ABSTRACT

Title of Dissertation: POPULATION DECLINES AND GENETIC VARIATION: EFFECTS OF SERIAL BOTTLENECKS

Taylor Eilers Callicrate, Doctor of Philosophy, 2015

Dissertation directed by: Associate Professor Jiuzhou Song
Department of Animal & Avian Sciences

Adjunct Professor Robert C. Fleischer
Department of Biology

Islands foster unique biodiversity, yet also present biogeographic limitations that impose increased risk for population extinction through demographic and genetic constraints and decreased probability of surviving a catastrophe. Of particular interest, especially with regard to endangered species, is the genetic response of insular species to severe population declines or translocations. Both types of events, considered population bottlenecks, are expected to reduce genetic variation, and correspondingly, adaptive potential. For these reasons, it is important to understand how bottlenecks interact with insular population dynamics to affect genetic diversity. I used a combination of a laboratory model experiment and population genetics study of an *in situ* bottleneck in an endangered species to investigate how quantitative and molecular genetic variation are affected during bottlenecks. I used a laboratory animal model (red flour beetle, *Tribolium castaneum*) to compare how quantitative genetic variation is affected if a serial bottleneck

occurs in a novel versus familiar environment. The experiment was designed to model a founder event or translocation to a new island with a novel environment. I found that phenotypic and additive variance for a quantitative trait were larger following a bottleneck occurring in the novel environment, suggesting that the novel environment could improve adaptive potential in bottlenecked populations. Next, I used molecular genetic markers to assess variation and signatures of selection in the Laysan finch (*Telespiza cantans*), a Hawaiian honeycreeper endemic to a small Northwestern Hawaiian island. Laysan finches experienced a major bottleneck on Laysan in the early 20th century, followed by a translocation and series of founder events as populations were established on the islets of Pearl and Hermes Reef (PHR) in the 1960s – 70s. I found that, contrary to expectation, bottlenecked Laysan finch populations did not show declines in genetic variation and were not differentiated as a result of genetic drift. These results are potentially caused by insular demographic dynamics. I identified loci with extreme differentiation between modern populations, potentially indicating genomic signals of selection. These regions could be important for adaptation to the novel environment on PHR and are candidates for future study.

POPULATION DECLINES AND GENETIC VARIATION: EFFECTS OF SERIAL BOTTLENECKS

by

Taylor Eilers Callicrate

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee

Professor Jiuzhou Song, Chair
Professor Michael Cummings
Professor Robert C. Fleischer
Professor Thomas D. Kocher
Dr. Frank Siewerdt, Special Faculty

© Copyright by

Taylor Eilers Callicrate

2015

Acknowledgements

Thanks to my committee members (Thomas Kocher, Michael Cummings, Jiuzhou Song, Frank Siewerdt, and Rob Fleischer) for your advice and feedback and patience throughout this process.

Thanks to all our collaborators: Sheila Conant, who shared Laysan finch data and expertise; and Jim Thomas, Jim Mullikin, and Rebecca Dikow who were all essential for the amakihi genome paper.

Thanks to all the lab members at CCEG, the Siewerdt 'labbers', and the Song lab for camaraderie and generous help when needed. Thanks for keeping the lab running smoothly, Nancy!

Thanks to the University of Maryland and Smithsonian for funding and invaluable training, both for me and for this project.

Thanks to the museums who provided Laysan finch samples for this project: National Museum of Natural History, American Museum of Natural History, Field Museum of Natural History, Royal Ontario Museum, University of Michigan Museum of Zoology, and California Academy of Sciences.

These particular people helped in too many ways to list, but I want to make sure they know I am grateful:

Rob Fleischer, thanks for giving me the opportunity to work at the Smithsonian and for the introduction to Hawaii and its wonderful birds.

Frank Siewerdt, thanks for being a patient and encouraging teacher and giving me a start in quantitative genetics.

Jiuzhou Song, thanks for welcoming me into your lab and supporting me at University of Maryland.

Joel Callicrate, thanks for putting up with and even encouraging all of this.

Janess Eilers and Bruce Boatwright, the details are in the soup.

Table of Contents

Acknowledgements ii

Table of Contents iv

List of Tables vi

List of Figures vii

General Introduction 1

Chapter 1 Bottlenecks, Genetic Diversity, and Adaptive Potential 6

- **Bottleneck effects on genetic diversity** 6
- **Bottleneck effects on fitness** 9
- **Adaptive potential** 10
- **Adaptation following bottlenecks** 12

Chapter 2 Exposure to a novel environment in conjunction with serial bottlenecks increases phenotypic and additive variation of a quantitative trait 15

- **Abstract** 15
- **Introduction** 16
- **Materials and Methods** 18
- **Results** 24
- **Discussion** 25

Chapter 3 Genomic Resources for the endangered Hawaiian honeycreepers 31

- **Abstract** 31
- **Introduction** 32

Materials and Methods 35

Results 42

Discussion 57

Chapter 4 Serial bottlenecks in an endangered insular passerine, the Laysan finch, *Telespiza cantans* 61

Abstract 61

Introduction 62

Methods 67

Results 79

Discussion 83

Supplementary Material 92

Chapter 5 Identification of outlier SNP variants in populations of an endangered Hawaiian honeycreeper, the Laysan finch 97

Abstract 97

Introduction 97

Materials & Methods 101

Results 108

Discussion 115

General Conclusion 119

Appendix A. Sample information 122

References 135

List of Tables

Table 2-1. Placement criteria for each generation. 21

Table 2-2. Variance component estimates from the REML analysis. 25

Table 2-3. Nutritional information for bran and flour. 26

Table 3-1. Summary of input for genome assembly. 42

Table 3-2 Alignment statistics for zebra finch and amakihi against amakihi genome. 44

Table 3-3. Nucleotide diversity by chromosome. 47

Table 3-4. RAD read counts. 52

Table 3-5. Stacks results after quality filtering. 53

Table 3-6. SNP sites discovered by comparison to the honeycreeper reference. 56

Table 4-1. Laysan finch museum samples passing quality filters. 72

Table 4-2. Laysan finch modern samples passing quality filters. 73

Table 4-3. Sample size and heterozygosity for Laysan finch modern populations and museum samples.. 79

Table 4-4. Pairwise population Fst values. All P > 0.05.. 83

Table 4-5. Bayesian inference estimates of Θ and migration using 173 DNA sequences from Laysan finch capture. 83

Table 5-1. Blast hits for outlier loci detected by Lositan... 114

Table A-1. Modern Laysan finch samples. 122

Table A-2. Museum Laysan finch samples.. 131

List of Figures

Figure 3-1. RAD coverage of amakihi chromosomes. 46

Figure 3-2. Patterns of heterozygosity across amakihi chromosomes.. 51

Figure 3-3. Relationship between relative chromosome size and RAD density............... 55

Figure 3-4. Reconstructed maximum likelihood tree of relationships of the five study species based on RAD sequences. 57

Figure 4-1. Colonization of Pearl & Hermes Reef Islets.. 65

Figure 5-1. LOSITAN outlier detection.. 105

Figure 5-2. Fst values for 11,527 SNPs across the Laysan finch genome..................... 109

Figure 5-4. Plot of the first two eigenvectors from DAPC for 51 directionally selected SNPs.. 110

Figure 5-6. Individual loadings for the first three principal components from PCA using 51 directionally selected SNPs.. 112

General Introduction

Demographic bottlenecks, or population reductions, can result in loss of genetic diversity and reduced fitness through both chance fixation of detrimental alleles and inbreeding depression as a result of reducing the number of breeding individuals in a population. Genetic diversity is reduced due to the effects of random sampling, or genetic drift, resulting in decreased opportunity for adaptation through natural selection. Bottlenecks affect not only molecular variation but also can impact phenotypic trait diversity, the additive, epistatic and dominance variation balance of specific traits, and adaptive or selection potential of the population (Mayr 1966, Nei et al. 1975, Goodnight 1987, Bryant and Meffert 1988, Carson and Wisotzkey 1989). Loss of genetic diversity is especially important when considering the need to adapt to novel circumstances (Fisher 1958), making bottlenecks a significant issue for conservation of endangered species (Bouzat 2010).

Although theoretical and laboratory studies have been used to predict the genetic diversity outcomes of bottlenecks, study of naturally bottlenecked populations is important because of the complex ecological context in which natural bottlenecks occur (Keller et al. 2001). Environmental pressures may increase the effects of selection resulting in non-random survival of individuals, and factors such as dispersal and generation interval can affect the magnitude of genetic diversity lost (Nei et al. 1975). Therefore, the genetic effects of bottlenecks in natural populations may deviate from general theoretical and experimental predictions (Dinerstein and McCracken 1990, Hartl

and Hell 1994). Because of the chance nature of loss of genetic variation in a bottleneck, the consequences will be unique for each population (Mayr 1966), depending on which alleles and haplotypes remain and their effects on the adaptive potential of the population. Changes in allele frequencies due to drift and inbreeding can alter epistasis and dominance interactions for affected traits (van Buskirk and Willi 2006). An additional consideration is whether the bottleneck takes place within the existing environment, as in a population crash, or involves a founder event to a new environment with different selective pressures. Because adaptation is likely to occur largely as a result of standing genetic variation (Barrett and Schluter 2008), it is critical to understand how standing variation is affected by bottlenecks.

Bottleneck effects on genetic diversity have frequently been studied using neutral molecular markers such as allozymes, microsatellites and single nucleotide polymorphisms (SNPs) to determine how bottlenecks may contribute to population divergence or reduce adaptability. Diversity in neutral markers, measured as heterozygosity or allelic diversity, may be proportional to variation in quantitative trait loci (Falconer and Mackay 1996), which is strongly responsible for adaptability (Franklin 1980, Frankham et al. 2002). Comparison between differentiation in neutral and quantitative or adaptive variation can also be used to determine whether selection or random processes such as drift have had a greater role in population differentiation. Multilocus heterozygosity measured with markers such as microsatellites has been shown to be correlated with fitness traits such as number of offspring produced (Charpentier et al. 2005), territory size and seasonal reproductive success (Seddon et al. 2004), as well as many others (Mitton 1997). Neutral variation is also important because loci or traits

which were neutral in one environment or situation may become adaptive because of the effects of a bottleneck or translocation (Lande 1988). Additionally, when neutral markers are evenly dispersed throughout the genome, they can be used to detect selection or for genome-wide association studies (GWAS), helping researchers find genomic regions which are ecologically important. Such studies can be an important first step in understanding how adaptation takes place and how bottlenecks or founder events affect adaptive variation.

Insular species are subject to a unique set of limitations that could interact with bottleneck effects to lead to different outcomes than for mainland species. In particular, islands typically support smaller population sizes (Pimm et al. 1988), which has implications for the magnitude of the effect of genetic drift and the effectiveness of natural selection (Frankham 1998). Since some of the most prominent genetic effects of bottlenecks are due to genetic drift- chance fixation of detrimental alleles and loss of genetic diversity- these effects could be exacerbated in an island population. Island populations are known to be at increased risk of extinction due to stochastic factors, both environmental and genetic (Steadman 2006, Jones and Merton 2012); the occurrence of a demographic bottleneck would likely increase risk from both types of factors. Perhaps paradoxically, islands are also home to adaptive radiations such as the Hawaiian honeycreepers (Amadon 1950), a testament to the magnitude of diversity that can come out of a very small founder group (Raikow 1977, Johnson et al. 1989) (and proportionately small pool of genetic variation). Understanding how genetic variation and adaptive potential of insular species respond to population decline will contribute to our

understanding of evolutionary mechanisms on islands and will help guide conservation strategies for these species.

Here, we investigate bottleneck effects on quantitative and molecular genetic variation of insular species using a combination of experimental and *in situ* bottleneck studies. An animal model is used to determine bottleneck effects on quantitative trait variation in serial bottlenecks with a novel environment (Callicrate et al. 2012). Using a laboratory model provides the opportunity for replication of the bottlenecked population and manipulation of the environmental conditions, and allows us to calculate how the additive genetic variance of a phenotypic trait is affected by the bottleneck. Using the red flour beetle (*Tribolium castaneum*) as a model allows for following specific populations or lines through many generations quickly (typically up to 13 generations per year) and inexpensively exploring the outcomes of various population demographics questions. In this model, pupae weight in replicate bottlenecked founder populations of beetles is measured to investigate the effects of the bottleneck on phenotypic mean and variance, and to determine how widely these effects vary between replicated source and founder populations. We model a sustained bottleneck with controlled population growth as might occur in an insular species and compare results for bottlenecks in a familiar and a novel environment. Genetic drift is measured as the variance among replicated populations established with individuals with the same initial degree of coancestry (Falconer and Mackay 1996). T. Callicrate participated in the experimental design and carried out all laboratory work, data analysis, and writing.

The *in situ* bottleneck component uses genomic markers (Callicrate et al. 2014) to quantify molecular genetic variation in a serially bottleneck endangered passerine, the

Laysan finch (*Telespiza cantans*). Genomic resources for Hawaiian honeycreepers, the adaptive radiation to which Laysan finches belong, were developed using whole-genome sequencing for one individual (Hawaii amakihi, *Hemignathus virens*) and reduced representation sequencing for six additional individuals of four species (palila *Loxioides bailleui*, Nihoa finch *Telespiza ultima*, apapane *Himatione sanguinea*, and iiwi *Vestiaria coccinea*). For development of honeycreeper genomic resources, T. Callicrate extracted DNA for whole genome sequencing, did all the laboratory work and data analysis for the reduced representation sequencing and honeycreeper SNP discovery, participated in analysis of the genome, and wrote the majority of the manuscript. The population of Laysan finches on Laysan Island, the only existing population known historically, experienced a very severe bottleneck in the late 19th century – early 20th century, followed by a series of translocations (founder events) when back-up populations were established on Pearl & Hermes Reef approximately 500 km away. Using museum specimens and modern samples, we compare genetic variation in Laysan finches through time and across the spread of translocated populations to identify bottleneck and founder event effects on genetic variation (Callicrate et al. n.d.). T. Callicrate participated in the research plan design and carried out the laboratory work, data analysis, and writing.

Chapter 1
Bottlenecks, Genetic Diversity, and Adaptive Potential

Bottleneck effects on genetic diversity

Genetic effects of bottlenecks have been well-examined through the use of computer or mathematical models and laboratory studies. Studies of free-living populations have resulted in support for many of the predictions. Comparison between modern and historical or ancient samples has been carried out for several species which experienced bottlenecks, for example black-footed ferrets (Wisely et al. 2002), Seychelles kestrels (Groombridge et al. 2009) and Mauritius kestrels (Nichols et al. 2001). These studies have successfully compared genetic diversity between pre-bottleneck and contemporary populations (with many finding a loss of variation post-bottleneck) or differentiation between historical and/or modern subpoulations. Historical or ancient DNA samples have also been used to infer prehistoric bottlenecks (Paxinos et al. 2002, Campos et al. 2010) and to estimate historic effective population size (Nichols et al. 2001) or range (Shepherd and Lambert 2008). Generally, the rate of loss of genetic diversity will depend on the effective size of the bottlenecked population, the duration of the restriction (number of generations), mutation rate and the population growth rate. In addition, it has been shown that bottlenecks affect the various measures of genetic diversity differently.

Average heterozygosity per locus is expected to decrease during a bottleneck (although relatively slowly compared to loss of alleles), and this effect depends primarily

on the bottleneck population size and population growth rate (Wright 1931, Chakraborty and Nei 1977, Maruyama and Fuerst 1985). Populations with a larger founder number should experience less overall loss of heterozygosity than those with fewer founders, unless they have a very low population growth rate, in which case a significant portion of original heterozygosity is still expected to be lost. Generally, populations with higher growth rates are expected to lose less of their original heterozygosity, even if they experience a severe bottleneck (Nei et al. 1975). The contemporary population of Mauritius kestrels (*Falco punctatus*), which has experienced a very severe (low founder number), sustained (low population growth) bottleneck, has 50% lower heterozygosity at microsatellite loci than the pre-bottleneck population (Groombridge et al. 2000). Once heterozygosity has been lost, it is predicted to take many generations for the original level to be attained, recovering more slowly than allelic diversity (Chakraborty and Nei 1977). However, random changes in allele frequencies (genetic drift) may result in increased heterozygosity following a bottleneck (Leberg 1992, Spencer et al. 2000). As the population increases after a bottleneck, new mutations are expected to result in increasing heterozygosity (Wright 1931).

Allelic diversity, or average number of alleles per locus, is more sensitive to bottleneck effects because alleles are lost more rapidly than heterozygosity, with the magnitude of loss depending primarily on number of founders or bottleneck survivors (Nei et al. 1975, Leberg 1992, Keller et al. 2001). For example, three populations of northern quoll (*Dasyurus hallucatus*) translocated to islands as 'insurance populations' showed lower levels of allelic richness compared to their source populations on the mainland, after three generations (Cardoso et al. 2009). Post-bottleneck populations are

expected to have distinctly different overall allele frequencies from their source and to exhibit a deficiency of alleles (and excess of heterozygotes) until the effects of mutation can increase diversity (Maruyama and Fuerst 1985, Hedrick 2005). Because of the accumulation of new mutations, allelic diversity is expected to increase as a population expands; during the growth phase, there may even appear to be an excess of alleles when compared to heterozygosity (Maruyama and Fuerst 1984). Loss of allelic diversity can be easy to see through multiallelic neutral markers (Amos and Harwood 1998), as in the case of the in the Galápagos hawk (*Buteo glapagoensis).* This species, comprising several small island populations, exhibits low allelic richness in minisatellite alleles, likely because the populations originated from a single founder event by a mainland hawk, followed by colonization of each island and random genetic drift (Bollmer et al. 2005).

Populations founded by a small number of individuals may differentiate genetically and phenotypically from their source population (Lande 1976, 1980). Genetic distance between source and founder populations is expected to initially increase rapidly after a founder event occurs, with a more pronounced effect when the bottleneck population size (founder size) is smaller. When populations have small founder numbers, founder effects are expected to play a major role in any genetic differentiation that occurs because of the random inclusion of only a few genomes in the new population, a process known as genetic drift (Frankham et al. 2002). The bottleneck effect on genetic distance is also dependent on the average heterozygosity of the founding population, with a larger effect for populations with lower heterozygosity. However, once populations reach an equilibrium size and heterozygosity levels rise, the bottleneck effect on genetic distance is predicted to disappear (Chakraborty and Nei 1977, Hedrick 2005).

In populations which have experienced sequential founder events, there may be an increase in genetic (Clegg et al. 2002a) and phenotypic (Bryant and Meffert 1996, Clegg et al. 2002b) differentiation as populations are separated by a greater number of founder events (in addition to a decrease in allelic diversity). Introduction of unrelated individuals and rotation of individuals between small, isolated populations may be implemented to help ensure that genetic variation is maintained between populations and that populations do not diverge to the point where they are no longer useful as mutual reserves (Frankham et al. 2002, Cardoso et al. 2009). Recently established small populations are at increased risk of extinction due to low initial genetic variation and lack of management strategies to maintain variation can exacerbate this risk (Frankham 2005, Bradshaw et al. 2007).

Bottleneck effects on fitness

Fitness can be affected by bottlenecks in several ways. Reduced genetic diversity may directly impact fitness, as in the case where the heterozygote has a selective advantage over homozygotes but heterozygosity of the population is lowered due to drift during a bottleneck (Amos and Balmford 2001). Reduced population size provides the opportunity for many detrimental alleles of small effect to become fixed by chance, decreasing mean population fitness. Additionally, inbreeding resulting from the population size reduction can exacerbate this effect by further increasing homozygosity (Hedrick and Kalinowski 2000). Reduced genetic variation lowers a population's ability to respond to selective pressure (Falconer and Mackay 1996), and more highly inbred populations or individuals may be less likely to survive in harsh or extreme situations (Amos and Balmford 2001).

In populations with large historical sizes, there can be a substantial build-up of small-effect detrimental alleles and a decrease in population size may cause these to drift to higher frequencies with direct fitness consequences, and potentially even leading to extinction (Hedrick 2001, 2004). On the other hand, detrimental alleles with large or medium effects on fitness may be purged from bottlenecked populations through natural selection due to differential survival or reproductive fitness, potentially lessening the effects of inbreeding depression (Templeton and Read 1984). Environmental conditions which the population is experiencing also can alter the effect of inbreeding depression on survival and fitness (Bouzat 2010).

Adaptive potential

Low levels of genetic diversity provide reduced material upon which adaptive selection may act, increasing a population's risk of extinction because of inadaptability (Fisher 1958, Frankham 2005). Reduced ability to respond to selection pressures has been demonstrated in laboratory experiments with *Drosophila melanogaster* in which populations undergoing more severe bottlenecks showed a more severely decreased response to artificial selection. The reduced selection response was likely related to low remaining levels of additive genetic variation (Swindell and Bouzat 2005). Phenotypic and quantitative traits are particularly susceptible to the random processes acting on the genome during a bottleneck event since such traits are affected by multiple loci and the disruption of genetic relationships brought about by a bottleneck can affect evolutionary potential (Bryant and Meffert 1988, 1990). The random nature of changes in allele frequencies or losses of polymorphism due to drift mean that the outcome of replicated

bottlenecks could include both positive and negative consequences for additive genetic diversity, adaptive potential, and fitness. Changes in dominance and epistatic interactions brought about by drift may result in increased additive variance, but such changes may also disrupt advantageous locus interactions that have been developed through selection in the pre-bottleneck population (Carson and Wisotzkey 1989, Carson 1990, van Buskirk and Willi 2006).

The adaptive potential of genetic variation is especially relevant when novel selective pressures are concerned, or when considering population growth subsequent to the bottleneck. Although variants which conferred a survival advantage in the original situation are unlikely to be lost from a population, even during a bottleneck, under new conditions, variants which had previously been neutral may become advantageous (Amos and Balmford 2001). Adaptation to novelty is critical for endangered species, which face survival challenges due to the effects of disease, introduced competitors, or new environmental conditions encountered because of translocation. Populations with greater genetic variation will have a better probability of surviving these novel obstacles (Hedrick 2001); those which have experienced extreme bottlenecks are less able to handle environmental challenges (Choiniere 2008). The importance of overall levels of variation can also be seen when taking population growth into account. When population size is small, as after a bottleneck, unless the selective advantage of a variant is very strong, drift is likely to have a larger impact than selection on the fate of the variant within the population. This renders many variants effectively neutral. However, as effective population size grows the effect of drift decreases and advantageous variants have the potential to increase in the population (Hedrick 2004).

