J., Bernatchez, L., Foster, J.T., Miller, K.M., Morin, P.A., Narum, S.R., O'Brien, S.J., Roffler, G., Templin, W.D., Sunnucks, P., Strait, J., Warheit, K.I., Seamons, T.R., Wenburg, J., Olsen, J., Luikart, G., 2016. Genomics in Conservation: Case Studies and Bridging the Gap between Data and Application. Trends Ecol. Evol. 31, 81–83. https://doi.org/10.1016/j.tree.2015.10.009

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.-L., 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10, 550. https://doi.org/10.1186/1471-2164-10-550

Ghani, M., Reitz, C., Cheng, R., Vardarajan, B.N., Jun, G., Sato, C., Naj, A., Rajbhandary, R., Wang, L.-S., Valladares, O., Lin, C.-F., Larson, E.B., Graff-Radford, N.R., Evans, D., De Jager, P.L., Crane, P.K., Buxbaum, J.D., Murrell, J.R., Raj, T., Ertekin-Taner, N., Logue, M., Baldwin, C.T., Green, R.C., Barnes, L.L., Cantwell, L.B., Fallin, M.D., Go, R.C.P., Griffith, P.A., Obisesan, T.O., Manly, J.J., Lunetta, K.L., Kamboh, M.I., Lopez, O.L., Bennett, D.A., Hendrie, H., Hall, K.S., Goate, A.M., Byrd, G.S., Kukull, W.A., Foroud, T.M., Haines, J.L., Farrer, L.A., Pericak-Vance, M.A., Lee, J.H., Schellenberg, G.D., St. George-Hyslop, P., Mayeux, R., Rogaeva, E., 2015. Association of Long Runs of Homozygosity With Alzheimer Disease Among African American Individuals. JAMA Neurol. 72, 1313. https://doi.org/10.1001/jamaneurol.2015.1700

Gibson, J., Morton, N.E., Collins, A., 2006. Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet. 15, 789–795. https://doi.org/10.1093/hmg/ddi493

Gill, P., d'Aloja, E., Dupuy, B., Eriksen, B., Jangblad, M., Johnsson, V., Kloosterman, A.D., Kratzer, A., Lareu, M.V., Mevag, B., Morling, N., Phillips, C., Pfitzinger, H., Rand, S., Sabatier, M., Scheithauer, R., Schmitter, H., Schneider, P., Skitsa, I., Vide, M.C., 1998. Report of the European DNA profiling group (EDNAP)-an investigation of the hypervariable STR loci ACTBP2, APOAI1 and D11S554 and the compound loci D12S391 and D1S1656. Forensic Sci. Int. 98, 193–200. https://doi.org/10.1016/S0379-0738(98)00155-8

Gil-Sánchez, J.M., McCain, E.B., 2011. Former range and decline of the Iberian lynx (Lynx pardinus) reconstructed using verified records. J. Mammal. 92, 1081–1090. https://doi.org/10.1644/10-MAMM-A-381.1

Graham, M., Veitch, D., Aguilar, G., Galbraith, M., 2013. Monitoring terrestrial bird populations on Tiritiri Matangi Island, Hauraki Gulf, New Zealand, 1987–2010. N. Z. J. Ecol. 37, 11.

Greenbaum, G., Templeton, A.R., Zarmi, Y., Bar-David, S., 2014. Allelic Richness following Population Founding Events – A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. PLoS ONE 9, e115203. https://doi.org/10.1371/journal.pone.0115203

Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D.M., Kim, H., Frantz, L.A.F., Caccamo, M., Ahn, H., Aken, B.L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C.W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R.C., Clee, C., Crooijmans, R.P.M.A., Dawson, H.D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G.J., Fowler, K.E., Fredholm, M., Fritz, E., Gilbert, J.G.R., Giuffra, E., Gorodkin, J., Griffin, D.K., Harrow, J.L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S.J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H.A., Li, Y., Liu, W., Loveland, J.E., Lu, Y., Lunney, J.K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M.P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J.M., Rodriguez-Zas, S., Rohrer, G.A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J.G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B.R., Sperber, G., Stadler, P., Sweedler, J.V., Tafer, H., Thomsen, B., Wali, R., Wang, Jian, Wang, Jun, White, S., Xu, X., Yerle, M., Zhang, G., Zhang, Jianguo, Zhang, Jie, Zhao, S., Rogers, J., Churcher, C., Schook, L.B., 2012. Analyses of pig genomes provide insight into porcine demography and evolution. Nature 491, 393–398. https://doi.org/10.1038/nature11622

Grossen, C., Biebach, I., Angelone-Alasaad, S., Keller, L.F., Croll, D., 2018. Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. Evol. Appl. 11, 123–139. https://doi.org/10.1111/eva.12490

Grueber, C.E., Jamieson, I.G., 2011. Low genetic diversity and small population size of Takahe *Porphyrio hochstetteri* on European arrival in New Zealand: Historical genetic diversity of Takahe. Ibis 153, 384–394. https://doi.org/10.1111/j.1474-919X.2011.01110.x

Grueber, C.E., Wallis, G.P., Jamieson, I.G., 2008. Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. Mol. Ecol. 17, 3978–3984. https://doi.org/10.1111/j.1365-294X.2008.03910.x

Gudiseva, H.V., Hansen, M., Gutierrez, L., Collins, D.W., He, J., Verkuil, L.D., Danford, I.D., Sagaser, A., Bowman, A.S., Salowe, R., Sankar, P.S., Miller-Ellis, E., Lehman, A., O'Brien, J.M., 2016. Saliva DNA quality and genotyping efficiency in a predominantly elderly population. BMC Med. Genomics 9, 17. https://doi.org/10.1186/s12920-016-0172-y

Gustafson, K.D., Vickers, T.W., Boyce, W.M., Ernest, H.B., 2017. A single migrant enhances the genetic diversity of an inbred puma population. R. Soc. Open Sci. 4, 170115. https://doi.org/10.1098/rsos.170115

Hadly, E.A., van Tuinen, M., Chan, Y., Heiman, K., 2003. Ancient DNA evidence of prolonged population persistence with negligible genetic diversity in an endemic Tuco-tuco (*Ctenomys sociabilis*). J. Mammal. 84, 403–417. https://doi.org/10.1644/1545-1542(2003)084<0403:ADEOPP>2.0.CO;2

Hagen, I.J., Billing, A.M., Rønning, B., Pedersen, S.A., Pärn, H., Slate, J., Jensen, H., 2013a. The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). Mol. Ecol. Resour. 13, 429–439. https://doi.org/10.1111/1755-0998.12088

Hagen, I.J., Billing, A.M., Rønning, B., Pedersen, S.A., Pärn, H., Slate, J., Jensen, H., 2013b. The easy road to genome-wide medium density SNP screening in a non-model species: development and application of a 10 K SNP-chip for the house sparrow (*Passer domesticus*). Mol. Ecol. Resour. 13, 429–439. https://doi.org/10.1111/1755-0998.12088

Hansson, B., Ljungqvist, M., Dawson, D.A., Mueller, J.C., Olano-Marin, J., Ellegren, H., Nilsson, J.-Å., 2010. Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). Heredity 104, 67–78. https://doi.org/10.1038/hdy.2009.107

Harding, L.E., Heffelfinger, J., Paetkau, D., Rubin, E., Dolphin, J., Aoude, A., 2016. Genetic management and setting recovery goals for Mexican wolves (*Canis lupus baileyi*) in the wild. Biol. Conserv. 203, 151–159. https://doi.org/10.1016/j.biocon.2016.09.018

Harris, R.B., Wall, W.A., Allendorf, F.W., 2002. Genetic Consequences of Hunting: What Do We Know and What Should We Do? Wildl. Soc. Bull. 30, 634–643.

Harrisson, K.A., Pavlova, A., Telonis-Scott, M., Sunnucks, P., 2014. Using genomics to characterize evolutionary potential for conservation of wild populations. Evol. Appl. 7, 1008–1025. https://doi.org/10.1111/eva.12149

Hartl, D.L., Clark, Andrew G., 1989. Principles of population genetics / Daniel L. Hartl, Andrew G. Clark, 2nd ed. ed. Sinauer Associates Sunderland, Mass.