Adaptation following bottlenecks

When environmental conditions after a bottleneck or founder event are drastically different, populations may begin to adapt to their new environments, provided that there is sufficient genetic variation upon which selection may act. The effect of selection is difficult to predict as it is contingent upon the specific situation (Barton and Mallet 1996). Populations subject to selection may show divergence in phenotypic or quantitative traits beyond that which is likely to be caused by genetic drift. In a series of recent island colonizations by silvereyes (*Zosterops lateralis*), drift alone was not sufficient to explain morphological changes between populations on islands and their mainland sources. Directional selection was determined to be a major factor driving morphological divergence (Clegg et al. 2002b). Selection can have an appreciable effect when a founder population's habitat is different from that of the source population, with environmental differences potentially causing significant genetic or phenotypic divergence between populations even if distance between them permits gene flow. For example, for two populations of blue tits (*Cyanistes*/*Parus caeruleus ogliastrae)* separated by 25 km on the island of Corsica, selection pressures due to their two different habitat types were such that the birds diverged with regard to onset of reproduction. This divergent selection was strong enough to outweigh the effects of gene flow and random factors even though there were no barriers to dispersal between the two habitats (Blondel et al. 1999). Environmental differences in the form of foraging niches can also be a factor driving differentiation despite gene flow. Two avian radiations, the Hawaiian honeycreepers and Darwin's finches in the Galapagos, have developed morphological differences in response to foraging niches (Grant 1994). Darwin's finches are of course a

classic example of how environmental conditions have been demonstrated to have a selective effect on both beak shape and body size (Grant and Grant 2002). The black-bellied African seedcracker (*Pyrenestes ostrinus*) is another case where there has been differentiation in beak morphology based on foraging niche, despite large amounts of interbreeding between three different beak morphs (Smith 1993).

Comparison of differentiation in quantitative traits to differentiation in neutral genetic markers can be used to determine the relative importance of selection and random processes in phenotypic or quantitative divergence between populations. This approach was taken for a study of yellow dung fly (*Scathophaga stercoraria*) populations across Europe (Demont et al. 2008). The authors compared variation in two molecular markers, allozymes and microsatellites, to variation in quantitative traits relevant to an environmental gradient across the populations and found that populations were differentiated when assessed by their quantitative traits, but not by neutral molecular markers. This result was interpreted to indicate that selection was a stronger driving force in population differentiation than random processes.

When compared to an appropriate null population genetics model, differences in allele frequencies between populations can be used to identify adaptive selection at specific loci or closely linked sites (Lewontin and Krakauer 1973). However, it is difficult to discriminate between differences which are due to random factors such as drift and differences which are actually due to selection. A measure such as F_{ST} can be used for such comparisons (Beaumont and Balding 2004). The distribution of F_{ST} values under neutral processes can be modeled so that there is a comparison point for observed differences in allele frequency between populations. If natural selection is acting on a

locus and favors a certain allele in one population over the others, F_{ST} will be higher than expected, whereas if the heterozygote is favored, F_{ST} will be lower than expected (Bowcock et al. 1991)(Bowcock et al. 1991). The ability to detect selection is dependent upon the selection coefficient, sample size, population size, and demographic parameters (Oleksyk et al. 2010). For example, fluctuations in population size may inflate variation in $F_{ST,}$ leading to detection of false positives for selection. Therefore, it is useful for a model to incorporate the evolutionary history of populations, if it is known (Bowcock et al. 1991, Narum and Hess 2011). Methods using Bayesian regression and summary statistics are reliable for detecting modeled loci undergoing positive selection, although balancing selection is more difficult to detect (Beaumont and Balding 2004, Foll and Gaggiotti 2008). Finding genomic regions which have recently been subject to selection could be a first step in investigating how genotype underlies phenotypic traits important for adaptation to a novel environment (Brumfield et al. 2003, Luikart et al. 2003, Morin et al. 2004, Mitchell-Olds et al. 2007).

Chapter 2
Exposure to a novel environment in conjunction with serial bottlenecks increases phenotypic and additive variation of a quantitative trait

Abstract

Demographic bottlenecks were replicated in laboratory conditions using *Tribolium castaneum* in order to determine how additive genetic variance, and therefore adaptive potential, may be affected by a novel environment. A two-level bottleneck was imposed. The first level was a 100-individual bottleneck (five replicates). After three generations at population size 100, the second bottleneck consisted of a single mating pair (15 sub-replicates stemming from each first level replicate). Two growth media (environments), standard (wheat flour) or novel (wheat bran), were used in the sub-replicates to simulate translocation to an environment with a different foraging substrate. Sub-replicates were managed for slow population growth during six generations to population size 50 and were thus maintained for nine additional generations.

Variance in pupa weight in generation 16 was affected by drift from the first and second level bottlenecks (flour: N=382, bran: N=470 pupae, both P<0.0001 at each bottleneck level). Total phenotypic variance was determined from the progeny of the single pair matings and was larger for bottlenecks occurring in bran than in flour, P<0.01 ($F_{469,381}$ = 1.29). Additive variance was also larger in bran than flour bottlenecks, P<0.01 ($F_{469,381}$ = 1.38). These results indicate that bottlenecks occurring in a novel environment could potentially increase the effectiveness of selection by bolstering additive variance.

Introduction

Reduced additive variation limits a population's ability to respond to selection and is therefore of great concern for conservation biologists (Lande 1988, Willi et al. 2006). The ability to adapt is highly relevant for endangered species facing threats of disease, changing climate and environmental degradation. These species are also likely to experience translocation, which may involve environmental changes such as novel food resources, requiring further adaptation. Understanding how the demographic challenges faced by endangered species affect their additive variation will help predict their potential to adapt and evolve and guide conservation decisions (Frankham 1999).

Endangered species have experienced demographic bottlenecks, or severe population declines, followed by recovery to a larger population size. Some species or populations also have experienced founder events when several individuals are translocated or taken into captivity; such founder events are effectively demographic bottlenecks. Theory predicts that severe bottlenecks will reduce additive variance within populations in proportion to the inbreeding generated by the bottleneck (Wright 1951, Crow and Kimura 1970), which is dependent upon the magnitude of the bottleneck and the population's growth rate (Nei et al. 1975). Reduced additive variance following bottlenecks has been supported by some laboratory experiments (Saccheri et al. 2001, Franklin and Siewerdt 2011), especially for non life-history traits such as color or behavior (reviewed in van Buskirk and Willi, 2006), while other studies have found that additive variance does not decrease as much as expected or even increases (Bryant et al. 1986, Lopez-Fanjul and Villaverde 1989, Wade et al. 1996, Cheverud et al. 1999). Reduced ability to respond to selection following a bottleneck or inbreeding has been

demonstrated (Wade et al. 1996, Reed et al. 2003, Swindell and Bouzat 2005), as has reduced heritability (Franklin and Siewerdt, 2011), which would lead to reduced ability to respond to selection. From the mixed results of these studies, it is clear that the effect of bottlenecks on quantitative variation and adaptive potential is situation-specific and depends on many factors, potentially including genetic architecture of the trait in question (Barton et al. 2004), level of inbreeding generated (Cheverud et al. 1999), bottleneck size (Swindell and Bouzat, 2005), and duration (Wade et al. 1996).

Demographic bottlenecks were replicated in laboratory conditions using *Tribolium castaneum* in order to determine how additive genetic variance, and therefore adaptive potential, may be affected by a novel environment. A two-level bottleneck was imposed. The first level was a 100-individual bottleneck (five replicates). After three generations at population size 100, the second bottleneck consisted of a single mating pair (15 sub-replicates stemming from each first level replicate). Two growth media (environments), standard (wheat flour) or novel (wheat bran), were used in the sub-replicates to simulate translocation to an environment with a different foraging substrate. Sub-replicates were managed for slow population growth during six generations to population size 50 and were thus maintained for nine additional generations.

Variance in pupa weight in generation 16 was affected by drift from the first and second level bottlenecks (flour: N=382, bran: N=470 pupae, both P<0.0001 at each bottleneck level). Total phenotypic variance was determined from the progeny of the single pair matings and was larger for bottlenecks occurring in bran than in flour, P<0.01 ($F_{469,381}$ = 1.29). Additive variance was also larger in bran than flour bottlenecks, P<0.01

($F_{469,381} = 1.38$). These results indicate that bottlenecks occurring in a novel environment could potentially increase the effectiveness of selection by bolstering additive variance.

Materials and Methods

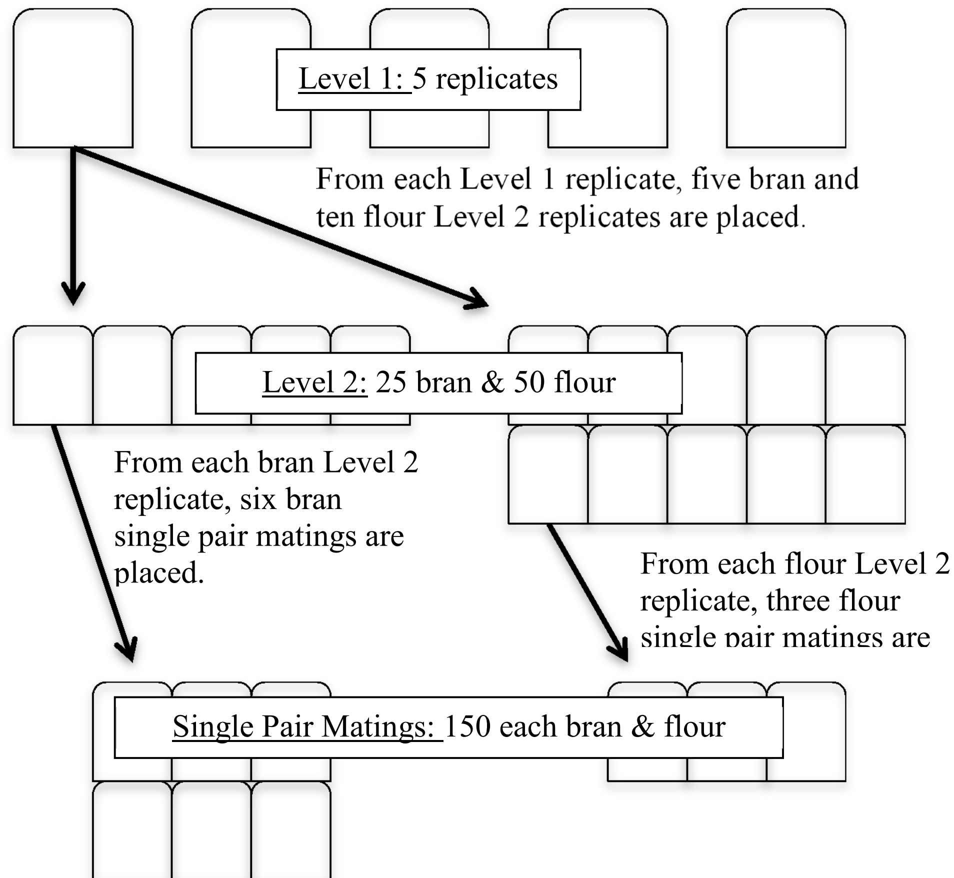
Throughout the study, all beetles were housed in a walk-in incubator kept at temperature 33 ± 1 °C and humidity of 70 ± 2%.

For the first level bottleneck beetles from four distinct, reproductively isolated populations were combined in equal proportions to create five replicate populations of 100 individuals (Figure 2-1). These populations were housed in canning jars containing three tablespoons of a medium composed of all-purpose flour (General Mills, MN, USA) mixed with 5% brewer's yeast (Twinlab, MI, USA) by weight. After four weeks, 100 F1 pupae from each population were selected at random and placed in a new jar. This step was repeated to place generations F2 and F3.

Randomly selected progeny of each F3 first level replicate were used to place second level bottleneck replicates. An additional 20 female, 20 male, and 75 straight run pupae were collected from the F3 progeny of each first level replicate. Each group (females, males or straight run) was housed in a small glass jar (three jars per first level replicate) and those individuals were used as migrants into second level bottleneck populations at generation 3. First level replicates were maintained at population size 100 for the remainder of the experiment, with a new placement every four weeks.

From each first level replicate, 15 second level bottleneck replicates were placed (Figure 2-1): ten in flour medium and five in bran (Bob's Red Mill, OR, USA) with 5% brewer's yeast by weight (bran medium), for a total of 75 second level replicates in the study. All second level replicates were housed in small glass jars with one teaspoon of medium (flour or bran). Each second level replicate started with a founding pair consisting of one male and one female pupa from the associated first level replicate. This was considered second level generation 1 (G1). New generations were placed every four weeks. However, as the experiment progressed, development of bran replicates began to lag behind flour replicates. Therefore, G15 flour replicates were placed 28 days after G14, while G15 bran replicates were placed 30 days after G14.

Figure 2-1. Bottleneck replication diagram for the *Tribolium* model.



The placement scheme was designed to mimic gradual population growth with new recruits only joining the breeding population after they reached a mature age. At the end of every generation, the next generation was placed using surviving adults first, then pupae if necessary to achieve the desired population size. Population growth was kept to a rate of one recruit per adult per generation (i.e., the number of recruits collected at the end of a given generation was the adult population size at that generation). Pupae were randomly chosen at the end of each generation to be recruits to their population and were kept in individual jars in either flour or bran medium for one generation before being added into their population with the placement of the next generation. Starting in G4, in some cases there were not enough pupa and larva to provide the correct number of recruits. When that situation occurred, all of the available pupa and larva were collected. Once the population reached the desired stable population size of 50 individuals (including adults and juveniles), collection of new recruits stopped and placement of each subsequent generation used adults first and pupae as necessary to achieve a population size of 50. Populations were kept at this size for nine generations to increase the relevance of the results to captive and translocated populations, which are often limited to a small size after a bottleneck event. When placing a new generation, if there were not enough individuals to achieve the target population size in a given replicate, that replicate would be placed using all available individuals.

Generations G1 and G2 were combined into one placement cycle because, according to the placement scheme, at the end of G1 only the two original second level adults would be placed for G2, and two pupae would be collected to join the population as recruits at the placement of G3. Because there was only one pair in the population for

the first two generations, the same effect was achieved by placing the two original adults, plus two pupae (recruits from G1) as G3, while keeping two additional pupae (progeny of G2) to add as recruits during G4. At the placement of G3, migrants were added from the three groups of pupae collected from the progeny of first level F3. One male, one female and four straight run were added to each second level replicate from the groups collected from the associated first level replicate. Population size at each generation and details of each generation's placement can be found in Table 2-1.

Table 2-1. Placement criteria for each generation.

---

1. Generations 1 & 2
    a. For each source population's replicates
        i. Add 1 male, 1 female to each of 10 flour replicates
        ii. Add 1 male, 1 female to each of 5 bran replicates
        b. For each source population also collect pupae to represent 6 immigrants to add during generation 3
    c. Adult population = 2; total population = 6
2. Generation 3
    a. Into each replicate add
        i. 2 adults from previous generation
        ii. 2 pupae from previous generation's offspring, representing offspring of generation 1
        iii. Add 6 random immigrants collected from source populations during previous generation (1 male, 1 female, 4 straight run)
        b. Collect 2 pupae to be added when generation 4 is placed
    c. Adult population = 10; total population = 12
3. Generation 4
    a. Into each replicate add
        i. 10 adults from previous generation
        ii. 2 individuals collected when generation 3 was placed
        b. Collect 10 pupae to be added when generation 5 is placed
    c. Adult population = 12; total population = 22
4. Generation 5
    a. Into each replicate add
        i. 12 adults from previous generation
        ii. 10 individuals collected when generation 4 was placed

        b. Collect 12 pupae to be added when generation 6 is placed
    c. Adult population = 22; total population = 34
5. Generation 6
    a. Into each replicate add
        i. 22 adults from previous generation
        ii. 12 individuals collected when generation 5 was placed
        iii. 16 pupae from previous generation
    b. Total population = 50
6. Generations 7 to 15
        a. Into each replicate add 50 adults
        b. Total population = 50

After nine generations of stable population size, progeny of G15 were used to set up single pair matings (SPM, Figure 2-1). For each flour medium replicate, three single pair replicates were placed in flour; for each bran replicate, six single pair replicates were placed in bran. If there were not enough pupae available from a given replicate to place the desired number of single pair matings, the maximum possible number were placed. Flour SPM were placed 28 days after G15 was placed and bran SPM were placed 30 days after G15 was placed.

Pupae weights were taken at the initial placement of second level bottleneck replicates (both pupae per replicate), at the end of the population growth phase (five G6 pupae per replicate) and at the end of the experiment (five G15 pupae per replicate). Four pupae chosen randomly from each single pair mating were also weighed, 28 days after the placement of SPM for flour and 30 days after for bran. If there were less than the desired number of pupae for a given replicate, all available pupae were weighed. If no pupae were present, a missing value was recorded for that replicate. Weights were measured to the one thousandth of one milligram.

Mean pupae weights per replicate were analyzed using SAS v.9 (SAS Institute, NC, USA). The Shapiro-Wilk test and visual examination of normal probability plots were used to verify whether pupae weights conformed to a normal distribution, which they did. A general linear model (GLM) with medium type as a fixed factor and first level replicate as a random factor was used to investigate whether either of these factors had an effect on mean pupae weight at second level G1 or at the end of the population growth phase (second level G6). Type I sum of squares from a GLM with first level replicate, second level replicate within first level replicate and single pair within second level replicate as random factors were used to evaluate the effect of drift caused by each of the two bottleneck levels on phenotypic variance for pupa weight. Weight of progeny of the single pair mating was the dependent variable.

A restricted maximum likelihood (REML) variance components analysis with first level replicate, second level replicate within first level replicate and single pair within second level replicate as factors was used to determine the magnitude of phenotypic and additive variance for each medium type. The REML was performed separately for each medium type and Type I estimates of variance components were summed to give an estimate of the phenotypic variance for that medium type. Pupae weight for progeny of the single pair matings was the dependent variable. To determine if there was a significant difference in phenotypic variance by medium type, a folded F test was used.

To calculate the additive variance for pupa weight in the progeny of the single pair matings, the Type I variance component estimate for single pair within second level replication was multiplied by two (Becker 1984, Lynch 1988, Falconer and Mackay

1996). This variance component is equivalent to the covariance of full sibs, which estimates one half of the additive variance.

Results

During the course of the experiment several second level replicates went extinct. Reasons for this included sexing errors, the death of one individual when placing single pairs, the death of all adults without pupae to replace them, or moldy medium killing pupae. Overall, 65 second level replicates remained at the end of the experiment (41 flour, 24 bran) and 852 progeny of single pair matings (382 flour, 470 bran) were weighed for analysis of phenotypic and additive variance effects of the bottlenecks.

There was no significant difference in mean pupa weights from different first level replicates at the beginning of the second level (N=75), at the end of the population growth phase (N=51) or at the end of the experiment (N=61). However, medium type had an effect on mean pupa weight at the end of the population growth phase (N=51, P<0.0001) and at the end of the experiment (N=61, P<0.0001) with pupa raised in flour being larger in both cases. Therefore, subsequent analysis was carried out separately for flour and bran replicates. For both flour and bran replicates, there were effects of drift on phenotypic variance for both first level and second level of the bottleneck (flour: N=382, bran: N=470; P<0.0001 for both medium types and bottleneck levels).

Total phenotypic variance was determined from the progeny of the single pair matings and was 0.07349 mg^2 for flour and 0.09507 mg^2 for bran. Variance component estimates for each factor in the model are shown in Table 2-2. The variance in pupa weight of single pair progeny for bran was larger than for flour at the P<0.01 level

($F_{469,381}$ = 1.29), as was the additive variance (flour V_A = 0.02058 mg^2; bran V_A = 0.02844 mg^2; $F_{469,381}$ = 1.38). Heritability for flour, calculated as V_A/V_P, was approximately 28% (SE: 0.10), while for bran it was approximately 30% (SE: 0.10).

Table 2-2. Variance component estimates from the REML analysis.

Variance Component	Estimate	Flour Values (mg^2)	Bran Values (mg^2)
L1	Drift, Bottleneck	0.0048	0.0019
L2 within L1	Drift, Bottleneck	0.0159	0.0071
SP within L2	½ VA	0.0103	0.0142
Error	Error	0.0426	0.0719

Discussion

Understanding whether environmental change concurrent with bottlenecks affects the outcome for additive variation provides information about evolutionary processes and is useful for management of endangered species. In this study, pupa weight, additive genetic variance and phenotypic variance were all significantly affected by an environmental change representing altered foraging substrate. However, novel medium did not alter the effect of drift at either bottleneck level. Phenotypic variances increased between replicates, as predicted (Wade et al. 1996), for both levels of the bottleneck and both medium types.

Novel medium type resulted in smaller mean pupa weight by the end of the population growth phase, and this effect was also present after nine more generations during which the population was maintained at constant small size. Another study found no significant difference in pupae weights for *Tribolium castaneum* raised on flour versus bran, although in that study only a single generation of non-bottlenecked pupae were

measured (White et al. 2000) as opposed to sixteen generations in the present study. It is unlikely that drift alone could have caused the size differential between flour and bran replicates, since it is expected to change trait means in random directions, resulting in no net change when many replicates are compared (Hill 1972). Smaller pupae weights could have been due to a growth constraint imposed by the bran over time, since in this study bottlenecks occurred in bran over a number of generations. Bran is less nutritionally dense than flour (Table 2-3), so an individual consuming the same volume of bran versus flour takes in less nutrients. Total amount of food did not appear to be a limiting factor as the authors never observed a case where all of the bran, or flour, had been consumed during a generation, but the possibility cannot be ruled out that bran-raised beetles consumed less total nutrients than flour-raised beetles.

Table 2-3. Nutritional information for bran and flour. Serving size for both items: ¼ cup.