Harvey, M.G., Bonter, D.N., Stenzler, L.M., Lovette, I.J., 2006. A comparison of plucked feathers versus blood samples as DNA sources for molecular sexing. J. Field Ornithol. 77, 136–140. https://doi.org/10.1111/j.1557-9263.2006.00033.x

Heber, S., Varsani, A., Kuhn, S., Girg, A., Kempenaers, B., Briskie, J., 2012. The genetic rescue of two bottlenecked South Island robin populations using translocations of inbred donors. Proc. R. Soc. B Biol. Sci. 280, 2228. https://doi.org/10.1098/rspb.2012.2228

Hedrick, P.W., Garcia-Dorado, A., 2016. Understanding Inbreeding Depression, Purging, and Genetic Rescue. Trends Ecol. Evol. 31, 940–952. https://doi.org/10.1016/j.tree.2016.09.005

Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E., Diopere, E., Carvalho, G.R., Nielsen, E.E., 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges: ANALYTICAL APPROACHES.pdf. Mol. Ecol. Resour. 11, 123–136. https://doi.org/10.1111/j.1755-0998.2010.02943.x

Hernández, S., Daley, R., Walker, T., Braccini, M., Varela, A., Francis, M.P., Ritchie, P.A., 2015. Demographic history and the South Pacific dispersal barrier for school shark (*Galeorhinus galeus*) inferred by mitochondrial DNA and microsatellite DNA mark. Fish. Res. 167, 132–142. https://doi.org/10.1016/j.fishres.2015.02.010

Hindar, K., Tufto, J., Sattem, L., Balstad, T., 2004. Conservation of genetic variation in harvested salmon populations. ICES J. Mar. Sci. 61, 1389–1397. https://doi.org/10.1016/j.icesjms.2004.08.011

Hodel, R.G.J., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A., Douglas, N.A., Germain-Aubrey, C.C., Chen, S., Soltis, D.E., Soltis, P.S., 2016. The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. Appl. Plant Sci. 4, 1600025. https://doi.org/10.3732/apps.1600025

Hoffman, J.I., Simpson, F., David, P., Rijks, J.M., Kuiken, T., Thorne, M.A.S., Lacy, R.C., Dasmahapatra, K.K., 2014. High-throughput sequencing reveals inbreeding depression in a natural population. Proc. Natl. Acad. Sci. 111, 3775–3780. https://doi.org/10.1073/pnas.1318945111

Höglund, J., 2009. Evolutionary Conservation Genetics. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199214211.001.0001

Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., Bishop, S.C., Archibald, A.L., Bron, J.E., Penman, D.J., Davassi, A., Brew, F., Tinch, A.E., Gharbi, K., Hamilton, A., 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics 15, 90. https://doi.org/10.1186/1471-2164-15-90

Howrigan, D.P., Simonson, M.A., Keller, M.C., 2011. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. BMC Genomics 12, 460. https://doi.org/10.1186/1471-2164-12-460

Human Genome Sequencing Consortium, I., 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945. https://doi.org/10.1038/nature03001

Husby, A., Kawakami, T., Ronnegard, L., Smeds, L., Ellegren, H., Qvarnstrom, A., 2015. Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. Proc. R. Soc. B Biol. Sci. 282, 156. https://doi.org/10.1098/rspb.2015.0156

IUCN/SSC, 2013. Guidelines for Reintroductions and Other Conservation Translocations, Version 1.0.

Jahner, J.P., Matocq, M.D., Malaney, J.L., Cox, M., Wolff, P., Gritts, M.A., Parchman, T.L., 2019. The genetic legacy of 50 years of desert bighorn sheep translocations. Evol. Appl. 12, 198–213. https://doi.org/10.1111/eva.12708

Jamieson, I.G., 2015. Significance of population genetics for managing small natural and reintroduced populations in New Zealand. N. Z. J. Ecol. 39, 18.

Jamieson, I.G., 2011. Founder Effects, Inbreeding, and Loss of Genetic Diversity in Four Avian Reintroduction Programs: Inbreeding in Reintroduction Programs. Conserv. Biol. 25, 115–123. https://doi.org/10.1111/j.1523-1739.2010.01574.x

Jamieson, I.G., 2009. Loss of genetic diversity and inbreeding in New Zealand's threatened bird species, Science for Conservation. New Zealand Department of Conservation.

Jangjoo, M., Matter, S.F., Roland, J., Keyghobadi, N., 2016. Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. Proc. Natl. Acad. Sci. 113, 10914–10919. https://doi.org/10.1073/pnas.1600865113

Jehle, R., Arntzen, J.W., 2002. Review: Microsatellite marker in amphibian conservation genetics. Herpetol. J. 12, 1–9.

Johnston, S.E., Bérénos, C., Slate, J., Pemberton, J.M., 2016. Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis aries*). Genetics 203, 583–598. https://doi.org/10.1534/genetics.115.185553

Johnston, S.E., Lindqvist, M., Niemelä, E., Orell, P., Erkinaro, J., Kent, M.P., Lien, S., Vähä, J.-P., Vasemägi, A., Primmer, C.R., 2013. Fish scales and SNP chips: SNP genotyping and allele frequency estimation in individual and pooled DNA from historical samples of Atlantic salmon (*Salmo salar*). BMC Genomics 14, 439. https://doi.org/10.1186/1471-2164-14-439

Johnston, S.E., McEWAN, J.C., Pickering, N.K., Kijas, J.W., Beraldi, D., Pilkington, J.G., Pemberton, J.M., Slate, J., 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population: gwas of sexual weaponry in a wild population. Mol. Ecol. 20, 2555–2566. https://doi.org/10.1111/j.1365-294X.2011.05076.x

Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 11, 94. https://doi.org/10.1186/1471-2156-11-94

Joshi, P.K., Esko, T., Mattsson, H., Eklund, N., Gandin, I., Nutile, T., Jackson, A.U., Schurmann, C., Smith, A.V., Zhang, W., Okada, Y., Stančáková, A., Faul, J.D., Zhao, W., Bartz, T.M., Concas, M.P., Franceschini, N., Enroth, S., Vitart, V., Trompet, S., Guo, X., Chasman, D.I., O'Connel, J.R., Corre, T., Nongmaithem, S.S., Chen, Y., Mangino, M., Ruggiero, D., Traglia, M., Farmaki, A.-E., Kacprowski, T., Bjonnes, A., van der Spek, A., Wu, Y., Giri, A.K., Yanek, L.R., Wang, L., Hofer, E., Rietveld, C.A., McLeod, O., Cornelis, M.C., Pattaro, C., Verweij, N., Baumbach, C., Abdellaoui, A., Warren, H.R., Vuckovic, D., Mei, H., Bouchard, C., Perry, J.R.B., Cappellani, S., Mirza, S.S., Benton, M.C., Broeckel, U., Medland, S.E., Lind, P.A., Malerba, G., Drong, A., Yengo, L., Bielak, L.F., Zhi, D., van der Most, P.J., Shriner, D., Mägi, R., Hemani, G., Karaderi, T., Wang, Z., Liu, T., Demuth, I., Zhao, J.H., Meng, W., Lataniotis, L., van der Laan, S.W., Bradfield, J.P., Wood, A.R., Bonnefond, A., Ahluwalia, T.S., Hall, L.M., Salvi, E., Yazar, S., Carstensen, L., de Haan, H.G., Abney, M., Afzal, U., Allison, M.A., Amin, N., Asselbergs, F.W., Bakker, S.J.L., Barr, R.G., Baumeister, S.E., Benjamin, D.J., Bergmann, S., Boerwinkle, E., Bottinger, E.P., Campbell, A., Chakravarti, A., Chan, Y., Chanock, S.J., Chen, C., Chen, Y.-D.I., Collins, F.S., Connell, J., Correa, A., Cupples, L.A., Smith, G.D., Davies, G., Dörr, M., Ehret, G., Ellis, S.B., Feenstra, B., Feitosa, M.F., Ford, I., Fox, C.S., Frayling, T.M., Friedrich, N., Geller, F., Scotland, G., Gillham-Nasenya, I., Gottesman, O., Graff, M., Grodstein, F., Gu, C., Haley, C., Hammond, C.J., Harris, S.E., Harris, T.B., Hastie, N.D., Heard-Costa, N.L., Heikkilä, K., Hocking, L.J., Homuth, G., Hottenga, J.-J., Huang, J., Huffman, J.E., Hysi, P.G., Ikram, M.A., Ingelsson, E., Joensuu, A., Johansson, Å., Jousilahti, P., Jukema, J.W., Kähönen, M., Kamatani, Y., Kanoni, S., Kerr, S.M., Khan, N.M., Koellinger, P., Koistinen, H.A., Kooner, M.K., Kubo, M., Kuusisto, J., Lahti, J., Launer, L.J., Lea, R.A., Lehne, B., Lehtimäki, T., Liewald, D.C.M., Lind, L., Loh, M., Lokki, M.-L., London, S.J., Loomis, S.J., Loukola, A., Lu, Y., Lumley, T., Lundqvist, A., Männistö, S., Marques-Vidal, P., Masciullo, C., Matchan, A., Mathias, R.A., Matsuda, K., Meigs, J.B., Meisinger, C., Meitinger, T., Menni, C., Mentch, F.D., Mihailov, E., Milani, L., Montasser, M.E., Montgomery, G.W., Morrison, A., Myers, R.H., Nadukuru, R., Navarro, P., Nelis, M., Nieminen, M.S., Nolte, I.M., O'Connor, G.T., Ogunniyi, A., Padmanabhan, S., Palmas, W.R., Pankow, J.S., Patarcic, I.,

Pavani, F., Peyser, P.A., Pietilainen, K., Poulter, N., Prokopenko, I., Ralhan, S., Redmond, P., Rich, S.S., Rissanen, H., Robino, A., Rose, L.M., Rose, R., Sala, C., Salako, B., Salomaa, V., Sarin, A.-P., Saxena, R., Schmidt, H., Scott, L.J., Scott, W.R., Sennblad, B., Seshadri, S., Sever, P., Shrestha, S., Smith, B.H., Smith, J.A., Soranzo, N., Sotoodehnia, N., Southam, L., Stanton, A.V., Stathopoulou, M.G., Strauch, K., Strawbridge, R.J., Suderman, M.J., Tandon, N., Tang, S.-T., Taylor, K.D., Tayo, B.O., Töglhofer, A.M., Tomaszewski, M., Tšernikova, N., Tuomilehto, J., Uitterlinden, A.G., Vaidya, D., van Hylckama Vlieg, A., van Setten, J., Vasankari, T., Vedantam, S., Vlachopoulou, E., Vozzi, D., Vuoksimaa, E., Waldenberger, M., Ware, E.B., Wentworth-Shields, W., Whitfield, J.B., Wild, S., Willemsen, G., Yajnik, C.S., Yao, J., Zaza, G., Zhu, X., Salem, R.M., Melbye, M., Bisgaard, H., Samani, N.J., Cusi, D., Mackey, D.A., Cooper, R.S., Froguel, P., Pasterkamp, G., Grant, S.F.A., Hakonarson, H., Ferrucci, L., Scott, R.A., Morris, A.D., Palmer, C.N.A., Dedoussis, G., Deloukas, P., Bertram, L., Lindenberger, U., Berndt, S.I., Lindgren, C.M., Timpson, N.J., Tönjes, A., Munroe, P.B., Sørensen, T.I.A., Rotimi, C.N., Arnett, D.K., Oldehinkel, A.J., Kardia, S.L.R., Balkau, B., Gambaro, G., Morris, A.P., Eriksson, J.G., Wright, M.J., Martin, N.G., Hunt, S.C., Starr, J.M., Deary, I.J., Griffiths, L.R., Tiemeier, H., Pirastu, N., Kaprio, J., Wareham, N.J., Pérusse, L., Wilson, J.G., Girotto, G., Caulfield, M.J., Raitakari, O., Boomsma, D.I., Gieger, C., van der Harst, P., Hicks, A.A., Kraft, P., Sinisalo, J., Knekt, P., Johannesson, M., Magnusson, P.K.E., Hamsten, A., Schmidt, R., Borecki, I.B., Vartiainen, E., Becker, D.M., Bharadwaj, D., Mohlke, K.L., Boehnke, M., van Duijn, C.M., Sanghera, D.K., Teumer, A., Zeggini, E., Metspalu, A., Gasparini, P., Ulivi, S., Ober, C., Toniolo, D., Rudan, I., Porteous, D.J., Ciullo, M., Spector, T.D., Hayward, C., Dupuis, J., Loos, R.J.F., Wright, A.F., Chandak, G.R., Vollenweider, P., Shuldiner, A.R., Ridker, P.M., Rotter, J.I., Sattar, N., Gyllensten, U., North, K.E., Pirastu, M., Psaty, B.M., Weir, D.R., Laakso, M., Gudnason, V., Takahashi, A., Chambers, J.C., Kooner, J.S., Strachan, D.P., Campbell, H., Hirschhorn, J.N., Perola, M., Polašek, O., Wilson, J.F., 2015. Directional dominance on stature and cognition in diverse human populations. Nature 523, 459–462. https://doi.org/10.1038/nature14618

Kapusta, A., Suh, A., Feschotte, C., 2017. Dynamics of genome size evolution in birds and mammals. Proc. Natl. Acad. Sci. 114, E1460–E1469. https://doi.org/10.1073/pnas.1616702114

Kardos, M., Husby, A., McFarlane, S.E., Qvarnström, A., Ellegren, H., 2016. Whole-genome resequencing of extreme phenotypes in collared flycatchers highlights the difficulty of detecting quantitative trait loci in natural populations. Mol. Ecol. Resour. 16, 727–741. https://doi.org/10.1111/1755-0998.12498

Kardos, M., Qvarnström, A., Ellegren, H., 2017. Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in *Ficedula* Flycatcher Genome Sequences. Genetics 205, 1319–1334. https://doi.org/10.1534/genetics.116.198861

Karlsson, S., Moen, T., Lien, S., Glover, K.A., Hindar, K., 2011. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip: snp genotyping and applications. Mol. Ecol. Resour. 11, 247–253. https://doi.org/10.1111/j.1755-0998.2010.02959.x

Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A.M., Ålund, M., Qvarnström, A., Ellegren, H., 2014a. Estimation of linkage disequilibrium and interspecific gene flow in *Ficedula* flycatchers by a newly developed 50k single-

nucleotide polymorphism array. Mol. Ecol. Resour. 14, 1248–1260. https://doi.org/10.1111/1755-0998.12270
Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., Olason, P., Ellegren, H., 2014b. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol. Ecol. 23, 4035–4058. https://doi.org/10.1111/mec.12810
Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W., Prodöhl, P.A., 2013. diveRsity : An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. 4, 782–788. https://doi.org/10.1111/2041-210X.12067
Keller, L., 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17, 230–241. https://doi.org/10.1016/S0169-5347(02)02489-8
Keller, M.C., Visscher, P.M., Goddard, M.E., 2011. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. Genetics 189, 237–249. https://doi.org/10.1534/genetics.111.130922
Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V.H., Nicholas, F.W., Raadsma, H., for the International Sheep Genomics Consortium, 2009. A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. PLoS ONE 4, e4668. https://doi.org/10.1371/journal.pone.0004668
Kim, J.-M., Santure, A.W., Barton, H.J., Quinn, J.L., Cole, E.F., Great Tit HapMap Consortium, Visser, M.E., Sheldon, B.C., Groenen, M.A.M., van Oers, K., Slate, J., 2018. A high-density SNP chip for genotyping great tit (*Parus major*) populations and its application to studying the genetic architecture of exploration behaviour. Mol. Ecol. Resour. 18, 877–891. https://doi.org/10.1111/1755-0998.12778
Kimura, Mot., 1955. Random Genetic Drift in Multi-Allelic Locus. Evolution 9, 419–435.
Kinziger, A.P., Nakamoto, R.J., Anderson, E.C., Harvey, B.C., 2011. Small founding number and low genetic diversity in an introduced species exhibiting limited invasion success (speckled dace, *Rhinichthys osculus*): Founder Number in Introduced Speckled Dace. Ecol. Evol. 1, 73–84. https://doi.org/10.1002/ece3.8
Kleinman-Ruiz, D., Martínez-Cruz, B., Soriano, L., Lucena-Perez, M., Cruz, F., Villanueva, B., Fernández, J., Godoy, J.A., 2017. Novel efficient genome-wide SNP panels for the conservation of the highly endangered Iberian lynx. BMC Genomics 18, 556. https://doi.org/10.1186/s12864-017-3946-5
Klicka, L.B., Kus, B.E., Title, P.O., Burns, K.J., 2016. Conservation genomics reveals multiple evolutionary units within Bell's Vireo (*Vireo bellii*). Conserv. Genet. 17, 455–471. https://doi.org/10.1007/s10592-015-0796-z
Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S.C., Kempenaers, B., Forstmeier, W., 2015. Quantifying realized inbreeding in wild and captive animal populations. Heredity 114, 397–403. https://doi.org/10.1038/hdy.2014.116
Kothiyal, P., Cox, S., Ebert, J., Aronow, B.J., Greinwald, J.H., Rehm, H.L., 2009. An Overview of Custom Array Sequencing, in: Haines, J.L., Korf, B.R., Morton, C.C., Seidman, C.E., Seidman, J.G., Smith, D.R. (Eds.), Current Protocols in Human Genetics. John Wiley & Sons, Inc., Hoboken, NJ, USA, unit 7, 17. https://doi.org/10.1002/0471142905.hg0717s61