Amount per Serving	Flour	Bran
Calories	100	50
Total fat	0 g	0.5 g
Total carbohydrates	22 g	10 g
Dietary fiber	1 g	6 g
Sugars	1 g	0 g
Protein	3 g	2 g

In the present study, the authors also observed that bran replicates had a slower development period in the later generations of the experiment, requiring 30 days to produce pupae by G15. This was two days longer than the time required for flour replicates. Although longer developmental period has been suggested to represent reduced fitness in *Tribolium castaneum*, this association was made in conjunction with increased inbreeding accumulation (Franklin and Siewerdt, 2011). Flour and bran

replicates were maintained at the same population sizes and population growth rates, so inbreeding accumulation was the same for both groups. The longer developmental period was likely not associated with inbreeding-induced fitness reductions. Additionally, a single generation study measuring development rates of *Tribolium* beetles from larva to adulthood in different grades of wheat products found no difference in developmental period for flour and bran, although they did note a strong feeding preference for flour over bran (Shafique et al. 2006). It is possible that the longer developmental period was an adaptive response that was elicited after a number of generations in the bran selective environment. Further study would be required to investigate this possibility and determine the adaptive optima in bran.

Following the bottlenecks and maintenance at a small population size for nine generations, both additive and phenotypic variances were higher when bottlenecks occurred in the novel medium type. Increased phenotypic variance without increased additive variance would indicate an increase in environmental variance. Such a change would not be useful for adaptation since selection can only act upon heritable variance. In this study, the purpose was not to compare additive variances before and after bottlenecks; rather, the intent was to examine the additive effects when bottlenecks occurred in a novel environment versus a standard one. Bottleneck events alone may increase additive variance in a population beyond predicted theoretical levels because drift alters allele frequencies and can change the additive interactions between loci. Additive effects of recessive alleles may increase if those alleles increase in frequency (Willis and Orr 1993), and changing allele frequencies combined with epistasis could increase overall additive effects (Goodnight 1987, 1988; Bryant and Meffert 1996,

Cheverud et al. 1999). These effects are difficult to predict (Barton et al. 2004), and depend on the genetic architecture of the trait in question (Cheverud et al. 1999), but increase in additive variance of quantitative traits following bottlenecks has been demonstrated experimentally (Bryant et al. 1986, Lopez-Fanjul and Villaverde 1989). In the present study, the observed larger additive variance for novel medium bottlenecks is likely due to a combination of changes in gene frequency caused by the bottlenecks and gene-environment interactions of the novel environment.

Reed et al. (2003) determined that *Drosophila* lineages which were inbred in a variably stressful environment exhibited increased response to selection for fitness following the period of inbreeding compared to lineages inbred in a benign environment. Although the authors did not measure quantitative variance, response to selection is dependent upon heritable variation (Falconer and Mackay, 1996) and they suggested that the variable condition maintained higher levels of genetic variation. In the present study, the novel environment, although not variable, resulted in higher levels of additive variation than the standard environment following bottlenecks. As suggested by Reed *et al.* (2003), this could be the result of the novel bottleneck environment preserving a greater portion of the original variance.

An increase or preservation of additive variance alone is not useful for adaptation in a static environment if there is a single optimum phenotype, because selection will rapidly remove non-advantageous variation. To understand whether an increase in additive variance for a given trait will increase adaptability of a bottlenecked population, it is necessary to understand the fitness landscape for that trait (Barton and Turelli 1989, Barton et al. 2004). In simulations, increasing the phenotypic variance of a trait has the

effect of flattening the mean population fitness function, which makes it easier for the population to move to a higher adaptive peak (Wright 1932, Kirkpatrick 1982). Although the fitness landscape for pupa weight is unknown, because bran represents a novel environment, it is possible that the fitness landscape in bran is different than it would be in the original flour medium. In this case, increased additive and phenotypic variance could increase adaptive potential. As an illustration, suppose there is a multimodal fitness distribution for pupa weight in bran and the population mean starts out near one peak, which is also an optimum in flour. The larger range of available phenotypes resultant from the increased additive variance of the bran bottleneck environment may include outlier individuals with phenotypes closer to another selective peak. At that point, divergent selection for both optimum phenotypes could contribute to increasing genetic variance in each subsequent generation. This effect is predicted to be at a maximum when the population is in the middle between two fitness peaks (Kirkpatrick 1982). In a natural population experiencing a bottleneck in a novel environment, increased phenotypic and additive variance could allow adaptation in such a manner, moving the population towards new optima in the altered fitness landscape (Whitlock 1995), and possibly increasing genetic variance in the process. In this study, because the fitness landscape for pupa weight is unknown, it cannot be determined whether higher additive variance in bran replicates was adaptively advantageous or contributed to by the effects of selection moving the populations to new fitness optima.

The present results are relevant to conservation management of endangered species, especially those that have been translocated or experienced founder events. Bottlenecks were modeled following population size and growth parameters that are

likely to be encountered by endangered populations. The quantitative effects in a novel environment with a different foraging substrate were compared with those in a standard environment. The results indicate that when bottlenecks occurred in a novel environment, higher levels of additive and phenotypic variance were present than for populations bottlenecked in a typical environment. However, heritabilities for bran and flour were not markedly different given that each had a standard error of 10%. The outcome for adaptive potential is dependent on the genetic architecture of a given trait and its fitness landscape in the novel environment. The fitness landscapes for biologically relevant traits in a translocation environment need to be evaluated to determine the adaptive significance of a novel environment preserving additive variance through demographic bottlenecks.

Chapter 3
Genomic Resources for the endangered Hawaiian honeycreepers

Abstract

The Hawaiian honeycreepers are an avian adaptive radiation containing many endangered and extinct species. They display a dramatic range of phenotypic variation and are a model system for studies of evolution, conservation, disease dynamics and population genetics. Development of a genome-scale resources for this group would augment the quality of research focusing on Hawaiian honeycreepers and facilitate comparative avian genomic research.

We assembled the genome sequence of a Hawaii amakihi (*Hemignathus virens*), and identified ~3.9 million single nucleotide polymorphisms (SNPs) in the genome. Using the amakihi genome as a reference, we also identified ~156,000 SNPs in RAD tag (restriction site associated DNA) sequencing of five honeycreeper species (palila [*Loxioides bailleui*], Nihoa finch [*Telespiza ultima*], iiwi [*Vestiaria coccinea*], apapane [*Himatione sanguinea*], and amakihi). SNPs are distributed throughout the amakihi genome, and the individual sequenced shows several large regions of low heterozygosity on chromosomes 1, 5, 6, 8 and 11. SNPs from RAD tag sequencing were also found throughout the genome but were found to be more densely located on microchromosomes, apparently a result of differential distribution of the particular site recognized by restriction enzyme *Bse*XI.

The amakihi genome sequence will be useful for comparative avian genomics research and provides a significant resource for studies in such areas as disease ecology,

evolution, and conservation genetics. The genome sequences will enable mapping of transcriptome data for honeycreepers and comparison of gene sequences between avian taxa. Researchers will be able to use the large number of SNP markers to genotype honeycreepers in regions of interest or across the whole genome. There are enough markers to enable use of methods such as genome-wide association studies (GWAS) that will allow researchers to make connections between phenotypic diversity of honeycreepers and specific genetic variants. Genome-wide markers will also help resolve phylogenetic and population genetic questions in honeycreepers.

Introduction

Avian genome sequences were first obtained for well-studied model systems for which there was a long history of multidisciplinary research, namely the chicken *Gallus gallus* (Genome Sequencing Center 2004) and zebra finch *Taeniopygia guttata* (Warren et al. 2010). But now genomes are starting to appear along lines of interest such as other agricultural species (turkey, *Meleagris gallopavo* (Dalloul et al. 2010)), members of adaptive radiations (Darwin's medium ground finch, *Geospiza magnirostris* (Rands et al. 2013)), species with traits of interest such as vocal learning (budgerigar, *Melopsittacus undulatus* (Ganapathy et al. 2014)) and systems with possible incipient speciation (*Ficedula* flycatchers (Ellegren et al. 2012)). Genome-scale resources for non-traditional model organisms have become a reality over a short period of time, due in a large part to the commercialization of sequencing-by-synthesis (also called next-generation sequencing) technology (Lerner and Fleischer 2010). Initial examinations of these genomes have revealed that there is a high degree of synteny among avian species,

confirming hypotheses from cytogenetic studies (Griffin et al. 2007). Although 40 million years of evolution separate chickens and turkeys, only 30 minor chromosome rearrangements were detected between the two and their karyotypes are strikingly similar (Dalloul et al. 2010). Chicken and zebra finch (perhaps 100 million years diverged (Hackett et al. 2008)) also exhibit a high degree of synteny and conservation of karyotype (Warren et al. 2010). However, recent work shows that small inversions may be common when comparing distantly-related avian taxa (Kawakami et al. 2014).

There are over 5,000 passerine species with many unique traits and adaptations (Barker et al. 2004). Each additional passerine genome (Warren et al. 2010, Ellegren et al. 2012, Rands et al. 2013) that is sequenced offers an opportunity to identify different genes under selection and to elucidate the mechanisms underlying avian adaptations (Rands et al. 2013). The Hawaiian honeycreepers are an endemic Hawaiian passerine adaptive radiation in the Cardueline finch subfamily Drepanidinae (Lerner et al. 2011), and display a tremendous diversity of plumages, beak shapes (some unique to this radiation) and niches (James and Olson 1991). Molecular analyses indicate that the radiation is sister to the Eurasian *Carpodacus* rosefinches, and dates to about 5.7 million years ago (Fleischer et al. 1998, Lerner et al. 2011). Adaptive radiations have long been recognized for their value as evolutionary case studies and their usefulness in understanding adaptive evolutionary processes. The Hawaiian honeycreepers have the special characteristic that the history of their radiation is integrated with the geological history of the Hawaiian Islands. Patterns in honeycreeper divergence appear to be linked to the pattern of island emergence (Lerner et al. 2011), which has been well-documented as part of a volcanic time series (Price and Clague 2002). Because this unusual geology

provides a well-defined timeline, honeycreepers are a good system for estimation of rates of molecular evolution (Fleischer et al. 1998).

Unfortunately, of the 33 described historical honeycreeper species (plus over 17 species known only from subfossil material) (James and Olson 1991), roughly two-thirds are now extinct, largely from human-related impacts such as habitat loss, introduced mammalian predators and vectored pathogens (Banko and Banko 2009). Study of the evolution of disease resistance is an area that will especially benefit from genome-wide markers. In particular, honeycreepers appear extremely susceptible to introduced diseases such as avian malaria (*Plasmodium relictum*) and avian poxvirus, both vectored by an introduced *Culex* mosquito (van Riper et al. 1986, Atkinson et al. 1995, Atkinson and Samuel 2010). Most extant honeycreepers are limited to higher elevations free from mosquitoes and disease (van Riper and Scott 2001). However, a few species, most notably the Hawaii amakihi (*Hemignathus virens*), can survive with chronic malaria infection, exhibiting tolerance or resistance to the disease (Atkinson et al. 2000, Jarvi et al. 2001, Woodworth et al. 2005). A few studies suggest that strong selective pressure from malaria resulted in rapid evolution of disease tolerance in certain low-elevation Hawaii amakihi populations and that resistance may be spreading amongst low-elevation amakihi, although it is unknown whether resistance arose once or simultaneously in multiple source populations (Foster et al. 2007). Understanding the source and mechanism of disease resistance in amakihi is a priority research area using the SNP markers. Such work is needed to improve our strategies for identifying and preserving the most viable populations of many species threatened by invasive pathogens.

Our objective in this study is to characterize the genome of a Hawaiian honeycreeper, the Hawaii amakihi (*Hemignathus virens*), and to develop and assess a set of genome-wide SNP markers to enable both phylogenetics-scale and fine-scale investigations about adaptive evolution and population genetics. We used two sequencing-by-synthesis approaches and then performed a hybrid assembly to create a draft Hawaii amakihi genome sequence. The Hawaii amakihi, in addition to being a member of the honeycreeper adaptive radiation, serves as an ecological model for disease transmission due to its variable responses to infection by avian malaria (Atkinson et al. 2000, Woodworth et al. 2005). The individual selected for the genome sequence had a high level of infection, but had been recaptured several times, indicating persistence despite a chronic, intense malaria infection. To increase the utility of markers for broader topics of study, we combined de-novo genome sequencing with a reduced representation sequencing method (restriction site-associated DNA, or RAD) to identify and map SNP polymorphisms isolated from four additional honeycreeper species. In addition to facilitating research into honeycreeper evolution and disease resistance, the draft amakihi genome will contribute to knowledge of avian genome biology and improve the pool of resources for comparative genomic study.

Materials and Methods

Study samples

A single female amakihi (*Hemignathus virens*) was sequenced for genome assembly (USGS aluminum band 1771-10606, sampled 22 February 2002 at Nanawale, Hawaii Island). Although it has been typically preferred to use an inbred individual for

genome sequencing to simplify assembly, the possibility of high-coverage sequencing-by-synthesis makes it possible to assemble even with potentially high levels of variation (Dalloul et al. 2010). Indeed, when SNP discovery is a major goal it is typically preferred to use an outbred individual. Seven Hawaiian honeycreeper samples were selected for RAD tag sequencing: one iiwi (*Vestiaria coccinea*; female RCF 2682, sampled 8 March 1987 at Kokee State Park, Kauai), two palila (*Loxioides bailleui*; bands 8031-75515 and 8031-75622, sampled in 1993 at Puu Laau, Hawaii Island), one apapane (*Himatione sanguinea*; 1540-45550 sampled at Waikimoi Preserve, Maui), one Hawaii amakihi (the same individual used for genome assembly), and two Nihoa finches (*Telespiza ultima*; bands 1381-62204 and 1381-62194 sampled on Nihoa Island, HI). This selection of honeycreepers covers much of the Drepanidine tree, and includes two redbird species (iiwi, apapane), two finchbill species (Nihoa finch, palila) and a greenbird (amakihi). For a recent phylogeny of Hawaiian honeycreepers, see Lerner et al. *Current Biology* 2011, 21:1838-1844.

DNA isolation

Genomic DNA was extracted from whole blood using proteinase K digestion followed by phenol:chloroform extraction and either ethanol precipitation (Nihoa finches one palila) or Amicon® Ultra-4 (Millipore, Billerica, MA) centrifugal dialysis (Slikas et al. 2000) (amakihi). Alternately, for iiwi, apapane, and the other palila, DNA was extracted using a Qiagen DNeasy Blood and Tissue kit (Qiagen, Germantown, MD). DNA quality and concentration were visualized using agarose gel electrophoresis and quantified using a NanoDrop 1000 spectrophotometer (NanoDrop, Wilmington, DE).

454 Library construction and sequencing

Library construction and sequencing

For 454 sequencing, ~10 ug of genomic DNA was fragmented using a HydroShear apparatus from Genomic Solutions Ltd, and 454 library preparation was done following manufacturer recommended protocols using the Titanium Rapid Library Preparation Kit, with insert sizes greater than 1000 bp. The libraries were then processed for shotgun Roche FLX+ sequencing in 4 lanes, to a total of 2.5X coverage. Average read length was 458 bp.

A total of 5 ug of input DNA was sheared by sonication (Covaris) and size-selected using a Pippin Prep (Sage Science). The fragmented DNA was end-repaired and ligated to Illumina adapters using a SPRI-TE robot and reagents (Beckman Coulter, Inc.). Illumina indexes were then added using 10-cycle PCR reaction performed in duplicate. The amplified library products were pooled and subjected to two rounds of Agencourt AMPure XP (Beckman Coulter, Inc.) bead clean up. The library was run on an Illumina MiSeq (v1 reagents) and two lanes of an Illumina HiSeq2000 (v3 reagents). The insert size of the library was subsequently determined by paired-end read mapping back to the genome assembly to be 392 +/- 29 bp.

RAD tag library construction and sequencing

For the samples involved in RAD tag development, DNA samples were prepared for RAD tag sequencing generally following the protocol of Baird et al. (2008), with modifications. These included the use of directional TruSeq-style adapters with 10 bp unique indices, and selecting a restriction enzyme with indeterminate bases at the cut site to accommodate requirements of Illumina HiSeq chemistry (Faircloth and Glenn 2012). Briefly, 2 ug of genomic DNA for each sample was digested with the *Bse*XI enzyme,

ligated to an adapter with a unique 10 bp index sequence, and sheared to approximately 300 – 500 bp fragments. A second adapter also containing the index sequence was ligated to the other end of the sheared fragments. Adapters were designed so that only fragments with adapters ligated to both ends would amplify. Each library was amplified using Phusion master mix (New England Biolabs, Ipswich, MA) for 15 – 18 cycles of PCR. Magnetic beads (Sera-Mag Speed Beads, Thermo Fisher Scientific, Waltham, MA) were used to purify libraries after amplification and filter out small fragments. Libraries were assessed for correct size and concentration using an Agilent BioAnalyzer. Samples were pooled in equimolar ratios and sequenced on an Illumina HiSeq with 100 bp paired-end reads (amakihi, iiwi, apapane and one palila) or MiSeq with 150 bp paired-end reads (both Nihoa finches and one palila). Paired-end sequencing generates two reads for each fragment, each starting from opposite ends of the fragment.

Genome assembly and comparative analysis

Quality filtered Illumina reads (>80% of bases in the read pair had quality scores > 20) corresponding to ~19-fold coverage (assuming a 1 Gb genome) and filtered 454 reads (reads with at least 300 bp of Q20 bases) corresponding to ~2-fold coverage were used for a genome assembly with phusion (Mullikin and Ning 2003). Chromosome level scaffolds were generated from the assembled contigs by merging position and orientation information about a subset of the reads in the amakihi contigs with their orthologous position in the zebra finch genome (taeGut1) (Warren et al. 2010) as determined by a megablast (Zhang et al. 2000) search. The amakihi chromosome level scaffolds were aligned to the zebra finch genome with Pecan (Paten et al. 2008) using the default

settings. The consensus sequences for each chromosome have been uploaded to NCBI (BioProject 252695) and will be available upon publication of this article.

SNP discovery in the amakihi genome

The Illumina reads were mapped to the amakihi genome assembly with Novoalign V2.08.02 (Novoalign short read mapper: http://www.novocraft.com/), duplicate read-pairs were removed using SAMtools (Li et al. 2009) and variants detected using MPG (Teer et al. 2010). For genome-wide statistics, single-nucleotide variants were filtered to include only heterozygous sites with an MPG score > =10 and a MPG score to read-depth ratio > = 0.5, and sites that had a read-depth less than approximately 2-fold the mean depth of coverage, i.e. <=100x on the autosomes and < =50x on the Z chromosome.

Sequence processing using RAD tags without a reference

Raw reads were evaluated for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were trimmed at the point where per-base quality score inter-quartile range dropped below a quality score of 20. The quality of most read two sequences deteriorated near the beginning of the read, so these sequences were not used. All read one sequences were trimmed to a length of 75 bp, the shortest length of any of the libraries before quality score dropped below 20. All reads were trimmed to this length because the Stacks RAD tag analysis software requires reads from all samples to be the same length. After they were trimmed, reads were filtered for quality using a python script (QualityFilterFastQ.py (Kircher 2012)) (amakihi, iiwi, apapane, both palila) or fastq_quality_filter from the FastX-toolkit

(http://hannonlab.cshl.edu/fastx_toolkit/index.html) (both Nihoa finches), both of which removed any read that had any base pair with a quality score below 20.

Stacks (Catchen et al. 2011) was used to assemble and call SNPs from RAD loci using the denovo_map.pl pipeline for samples without a reference genome. Several samples were first run individually using the populations mode of Stacks. Next, all samples were analyzed together using superparent mode. This mode is designed for test crosses and creates a catalog of possible loci based on the loci present in the parents. For non-cross samples, read one sequences are concatenated into a 'superparent' from which a catalog of stacks loci is developed, followed by alignment and genotyping of each sample at each catalog locus. Default parameters were used except as follows: minimum of three identical raw reads to create a stack and three mismatches allowed between loci when building the catalog of possible loci. The apapane read one file became corrupted during the compression process and was not used in analyses subsequent to individual Stacks runs. After running Stacks, Python scripts were used to filter the output to remove stacks that were found in the superparent catalog but not found in any progeny (samples; no progeny filter) or where one or more individuals had more than two genotypes for a given locus (bad genotypes filter). Stacks representing repetitive regions of the genome were removed by assembling the stacks consensus sequences with minimum overlap 70 bp and maximum read difference of 5% and then discarding stacks that assembled into contigs composed of greater than two sequences.

Using the quality-filtered Stacks consensus sequences only, we compared Stacks SNP calls for the amakihi with genotypes from the genome assembly (same amakihi). BWA (Li and Durbin 2009) was used to align Stacks consensus sequences to the genome

assembly. Next, custom Python and Perl scripts were used to match Stacks SNP genotypes with genome genotypes on a sample of 11 chromosomes selected to include various sizes (chromosomes 1, 5, 7, 9, 15, 20, 22, 23, 24, 26 and 28). These scripts are available upon request to the author.

Alignment of RAD reads to amakihi genome and SNP genotyping

Read one sequences from the RAD tag libraries were trimmed and quality filtered as for Stacks analysis, except reads from the MiSeq run (both Nihoa finch and one palila) were trimmed to 130 bp instead of 75 bp as there was no need to keep all sequences the same length for this part of the analysis. The amakihi genome assembly was indexed using the 'bwtsw' algorithm of BWA and the trimmed, quality-filtered read one sequences were aligned to the indexed reference using the 'samse' algorithm for single reads. The HaplotypeCaller function (DePristo et al. 2011) of the Genome Analysis Toolkit (GATK (McKenna et al. 2010)) was used to identify variable sites between the amakihi genome and aligned honeycreeper reads using the MalformedReadFilter and default parameters. The VariantFiltration function of GATK was used to filter variant sites, passing those with quality >30 and depth >6.

Interspecies comparisons

All RAD read one sequences were aligned to the amakihi reference sequence using Geneious and calls for each sample for all sites were generated using the GATK HaplotypeCaller function with the EMIT_ALL_CONFIDENT_SITES parameter. PyRAD v. 1.2 (Eaton 2014) was used to identify RAD sequences with 10X or higher coverage present in three or more (out of seven) taxa. These were clustered based on similarity of 0.9 in USEARCH (Edgar 2010). The total number of aligned base-pairs was

12,847. A maximum likelihood analysis in Garli v2.0 (Zwickl 2006) was performed on these data with 100 search replicates.

Results

Genome assembly

Our hybrid approach utilized both Roche/454 and Illumina technology (see Table 3-1). Illumina sequencing of the amakihi genome generated approximately 31 GB of data composed of over 300 million read pairs (174.24×10^6 2 × 101 bp, 4.08×10^6 2 × 151 bp and 152.67×10^6 101 × 88 bp pass-filter reads) and represented an approximately 60-fold coverage of the genome. The 454 data comprised 2 – 3x coverage, with 458 bp average read length. This is a substantially larger dataset than for the first avian genome, chicken, which was done using 11 million Sanger reads with 6.6- fold coverage (Genome Sequencing Center 2004).

Table 3-1. Summary of input for genome assembly.

Platform	Read Type	Reads /Read pairs
Illumina	2 × 151	3.93×10^6
Illumina	2 × 101	86.97×10^6
454	Fragment	3.64×10^6

The hybrid assembly used the full 2x 454 coverage and ~19x Illumina coverage (see Table 3-1), similar to the process for turkey which used ~5x 454 and ~25x Illumina GAII (Dalloul et al. 2010). We used only a portion of the total Illumina data to avoid overwhelming the information from the 454 reads; limiting the data volume was also necessary to stay within the memory limits of the computer used (512 GB RAM). Contigs were ordered and oriented and extended into scaffolds by aligning to the zebra

finch genome sequence. In this way, amakihi genotypes at each zebra finch genomic position were determined. Genotype calls were generated using only high-quality (Phred-like Q20 or above) bases in the mapped reads and that an MPG (Teer et al. 2010) score cutoff of ≥ 10 is expected to yield high-quality genotypes with >99.84% concordance with those from an Illumina Infinium genotyping assay (Paten et al. 2008).

The structure of avian genomes in general appears to be relatively undisturbed with regard to rearrangements, resulting in high degree of synteny among a variety of bird species (Burt et al. 1999). This property has been observed when comparing turkey (Dalloul et al. 2010) and *Ficedula* flycatcher to chicken (Backström et al. 2008). Our use of zebra finch as a template for aligning and assembling the amakihi genome is justified, in part, by the relatively recent divergence (33.5 million years) of the species (Jetz et al. 2012). In fact, the *Ficedula albicollis* genome shows remarkably strong synteny with chicken despite perhaps 100 million years of evolutionary distance (Backström et al. 2008). However, on a more localized scale, Ficedula flycatchers show many small rearrangements with respect to zebra finch (Kawakami et al. 2014). If similar rearrangements have occurred between zebra finch and amakihi, then our assembly could be different from the true amakihi genome sequence.

The N50 value of contigs from the hybrid assembly was 23 kb, and 50 kb for scaffolds. This value is smaller than for other recently published bird genomes; for example, Darwin's finch had a 382 kb scaffold N50 (Rands et al. 2013), and the value for flycatcher was 7.3 Mb (Ellegren et al. 2012). Additional sequencing libraries of larger insert sizes would perhaps have resulted in larger N50 values; however, this was effectively accomplished by ordering the contigs relative to the zebra finch genome.

Total assembly size of the amakihi genome was approximately 1 Gb, similar in size to other bird genome assemblies (for example, 1.05 Gb for chicken (Genome Sequencing Center 2004), 1.2 Gb for zebra finch (Warren et al. 2010), 1.1 Gb for turkey (Dalloul et al. 2010), 1.1 Gb for collared flycatcher (Ellegren et al. 2012), and 991 Mb (true size estimated to be 1.25 Gb) for Darwin's medium ground finch (Rands et al. 2013)). We believe that our amakihi genome is relatively complete because the assembly size is similar to other bird genomes. We further tested this assumption by aligning zebra finch sequences to selected portions of the honeycreeper assembly and determining the percentage that successfully aligned. Overall for the numbered chromosomes (not including random, chrM or chrUn), 86.33% of zebra finch sites could be aligned (mean: 77.26 ± 17.69; see Table 3-2).From this alignment we also calculated the genetic distance between amakihi and zebra finch as 0.0905 (Kimura two parameter model; see Table 3-2). It is possible that this value is underestimated since regions greatly diverged between amakihi and zebra finch may not have successfully mapped to the zebra finch reference.

Table 3-2 Alignment statistics for zebra finch and amakihi against amakihi genome.

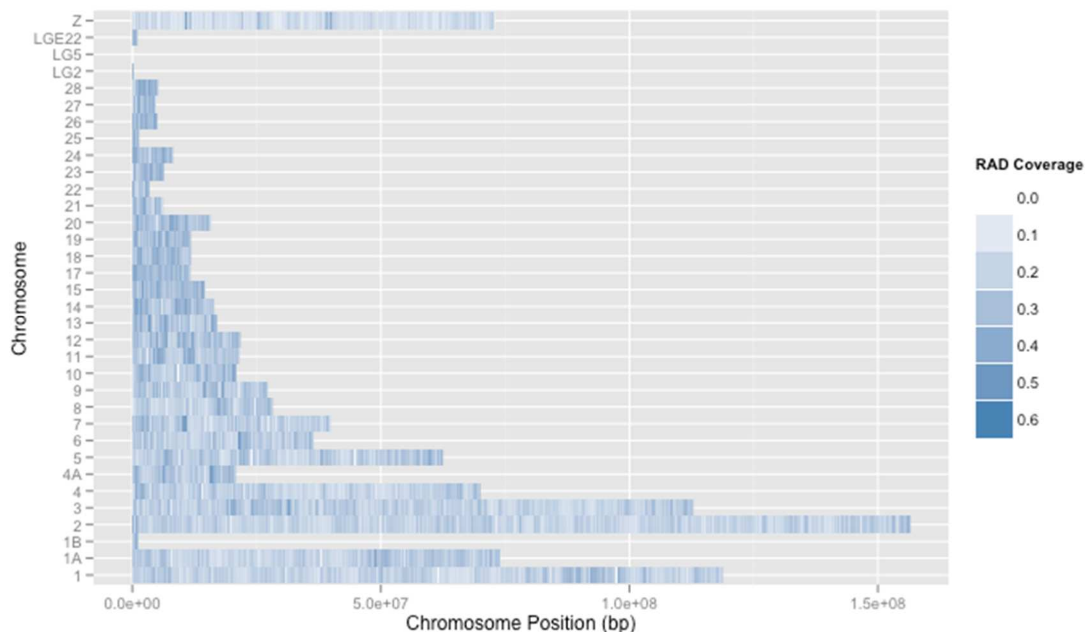
Chrom.	**% of zebra finch sites aligned (non-N)**	**% of amakihi sites aligned (non-N)**	**Uncorrected p-distance**	**Kimura two parameter model distance**
chr1	87.54	90.67	0.09	0.09
chr10	87.66	91.68	0.08	0.08
chr11	87.65	89.89	0.08	0.08
chr12	85.83	90.77	0.08	0.08
chr13	84.46	89.64	0.08	0.09
chr14	88.44	89.85	0.08	0.08
chr15	83.49	87.47	0.08	0.08
chr17	83.60	87.63	0.08	0.08

chr18	81.18	79.45	0.08	0.09
chr19	81.35	86.12	0.08	0.08
chr2	87.90	90.72	0.09	0.09
chr20	81.75	87.00	0.08	0.08
chr21	77.05	81.73	0.08	0.09
chr22	63.61	70.26	0.10	0.11
chr23	71.64	78.85	0.09	0.09
chr24	72.44	78.98	0.09	0.10
chr25	72.01	75.09	0.10	0.11
chr26	77.91	79.94	0.09	0.09
chr27	65.53	73.24	0.09	0.10
chr28	75.92	72.64	0.09	0.09
chr3	91.56	92.23	0.08	0.09
chr4	87.78	91.91	0.09	0.09
chr5	87.80	91.76	0.08	0.09
chr6	85.38	91.11	0.08	0.09
chr7	83.72	91.19	0.08	0.09
chr8	87.09	92.25	0.08	0.08
chr9	83.21	90.02	0.08	0.09
chr1A	88.13	90.70	0.09	0.09
chr1B	68.96	78.58	0.10	0.11
chr4A	84.31	90.13	0.08	0.09
chrLG2	19.65	64.27	0.15	0.17
chrLG5	5.49	62.13	0.15	0.17
chrLGE22	73.78	79.21	0.09	0.10
chrZ	82.55	84.25	0.10	0.11

A total of 1.04 Gb of the amakihi assembly was localized to 34 chromosomes by aligning contigs and scaffolds to zebra finch chromosomal sequences. Although previously assembled avian genomes have taken advantage of linkage maps from the same species for chromosome assignment (i.e., 93% assigned to chromosomes for turkey (Dalloul et al. 2010)), alignment to other genomes has also been used. For *Ficedula albicollis*, 73% of the genome sequence was assigned to chromosomes using the

flycatcher linkage map; by comparing conserved organization with zebra finch, a total of 89% could be assigned (Ellegren et al. 2012). As was the case for turkey (Dalloul et al. 2010) and chicken (Genome Sequencing Center 2004), most of the honeycreeper chromosomes are microchromosomes that cannot always be distinguished by size alone (see Figure 3-1, which shows relative chromosome lengths). The draft amakihi genome sequence is available in FASTA format in the NCBI repository, BioProject: PRJNA252695.

Figure 3-1. RAD coverage of amakihi chromosomes. Colors indicate proportion of 100 KB bins covered by at least 1 bp of RAD sequence.



After assembly, a larger number of Illumina reads were aligned back to the assembled genome to a depth of ~47.6x for the autosomes and ~25x for the Z chromosome to identify and call SNPs. Nucleotide diversity (π) on the autosomes ranged

from 0.0022 on chromosome LGE22_random to 0.0113 on chromosome LG5 (Table 3-3: summary of nucleotide diversity by chromosome).

Table 3-3. Nucleotide diversity by chromosome.

Chromosome	Homozygous Sites	Heterozygous Sites	π
chr1	112,544,959	485,712	0.0043
chr10	19,502,766	67,127	0.0034
chr10_random	181,773	748	0.0041
chr11	20,339,491	68,262	0.0033
chr11_random	205,478	795	0.0039
chr12	19,966,665	77,365	0.0039
chr12_random	142,337	627	0.0044
chr13	15,608,448	63,230	0.0040
chr13_random	2,273,196	6,948	0.0030
chr14	15,783,392	62,881	0.0040
chr14_random	119,916	586	0.0049
chr15	13,395,570	47,525	0.0035
chr15_random	336,675	1,356	0.0040
chr16_random	28,278	132	0.0046
chr17	10,789,469	43,477	0.0040
chr17_random	69,369	290	0.0042
chr18	11,093,387	30,844	0.0028
chr18_random	393,813	1,922	0.0049
chr19	10,638,978	41,233	0.0039
chr19_random	61,004	162	0.0026
chr1A	70,419,613	301,663	0.0043
chr1A_random	429,913	1,856	0.0043
chr1B	900,172	4,219	0.0047
chr1B_random	100,455	761	0.0075
chr1_random	150,801	806	0.0053
chr2	149,097,369	652,060	0.0044
chr20	14,291,352	57,186	0.0040
chr20_random	138,194	682	0.0049
chr21	5,425,030	24,579	0.0045

chr21_random	1,777,800	4,856	0.0027
chr22	2,908,707	11,322	0.0039
chr22_random	657,788	3,832	0.0058
chr23	5,370,519	23,530	0.0044
chr23_random	370,728	2,169	0.0058
chr24	7,044,699	31,458	0.0044
chr24_random	74,717	253	0.0034
chr25	1,142,233	4,993	0.0044
chr25_random	345,747	2,115	0.0061
chr26	4,582,739	19,099	0.0042
chr26_random	1,375,049	7,314	0.0053
chr27	3,929,203	14,589	0.0037
chr27_random	187,008	875	0.0047
chr28	4,923,374	18,553	0.0038
chr28_random	158,967	1,285	0.0080
chr2_random	408,633	1,750	0.0043
chr3	110,159,365	497,976	0.0045
chr3_random	850,964	4,677	0.0055
chr4	65,570,862	294,828	0.0045
chr4A	18,959,367	64,240	0.0034
chr4A_random	68,624	262	0.0038
chr4_random	4,413,118	21,779	0.0049
chr5	58,574,618	240,015	0.0041
chr5_random	1,912,995	11,030	0.0057
chr6	33,425,145	73,863	0.0022
chr6_random	1,513,279	8,220	0.0054
chr7	35,848,910	146,136	0.0041
chr7_random	205,374	1,023	0.0050
chr8	25,953,825	94,369	0.0036
chr8_random	4,504,789	14,345	0.0032
chr9	24,645,966	100,108	0.0040
chr9_random	121,289	669	0.0055
chrLG2	27,825	241	0.0086
chrLG5	1,309	15	0.0113
chrLGE22	781,559	2,626	0.0033
chrLGE22_random	75,448	170	0.0022
chrUn	7,431,999	41,937	0.0056

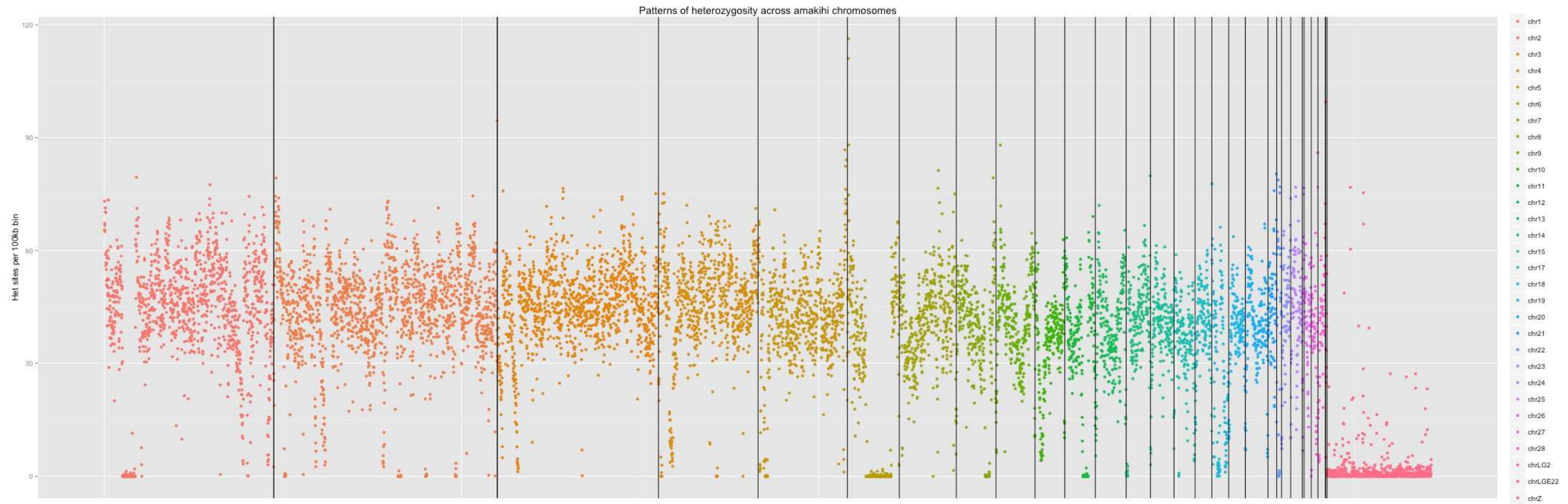
chrZ	68,235,778	9,906	0.0001
chrZ_random	2,178,358	2,162	0.0010
chrUn2	21,894,126	153,065	0.0069

Because in birds females are the heterogametic sex (we sequenced a female) chromosome Z should in theory have no heterozygous sites except in pseudo autosomal regions. Our data show about 0.017% of the total sequence sites assigned to Z and Z random are heterozygous (9,906 heterozygous sites on Z and 2,162 on Z random) versus 0.417% for sites on autosomal chromosomes. These false positives on the Z could be attributed to mismapping of paralagous reads or misassignment of autosomal segments to the Z and Z random chromosomes. The false positive rate on Z/Z random is an approximate indicator of the false positive rate elsewhere in the genome because mismapping of paralagous sites could have occurred for autosomal chromosomes as well.

Approximately 3.9 million SNP sites were discovered in the assembled amakihi genome, or approximately one SNP every 256 bp. This is similar to results for the flycatcher, where 3.66 million SNPs (one per 330 bp) were identified in one individual (Ellegren et al. 2012). Heterozygosity was characterized for each chromosome by counting the number of heterozygous sites in 100 kb bins along each chromosome (Figure 3-2). Large stretches of extremely low variability (nearly zero heterozygosity) were observed on five chromosomes (1, 5, 6, 8 and 11). Coverage for these regions was not different than for other sites in the genome. They ranged in size from 2 Mb on chromosome 5 to 17.9 Mb on chromosome 6 and together made up 3.51% of the genome sequence (Figure 3-2). Large stretches of low heterozygosity were also observed on turkey chromosomes 1 and 3 and were interpreted as IBD (identical by descent; having

come from a recent common ancestor) haplotypes (Aslam et al. 2012). The turkeys described in that study were from domestic lines that had been subjected to many generations of artificial selection, so finding IBD regions was not unexpected. In the case of the amakihi, which has a relatively large population size, inbreeding is not expected. For inbreeding between first order relatives (i.e., parent-child) approximately 25% of the genome would be expected to show large homozygous stretches, while inbreeding of second order relatives (such as uncle-niece/aunt-nephew) would result in about 12.5%. To differentiate between the effects of inbreeding and selection, we would need to determine the probability of SNP loci in the low heterozygosity regions being IBD or identical by state (IBS; sharing the allele by chance rather than inheriting it from the same ancestor). As we obtain more data from other amakihi, we will be able to calculate allele frequencies for the loci in question and be able to calculate IBD/IBS probability for the low heterozygosity regions. These regions could possibly represent signatures of selective sweeps in the evolutionary history of the amakihi, or be the result of inbreeding, although the latter may be less likely given the relatively high variation found in amakihi from the same locality as 1771-10606, the individual whose genome is presented here . We compared gene classifications within each homozygous region to those on the rest of each respective chromosome using Ensembl annotations for the zebra finch (http://www.ensembl.org/Taeniopygia_guttata/Info/Index). No substantial difference was observed.

Figure 3-2. Patterns of heterozygosity across amakihi chromosomes. Each dot represents the count of heterozygous sites in a 100 kb bin. Colors represent different chromosomes. Note stretches of low heterozygosity on chromosomes 1, 5, 6, 8, and 11.



RAD data

The RAD tag method involves digesting genomic DNA with a restriction enzyme and sequencing fragments (tags) of DNA adjacent to restriction sites (Baird et al. 2008). We sequenced RAD tags for six individuals of four honeycreeper species in addition to the same amakihi for which we obtained the genome. This method yielded a wide range of sequences per individual, with an average of 7,596,336 post quality filtering (range: 319,559 – 24,263,032; see Table 3-4.). We attribute the large range of number of reads to stochastic factors and variable sample DNA quality, as all other parameters (DNA quantity, library preparation protocol, pooled for sequencing in equimolar ratios) were the same between samples. RAD sequences were analyzed following two protocols: without a reference genome, using the Stacks pipeline, or utilizing the amakihi sequence as a reference for assembly and genotype calling. Raw reads for each individual in FASTQ format have been uploaded to NCBI (BioProject 252695) and will be available after publication of this article.

Table 3-4. RAD read counts.

Sample	Raw Read 1 Sequences	Quality-filtered Read 1 Sequences
Apapane	16,017,647	7,331,429
Iiwi	43,630,304	16,380,184
Amakihi	50,069,524	24,263,032
Palila_Tag1	6,869,311	3,023,108
Palila_2	652,854	319,559
Nihoa_Finch_1	1,098,237	772,863
Nihoa_Finch_2	1,503,204	1,084,183

By using Stacks to assemble and genotype RAD sequences, we found 309,957 loci with 173,553 passing our filters, 17,513 of which were variable loci containing at least one SNP site within or between individuals (see Table 3-5). There were, on average,

40,270 loci per species passing our filters (range: 2,351 – 123,623) and 3,996 SNPs per species (range: 515 – 12,422); i.e., about 10% of loci contained SNP(s). Only 473 stacks with 109 total SNPs were shared by at least three of the honeycreeper species.

Table 3-5. Stacks results after quality filtering.

Species	Number of Stacks Loci	Number of SNPs	Number of Variable Loci
Apapane	17,357	680	573
Nihoa Finch	3,004	841	577
Palila	2,351	515	354
Iiwi	55,014	5,523	4,197
Amakihi	123,623	12,422	9,536

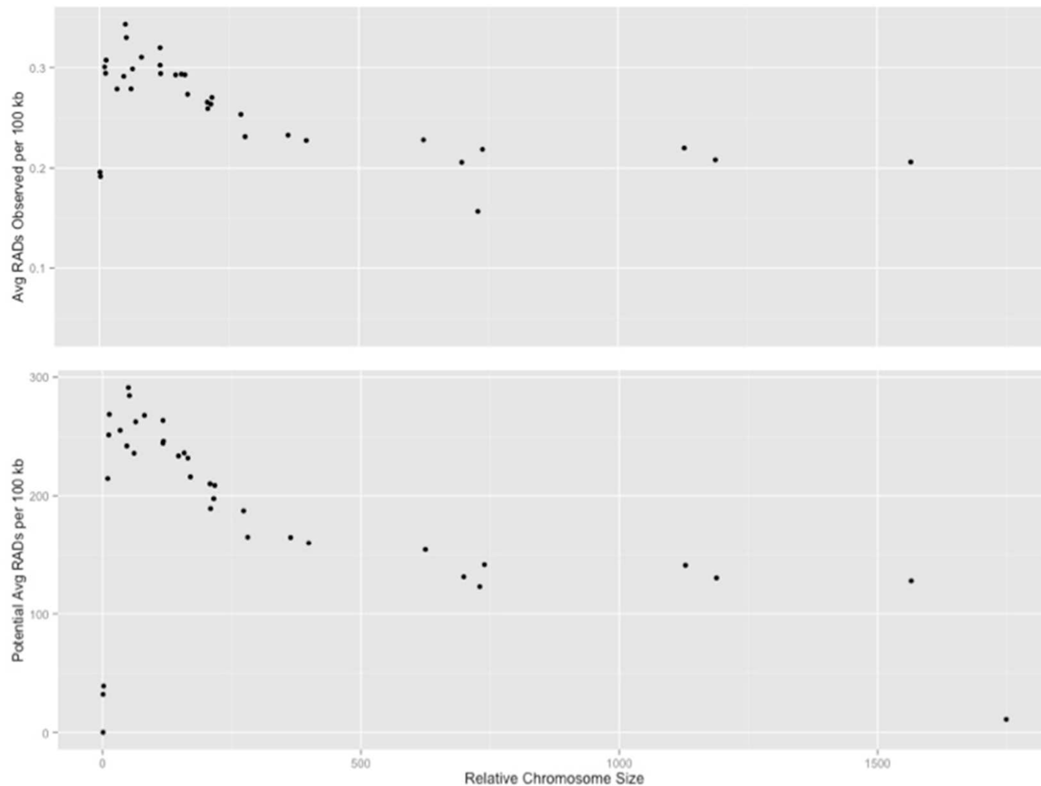
Since we had both RAD and genome data for the same individual amakihi, we compared genotype calls from Stacks to known values from the genome sequence. With a minimum stack depth requirement of nine, only 0.8% of Stacks SNP calls differed from the genome value.