Kreziou, A., de Boer, H., Gravendeel, B., 2016. Harvesting of salep orchids in north-western Greece continues to threaten natural populations. Oryx 50, 393–396. https://doi.org/10.1017/S0030605315000265

Kuningas, M., McQuillan, R., Wilson, J.F., Hofman, A., van Duijn, C.M., Uitterlinden, A.G., Tiemeier, H., 2011. Runs of Homozygosity Do Not Influence Survival to Old Age. PLoS ONE 6, e22580. https://doi.org/10.1371/journal.pone.0022580

Kuparinen, A., Festa-Bianchet, M., 2017. Harvest-induced evolution: insights from aquatic and terrestrial systems. Philos. Trans. R. Soc. B Biol. Sci. 372, 20160036. https://doi.org/10.1098/rstb.2016.0036

LaHood, E.S., Moran, P., Olsen, J., Grant, S., Park, L., 2002. Microsatellite allele ladders in two species of Pacific salmon: preparation and field-test results. Mol. Ecol. Notes 2, 187–190.

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O., Verhoeven, K.J.F., de Jager, V., Megens, H.-J., Warren, W.C., Minx, P., Crooijmans, R.P.M.A., Corcoran, P., Adriaensen, F., Belda, E., Bushuev, A., Cichon, M., Charmantier, A., Dingemanse, N., Doligez, B., Eeva, T., Erikstad, K.E., Fedorov, S., Hau, M., Hille, S., Hinde, C., Kempenaers, B., Kerimov, A., Krist, M., Mand, R., Matthysen, E., Nager, R., Norte, C., Orell, M., Richner, H., Slagsvold, T., Tilgar, V., Tinbergen, J., Torok, J., Tschirren, B., Yuta, T., Sheldon, B.C., Slate, J., Zeng, K., van Oers, K., Visser, M.E., Groenen, M.A.M., 2016. Evolutionary signals of selection on cognition from the great tit genome and methylome. Nat. Commun. 7, 10474. https://doi.org/10.1038/ncomms10474

Lambert, D.M., King, T., Shepherd, L.D., Livingston, A., Anderson, S., Craig, J.L., 2005. Serial population bottlenecks and genetic variation: Translocated populations of the New Zealand Saddleback (*Philesturnus carunculatus rufusater*). Conserv. Genet. 6, 1–14. https://doi.org/10.1007/s10592-004-7857-z

Laugier, G.J.M., Le Moguédec, G., Su, W., Tayeh, A., Soldati, L., Serrate, B., Estoup, A., Facon, B., 2016. Reduced population size can induce quick evolution of inbreeding depression in the invasive ladybird *Harmonia axyridis*. Biol. Invasions 18, 2871–2881. https://doi.org/10.1007/s10530-016-1179-1

Lawson, M.J., Zhang, L., 2006. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol. 11, 7:R14.

Leclerc, M., Zedrosser, A., Pelletier, F., 2017. Harvesting as a potential selective pressure on behavioural traits. J. Appl. Ecol. 54, 1941–1945. https://doi.org/10.1111/1365-2664.12893

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc. Natl. Acad. Sci. 104, 19942–19947. https://doi.org/10.1073/pnas.0710021104

Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496. https://doi.org/10.1038/nature10231

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009.

The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
Low, M., 2004. Female weight predicts the timing of forced copulation attempts in stitchbirds, *Notiomystis cincta*. Anim. Behav. 68, 637–644. https://doi.org/10.1016/j.anbehav.2004.01.006
Low, M., Alley, M.R., Scott, I., 2007a. Pruritic Facial Dermatitis in a Population of Free-living Stitchbirds. J. Wildl. Dis. 43, 262–268. https://doi.org/10.7589/0090-3558-43.2.262
Low, M., Castro, I., Berggren, Å., 2005. Cloacal erection promotes vent apposition during forced copulation in the New Zealand stitchbird (hihi): implications for copulation efficiency in other species. Behav. Ecol. Sociobiol. 58, 247–255. https://doi.org/10.1007/s00265-005-0935-5
Low, M., Pärt, T., 2009. Patterns of mortality for each life-history stage in a population of the endangered New Zealand stitchbird. J. Anim. Ecol. 78, 761–771. https://doi.org/10.1111/j.1365-2656.2009.01543.x
Low, M., PäRt, T., Forslund, P., 2007b. Age-specific variation in reproduction is largely explained by the timing of territory establishment in the New Zealand stitchbird *Notiomystis cincta.* J. Anim. Ecol. 76, 459–470. https://doi.org/10.1111/j.1365-2656.2007.01234.x
Luikart, G., 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. J. Hered. 89, 238–247. https://doi.org/10.1093/jhered/89.3.238
Lundregan, S.L., Hagen, I.J., Gohli, J., Niskanen, A.K., Kemppainen, P., Ringsby, T.H., Kvalnes, T., Pärn, H., Rønning, B., Holand, H., Ranke, P.S., Båtnes, A.S., Selvik, L.-K., Lien, S., Saether, B.-E., Husby, A., Jensen, H., 2018. Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. Mol. Ecol. 27, 3498–3514. https://doi.org/10.1111/mec.14811
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Yunjie, Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Yong, Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, Jian, Lam, T.-W., Wang, Jun, 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1:18. https://doi.org/10.1186/2047-217X-1-18
MacLeod, I.M., Larkin, D.M., Lewin, H.A., Hayes, B.J., Goddard, M.E., 2013. Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. Mol. Biol. Evol. 30, 2209–2223. https://doi.org/10.1093/molbev/mst125
Malenfant, R.M., Coltman, D.W., Davis, C.S., 2015. Design of a 9K illumina BeadChip for polar bears (*Ursus maritimus*) from RAD and transcriptome sequencing. Mol. Ecol. Resour. 15, 587–600. https://doi.org/10.1111/1755-0998.12327
Malenfant, R.M., Coltman, D.W., Richardson, E.S., Lunn, N.J., Stirling, I., Adamowicz, E., Davis, C.S., 2016. Evidence of adoption, monozygotic twinning, and low inbreeding rates in a large genetic pedigree of polar bears. Polar Biol. 39, 1455–1465. https://doi.org/10.1007/s00300-015-1871-0
Malenfant, R.M., Davis, C.S., Richardson, E.S., Lunn, N.J., Coltman, D.W., 2018. Heritability of body size in the polar bears of Western Hudson Bay. Mol. Ecol. Resour. 18, 854–866. https://doi.org/10.1111/1755-0998.12889
Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J.L., Macciotta, N.P.P., 2015. Analysis of runs of homozygosity and