RADs with a reference

We also analyzed RAD data with the benefit of the amakihi reference sequence. Restriction cut sites, and therefore RAD sequences, are expected to be randomly, not evenly, distributed across the genome (Davey et al. 2013). When aligning honeycreeper RAD sequences to the amakihi genome, we observed a denser distribution of RADs on the microchromosomes (Figures 3-1 and 3-3). We found the same pattern of non-random distribution of restriction sites based on an *in silico* restriction digest of the amakihi genome (Figure 3-3). One possible explanation for this is that the microchromosomes of avian species are commonly more gene-dense than the macrochromosomes, with a higher GC content (McQueen et al. 1996, Smith et al. 2000, Federico et al. 2005), and restriction

enzymes tend to have a high proportion of GC content in their binding site (Nikolajewa 2005). The enzyme used in this study, BseXI, contains 80% GC in its 5 bp recognition site, making this a plausible explanation. Alternatively, there may be more repetitive DNA sequences in macrochromosomes, and the repetitive sequences might not contain the BseXI recognition site. Being able to align RADs to a reference provides an advantage for researchers who may wish to select a smaller number of RAD SNP sites for genotyping, as the spacing and location of specific markers makes it easier to narrow down to only the necessary ones.

Figure 3-3. Relationship between relative chromosome size and RAD density. Top panel shows the density of RADs based on our RAD sequencing; bottom panel shows the density of restriction sites and potential RADs based on in silico digest of the amakihi genome.



We used the Burrows-Wheeler Aligner (BWA (Li and Durbin 2009)) and the Genome Analysis Toolkit (GATK (McKenna et al. 2010)) in conjunction with the amakihi reference sequence to identify inter- and intraspecific SNPs using the RAD sequences. Using this method, we identified 172,085 SNP sites with 156,486 passing quality filters (See Table 3-6). After filtering, there were, on average, 52,348 sites with a known genotype identified per sample (range: 15,800 – 110,844) including an average of 1,727 heterozygous sites per sample (range: 291 – 4,137). 9,714 non-reference sites were shared by at least four samples.

Table 3-6. SNP sites discovered by comparison to the honeycreeper reference. Filtered for Qual > 30 and Depth > 6.

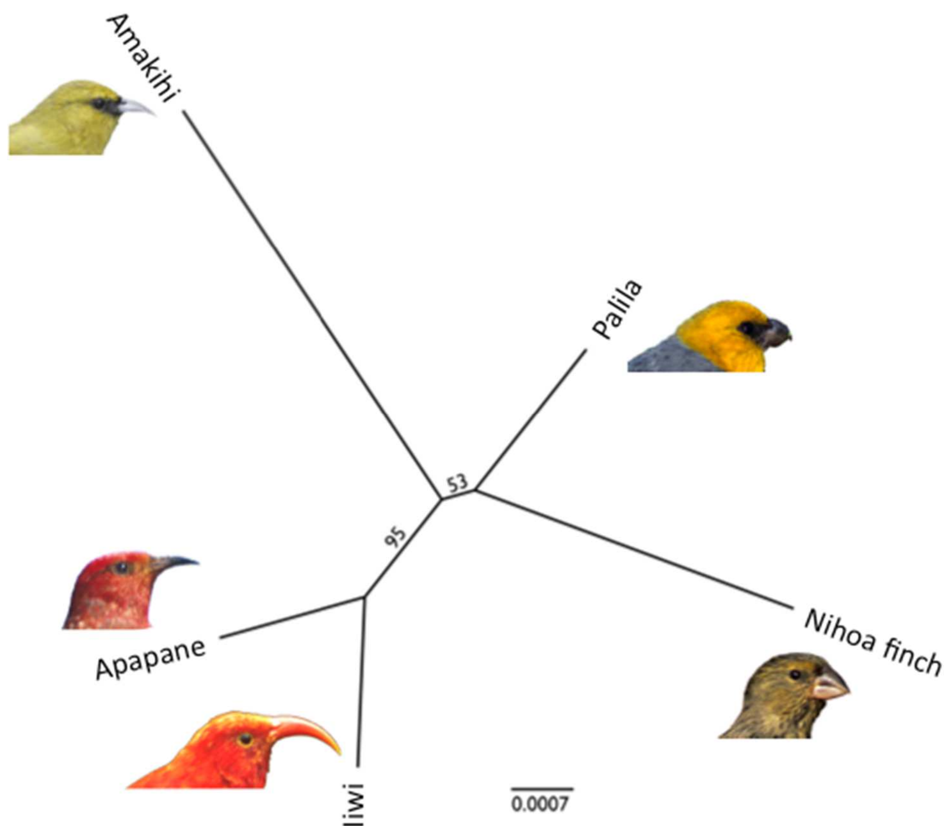
Sample	Positions with Known Genotype	Heterozygous Sites	Sites with Non-Reference Allele	Private Non-Reference Alleles
Nihoa_Finch_1	93,646	2,864	91,038	15,715
Nihoa_Finch_2	110,844	4,137	108,297	30,168
Iiwi	15,800	291	12,685	524
Palila_1	17,511	571	14,580	841
Amakihi	50,489	2,202	22029	3,587
Palila_2	25,795	299	22,529	1,664

Compared to analyzing without a reference, the BWA-GATK pipeline resulted in more SNPs identified for Nihoa finch, fewer for iiwi, about the same for palila, and fewer for amakihi.

Interspecies comparisons

We performed a phylogenetic analysis to demonstrate the utility of RAD sequences for determining relationships amongst taxa. PyRAD (Eaton 2014) was used to identify and homologize RAD sequences with 10X or higher coverage present in three or more taxa, which produced 38,889 bp. A maximum likelihood analysis was performed on these data in Garli (Zwickl 2006) with 1,000 bootstrap replicates and the relationships of the five species are shown (Figure 3-4). This analysis recovered the expected topology with good support for the iiwi/apapane relationship. Support for the palila/Nihoa finch node was low, perhaps as a result of the deeper divergence between these species than between iiwi and apapane, and the shorter internode between this clade and the amakihi clade (Lerner et al. 2011).

Figure 3-4. Reconstructed maximum likelihood tree of relationships of the five study species based on RAD sequences.



Discussion

Herein, we describe a draft genome sequence for the Hawaii amakihi and associated genomic resources for Hawaiian honeycreepers including approximately 3.9 million SNPs within the amakihi genome and over 150,000 SNPs within and between amakihi and four other honeycreeper species. Honeycreepers are an important model system for many questions in evolutionary biology, and the SNP markers will facilitate a wide range of future studies in ongoing and new research areas. Being genome-enabled

both enhances the resolution of current research methods (for example, fully resolving the honeycreeper phylogeny) and also opens up new analyses that weren't possible before (such as GWAS for malaria tolerance). Some of the important questions which may be addressed include: how do rates of sequence evolution vary among different classes of DNA; what genes or genome regions are involved in speciation, adaptation or evolution of tolerance or resistance to disease; and how much adaptive potential exists in a population after demographic decline or fragmentation?

Studies of the evolutionary relationships of honeycreepers (Amadon 1950, Richards and Bock 1973, Raikow 1977) have been limited by available technology and methods, as well as by rapid speciation and low levels of sequence divergence. Early molecular studies used allozyme electrophoresis (Johnson et al. 1989, Fleischer et al. 1998), restriction fragment length polymorphism of mitochondrial DNA (Tarr and Fleischer 1993), and relatively short DNA sequences (Fleischer et al. 1998, Reding et al. 2009, 2010) to only marginally resolve nucleotide substitution rates and relationships within the honeycreepers. Larger molecular datasets, such as one with entire mitochondrial genomes and 13 nuclear loci (>15 Kb) more adequately resolved the phylogeny, and estimated rates of sequence evolution and a split from a cardueline finch lineage at 5.7 Mya (Lerner et al. 2011). Re-evaluating the honeycreeper phylogeny with a larger, more comprehensive dataset will allow researchers to investigate the pattern and tempo of evolution in this radiation. With genome-wide markers, it will be possible to connect genomic regions with specific adaptive traits across the phylogeny. Because precise geological information about the Hawaiian Islands provides a framework for dating evolutionary events, the honeycreeper radiation can provide unique insights into

the evolutionary process. What is learned from honeycreepers can also be compared with other avian adaptive radiations such as Darwin's finches (Rands et al. 2013) to further our understanding of the evolutionary process overall.

The ability to use analytical tools that connect genotypes to traits, such as GWAS (Orr et al. 2010, Jones et al. 2012)) is a key benefit of the honeycreeper genomic marker set. These methods require large numbers of markers and were previously only useful for genome-enabled model organisms. Such techniques may allow identification of genes or regions implicated in disease resistance or specific adaptive traits; when such information is combined with results in other taxa, it contributes to our overall understanding of molecular mechanisms. This is also a first step towards investigating what happens to the genetic diversity in adaptively important genes or regions when species decline and become endangered. Identifying key genomic regions for disease resistance or adaptation could help focus conservation efforts towards preserving genetic variation in those areas and provide guidance for genetically-based population management decisions.

Hawaiian honeycreepers are also a model to investigate the response of genetic variation to human caused population decline, fragmentation and founder effects. For example, the Hawaii akepa (*Loxops coccineus coccineus*) occupies < 10% of its historical range in fragmented habitat and is a magnitude less populous than before its decline, yet contemporary samples show the same level of mitochondrial genetic diversity as in specimens sampled > 100 years ago and no significant differentiation between fragmented populations is detected (Reding et al. 2010). In another case, several founder populations of Laysan finch (*Telespiza cantans*) have been established on Pearl & Hermes reef and microsatellite data reveal that these have become genetically

differentiated from the Laysan population and, to some extent, from each other (Tarr et al. 1998). Finally, Hawaii amakihi, which have a relatively large population size, exhibit a rather unique elevational structuring, with populations from high elevation genetically differentiated from those at low elevation; data from museum skins suggest that this was also true historically. This elevational pattern is not found in contemporary iiwi (*Vestiaria coccinea*) or apapane (*Himatione sanguinea*) populations (Foster et al. 2007). Using the more comprehensive SNP marker set will provide the power to start looking at selection and adaptation to anthropogenic caused change in these species.

Our results provide a set of genomic resources for Hawaiian honeycreepers that will facilitate research on disease interactions, metapopulation dynamics, adaptive radiations, and genome evolution. The amakihi genome sequence will enable comparative studies of avian genomes and is an important contribution as it represents one of the more than 5,000 passeriform species. The results yield a large number of genome wide markers, both from heterozygous sites in the sequenced individual and discovered using RAD tags with other honeycreeper species. We have demonstrated their potential phylogenetic utility based on a tree of relationships between honeycreeper species used in our RAD analysis that matches expectation based on previous molecular phylogenetic analyses (Lerner et al. 2011). Heterozygosity measures for the individual sequenced, a malaria-resistant amakihi, indicate some regions of potential selective sweeps that could be of interest for study of malaria resistance. These regions are being targeted for resequencing in populations of malaria resistant and susceptible amakihi. The markers could also be used to identify regions of divergence among honeycreeper species to help elucidate the speciation process (Ellegren et al. 2012).

Chapter 4
Serial bottlenecks in an endangered insular passerine, the Laysan finch, *Telespiza cantans*

Abstract

Islands foster unique biodiversity but also impose restrictions on population size and increased stochastic risk. A common characteristic is reduced genetic variation compared to mainland counterparts. Population crashes (bottlenecks) or founder events may exacerbate the effects of genetic drift and associated problems for insular species. The Laysan finch, *Telespiza cantans,* is an endangered Hawaiian honeycreeper endemic to Laysan Island in the Northwestern Hawaiian Islands. Its history includes a severe bottleneck in the early 20th century caused by introduced rabbits, as well as several founder events when Laysan finches were translocated to the islets of Pearl and Hermes Reef (PHR). We used a DNA sequence capture approach to obtain SNP genotypes for museum samples of Laysan finches prior to the 20th century bottleneck and for modern samples from Laysan and three translocated populations on different islets of PHR. We found no difference in heterozygosity or number of fixed alleles or private loci between any of the populations. We investigated population structure using clustering, principle components, discriminant function of principle components, F_{ST}, and AMOVA. No discernible population structure was detected, including comparison between samples collected before and after the major 20th century bottleneck.

Introduction

The evolutionary dynamics of living on islands have fascinated biologists for centuries (Darwin 1859, Wallace 1880). Islands can be hotspots of biodiversity due to the adaptive radiations that tend to form. An excellent example is avian species on Hawaii that have filled many niches that might otherwise have been filled by mammals (Amadon 1950). However, extinction risk is higher on islands than on continents (Steadman 2006, Jones and Merton 2012). With the entire population or species localized in a relatively small area, the probably of extinction by natural or anthropogenic disaster is higher. Small islands generally support smaller population sizes, and small populations are more susceptible to loss of genetic variation through drift, reducing adaptive potential and increasing the chance of inbreeding depression (Frankham 1997, 1998, 2005; Keller and Waller 2002). A population with low diversity may be more susceptible to disease (Spielman et al. 2004) or less able to cope with environmental change (Willi et al. 2006). Tragic examples of how introduced disease and predators have decimated irreplaceable endemic island fauna can be found around the world. Examples include avian malaria in Hawaii (Warner 1968, van Riper et al. 1986) or tree snakes in Guam (Savidge 1987).

Studying insular species that have experienced population bottlenecks can help us understand how island population dynamics interact with these significant demographic events. The equilibrium level of genetic variation in island populations is frequently lower than their mainland counterparts (Frankham 1997, 2005), and because bottlenecks (and founder events) are predicted to reduce genetic diversity (Kimura and Ohta 1969, Nei et al. 1975), there is potential for significantly low genetic variation in bottlenecked insular species, and correspondingly increased risk of extinction. However, the outcome

of bottlenecks for genetic variation is not always predictable (Bouzat 2010). There have been several reports of island endemics recovering from very severe bottlenecks (Ardern and Lambert 1997, Groombridge et al. 2009), and some have suggested that chronically low population sizes on islands could reduce the genetic load (Bataillon and Kirkpatrick 2000). One intriguing study of an island endemic, the Mauritius kestrel (*Falco punctatus*), which survived an extreme bottleneck compared pre- and post- bottleneck genetic variation and reported much higher variation than predicted prior to the bottleneck, possibly because of population fragmentation (Groombridge et al. 2000). An alternative example is that of the nene (*Branta sandvicensis*), which showed similarly low levels of genetic variation before and after a historical bottleneck during the 1800-1900s. Researchers used ancient DNA and modeling to determine that nene had lost much of their genetic variation in an earlier bottleneck coinciding with a time of prehistoric human population growth, creating a situation where nearly all of the existing variation was preserved during the historical bottleneck (Paxinos et al. 2002). Clearly, much remains to be discovered about how genetic diversity is impacted by bottlenecks in systems where diversity is restricted by insular evolutionary history. The more we understand how bottlenecks or founder events are likely to affect genetic diversity of island endemics, the better we will be able to provide targeted assistance to endangered island species (Lambert et al. 2005, Bouzat 2010).

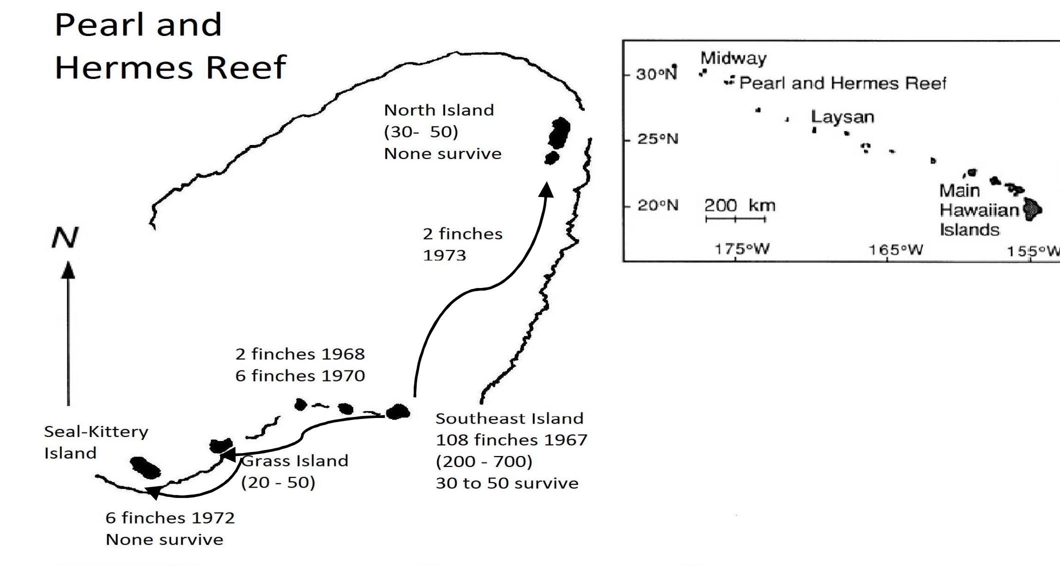
The Laysan finch (*Telespiza cantans*), an endangered Hawaiian honeycreeper (Carduelini: Drepanidini), is an island endemic with documented population declines. Laysan is a very small island (~4.11 km^2) in the Pacific located approximately halfway along the Northwestern Hawaiian islands, between the main Hawaiian islands and

Midway island. Laysan was home to five endemic bird taxa (four unique species and one subspecies), as well as endemic flora and insects, as of the late 19th century. Observers reported little human impact on the flora and fauna of Laysan prior to 1896, and early reports indicated Laysan finches to be "everywhere in abundance" (Ely and Clapp 1973b). Unfortunately, in 1902 rabbits were introduced to Laysan, with devastating effects Laysan's vegetation and thus on the native and endemic species (Ely and Clapp 1973b). Three of the five endemic birds went extinct. Laysan finch numbers fell in response to the degradation of Laysan's plant life. From 1911 to 1913 there were between 2,700 (Dill and Bryan 1912) and 4,000 Laysan finches (Munter 1915). An expedition sent to exterminate the rabbits in the winter of 1912-1913 removed many thousands, but not all of them. However, the rabbits contributed to their own demise by consuming all of their own food resources. By 1923, only a few hundred rabbits remained, and these were destroyed by the Tanager Expedition. At that point, only an estimated 100 Laysan finches remained (Wetmore 1925, Ely and Clapp 1973b, Olson 1996). The Laysan finch population rebounded, and by 1957, censuses estimated 5,000 individuals (Woodside and Kramer 1961). In 1959, 36 years after the rabbits were eradicated, there were an estimated 10,000 Laysan finches on Laysan (Warner 1959).

In an attempt to mitigate the risk of extinction for the Laysan finch, the United States Fish and Wildlife Service decided to found a second population (Sincock and Kridler 1977). In 1967, they translocated 108 individuals 500 km to Southeast Island at Pearl and Hermes Reef (PHR), about 44 years after rabbits were removed from Laysan. A series of founder events followed as finches colonized the other islets of PHR (Figure 4-1). Recently, Laysan finch populations on PHR (329 total individuals in 2004 (Sprague

2004)) were described as in decline and at risk of extinction (McClung 2005). A major factor driving current extinction risk is an invasive annual plant species, *Verbesina encelioides,* which has supplanted the native vegetation (Sprague 2004). The native vegetation provided year-round foraging and nesting habitat for the finch, but *V. enceliodies* dies out in the winter, leaving the finch with limited foraging resources. Finally, climate change is putting both the Laysan and PHR populations at increased jeopardy due to sea level rise and the increasing frequency of storms (USFWS 2008).

Figure 4-1. Colonization of Pearl & Hermes Reef Islets. Numbers in parentheses are population size ranges observed in years subsequent to translocations. Adapted from Tarr et al 1998.



The goals of this study were to: (1) determine the baseline historic level of genetic variation in the Laysan finch on Laysan island and compare that with the level of variation in modern populations; and (2) identify population differentiation or structure between the modern populations. Previous research has reported low genetic variation for

allozymes (Fleischer et al. 1991) and microsatellites (Tarr et al. 1998) in modern Laysan finch populations. For both studies, F_{ST} values suggested differentiation between PHR populations and Laysan, likely due to the effects of genetic drift. No study to date has ascertained genetic diversity prior to the introduction of rabbits to Laysan. Study of historic genetic diversity can help managers evaluate the relative genetic health of the modern populations as has been done, for example, for other endangered Hawaiian birds (Paxinos et al. 2002, Fleischer et al. 2007, Reding et al. 2010, Mounce et al. 2014). In this study, we used museum specimens collected prior to 1923 to determine the level of genetic diversity present in historic times and to compare that with modern specimens to determine the magnitude of genetic diversity lost during the major 20th century Laysan bottleneck. Comparison of historic and modern samples also helps us understand how stochastic events like the 20th century bottleneck interact with insular demographic/genetic dynamics to affect effective population size and genetic variation through time. Finally, determination of structure between modern populations on the islets of Pearl and Hermes Reef (PHR) and Laysan provides insight into how founder event bottlenecks affect genetic diversity and differentiation in an island system.

We used a capture approach to sequence targeted DNA fragments from museum samples collected on Laysan and modern samples collected on Laysan and PHR in the 1980s and 1990s. We called single nucleotide variations (SNPs) from the sequence data and used these to evaluate genetic diversity and population structure. We predicted that: (1) the historic Laysan finch population would show greater genetic diversity than the modern populations; (2) PHR populations founded by fewer Laysan finches (Grass and North islands) would show less genetic diversity than Southeast Island (founded by 108)

or modern Laysan; and (3) population structure between Laysan and PHR populations would be detectable.