their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46, 110–121. https://doi.org/10.1111/age.12259
Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S., Van Tassell, C.P., 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE 4, e5350. https://doi.org/10.1371/journal.pone.0005350
Maurer, G., Beck, N., Double, M.C., 2010. A 'feather-trap' for collecting DNA samples from birds: TECHNICAL ADVANCES. Mol. Ecol. Resour. 10, 129–134. https://doi.org/10.1111/j.1755-0998.2009.02711.x
McEachern, M.B., Van Vuren, D.H., Floyd, C.H., May, B., Eadie, J.M., 2011. Bottlenecks and rescue effects in a fluctuating population of golden-mantled ground squirrels (*Spermophilus lateralis*). Conserv. Genet. 12, 285–296. https://doi.org/10.1007/s10592-010-0139-z
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110
McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., MacLeod, A.K., Farrington, S.M., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S.H., Dunlop, M.G., Wright, A.F., Campbell, H., Wilson, J.F., 2008. Runs of Homozygosity in European Populations. Am. J. Hum. Genet. 83, 359–372. https://doi.org/10.1016/j.ajhg.2008.08.007
Mead, S., Poulter, M., Beck, J., Uphill, J., Jones, C., Ang, C.E., Mein, C.A., Collinge, J., 2008. Successful amplification of degraded DNA for use with high-throughput SNP genotyping platforms. Hum. Mutat. 29, 1452–1458. https://doi.org/10.1002/humu.20782
Meirmans, P.G., Hedrick, P.W., 2011. Assessing population structure: FST and related measures: invited technical review. Mol. Ecol. Resour. 11, 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x
Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., Gut, I.G., Distl, O., 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics 16:764. https://doi.org/10.1186/s12864-015-1977-3
Meyer-Lucht, Y., Mulder, K.P., James, M.C., McMahon, B.J., Buckley, K., Piertney, S.B., Höglund, J., 2016. Adaptive and neutral genetic differentiation among Scottish and endangered Irish red grouse (*Lagopus lagopus scotica*). Conserv. Genet. 17, 615–630. https://doi.org/10.1007/s10592-016-0810-0
Miller, J.M., Festa-Bianchet, M., Coltman, D.W., 2018. Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium ® HD SNP BeadChip. PeerJ 6, e4364. https://doi.org/10.7717/peerj.4364
Miller, K.A., Nelson, N.J., Smith, H.G., Moore, J.A., 2009. How do reproductive skew and founder group size affect genetic diversity in reintroduced populations? Mol. Ecol. 18, 3792–3802. https://doi.org/10.1111/j.1365-294X.2009.04315.x
Mills, L.S., Allendorf, F.W., 1996. The One-Migrant-per-Generation Rule in Conservation and Management. Conserv. Biol. 10, 1509–1518. https://doi.org/10.1046/j.1523-1739.1996.10061509.x

Milot, E., Weimerskirch, H., Duchesne, P., Bernatchez, L., 2007. Surviving with low genetic diversity: the case of albatrosses. Proc. R. Soc. B Biol. Sci. 274, 779–787. https://doi.org/10.1098/rspb.2006.0221

Miskelly, C.M., Powlesland, R.G., 2013. Conservation translocations of New Zealand birds, 1863-2012. Notornis 60, 3–28.

Mock, K.E., Latch, E.K., Rhodes, O.E., 2004. Assessing losses of genetic diversity due to translocation: long-term case histories in Merriam's turkey (*Meleagris gallopavo merriami)*. Conserv. Genet. 5, 631–645. https://doi.org/10.1007/s10592-004-1849-x

Morin, P.A., Luikart, G., Wayne, R.K., the SNP workshop group, 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19, 208–216. https://doi.org/10.1016/j.tree.2004.01.009

Moura, R.F., Dawson, D.A., Nogueira, D.M., 2017. The use of microsatellite markers in Neotropical studies of wild birds: a literature review. An. Acad. Bras. Ciênc. 89, 145–154. https://doi.org/10.1590/0001-3765201620160378

Narum, S.R., Banks, M., Beacham, T.D., Bellinger, M.R., Campbell, M.R., Dekoning, J., Elz, A., Guthrieiii, C.M., Kozfkay, C., Miller, K.M., Moran, P., Phillips, R., Seeb, L.W., Smith, C.T., Warheit, K., Young, S.F., Garza, J.C., 2008. Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol. Ecol. https://doi.org/10.1111/j.1365-294X.2008.03851.x

Natusch, D.J.D., Lyons, J.A., Mumpuni, Riyanto, A., Shine, R., 2016. Jungle Giants: Assessing Sustainable Harvesting in a Difficult-to-Survey Species (*Python reticulatus*). PLOS ONE 11, e0158397. https://doi.org/10.1371/journal.pone.0158397

Nei, M., 1973. Analysis of Gene Diversity in Subdivided Populations. Proc. Natl. Acad. Sci. 70, 3321–3323.

Nei, M., Maruyama, T., Chakraborty, R., 1975. The Bottleneck Effect and Genetic Variability in Populations. Evolution 27, 1–10.

Nietlisbach, P., Keller, L.F., Camenisch, G., Guillaume, F., Arcese, P., Reid, J.M., Postma, E., 2017. Pedigree-based inbreeding coefficient explains more variation in fitness than heterozygosity at 160 microsatellites in a wild bird population. Proc. R. Soc. B Biol. Sci. 284, 20162763. https://doi.org/10.1098/rspb.2016.2763

Nothnagel, M., Lu, T.T., Kayser, M., Krawczak, M., 2010. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum. Mol. Genet. 19, 2927–2935. https://doi.org/10.1093/hmg/ddq198

Nussey, D.H., Coltman, D.W., Coulson, T., Kruuk, L.E.B., Donald, A., Morris, S.J., Clutton-Brock, T.H., Pemberton, J., 2005. Rapidly declining fine-scale spatial genetic structure in female red deer. Mol. Ecol. 14, 3395–3405. https://doi.org/10.1111/j.1365-294X.2005.02692.x

Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R. (Kuke), Hedrick, P.W., 2010. Conservation genetics in transition to conservation genomics. Trends Genet. 26, 177–187. https://doi.org/10.1016/j.tig.2010.01.001

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., Dalén, L., 2015. Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Curr. Biol. 25, 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007

Panfylova, J., Bemelmans, E., Devine, C., Frost, P., Armstrong, D., 2016. Post-release effects on reintroduced populations of hihi: Post-Release Effects in Reintroduced Populations. J. Wildl. Manag. 80, 970–977. https://doi.org/10.1002/jwmg.21090

Paradis, E., Gosselin, T., Goudet, J., Jombart, T., Schliep, K., 2017. Linking genomics and population genetics with R. Mol. Ecol. Resour. 17, 54–66. https://doi.org/10.1111/1755-0998.12577

Pemberton, T.J., Absher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A., Li, J.Z., 2012. Genomic Patterns of Homozygosity in Worldwide Human Populations. Am. J. Hum. Genet. 91, 275–292. https://doi.org/10.1016/j.ajhg.2012.06.014

Perina, A., González-Tizón, A.M., Meilán, I.F., Martínez-Lage, A., 2017. De novo transcriptome assembly of shrimp *Palaemon serratus*. Genomics Data 11, 89–91. https://doi.org/10.1016/j.gdata.2016.12.009

Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R., Baldi, F., 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48, 255–271. https://doi.org/10.1111/age.12526

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., Panetto, J.C. do C., Ventura, R.V., Baldi, F., da Silva, M.V.G.B., 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. BMC Genomics 19(1):34. https://doi.org/10.1186/s12864-017-4365-3

Pozo, R.A., Schindler, S., Cubaynes, S., Cusack, J.J., Coulson, T., Malo, A.F., 2016. Modeling the impact of selective harvesting on red deer antlers: Hunting Effects on Red Deer Antler Size. J. Wildl. Manag. 80, 978–989. https://doi.org/10.1002/jwmg.21089

Price, M.R., Hadfield, M.G., 2014. Population Genetics and the Effects of a Severe Bottleneck in an Ex Situ Population of Critically Endangered Hawaiian Tree Snails. PLoS ONE 9, e114377. https://doi.org/10.1371/journal.pone.0114377

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795

Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 70. https://doi.org/10.1186/1471-2156-13-70

Purfield, D.C., McParland, S., Wall, E., Berry, D.P., 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLOS ONE 12, e0176780. https://doi.org/10.1371/journal.pone.0176780

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rajora, O.P., Rahman, M.H., Buchert, G.P., Dancik, B.P., 2000. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. Mol. Ecol. 9, 339–348.