Methods

Samples

Blood samples from Laysan (*N*=33) and each PHR island (Grass *N*=33, Southeast *N*=34, North *N*=33) were collected as part of previous work (Conant 1988, Tarr et al. 1998) and are listed in Appendix A. Samples were stored at -20°C or -80°C prior to DNA extraction. Toepads (Mundy et al. 1997) from Laysan finch museum specimens (*N*=38) were provided by various museums (Appendix A). Museum samples used in this study were collected between 1891 (prior to the introduction of rabbits to Laysan) and 1913, when rabbits had been on Laysan for approximately 10 years. At that point, the Laysan finch population was estimated to still be quite large, between approximately 2,700 and 4,000 individuals (Ely and Clapp 1973b).

DNA Extraction and Library Preparation

All pre-PCR work with museum samples was performed in a physically separate laboratory dedicated to ancient DNA work. Extractions were performed alternating every-other-sample with either an extraction blank or a sample from a different species to enable detection of cross-contamination. At least one extraction blank was used for each batch of extractions. Museum samples were extracted using phenol:chloroform extraction as described in (Fleischer et al. 2000) with a final purification step using Amicon Ultra centrifugal filters (Millipore).

For modern samples, genomic DNA was extracted from blood samples using a BioSprint or DNEasy Extraction kit and quality was assessed by agarose gel electrophoresis. At least one extraction blank was included with each batch of extractions.

Modern samples were quantified using a Qubit and 500 ng of each sample was fragmented using a QSonica sonicator for 2 – 6 minutes depending on range of fragment sizes present in order to obtain a mean fragment size of 300 – 500 bp. Sonicated samples were end-repaired using DNA Polymerase I, Large (Klenow) Fragment or NEBNext End-Repair Kit (both New England Biolabs). A cytosine was added to the end of fragments using Klenow Fragment (3' – 5' exo; New England Biolabs) and NEB buffer 2 with dCTP added in order to facilitate ligation of Nextera-style stubby adapter in a subsequent step using NEB Quick Ligation Kit (New England Biolabs). After stubby adapter ligation, unique dual 8-bp Nextera-style indices were added to each sample through PCR amplification using Kapa HiFi Polymerase (5 – 10 cycles for modern samples; 12 – 16 for museum samples). Success of library preparation was determined by running each post-PCR library on an agarose gel. Library preparation for museum samples was carried out as described above for modern samples, with the following modifications at the beginning of the protocol. The fragmentation step was skipped, as museum specimen DNA is already fragmented. Prior to the end repair step, 50 ul of DNA extract was subjected uracil-DNA glycoylase (UDG) treatment following manufacturer recommendations (New England Biolabs) to reduce the presence of cytosine deamination artifacts (Hofreiter 2001).

SNP Capture and Sequencing

In order to select SNP-containing fragments for sequencing, we used a custom in-solution array (MYcroarray MYbaits) designed to capture ~40,000 honeycreeper SNP loci (Callicrate et al. 2014). The capture array was designed using four honeycreepers related to Laysan finches rather than Laysan finch individuals in order to avoid ascertainment bias (Albrechtsen et al. 2010, Lachance and Tishkoff 2013). Amplified libraries were quantified using QBit and split into groups of seven or eight for pooled capture. Museum samples were pooled separately from modern samples (3 – 5 per pool). Each pool consisted of 600 ng total library DNA and no adapter indices were shared by samples in each pool. MYcroarray Protocol 1.3.7 was followed, except Block #1 (Human Cot-1) was replaced with Chicken Cot-1 to increase blocking effectiveness. Following capture, pools were amplified for 10 -12 cycles with Illumina primers and quantified using qPCR (Stratagene) with a Kapa Illumina Quantification Kit. Three samples were captured, pooled, and sequenced on an Illumina MiSeq to validate the capture procedure. Subsequently, all capture pools were pooled together and sequenced on two lanes of an Illumina HiSeq.

Sequence Alignment and SNP Calling

Each sample's reads from both lanes were pooled (for three samples which had also been run on the MiSeq, those reads were pooled also) and aligned to the amakihi genome sequence (Callicrate et al. 2014) using BWA (Li and Durbin 2009). PCR duplicates were marked using Picard Tools (http://broadinstitute.github.io/picard), and indels were identified and realigned using the Genome Analysis Toolkit (GATK) RealignerTargetCreator and IndelRealigner (McKenna et al. 2010, DePristo et al. 2011).

GATK UnifiedGenotyper was used to call variant sites in all samples simultaneously, using parameters --min_base_quality_score 20, --standard_min_confidence_threshold_for_calling 20, and --standard_min_confidence_threshold_for_emitting 20. The resulting variant file was subjected to hard filtering using the FilterVariants GATK tool as recommended by the Broad Institute (Van der Auwera et al. 2013) when standardized reference data (such as HapMap) are unavailable (filter expression: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0"; see http://gatkforums.broadinstitute.org/discussion/2806/howto-apply-hard-filters-to-a-call-set). A list of variants passing coverage filters was created using the CoveredByNSamples sites tool with parameters --minCoverage 9 (excludes a locus for a given sample if that sample's coverage is below 9) and --percentageOfSamples 0.3 (excludes loci from the set if they are found in fewer than 30% of all samples). Finally, sites meeting the following set of conditions were selected using the GATK SelectVariants tool: SNP only (e.g., excludes indels), variable when considering the entire set of samples, passed the hard filter, and passed the coverage filter. We further reduced the dataset to remove SNP sites and individuals with high missingness. First, VCFtools (Danecek et al. 2011) was used to remove sites with >50% missing data, resulting in 11,527 loci remaining. Next, any individuals with >10% missing data in this 11,527 locus set were removed, leaving 72 individuals (of 150 sequenced) which were used in all subsequent analyses (museum samples: *N*=13, Table 4-1; modern samples: Laysan *N*=12, Grass *N*=15, Southeast *N*=18, North *N*=14, Table 4-2). We assessed dropout related to sequencing coverage by calculating the percentage of quality-filtered

genotyped sites that were heterozygous for each of the 72 individuals for a range of minimum coverage values (9-20). Only very minimal differences in percentage of heterozygous sites were observed for this range of coverage, which we interpreted to mean that the amount of false homozygous calls was very low, and minimum coverage of nine was an acceptable value to include a genotype in the dataset.

Table 4-1. Laysan finch museum samples passing quality filters.

Sample	Museum	Collection Date	Collector	Total Reads	On-Target Reads[1]	Baits with Reads[2]	SNPs[3]
AMNH 788367	American Museum of Natural History	1911		349,232	29,839	9,708	11,052
CAS 83312	California Academy of Sciences	23 May 1902	Fisher, WK	33,552,892	insufficient memory		11,522
CAS 83315	California Academy of Sciences	23 May 1902	Fisher, WK	366,101	47,565	10,054	10,755
FMNH 188929	Field Museum of Natural History	1913		971,511	133,562	23,560	11,480
ROM 62812	Royal Ontario Museum	23Jun1891	Palmer	396,525	28,580	8,951	11,005
ROM 62814	Royal Ontario Museum	16Jun1891	Palmer	983,059	164,200	25,055	11,135
ROM 62816	Royal Ontario Museum	19Jun1891	Palmer	448,244	27,177	7,336	10,789
ROM 62820	Royal Ontario Museum	1Jan1913	Willet	379,740	23,819	7,229	10,974
UMMZ 121979	University of Michigan Museum of Zoology	Feb 6 1913	Bailey, Alfred	1,553,987	202,252	25,468	11,479
UMMZ 121980	University of Michigan Museum of Zoology	Feb 6 1913	Bailey, Alfred	253,079	33,015	9,898	10,423
UMMZ 121983	University of Michigan Museum of Zoology	Feb 22 1913	Bailey, Alfred	989,296	100,820	17,895	11,476
UMMZ 121986	University of Michigan Museum of Zoology	Mar 1 1913		979,103	136,551	23,191	11,421
UMMZ 70838	University of Michigan Museum of Zoology	Jan 2 1913	Willet	1,368,294	184,842	26,583	11,478

1: Indicates number of reads aligning to capture baits. This value could not be calculated for samples with very large numbers of reads due to program limitations. 2: Number of baits with at least one read aligning. 3: How many of the 11,527 final filtered SNP set were genotyped in this sample.

Table 4-2. Laysan finch modern samples passing quality filters.

Sample	Sample Type	Island	Total Illumina Reads	On-Target Reads[1]	Baits with Reads[2]	SNPs[3]
84444	Erythrocytes	Grass	296,380	21,457	6,991	10,665
84445	Erythrocytes	Grass	614,351	40,674	11,439	11,249
84449	Erythrocytes	Grass	1,062,886	149,283	23,817	11,362
84450	Erythrocytes	Grass	351,455	20,589	5,661	10,583
84452	Erythrocytes	Grass	493,886	61,266	14,112	11,229
98184448	Erythrocytes	Grass	637,342	72,071	16,274	11,285
800135046	Plasma and erythrocytes	Grass	336,005	23,076	7,127	11,016
800135047	Plasma and erythrocytes	Grass	391,243	19,950	4,903	10,406
800135054	Plasma and erythrocytes	Grass	34,199,645	insufficient memory		11,525
800135057	Plasma and erythrocytes	Grass	1,127,072	173,359	26,442	11,515
800135059	Plasma and erythrocytes	Grass	2,425,000	343,044	30,718	11,338
804110330	Blood	Grass	614,151	28,313	8,610	10,639
804110334	Blood	Grass	635,752	36,633	10,134	11,207
804110403	Blood	Grass	1,036,692	119,589	22,274	11,083
841626573	Erythrocytes	Grass	296,909	20,597	6,312	10,838
800135227	Plasma and erythrocytes	Laysan	4,046,903	347,557	23,749	11,522
800135228	Plasma and erythrocytes	Laysan	1,008,405	98,488	19,182	11,403
800135233	Plasma and erythrocytes	Laysan	2,658,600	241,911	26,219	11,475
800135240	Plasma and erythrocytes	Laysan	1,016,178	92,537	18,358	10,925
800135802	Blood	Laysan	620,558	34,689	9,327	11,249
806185525	Blood	Laysan	287,440	31,868	9,209	10,712

806185552	Blood	Laysan	545,695	32,869	9,765	10,992
806185560	Blood	Laysan	14,922,668	insufficient memory		11,526
806185561	Blood	Laysan	402,896	29,989	7,518	10,955
806185568	Blood	Laysan	329,814	19,301	5,493	10,396
806185579	Blood	Laysan	612,472	30,630	8,055	10,751
#7	Blood	Laysan	814,290	117,513	19,328	11,336
83403	Erythrocytes	North	370,609	41,838	9,014	10,969
84346	Erythrocytes	North	16,267,439	insufficient memory		11,524
84348	Erythrocytes	North	5,464,336	insufficient memory		11,474
84350	Erythrocytes	North	310,060	24,328	8,046	10,866
84351	Erythrocytes	North	1,121,028	43,262	9,798	10,399
84406	Erythrocytes	North	1,579,366	210,752	25,837	11,468
84407	Erythrocytes	North	277,924	20,716	4,795	10,726
84408	Plasma and erythrocytes	North	335,206	209,722	29,079	10,906
84408	Erythrocytes	North	212,222	23,366	7,376	11,394
84126525	Erythrocytes	North	1,304,308	207,693	27,839	11,003
98184455	Erythrocytes	North	549,613	62,607	15,390	11,313
800135173	Plasma and erythrocytes	North	688,052	75,668	16,159	11,339
800135675	Blood	North	4,586,960	415,629	23,625	11,468
800135688	Blood	North	715,005	95,989	17,823	11,441
9818449	Erythrocytes	Southeast	1,448,922	151,096	23,208	11,493
97176172	Plasma and erythrocytes	Southeast	357,601	25,332	5,453	11,021
98184493	Erythrocytes	Southeast	648,091	49,049	12,502	11,350

98184494	Erythrocytes	Southeast	236,507	30,514	10,722	10,511
98184500	Erythrocytes	Southeast	1,763,190	190,593	25,719	11,508
99167303	Erythrocytes	Southeast	27,253,040	insufficient memory		11,527
99167305	Erythrocytes	Southeast	233,770	17,909	6,175	10,568
99167309	Erythrocytes	Southeast	381,805	40,945	10,967	10,936
99167311	Erythrocytes	Southeast	475,175	32,216	5,924	11,165
99167321	Erythrocytes	Southeast	1,525,499	223,511	25,971	11,486
99167332	Erythrocytes	Southeast	449,849	28,503	8,546	11,070
99167334	Erythrocytes	Southeast	455,236	51,173	15,158	11,025
800135720	Blood	Southeast	1,109,455	111,652	18,605	11,404
804110322	Blood	Southeast	884,313	96,123	17,844	11,376
804110326	Blood	Southeast	2,867,636	288,026	28,931	11,518
804110327	Blood	Southeast	1,137,972	220,679	27,075	11,221
804110342	Blood	Southeast	512,917	30,283	9,486	11,076
804110377	Blood	Southeast	1,554,039	85,879	16,312	11,494

1: Indicates number of reads aligning to capture baits. This value could not be calcuated for samples with very large numbers of reads due to program limitations. 2: Number of baits with at least one read aligning. 3: How many of the 11,527 final filtered SNP set were genotyped in this sample.

Population Genetic Analysis

Heterozygosity was assessed by obtaining for each individual the total number of genotyped SNP loci and the number of those that were heterozygous, then calculating the percentage of genotyped sites that were heterozygous. Significant differences in heterozygosity between populations were tested using ANOVA.

We used VCFtools to calculate average allele frequencies for each locus in each population and a custom python script to calculate the number of fixed loci and private alleles in each population. The *G*-test of independence was used to compare counts of private alleles fixed loci between each source and founder population, using a Bonferroni correction for multiple comparisons. Because these values were sensitive to sample size (Groombridge et al. 2009), we limited the sample size for these comparisons to 12 individuals from each population because there were only 12 individuals from modern Laysan with < 10% missing data in the 11,527 SNP set. In this set, considering all populations, 10,646 SNPs were biallelic, 860 SNPs had three alleles, and 21 had four.

Loci that were out of Hardy-Weinberg equilibrium (HWE) for each population were identified using an exact test suitable for many loci (Wigginton et al. 2005) in PLINK (Purcell et al. 2007) with a $P < 0.05$ cutoff. Loci that were out of HWE were removed for subsequent population genetic analyses so that only neutral loci were used to estimate neutral processes. We used several methods to detect population structure or differentiation, including principal components analysis using SmartPCA in EIGENSOFT 6.0.1 (Patterson et al. 2006, Price et al. 2006), Bayesian clustering with fastSTRUCTURE v. 1.0, a modification to the STRUCTURE program designed to run efficiently with genome-wide SNP data (Raj et al. 2014), and discriminant analysis of

principal components (DAPC) analysis in adegenet v. 1.4-2 (Jombart and Ahmed 2011). Hierarchical population structure was tested with AMOVA and pairwise population F_{ST} values were calculated in Arlequin v. 3.5 (Excoffier and Lischer 2010) with significance being tested with 10,000 permutations of the data for both. We ran AMOVA with three different structure settings: with all populations in the same group, with three groups of populations defined as follows: Laysan museum, Laysan modern, and PHR (Southeast, Grass, and North), and with three groups of populations defined as follows: Laysan museum, Laysan modern and Southeast, and Grass and North. The dataset was converted to the appropriate format for each software using PGDSpider (Lischer and Excoffier 2012) or PLINK. Museum samples that were included in population structure analysis were collected on Laysan in 1891, 1902, 1911, and 1913. Rabbits were introduced to Laysan in 1902 and were completely eradicated in 1923, so the 1911 and 1913 samples were collected during the period of the rabbits. Considering that the Laysan population probably regularly fluctuates between ~5,000 and ~10,000 finches (Morin and Conant 1994) and that the census estimate in 1913 was 4,000 individuals, it was still very close to the 'normal' range (and suggests that the rabbits had not significantly impacted the Laysan habitats by then). Therefore, the 1911 and 1913 samples could reasonably be considered pre-20th century-bottleneck. To verify this assumption, we ran the 1891/1902 and 1911/1913 samples through SmartPCA clustering analysis as two separate populations to determine if the two groups were homogenous. We did not detect any differentiation, so all museum samples were treated as one group.

Mutation-scaled effective population sizes ($\Theta = 4N_e \mu$) and migration rates ($M = m/\mu$) were estimated using a Bayesian inference strategy in the coalescence-based

program MIGRATE-n v. 3.6.11 (Beerli and Felsenstein 2001, Beerli and Palczewski 2010). We used five replicates of one long chain and four heated chains with an increment of 100 and sampling every 5,000 steps. The first 10,000 steps were discarded as burn-in. It is recommended to use DNA sequences of 100 bp or longer rather than SNPs for MIGRATE, so we filtered our BWA alignments to find regions of this length or longer with high quality genotype calls and good coverage amongst our samples. First, SAM files were filtered using SAMtools (Li et al. 2009) to select only those with mapping quality ≥ 60 and exclude filter flags indicating segment unmapped, next segment in the template unmapped, secondary alignment, not passing quality controls, PCR or optical duplicate, or supplementary alignment (command: SAMtools view -h -o sample_out.sam -q 60 -F 0xF0C -S sample.sam). Next, SAM files were converted to BAM using SAMtools and then to BED using BEDtools (Quinlan and Hall 2010), excluding split alignments. BEDtools multiinter with the cluster option was used to find overlapping sites between the 72 individuals, and these were filtered using awk to only include sites ≥ 100 bp and found in all 72 samples. Genotypes for these regions were called using GATK UnifiedGenotyper, filtered to remove indels, converted into tab format using VCFtools. A custom python script (Supplementary Material) was used to convert the tab format into migrate format. IUPAC codes were used to represent heterozygous genotypes. Although we had originally intended to only genotype continguous sequences, removal of indels resulted in some gaps. Therefore, any genotypes that were within 100 bp were concatenated into a single sequence, resulting in 173 total sequences for MIGRATE analysis.

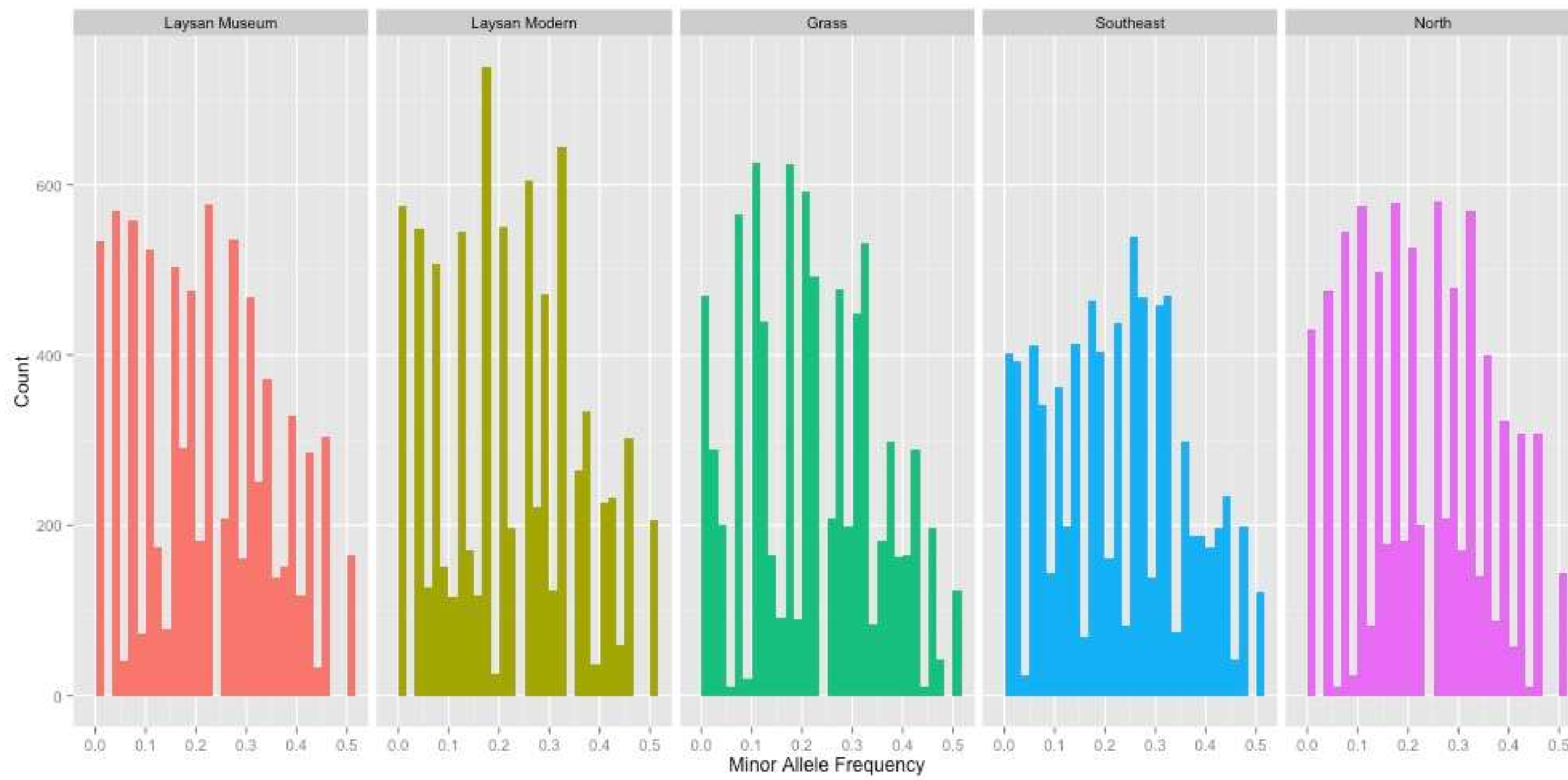
Results

A total of 11,527 variant sites which passed quality, coverage, and missing data filters were called from our data set. Table 4-3 shows mean heterozygosity in each population; there was no significant difference in heterozygosity between any of the populations ($P > 0.05$). Considering all 11,527 loci, neither number of fixed loci (mean 512.8 ± SD 23.2) nor number of private alleles (mean 218.6 ± SD 5.68) was significantly different for any founder-source population comparison at the Bonferroni-corrected level.

Table 4-3. Sample size and heterozygosity for Laysan finch modern populations and museum samples. First heterozygosity column is for 11,527 quality filtered loci; second column is for 8,095 HWE loci. Heterozygosity calculated as percent genotyped SNPs that were heterozygous. All samples had < 10% missing data.