Ramstad, K.M., Colbourne, R.M., Robertson, H.A., Allendorf, F.W., Daugherty, C.H., 2013. Genetic consequences of a century of protection: serial founder events and survival of the little spotted kiwi (*Apteryx owenii*). Proc. R. Soc. B Biol. Sci. 280, 576. https://doi.org/10.1098/rspb.2013.0576

Raymond, M., Rousset, F., 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. J. Hered. 86, 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Rebelato, A.B., Caetano, A.R., 2018. Runs of homozygosity for autozygosity estimation and genomic analysis in production animals. Pesqui. Agropecuária Bras. 53, 975–984. https://doi.org/10.1590/s0100-204x2018000900001
Reed, D.H., Frankham, R., 2003. Correlation between Fitness and Genetic Diversity. Conserv. Biol. 17, 230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x
Reid, J.M., Arcese, P., Sardell, R.J., Keller, L.F., 2011. Heritability of female extra-pair paternity rate in song sparrows (*Melospiza melodia)*. Proc. R. Soc. B Biol. Sci. 278, 1114–1120. https://doi.org/10.1098/rspb.2010.1704
Richardson, K., Hauber, M., Ewen, J., Armstrong, D., 2010. Sex-specific shifts in natal dispersal dynamics in a reintroduced hihi population. Behaviour 147, 1517–1532. https://doi.org/10.1163/000579510X531284
Richardson, K.M., Ewen, J.G., 2016. Habitat selection in a reintroduced population: social effects differ between natal and post-release dispersal. Anim. Conserv. 19, 413–421. https://doi.org/10.1111/acv.12257
Robertson, H.A., Heather, B.D., Onley, D.J., 2015. The hand guide to the birds of New Zealand, Revised and updated third edition. ed. Penguin Books, New Zealand.
Robinson, M.R., Santure, A.W., DeCauwer, I., Sheldon, B.C., Slate, J., 2013. Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. Mol. Ecol. 22, 3963–3980. https://doi.org/10.1111/mec.12375
Roitman, M., Gardner, M.G., New, T.R., Nguyen, T.T.T., Roycroft, E.J., Sunnucks, P., Yen, A.L., Harrisson, K.A., 2017. Assessing the scope for genetic rescue of an endangered butterfly: the case of the Eltham copper. Insect Conserv. Divers. 10, 399–414. https://doi.org/10.1111/icad.12235
Rousset, F., 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x
Rowe, G., Beebee, T.J.C., Burke, T., 1998. Phylogeography of the natterjack toad *Bufo calamita* in Britain: genetic differentiation of native and translocated populations. Mol. Ecol. 7, 751–760. https://doi.org/10.1046/j.1365-294x.1998.00387.x
Rutschmann, A., de Villemereuil, P., Brekke, P., Ewen, J.G., Anderson, N., Santure, A.W., in prep. Consequences of space sharing on individual phenotypes in the New Zealand hihi.
Santure, A.W., De Cauwer, I., Robinson, M.R., Poissant, J., Sheldon, B.C., Slate, J., 2013. Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. Mol. Ecol. 22, 3949–3962. https://doi.org/10.1111/mec.12376
Santure, A.W., Poissant, J., De Cauwer, I., van Oers, K., Robinson, M.R., Quinn, J.L., Groenen, M.A.M., Visser, M.E., Sheldon, B.C., Slate, J., 2015. Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. Mol. Ecol. 24, 6148–6162. https://doi.org/10.1111/mec.13452
Santure, A.W., Stapley, J., Ball, A.D., Birkhead, T.R., Burke, T., Slate, J., 2010. On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. Mol. Ecol. 19, 1439–1451. https://doi.org/10.1111/j.1365-294X.2010.04554.x
Sardell, R.J., Keller, L.F., Arcese, P., Bucher, T., Reid, J.M., 2010. Comprehensive paternity assignment: genotype, spatial location and social status in song sparrows, Melospiza Melodia: extra-pair paternity, location and status. Mol. Ecol. 19, 4352–4364. https://doi.org/10.1111/j.1365-294X.2010.04805.x
Schaberg, P.G., DeHayes, D.H., Hawley, G.J., Nijensohn, S.E., 2008. Anthropogenic alterations of genetic diversity within tree populations: Implications for forest

ecosystem resilience. For. Ecol. Manag. 256, 855–862. https://doi.org/10.1016/j.foreco.2008.06.038
Schmitt, T., Hewitt, G.M., 2004. The genetic pattern of population threat and loss: a case study of butterflies. Mol. Ecol. 13, 21–31. https://doi.org/10.1046/j.1365-294X.2004.02020.x
Seddon, J.M., Parker, H.G., Ostrander, E.A., Ellegren, H., 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population: SNPs in ecology - the Scandinavian wolves. Mol. Ecol. 14, 503–511. https://doi.org/10.1111/j.1365-294X.2005.02435.x
Seddon, P.J., Griffiths, C.J., Soorae, P.S., Armstrong, D.P., 2014. Reversing defaunation: Restoring species in a changing world. Science 345, 406–412. https://doi.org/10.1126/science.1251818
Seddon, P.J., Strauss, M.W., Innes, J., 2012. Animal translocations: What are they and why do we do them?, in: Reintroduction Biology Edited by John G. Ewen, Doug P. Armstrong, Kevin A. Parker and Philip J. Seddon. Blackwell Publishing Ltd., pp. 1–32.
Shackleton, C.M. (Ed.), 2015. Ecological sustainability for non-timber forest products: dynamics and case studies of harvesting, People and plants international conservation. Routledge, London.
Shafer, A.B.A., Wolf, J.B.W., Alves, P.C., Bergström, L., Bruford, M.W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K.D., Fior, S., Hajibabaei, M., Hill, J.A., Hoezel, A.R., Höglund, J., Jensen, E.L., Krause, J., Kristensen, T.N., Krützen, M., McKay, J.K., Norman, A.J., Ogden, R., Österling, E.M., Ouborg, N.J., Piccolo, J., Popović, D., Primmer, C.R., Reed, F.A., Roumet, M., Salmona, J., Schenekar, T., Schwartz, M.K., Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vergeer, P., Vijay, N., Vilà, C., Weissensteiner, M., Wennerström, L., Wheat, C.W., Zieliński, P., 2015. Genomics and the challenging translation into conservation practice. Trends Ecol. Evol. 30, 78–87. https://doi.org/10.1016/j.tree.2014.11.009
Sherley, G., Stringer, I., Parrish, G.R., New Zealand, Department of Conservation, 2010. Summary of native bat, reptile, amphibian and terrestrial invertebrate translocations in New Zealand. Pub. Team, Dept. of Conservation, Wellington, N.Z.
Silva, C.N.S., McFarlane, S.E., Hagen, I.J., Rönnegård, L., Billing, A.M., Kvalnes, T., Kemppainen, P., Rønning, B., Ringsby, T.H., Sæther, B.-E., Qvarnström, A., Ellegren, H., Jensen, H., Husby, A., 2017. Insights into the genetic architecture of morphological traits in two passerine bird species. Heredity 119, 197–205. https://doi.org/10.1038/hdy.2017.29
Slate, J., Santure, A.W., Feulner, P.G.D., Brown, E.A., Ball, A.D., Johnston, S.E., Gratten, J., 2010. Genome mapping in intensively studied wild vertebrate populations. Trends Genet. 26, 275–284. https://doi.org/10.1016/j.tig.2010.03.005
Slatkin, M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457–462.
Smith, M.J., Pascal, C.E., Grauvogel, Z., Habicht, C., Seeb, J.E., Seeb, L.W., 2011. Multiplex preamplification PCR and microsatellite validation enables accurate single nucleotide polymorphism genotyping of historical fish scales: SNP genotyping and applications. Mol. Ecol. Resour. 11, 268–277. https://doi.org/10.1111/j.1755-0998.2010.02965.x