Population	*N*	Heterozygosity (SD), All Loci	Heterozygosity (SD), HWE Loci
Laysan Museum	13	0.547 (0.090)	0.402 (0.101)
Laysan Modern	12	0.540 (0.089)	0.397 (0.097)
Southeast	18	0.559 (0.080)	0.414 (0.090)
Grass	15	0.531 (0.087)	0.386 (0.097)
North	14	0.555 (0.088)	0.412 (0.101)

Figure 4-2. Minor allele frequencies. 8,095 HWE SNPs.



3,432 loci were out of HWE for two or more populations, leaving 8,095 putatively neutral loci for population genetic analyses. Heterozygosity for these loci is shown in Table 4-3; as for the full SNP set, there was no significant difference in heterozygosity between any of the populations for neutral loci ($P > 0.05$). Minor allele frequencies are shown in Figure 4-2. No population structure was detected using the neutral locus set. Plots of the first three principle components and two discriminant functions for DAPC are shown in Figures 4-3 and 4-4, respectively. Tracy-Widom statistics indicated that the first 5 principal components were significant and they explained 57.7% of the variation in the data (Patterson et al. 2006). No pairwise population F_{ST} values were significantly different from zero (Table 4-4). Bayesian inference estimates of Θ ($4N_e \mu$) and long term effective migration rate are shown in Table 4-5.

Figure 4-3. Plot of individual loadings for the first three principal components from smartPCA analysis.

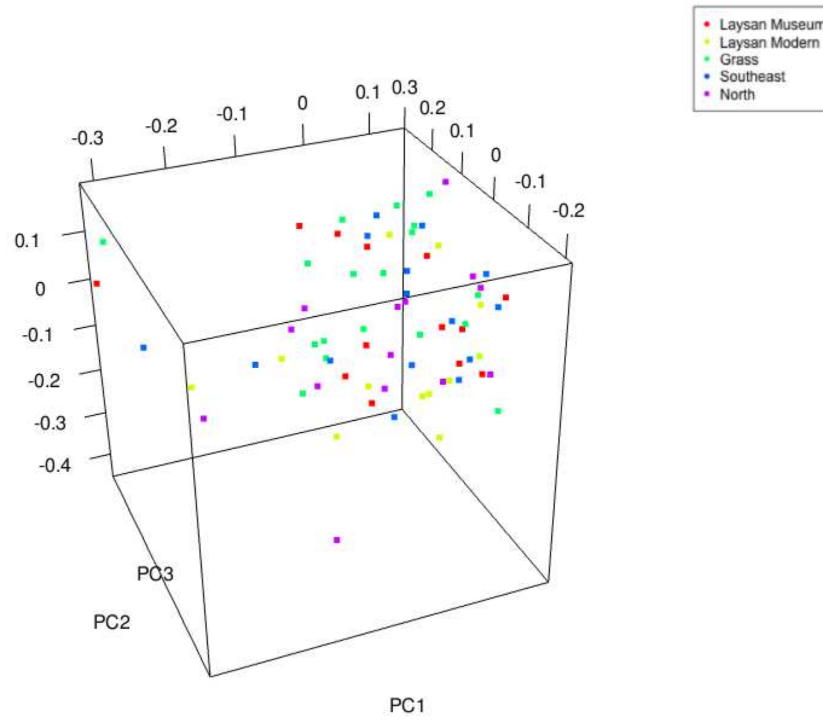


Figure 4-4. Plot of the first two discriminant functions from DAPC analysis.

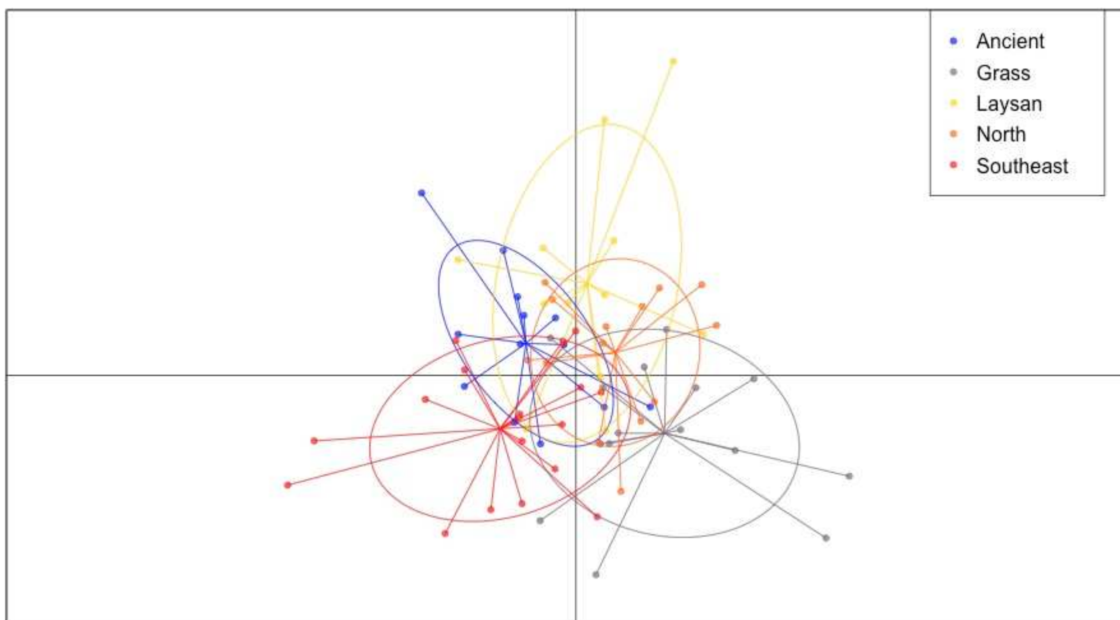


Table 4-4. Pairwise population Fst values. All P > 0.05.

	Laysan museum	Laysan modern	Grass	Southeast
Laysan museum	0.0			
Laysan modern	-0.015	0.0		
Grass	-0.012	-0.011	0.0	
Southeast	-0.012	-0.011	-0.010	0.0
North	-0.014	-0.015	-0.012	-0.010

Table 4-5. Bayesian inference estimates of Θ and migration using 173 DNA sequences from Laysan finch capture. Values are mode (25th, 75th percentile) of the posterior distribution.

Parameter	Laysan Museum	Laysan Modern	Southeast	Grass	North
Θ	0.04623 (0.04467, 0.0478)	0.04677 (0.0454, 0.04813)	0.04877 (0.04707, 0.05013)	0.04857 (0.04713, 0.04987)	0.04877 (0.04747, 0.05007)
Migration from Laysan		-	751 (739.9, 762.7)	754.3 (742, 768)	769 (756, 780.7)
Migration from Southeast		745 (732.7, 756)	-	759 (747.3, 770.7)	757.7 (746.7, 768)
Migration from Grass		762.3 (750.7, 773.3)	745 (732.7, 756.7)	-	750.3 (738, 762.7)
Migration from North		747 (746, 766.7)	743 (732, 752.7)	759 (748.7, 769.3)	-

Discussion

Island biogeographic theory predicts that populations on small islands like Laysan or PHR are subject to strong effects of drift and are more likely to go extinct due to stochastic events or catastrophes (Pimm et al. 1988). The well-documented history of the Laysan finch includes several bottleneck events which could be expected to exacerbate the existing effects of drift related to small population size: an episode of invasive rabbits

on Laysan in the early 20th century that brought the census size down to an estimated 100 individuals, the translocation of 108 finches from Laysan to Southeast island in 1967, and dispersal from Southeast to the other islets of PHR. Due to the random sampling nature of bottlenecks, each post-bottleneck or founder group should only include a subset of the variation present in its founding population; rare variants are unlikely to be sampled and have a high probability of being lost in each bottleneck or founder event. In this study, we expected to find decreasing genetic variation when comparing each founder population to its source, and discernible structure because of population differentiation caused by drift. However, our results show no discernible loss of diversity between museum samples collected before the 20th century bottleneck and modern samples and a distinct lack of differentiation between any of the Laysan finch populations. Several factors are likely to have contributed to the patterns we found, including demographic trends in Laysan finches, the properties of SNP loci, and the parameters of the bottlenecks investigated in our study.

Previous research on the modern Laysan finch populations showed that the loss of allelic diversity between source and founder populations was not as dramatic as expected, but was significant when comparing Grass and North to Laysan (Tarr et al. 1998). The authors suggested that loss of diversity in the 20th century Laysan bottleneck could have made it possible for most of Laysan's allelic diversity to be sampled during the founding of the PHR populations. They reasoned that rare alleles would have been lost in the bottleneck, leaving only common alleles that could then have more easily been sampled in the subsequent founder events on PHR. This hypothesis, with some modification, is supported by our results. In addition to the effects of the 20th century Laysan bottleneck,

the demographic dynamics of living on a tiny island could have played a major role in limiting genetic variation on Laysan historically, shaping how future bottlenecks and founder events would impact genetic diversity. Evidence from regular censuses taken from 1968 through 1988 suggests that the Laysan finch population on Laysan typically experiences "large, erratic fluctuations" in size (range: 5,000 – 20,000 individuals) as opposed to maintaining a stable population level (Morin and Conant 1994). Variable or fluctuating population size has been proposed as a better model than stable population size for bird species on the British isles as well (Pimm and Redfearn 1988). Such patterns have implications for patterns of genetic variation. Fluctuating population size results in an effective population size (N_e) much lower than the census size, even in otherwise ideal populations (Vucetich et al. 1997). Low N_e reduces the time for new mutations to be lost to drift (Kimura and Ohta 1969). In the case of Laysan, chronically low N_e could mean that most rare alleles that emerge through mutation are continually being lost through the effects of drift. Then very few rare alleles would have been present at the time of the 20th century bottleneck. Although we did observe some private alleles in each population, this effect would explain why there were not a greater number of private alleles in the museum samples than the modern samples.

Ironically, chronically small N_e and loss of rare alleles fostered by the unusual demographic pattern on Laysan could have maintained an allele frequency distribution with favorable odds for preserving genetic diversity through a bottleneck. Subdivision of island populations of the Mauritius kestrel appears to have preserved genetic diversity in that species (Nichols et al. 2001). With highly polymorphic loci like microsatellites, bottlenecks can results in significant loss of alleles even if heterozygosity is maintained,

and this effect can drive population differentiation (Chakraborty and Nei 1977, Hedrick 2005). However, for SNPs, which in most cases have only two alleles, even if the minor allele frequency is just 5%, a bottleneck survivor group of 30 has a 95% chance of including both alleles (Hedrick 2005). If allele frequencies are relatively moderate, then even with a small number of bottleneck survivors, the probability of sampling both alleles is high. Loci with intermediate allele frequencies have the highest probability of maintaining variation through a bottleneck or founder event. In our museum sample data, 75% of loci had minor allele frequencies (MAF) above 11%, and half were above 19% (see Figure 4-2). At its lowest census, in 1923 (the year the rabbits were removed by the Tanager expedition), the Laysan finch population included roughly 100 individuals. This bottleneck size, combined with the allele frequency distribution observed in our museum data, provides an explanation for why most of the 20th century genetic variation survived in the modern samples. It also supports the hypothesis of Tarr et al. (1998) that most of the genetic variation of Laysan could have been sampled in the translocation to PHR. Interestingly, comparison of modern and ancient samples of another Hawaiian bird, the nene, also showed no loss of genetic diversity despite population decline, likely due to a much older bottleneck event having already depleted variation and rarer alleles (Paxinos et al. 2002).

Our results also indicate that most of the variation on Laysan was captured in the sample of 108 individuals transported to Southeast Island in 1967. This finding is consistent with previous research, which has shown allelic diversity and heterozygosity on Southeast to be the same (Tarr et al. 1998) when compared with Laysan. When discussing the absence of a loss of heterozygosity for the founder population on

Southeast, previous studies have mentioned that drift likely resulted in wide fluctuations in allele frequencies for the multiallelic loci used, creating statistical noise and potentially obscuring the signal of the bottleneck on heterozygosity because few loci were sampled (Fleischer et al. 1991, Tarr et al. 1998). In this study, we observed no difference in heterozygosity between Laysan and Southeast when examining over 5,000 SNP loci. The founder size of 108 individuals and the distribution of allele frequencies we observed for modern Laysan samples, similar to that of the museum Laysan samples, suggests a high probability of most of the variants being sampled in a founder group of 108 finches. The rapid population growth on Southeast following the founder event would also have minimized loss of variation, as opposed to sustained small population size.

Grass and North Islands, both founded by small numbers of migrants from Southeast ($N = 2$ for North; $N = 8$ for Grass, which includes six individuals moved there by USFWS), provide the most anomalous results from our data. Both could reasonably be expected to have lower genetic diversity and to have differentiated from Southeast and Laysan due to the effects of drift because of their small founder numbers and small population sizes (Lande 1980, Frankham et al. 2002). However, our results show no loss of heterozygosity or differentiation from other populations. Although our estimates of Θ for modern Laysan finch populations suggest a larger effective population size for Laysan than for the PHR populations, both Grass (to a small degree) and North (by a wider margin) have higher estimates than their source population, Southeast.

The results of previous studies have also been somewhat contrary to expectations, finding these populations to harbor more genetic variation than predicted by their small founder sizes, especially for North. Fleischer et al. (1991) reported no difference in

heterozygosity amongst PHR islets using allozymes, while Tarr et al. (1998) reported that while average microsatellite heterozygosity was lower in Grass and North compared to Southeast and Laysan, the difference was not as large as expected. Non-significant differences between source and founder populations has also been observed for some populations of the New Zealand saddleback (Lambert et al. 2005).

Bottleneck theory predicts that there is a higher probability that alleles at a locus will be lost when there are more alleles at the locus (Allendorf 1986, Hedrick 2005), and the microsatellite diversity on North and Grass generally followed this prediction (Tarr et al. 1998). Both North and Grass retained both alleles for the two microsatellite loci that only had two. For the two loci with the most alleles (5), only 1 – 3 alleles were retained. In this study using (mostly) biallelic SNP loci, we did not see a loss of 'allelic diversity'- most polymorphism was maintained, as in the two-allele microsatellite loci. However, despite the better probability of retaining polymorphism when there are only two loci and allele frequencies are moderate, we would expect extremely small founder size to reduce SNP heterozygosity. Taking North Island as an example, with a founder size of two, the probability of retaining polymorphism at a biallelic locus ranges from approximately 30% (for minor allele frequency 0.1) to 90% (for MAF 0.5) (Hedrick 2005).

Rapid population growth on North Island following the translocation (Tarr et al. 2000) may have been a factor in the level of genetic diversity remaining, as was postulated for New Zealand saddleback founder populations (Lambert et al. 2005). Our estimate of Θ does suggest a larger effective population size for North and Grass compared to Southeast.

Another factor that could be influencing the results is undocumented migration between the PHR islets. Tropical island birds are thought to be sedentary compared to mainland birds (Pratt 2009), and observers on Laysan indicate that the Laysan finches don't move much within their island (Morin and Conant 1994). Corroborating these observations are the significant F_{ST} values of Tarr et al. (1998) and Fleischer et al. (1991), which appear to support a lack of migration between PHR islets. However, our F_{ST} values show very low genetic differentiation between the PHR islets. Laysan finches did colonize Grass, North, and Seal-Kittery Islands (with a supplemental translocation of six individuals to Grass), so migration within PHR cannot be completely ruled out as a mechanism contributing to the maintenance of high heterozygosity and lack of population differentiation for Grass and North.

A final factor which could be contributing to the maintenance of heterozygosity in Laysan finches is selective pressure which favors heterozygotes. This could come from inbreeding depression if homozygotes are less fit, or there could possibly be a heterozygote advantage. However, it is difficult to comment on such possibilities without more information about fitness – genotype associations. Our results underscore the importance of conducting *in situ* studies of genetic diversity in island populations. The Laysan finch superficially appears to be a straightforward case of successive bottlenecks or founder events and corresponding loss of genetic diversity. However, our data for over 5,000 SNP loci show a surprising consistency of genetic diversity across a history of bottlenecks and founder events. A possible explanation is that genetic diversity of the Laysan finch has been shaped through time by unique demographic patterns enforced by insular life. Wide fluctuations in population size maintain chronically low N_e, which

quickly removes rare alleles. This pattern may have resulted in moderate allele frequency distributions for loci that remain polymorphic, since alleles with extreme frequencies would either rapidly become fixed or disappear due to the effects of drift. Alleles at moderate frequencies, especially for biallelic loci, have a better probability of persisting through bottlenecks. Comparison of our results with conclusions from previous population genetic studies of the Laysan finch highlights the differing properties of multiallelic loci and SNPs in a system where chronically low N_e has interacted with bottleneck events. SNPs, with only two alleles, are likely to remain polymorphic even with small founder events or bottlenecks (Hedrick 2005), as evidenced by the consistent levels of heterozygosity and lack of population structure between all populations of the Laysan finch in this study. Microsatellites or allozymes are likely to lose allelic diversity when N_e is low, an effect that increases with increasing number of alleles (Allendorf 1986) and can result in large changes in genetic distance in response to bottlenecks (Chakraborty and Nei 1977, Hedrick 1999). In Laysan finches, which showed relatively low alleles per locus, this effect resulted in a great deal of variation in heterozygosity and allele frequency distributions after the founder events (Fleischer et al. 1991, Tarr et al. 1998).

Although our results provide a good outlook for preservation of genetic diversity of the Laysan finch through bottlenecks and founder events, it seems that new variation is unlikely to survive in these populations. This could have implications for our understanding of how adaptation occurs in this species, possibly resulting in a heavy reliance on standing variation. Future studies should focus on determining how adaptation has occurred in the Laysan finch (Stapley et al. 2010), because predicting how

a population responds to changing environments is dependent on understanding its source of genetic variation (Barrett and Schluter 2008). Understanding adaptive potential in the Laysan finch could be critical in the near future: further translocations may become urgently necessary as sea levels rise (Thomas et al. 2004), requiring adaptation to novel environments, and possibly also novel diseases, competitors, and predators (Atkinson and LaPointe 2009). For these reasons, future study of adaptation in the Laysan finch is advisable, as is monitoring of genetic variation in this species.

Supplementary Material

vcf-tab-to-migrate.py

Usage:

```
python vcf-tab-to-migrate.py sampleFile.tab
```

This script takes as input a vcf-to-tab file from VCFtools (Danecek et al. 2011) for a single population that is in single-letter IUPAC codes and converts it to sequences for MIGRATE (Beerli and Palczewski 2010). Run it once with each population's file, and then `cat` the files together to create one file with all your data. Any locus within 100 bp of the previous locus is concatenated to the previous one in order to reduce the chances of non-independent loci for migrate. Therefore, the script is most useful if you know that your vcf files contain calls for several independent contiguous sequences. MIGRATE requires some information about the number of populations, loci, samples, and locus lengths. There are several optional lines in this script (described below) that you can use to add this information to your output (or generate it, in the case of locus lengths). You will need to read the MIGRATE documentation to see what kind of information you need to add and modify the script accordingly.

When running the script with the first population, uncomment and modify the `headerLine` in order to add information about the number of populations, etc. required by MIGRATE to the output. You can also add locus length info to the output as follows: supply a tab-delimited file with the lengths of the loci and add this filename to the `lengthFile` line. Uncomment the `lengthFile` line and the two associated lines (marked with comments below) to add locus lengths to the output.

There are three lines at the end of the file beginning with **print** that you can use with trial runs of the script to get information about your data. If you don't know the number of loci, run this script one time through and uncomment the line `# print "number of loci:", len(locusLengths)`. Likewise, `# print locusLengths` will print the lengths of each locus in the data set and `# print locusInfo` will print their locations in the genome. See MIGRATE documentation for more information about the required format items, including numbers of populations, samples, locus lengths, etc. This script is offered free for use as-is with no guarantees.

```
#!/usr/bin/python

import csv
import sys

infile = sys.argv[1]
# lengthFile = 'filename.tab'

firstChr = 'no'
lastChr = 'no'
firstPos = 0
lastPos = 0
newSeq = True
samples = dict()
locusCounter = 0
locusLengths = []
locusInfo = []
minLength = 36

#headerLine = 'add text in here depending on what you want
to print'
#print headerLine

# Uncomment the next two lines if including a lengthFile
# with open(lengthFile, 'rU') as l:
# print l.read()

with open(infile, 'rU') as thisFile:
```

```
    lastRow = sum(1 for row in thisFile)

    # the number of rows is -1 for a header
    lastRow -= 1


with open(infile, 'rU') as f:
    reader = csv.reader(f, delimiter='\t')
    headers = reader.next()
    counter = 1

    for line in reader:
        fields = len(line)

        ourChr = line[0]
        ourPos = int(line[1])

        # is it a new locus or a continuation?
        if ourChr == lastChr and ourPos <= (lastPos + 100):
            newSeq = False

        else:
            newSeq = True

        # if this is a new locus
        if newSeq is True:

            # only print out results if this isn't the
first line in the file
            if newSeq is True and counter > 1:

                # see if it meets minimum length
requirement
                if len(samples[3]) >= minLength:

                    # record stats for previous locus
                    locusCounter += 1
                    locusInfo.append(str(lastChr + ':' +
str(firstPos) + '-' + str(lastPos) + '\t' + str((lastPos -
firstPos) + 1)))
                    locusLengths.append(len(samples[3]))

                    for x in range(3, fields):
                        print (headers[x] + ('x' * (10 -
len(headers[x])))) + '\t' + samples[x]
```

```
            # reset data structures for new loci
            firstChr = ourChr
            firstPos = ourPos

            lastChr = ourChr
            lastPos = ourPos

            # Add data for samples for new locus
            samples = dict()
            for x in range(3, fields):
                if line[x] != '.':
                    samples[x] = line[x]
                else:
                    samples[x] = "N"

        # If it's a continuation of a previous locus...
(newSequence = false)
        else:
            lastPos = ourPos
            lastChr = ourChr

            # ...add sample data to dictionary for this
locus
            for x in range(3, fields):
                if line[x] is not ".":
                    samples[x] += line[x]
                else:
                    samples[x] += "N"


        # If it's the very last row, print data for current
locus
        if counter == lastRow:
            # see if it meets minimum length requirement
            if len(samples[3]) >= minLength:

                locusCounter += 1
                locusInfo.append(str(lastChr + ':' +
str(firstPos) + '-' + str(lastPos) + '\t' + str((lastPos -
firstPos) + 1)))
                locusLengths.append(len(samples[3]))
                for x in range(3, fields):

                    print (headers[x] + ('x' * (10 -
len(headers[x])))) + '\t' + samples[x]
```

```
    counter += 1

# print "number of loci:", len(locusLengths)
# print locusLengths
# print locusInfo
```

Chapter 5
Identification of outlier SNP variants in populations of an endangered Hawaiian honeycreeper, the Laysan finch

Abstract

Assessment of genetic variation in endangered species is rapidly expanding to include both neutral and adaptive variation. Identification of variants under selection contributes to our understanding of evolutionary mechanisms and can help guide conservation management decisions. The Laysan finch, *Telespiza cantans,* is an endangered Hawaiian honeycreeper endemic to Laysan Island in the Northwestern Hawaiian Islands. Its demographic history includes bottlenecks and founder events, including translocation from its native Laysan island to the islets of Pearl and Hermes Reef (PHR) approximately 300 km away. We used a DNA sequence capture approach to obtain SNP genotypes for Laysan finches from Laysan and three translocated populations on PHR and identified 51 SNP loci putatively under directional selection. In contrast to results for neutral SNP loci, Laysan finch populations show differentiation at directionally selected loci. We identified functional annotations near these loci, including genes related to immune function, cilia mobility, calcium binding, and olfactory receptors.