Sovic, M., Fries, A., Martin, S.A., Gibbs, H.L., 2018. Genetic signatures of small effective population sizes and demographic declines in an endangered rattlesnake, *Sistrurus catenatus*. Evol. Appl. https://doi.org/10.1111/eva.12731

Stapley, J., Birkhead, T.R., Burke, T., Slate, J., 2008. A Linkage Map of the Zebra Finch *Taeniopygia guttata* Provides New Insights Into Avian Genome Evolution. Genetics 179, 651–667. https://doi.org/10.1534/genetics.107.086264

Steinbach, P., Heddergott, M., Weigand, H., Weigand, A.M., Wilwert, E., Stubbe, M., Helm, B., Campbell, R.E., Stubbe, A., Frantz, A.C., 2018. Rare migrants suffice to maintain high genetic diversity in an introduced island population of roe deer (*Capreolus capreolus*): Evidence from molecular data and simulations. Mamm. Biol. 88, 64–71. https://doi.org/10.1016/j.mambio.2017.11.009

Stockley, P., Searle, J.B., MacDonald, D.W., Jones, C.S., 1993. Female multiple mating behaviour in the common shrew as a strategy to reduce inbreeding. Proc. R. Soc. Lond. B Biol. Sci. 254, 173–179. https://doi.org/10.1098/rspb.1993.0143

Stockwell, B.L., Larson, W.A., Waples, R.K., Abesamis, R.A., Seeb, L.W., Carpenter, K.E., 2016. The application of genomics to inform conservation of a functionally important reef fish (*Scarus niger*) in the Philippines. Conserv. Genet. 17, 239–249. https://doi.org/10.1007/s10592-015-0776-3

Swain, D.P., Sinclair, A.F., Mark Hanson, J., 2007. Evolutionary response to size-selective mortality in an exploited fish population. Proc. R. Soc. B Biol. Sci. 274, 1015–1022. https://doi.org/10.1098/rspb.2006.0275

Swan, K.D., Lloyd, N.A., Moehrenschlager, A., 2018. Projecting further increases in conservation translocations: A Canadian case study. Biol. Conserv. 228, 175–182. https://doi.org/10.1016/j.biocon.2018.10.026

Syvänen, A.-C., 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat. Rev. Genet. 2, 930–942. https://doi.org/10.1038/35103535

Szulkin, M., Sheldon, B.C., 2007. The Environmental Dependence of Inbreeding Depression in a Wild Bird Population. PLoS ONE 2, e1027. https://doi.org/10.1371/journal.pone.0001027

Talbot, S.L., Pearce, J.M., Pierson, B.J., Derksen, D.V., Scribner, K.T., 2003. Molecular status of the dusky Canada goose (*Branta canadensis occidentalis*): A genetic assessment of a translocation effort. Conserv. Genet. 4, 367–381.

Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17, 6463–6471. https://doi.org/10.1093/nar/17.16.6463

Taylor, H.R., 2015. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. Ecol. Evol. 5, 3140–3150. https://doi.org/10.1002/ece3.1541

Taylor, H.R., Dussex, N., van Heezik, Y., 2017. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. Glob. Ecol. Conserv. 10, 231–242. https://doi.org/10.1016/j.gecco.2017.04.001

The Avian Genome Consortium, Zhang, G., Li, B., Li, C., Gilbert, M.T.P., Jarvis, E.D., Wang, J., 2014. Comparative genomic data of the Avian Phylogenomics Project. GigaScience 3:26. https://doi.org/10.1186/2047-217X-3-26

Thomson, M.J., Singh, N., Dwiyanti, M.S., Wang, D.R., Wright, M.H., Perez, F.A., DeClerck, G., Chin, J.H., Malitic-Layaoen, G.A., Juanillas, V.M., Dilla-Ermita, C.J., Mauleon, R., Kretzschmar, T., McCouch, S.R., 2017. Large-scale deployment of a rice 6 K SNP array for genetics and breeding applications. Rice 10. https://doi.org/10.1186/s12284-017-0181-2

Thorogood, R., Armstrong, D.P., Low, M., Brekke, P., Ewen, J.G., 2013. The value of long-term ecological research: integrating knowledge for conservation of hihi on Tiritiri Matangi Island. N. Z. J. Ecol. 37, 9.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H.C.M., Jamli, S., Jiken, A.J., Klopp, C., Lawley, C.T., McEwan, J., Martin, P., Moreno, C.R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhière, I., Rupp, R., Sarry, J., Sayre, B.L., Tircazes, A., Jun Wang, Wang, W., Zhang, W., and the International Goat Genome Consortium, 2014. Design and Characterization of a 52K SNP Chip for Goats. PLoS ONE 9, e86227. https://doi.org/10.1371/journal.pone.0086227

Toy, R., Greene, T., Greene, B., Warren, A., Griffiths, R., 2018. Changes in density of hihi (*Notiomystis cincta*), tīeke (*Philesturnus rufusater*) and tūī (*Prosthemadera novaeseelandiae*) on Little Barrier Island (*Te Hauturu-o-Toi*), Hauraki Gulf, Auckland, 2005–2013. N. Z. J. Ecol. 42:24. https://doi.org/10.20417/nzjecol.42.24

Uusi-Heikkilä, S., Sävilammi, T., Leder, E., Arlinghaus, R., Primmer, C.R., 2017. Rapid, broad-scale gene expression evolution in experimentally harvested fish populations. Mol. Ecol. 26, 3954–3967. https://doi.org/10.1111/mec.14179

Väli, Ü., Einarsson, A., Waits, L., Ellegren, H., 2008. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Mol. Ecol. 17, 3808–3817. https://doi.org/10.1111/j.1365-294X.2008.03876.x

van Bers, N.E.M., Santure, A.W., van Oers, K., De Cauwer, I., Dibbits, B.W., Mateman, C., Crooijmans, R.P.M.A., Sheldon, B.C., Visser, M.E., Groenen, M.A.M., Slate, J., 2012. The design and cross-population application of a genome-wide SNP chip for the great tit Parus major: Design and application of a great tit 10K SNP chip. Mol. Ecol. Resour. 12, 753–770. https://doi.org/10.1111/j.1755-0998.2012.03141.x

van Oers, K., Santure, A.W., De Cauwer, I., van Bers, N.E., Crooijmans, R.P., Sheldon, B.C., Visser, M.E., Slate, J., Groenen, M.A., 2014. Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. Heredity 112, 307–316. https://doi.org/10.1038/hdy.2013.107

Vilas, A., Pérez-Figueroa, A., Quesada, H., Caballero, A., 2015. Allelic diversity for neutral markers retains a higher adaptive potential for quantitative traits than expected heterozygosity. Mol. Ecol. 24, 4419–4432. https://doi.org/10.1111/mec.13334

Walker, L.K., Armstrong, D.P., Brekke, P., Chauvenet, A.L.M., Kilner, R.M., Ewen, J.G., 2013. Giving hihi a helping hand: assessment of alternative rearing diets in food supplemented populations of an endangered bird: Alternative rearing diets in hihi. Anim. Conserv. 16, 538–545. https://doi.org/10.1111/acv.12026

Walsh, M.R., Munch, S.B., Chiba, S., Conover, D.O., 2006. Maladaptive changes in multiple traits caused by fishing: impediments to population recovery: Evolutionary consequences of harvesting. Ecol. Lett. 9, 142–148. https://doi.org/10.1111/j.1461-0248.2005.00858.x

Wan, Q.-H., Wu, H., Fujihara, T., Fang, S.-G., 2004. Which genetic marker for which conservation genetics issue? Electrophoresis 25, 2165–2176. https://doi.org/10.1002/elps.200305922

Wang, J., Brekke, P., Huchard, E., Knapp, L.A., Cowlishaw, G., 2010. Estimation of parameters of inbreeding and genetic drift in populations with overlapping generations. Evolution 64, 1704–1718. https://doi.org/10.1111/j.1558-5646.2010.00953.x

Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., Yu, D., 2016. Development and application of a novel genome-wide SNP array reveals domestication history in soybean. Sci. Rep. 6:20728. https://doi.org/10.1038/srep20728

Wang, W., Yin, X., Soo Pyon, Y., Hayes, M., Li, J., 2013. Rare variant discovery and calling by sequencing pooled samples with overlaps. Bioinformatics 29, 29–38. https://doi.org/10.1093/bioinformatics/bts645

Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., Heger, A., Kong, L., Ponting, C.P., Jarvis, E.D., Mello, C.V., Minx, P., Lovell, P., Velho, T.A.F., Ferris, M., Balakrishnan, C.N., Sinha, S., Blatti, C., London, S.E., Li, Y., Lin, Y.-C., George, J., Sweedler, J., Southey, B., Gunaratne, P., Watson, M., Nam, K., Backström, N., Smeds, L., Nabholz, B., Itoh, Y., Whitney, O., Pfenning, A.R., Howard, J., Völker, M., Skinner, B.M., Griffin, D.K., Ye, L., McLaren, W.M., Flicek, P., Quesada, V., Velasco, G., Lopez-Otin, C., Puente, X.S., Olender, T., Lancet, D., Smit, A.F.A., Hubley, R., Konkel, M.K., Walker, J.A., Batzer, M.A., Gu, W., Pollock, D.D., Chen, L., Cheng, Z., Eichler, E.E., Stapley, J., Slate, J., Ekblom, R., Birkhead, T., Burke, T., Burt, D., Scharff, C., Adam, I., Richard, H., Sultan, M., Soldatov, A., Lehrach, H., Edwards, S.V., Yang, S.-P., Li, X., Graves, T., Fulton, L., Nelson, J., Chinwalla, A., Hou, S., Mardis, E.R., Wilson, R.K., 2010. The genome of a songbird. Nature 464, 757–762. https://doi.org/10.1038/nature08819

Weeks, A.R., Sgro, C.M., Young, A.G., Frankham, R., Mitchell, N.J., Miller, K.A., Byrne, M., Coates, D.J., Eldridge, M.D.B., Sunnucks, P., Breed, M.F., James, E.A., Hoffmann, A.A., 2011. Assessing the benefits and risks of translocations in changing environments: a genetic perspective: Translocations in changing environments. Evol. Appl. 4, 709–725. https://doi.org/10.1111/j.1752-4571.2011.00192.x

Weir, B.S., Cockerham, C.C., 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38, 14.

Weller, P., Jeffreys, A.J., Wilson, V., Blanchetot, A., 1984. Organization of the human myoglobin gene. EMBO J. 3, 439–446.

Whitehorn, P.R., Tinsley, M.C., Brown, M.J.F., Darvill, B., Goulson, D., 2011. Genetic diversity, parasite prevalence and immunity in wild bumblebees. Proc. R. Soc. B Biol. Sci. 278, 1195–1202. https://doi.org/10.1098/rspb.2010.1550

Whiteley, A.R., Fitzpatrick, S.W., Funk, W.C., Tallmon, D.A., 2015. Genetic rescue to the rescue. Trends Ecol. Evol. 30, 42–49. https://doi.org/10.1016/j.tree.2014.10.009

Wickham, H., 2016. ggplot2: elegant graphics for data analysis, Second edition. ed, Use R! Springer, Cham.

Willi, Y., Van Buskirk, J., Hoffmann, A.A., 2006. Limits to the Adaptive Potential of Small Populations. Annu. Rev. Ecol. Evol. Syst. 37, 433–458. https://doi.org/10.1146/annurev.ecolsys.37.091305.110145

Wright, B., Morris, K., Grueber, C.E., Willet, C.E., Gooley, R., Hogg, C.J., O'Meally, D., Hamede, R., Jones, M., Wade, C., Belov, K., 2015. Development of a SNP-based assay for measuring genetic diversity in the Tasmanian devil insurance population. BMC Genomics 16:791. https://doi.org/10.1186/s12864-015-2020-4

Wright, S., 1965. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution 19, 395–420.

Wright, S., 1951. The genetical structure of populations. Ann. Eugen. 15, 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Wright, S., 1943. Isolation by distance. Genetics 28, 114.
Wright, S., 1931. Evloution in Mendelian populations. Genetics 16, 97–159.
Wright, S., 1922. Coefficients of Inbreeding and Relationship. Am. Nat. 56, 330–338. https://doi.org/10.1086/279872
Xue, Y., Prado-Martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D.N., de Manuel, M., Hernandez-Rodriguez, J., Lobon, I., Siegismund, H.R., Pagani, L., Quail, M.A., Hvilsom, C., Mudakikwa, A., Eichler, E.E., Cranfield, M.R., Marques-Bonet, T., Tyler-Smith, C., Scally, A., 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. Science, 348, 242-245. DOI: 10.1126/science.aaa3952
Yasui, Y., 1998. The 'genetic benefits' of female multiple mating reconsidered 13, 5.
Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A., Flicek, P., 2018. Ensembl 2018. Nucleic Acids Res. 46, D754–D761. https://doi.org/10.1093/nar/gkx1098
Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., Odeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., Yang, W., Hu, J., Xiao, J., Yang, Z., Liu, Y., Xie, Q., Yu, H., Lian, J., Wen, P., Zhang, F., Li, H., Zeng, Y., Xiong, Z., Liu, S., Zhou, L., Huang, Z., An, N., Wang, J., Zheng, Q., Xiong, Y., Wang, G., Wang, B., Wang, J., Fan, Y., da Fonseca, R.R., Alfaro-Nunez, A., Schubert, M., Orlando, L., Mourier, T., Howard, J.T., Ganapathy, G., Pfenning, A., Whitney, O., Rivas, M.V., Hara, E., Smith, J., Farre, M., Narayan, J., Slavov, G., Romanov, M.N., Borges, R., Machado, J.P., Khan, I., Springer, M.S., Gatesy, J., Hoffmann, F.G., Opazo, J.C., Hastad, O., Sawyer, R.H., Kim, H., Kim, K.-W., Kim, H.J., Cho, S., Li, N., Huang, Y., Bruford, M.W., Zhan, X., Dixon, A., Bertelsen, M.F., Derryberry, E., Warren, W., Wilson, R.K., Li, S., Ray, D.A., Green, R.E., O'Brien, S.J., Griffin, D., Johnson, W.E., Haussler, D., Ryder, O.A., Willerslev, E., Graves, G.R., Alstrom, P., Fjeldsa, J., Mindell, D.P., Edwards, S.V., Braun, E.L., Rahbek, C., Burt, D.W., Houde, P., Zhang, Y., Yang, H., Wang, J., Avian Genome Consortium, Jarvis, E.D., Gilbert, M.T.P., Wang, J., Ye, C., Liang, S., Yan, Z., Zepeda, M.L., Campos, P.F., Velazquez, A.M.V., Samaniego, J.A., Avila-Arcos, M., Martin, M.D., Barnett, R., Ribeiro, A.M., Mello, C.V., Lovell, P.V., Almeida, D., Maldonado, E., Pereira, J., Sunagar, K., Philip, S., Dominguez-Bello, M.G., Bunce, M., Lambert, D., Brumfield, R.T., Sheldon, F.H., Holmes, E.C., Gardner, P.P., Steeves, T.E., Stadler, P.F., Burge, S.W., Lyons, E., Smith, J., McCarthy, F., Pitel, F., Rhoads, D., Froman, D.P., 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346, 1311–1320. https://doi.org/10.1126/science.1251385
Zhang, Q., Calus, M.P., Guldbrandtsen, B., Lund, M.S., Sahana, G., 2015. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in

three cattle breeds. BMC Genet. 16, 18. https://doi.org/10.1186/s12863-015-0227-7

Zhen, Y., Harrigan, R.J., Ruegg, K.C., Anderson, E.C., Ng, T.C., Lao, S., Lohmueller, K.E., Smith, T.B., 2017. Genomic divergence across ecological gradients in the Central African rainforest songbird (*Andropadus virens*). Mol. Ecol. 26, 4966–4977. https://doi.org/10.1111/mec.14270

Zimin, A.V., Delcher, A.L., Florea, L., Kelley, D.R., Schatz, M.C., Puiu, D., Hanrahan, F., Pertea, G., Van Tassell, C.P., Sonstegard, T.S., Marçais, G., Roberts, M., Subramanian, P., Yorke, J.A., Salzberg, S.L., 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10, R42. https://doi.org/10.1186/gb-2009-10-4-r42