Introduction

Conservation programs aim to preserve as much genetic variation as possible, with motivations including preserving local adaptations, preventing inbreeding

depression, and preserving adaptive potential (Soulé et al. 1986, Ballou and Lacy 1995). Therefore, it is important to assess adaptive variation in endangered species where possible (Bichet et al. 2015). This is becoming easier to do with greater numbers of genetic markers being available even in non-model organisms (Lerner and Fleischer 2010). Using genome-wide markers such as SNPs enables detection of regions of the genome undergoing selection and provides a step towards understanding how genetic variation enables adaptation in endangered species. However, how to most effectively incorporate genomic data into management strategies is an emerging area of research. Genomic markers could be used within current management strategies by accurately identifying unknown relationships among population members and providing accurate mean kinship values (W. Miller, K. Ralls, and J. Ballou, personal communication), but the assessment of population structure due to local adaptation and identification of selectively advantageous variants are relatively new strategies for conservation.

Assessment of adaptive molecular variation can lead to different conclusions regarding population structure than when purely neutral markers are used (Ackerman et al. 2013). This consideration is critical for conservation practitioners tasked with defining conservation units or determining how to sample populations for captive breeding so that all local adaptations are preserved. Spatially disparate populations, such as those on different islands, may diverge due to the neutral effects of genetic drift (Barton and Mallet 1996, Frankham 1997, Clegg et al. 2002b). However, variable selective pressures between different populations can also drive differentiation (Grant and Grant 2002, Stockwell et al. 2003, Jones et al. 2012). If selective pressures are strong enough, divergence due to differential selection may occur even in the presence of gene flow

(Blondel et al. 2006, Hess et al. 2013), which could result in divergent phenotypes and genetic differentiation at selected, but not neutral loci (Charlesworth et al. 1997, de León et al. 2010). For example, the superb fairy wren (*Malurus cyaneus*) population on Australia's Kangaroo island does not show divergence from the mainland population in microsatellite markers or mitochondrial DNA, but the two groups are morphologically divergent (Dudaniec et al. 2011). Overall, patterns of divergence at neutral versus selected loci will depend on the balance between gene flow, genetic drift, and the type of selective pressure.

The Laysan finch (*Telespiza cantans*), an endangered insular passerine, is an attractive candidate for investigation of molecular adaptive variation *in situ*. This species exists as a group of isolated populations on remote islands in the Pacific (Ely and Clapp 1973a, USFWS 2008) which have been demonstrated to show morphological divergence (Conant 1988). However, the results of neutral molecular studies regarding divergence are equivocal (Callicrate et al. n.d., Fleischer et al. 1991, Tarr et al. 1998). The demographic history of the Laysan finch includes several events which should limit its genetic diversity, and consequently, adaptive potential. A severe bottleneck on Laysan occurred in the early 20th century when rabbits introduced to Laysan destroyed most of the island's vegetation, bringing the Laysan finch population down to approximately 100 individuals and contributing to the extinction of other bird species on Laysan (Munter 1915, Ely and Clapp 1973a). A series of founder events occurred after 108 Laysan finches were translocated by the United States Fish and Wildlife Service to Southeast Island of Pearl and Hermes Reef (PHR) about 300 km away and subsequently colonized three other islets of PHR (Fleischer et al. 1991). The environment on PHR is quite

different from Laysan, including a different primary food source, the *Tribulus cistoides* seedpod, which is only found in a very small portion of the diet on Laysan. Differential beak morphology between Laysan and PHR has been reported, and size of Laysan finch bills on PHR parallels the size of *Tribulus* seedpods (Conant 1988). Bill divergence related to foraging has been well-described in another insular bird, the well-known group of Darwin's finches (Grant and Grant 2002, Abzhanov et al. 2006).

The demographic history of the Laysan finch is reflected in studies of neutral molecular diversity, which show low levels of allelic diversity (Fleischer et al. 1991, Tarr et al. 1998) and heterozygosity (Fleischer et al. 1991) and few rare alleles (Callicrate et al. n.d.). However, adaptive molecular variation has not been assessed in this species. Especially in light of the morphological differences between Laysan and PHR, it would be valuable to understand how translocation has affected adaptive variation and divergence. Studies of adaptive variation in non-model organisms are relatively new and have been boosted by the accessibility of genome-scale data (Nosil et al. 2009, Hess et al. 2013). Besides contributing to our understanding of evolutionary mechanisms, such studies are important from a conservation standpoint. In the case of the Laysan finch, for example, rising sea levels due to climate change and the spread of invasive plants on PHR could necessitate further translocations or captive breeding (McClung 2005, USFWS 2008). If captive breeding becomes necessary, knowledge of advantageous variants in the wild could help combat the effects of captive selection. The captive environment presents a different selective landscape from the wild and this can shift the captive population's genetic profile away from the optimal configuration under natural

selection, lowering fitness when captive individuals are released (Ford 2002, Woodworth et al. 2002).

In this study, we investigated adaptive variation in Laysan finch populations. We identified SNP markers exhibiting signatures of selection and investigated population differentiation based on these potentially adaptive loci. In order to begin to understand how adaptation is driving divergence in this species, we looked for annotations related to loci found to be under directional selection. We identified several loci showing signals of selection that were potentially associated with protein coding genes, providing a start into understanding how adaptation is occurring in the Laysan finch. Although we did not find any selective signal associated with morphological features previously found to be divergent between Laysan and PHR populations of the Laysan finch, our results emphasize the importance of investigating both neutral and adaptive variation when assessing genetic diversity and divergence in endangered species.

Materials & Methods

DNA extraction and sequencing library preparation

Blood samples from Laysan (N=33) and each PHR island (Grass N =33, Southeast N =34, North N =33) were collected as part of previous work (Conant 1988, Tarr et al. 1998; see Appendix A.). Samples were stored at -20°C prior to DNA extraction. Genomic DNA was extracted from blood using a BioSprint or DNEasy Extraction kit and quality was assessed by agarose gel electrophoresis. Samples were quantified using a Qubit and 500 ng of each sample was sonicated using a QSonica for 2 – 6 minutes depending on range of fragment sizes present in order to obtain a mean fragment size of

300 – 500 bp. Sonicated samples were end-repaired using DNA Polymerase I, Large (Klenow) Fragment or NEBNext End-Repair Kit. A cytosine was added to the end of fragments using Klenow Fragment (3' – 5' exo) and NEB buffer 2 with dCTP added in order to facilitate ligation of Nextera-style stubby adapter in a subsequent step using NEB Quick Ligation Kit. After stubby adapter ligation, unique dual 8-bp Nextera-style indices were added to each sample through PCR amplification using Kapa HiFi Polymerase. Success of library preparation was determined by running each post-PCR library on an agarose gel.

Sequence capture and Illumina sequencing

We used a custom in-solution array (MYcroarray MYbaits) designed to capture ~40,000 honeycreeper SNP loci (Callicrate et al. 2014) followed by Illumina sequencing to obtain SNP genotypes for this study. The capture baits were designed using other Hawaiian honeycreeper species to avoid ascertainment bias caused by developing a SNP capture array using the focal population(s) (Albrechtsen et al. 2010, Lachance and Tishkoff 2013). Although the capture array was designed to target ~40,000 SNPs, many off-target sequences are also captured and resulted in additional genotyped loci (see below).

We pooled individually indexed samples prior to capture in order to use resources efficiently (Hawkins et al. 2015). Amplified libraries were quantified using Qubit and split into groups of seven or eight individuals per pool. Each pool consisted of 600 ng total library DNA (split equally amongst included samples); no adapter indices were shared by samples within a pool. MYcroarray Protocol 1.3.7 was followed, except Block #1 (Human Cot-1) was replaced with Chicken Cot-1 to increase blocking effectiveness.

Following capture, pools were amplified for 10 – 16 cycles with Illumina primers and quantified using qPCR (Stratagene) with a Kapa Illumina Quantification Kit. Subsequently, all capture pools were pooled together in equimolar ratios and sequenced on two lanes of an Illumina HiSeq. Prior to processing the whole group of samples, three samples were captured, pooled, and sequenced on an Illumina MiSeq to validate the capture procedure.

Data processing

Each sample's reads from both lanes were pooled (for three samples which had also been run on the MiSeq, those reads were pooled also) and aligned to the amakihi draft genome sequence (Callicrate et al. 2014) using BWA (Li and Durbin 2009). PCR duplicates were marked using Picard Tools (http://broadinstitute.github.io/picard), and indels were identified and realigned using the Genome Analysis Toolkit (GATK) RealignerTargetCreator and IndelRealigner (McKenna et al. 2010, DePristo et al. 2011). GATK UnifiedGenotyper was used to call variant sites in all samples simultaneously, using parameters --min_base_quality_score 20, --standared_min_confidence_threshold_for_calling 20, and --standared_min_confidence_threshold_for_emitting 20. The resulting variant file was subjected to hard filtering using the FilterVariants GATK tool as recommended by the Broad Institute (Van der Auwera et al. 2013) when standardized reference data (such as HapMap) are unavailable (filter expression: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || HaplotypeScore > 13.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0"; see http://gatkforums.broadinstitute.org/discussion/2806/howto-apply-hard-filters-to-a-call-set). A list of variants passing coverage filters was created using the

CoveredByNSamples sites tool with parameters --minCoverage 9 (excludes a locus for a given sample if that sample's coverage is below 9) and --percentageOfSamples 0.3 (excludes loci from the set if they are found in fewer than 30% of all samples).

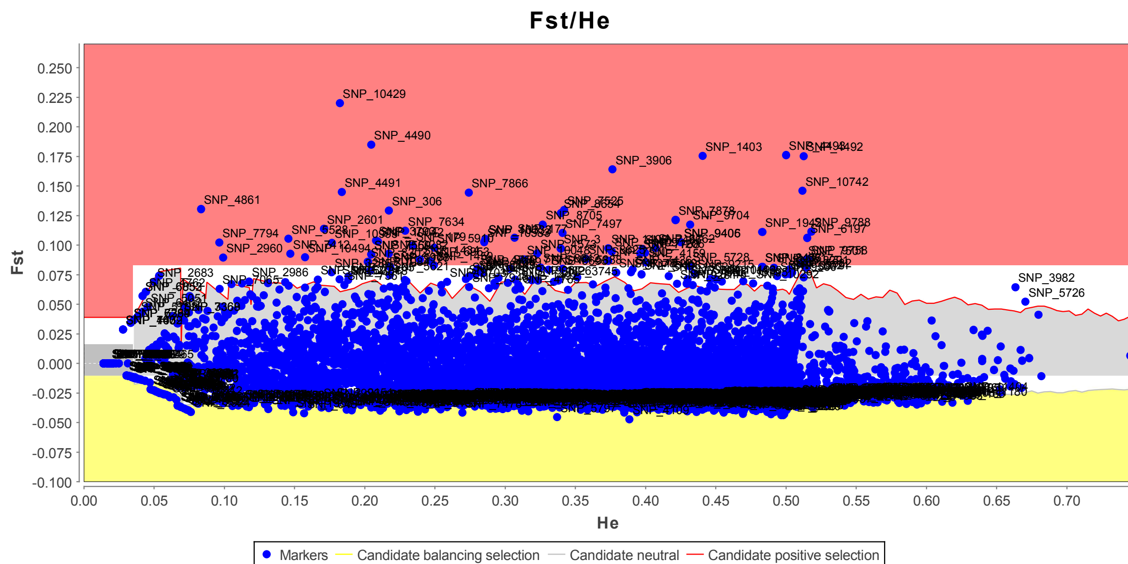
Finally, sites meeting the following set of conditions were selected using the GATK SelectVariants tool: SNP only (e.g., excludes indels), exclude non-variant sites, passed the hard filter, and passed the coverage filter. We further filtered the dataset to remove SNP sites and individuals with high missingness. First, VCFtools (Danecek et al. 2011) was used to remove sites with >50% missing data, resulting in 11,527 loci remaining. Next, any individuals with >10% missing data in this 11,527 locus set were removed, leaving 59 individuals which were used in all subsequent analyses (Laysan *N*=12, Grass *N*=15, Southeast *N*=18, North *N*=14). We assessed dropout related to sequencing coverage by calculating the percentage of quality-filtered genotyped sites that were heterozygous for each of the individuals for a range of minimum coverage values (9-20). Only very minimal differences in percentage of heterozygous sites were observed for this range of coverage, which we interpreted to mean that the amount of false homozygous calls was very low, and minimum coverage of 9 was an acceptable value to include a genotype in the dataset.

Detection of selected outliers

We used BayeScan v. 2.1 and LOSITAN to detect SNP loci under selection. BayeScan uses a Bayesian posterior odds approach to determine, for each locus, if a model including selection is more probable than a neutral model, given the allele frequencies observed in the data (Foll and Gaggiotti 2008). BayeScan was run with default parameters except for thinning set to 100 and 10,000 pilot runs. LOSITAN uses

the fdist method to identify loci with excessively high or low F_{ST} compared to neutral expectation (Beaumont and Nichols 1996, Beaumont and Balding 2004, Antao et al. 2008). F_{ST} -based tests work well in situations where there is gene flow because the background isn't a problem- there shouldn't be high F_{ST} for neutral loci as there would be because of drift if gene flow was not occurring, so only loci with differential selective pressure should have high F_{ST}. LOSITAN was run with 100,000 replicates and false discovery rate (FDR) of 0.01, with a first run to estimate the neutral mean F_{ST} and then a second one to detect outlier loci (Figure 5-1). Per-locus F_{ST} (Weir and Cockerham 1984) was calculated using VCFtools.

Figure 5-1. LOSITAN outlier detection. Loci in red region are under directional selection; loci in yellow region are under balancing selection.



We employed several strategies to help identify the nature of adaptive variation in Laysan finches. First, we blasted (Altschul et al. 1990, Morgulis et al. 2008) 200 bp

sequences surrounding directionally selected SNPs identified by Bayescan or LOSITAN against the NCBI nucleotide database in order to determine their relationship to genes or other annotations of interest. Because genes can be quite far apart from associated regulatory elements (Kleinjan and van Heyningen 2005), it is likely that we do not have a SNP within or very close to all genes that are important for adaptive variation in Laysan finches, but SNPs may still reflect the signal of selection acting on elements nearby if they are in linkage disequilibrium. Therefore, in addition to blasting the sequence immediately surrounding outlier SNPs, we also searched for annotations in the UCSC genome browser for zebra finch, *Taeniopygia guttata,* (Warren et al. 2010) in 15 kb regions centered on each focal SNP. Because stochastic demographic processes can create noise in the signal of individual SNPs, we also identified extreme F_{ST} values for 15 kb non-overlapping bins across the genome (Lamichhaney et al. 2015) and searched for annotations in these regions. Although differentiation of such regions does not by itself indicate adaptive significance, the combined localization of outlier SNP loci and highly differentiated genomic regions can help narrow the search for adaptive variation. F_{ST} for bins was calculated using VCFtools (Danecek et al. 2011) and these values were standardized using the formula $\frac{(F_{ST}-\mu F_{ST})}{\sigma F_{ST}}$ to obtain ZF_{ST} values. Regions with a ZF_{ST} of 5 or greater (Rubin et al. 2010, Axelsson et al. 2013) were examined in the UCSC genome browser for zebra finch, *Taeniopygia guttata,* (Warren et al. 2010) to identify any annotations of interest. For SNPs which were identified as outliers by LOSITAN but which were not located within a $ZF_{ST} > 5$ bin, we also checked the UCSC zebra finch browser for annotations in a 15 kb region centered on the SNP.

Our locus locations in this study are based on the draft amakihi genome (Callicrate et al. 2014). In the draft assembly, in which scaffolds were aligned to zebra finch chromosome sequences, contigs that could not be localized to zebra finch chromosomes were concatenated (separated by strings of Ns) to form 'chromosome' Un2 (named after chromosome Un of the zebra finch assembly). Since contigs within a given 15 kb stretch of Un2 sequence may not be related to each other, we did not calculate F_{ST} for bins across Un2 or search 15 kb regions surrounding outlier SNPs located in Un2.

The honeycreeper reference genome was assembled based on alignment to zebra finch but the coordinate system is slightly different. Therefore, before using the zebra finch genome browser, we converted the coordinates of each region of interest from amakihi to zebra finch using a custom perl script, MMLO.pl (written by James Thomas; see Appendix 1). Finally, because previous research has shown morphological differentiation in beaks between Laysan and PHR populations of the Laysan finch (Conant 1988), we looked for overlaps between our filtered SNP set and locations of genes known to influence beak morphology. These include bmp4 (Abzhanov et al. 2004), calmodulin (Abzhanov et al. 2006) and ALX1 (Lamichhaney et al. 2015). We identified the zebra finch locations for these genes using the UCSC genome browser and converted them to amakihi coordinates using MMLO.pl. BEDtools intersectbed (Quinlan and Hall 2010) was used determine whether any of the SNPs we genotyped were found in the range from 1,000 bp before to 1,000 bp after these genes.

Population differentiation due to selection

Using the directionally selected SNPs, we assessed population structure due to adaptive variation using a discriminant analysis of principle components (DAPC) in the

adegenet v. 1.4-2 package for R (Jombart and Ahmed 2011) and using principle components analysis (PCA) using SmartPCA in EIGENSOFT 6.0.1 (Patterson et al. 2006, Price et al. 2006). We also used the directionally selected SNPs to cluster individuals, disregarding population of origin, using adegenet.

Results

Detection of outlier loci and divergent genome regions

A plot of F_{ST} for each of the 11,527 SNP loci is shown in Figure 5-2, with loci in the 99th and 99.9th percentiles shown in dark blue and red, respectively. BayeScan found no loci with a false discovery rate below approximately 89%, so no loci were determined to be under selection using BayeScan. A possible explanation is that BayeScan is less able to detect weak effects of selection in SNPs than microsatellites and tends to work better with multiallelic markers (Foll and Gaggiotti 2008). In LOSITAN, loci with $P \leq 0.01$ or $P \geq 0.99$ were identified as under balancing (821 loci) or directional (51 loci) selection, respectively, while the remainder (10,653 loci) were considered to be neutral (Figure 5-1). SNPs identified as outliers by LOSITAN are shown as triangles in Figure 5-2.

Two 15 kb regions had ZF_{ST} scores above five, on chromosomes 4:10,035,001-10,050,000 and 7:8,670,001-8,685,000 (Figure 5-3). Two SNPs were located in the bin on chromosome 4 (10,037,743 and 10,037,818), and two were located in the bin on chromosome 7 (8,683,890 and 8,683,955). The SNPs at 4: 10,037,743 and 7: 8,683,955 were identified as outliers by LOSITAN.

Figure 5-2. Fst values for 11,527 SNPs across the Laysan finch genome. Values in the 99th percentile are shown in dark blue; values in the 99.9th percentile are shown in red; loci identified as outliers by LOSITAN are shown as triangles.

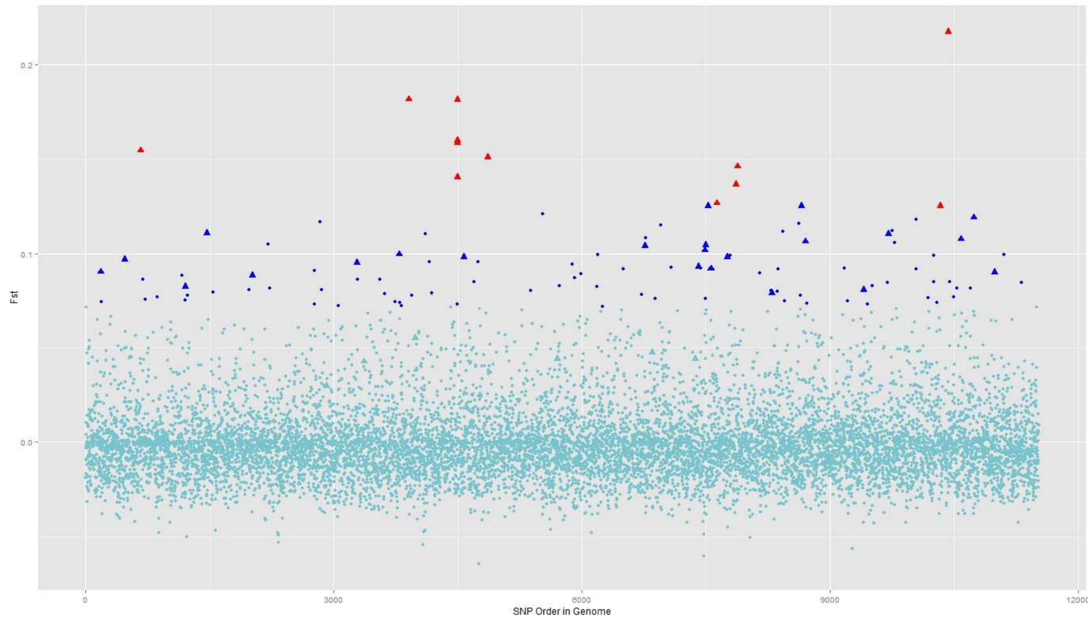
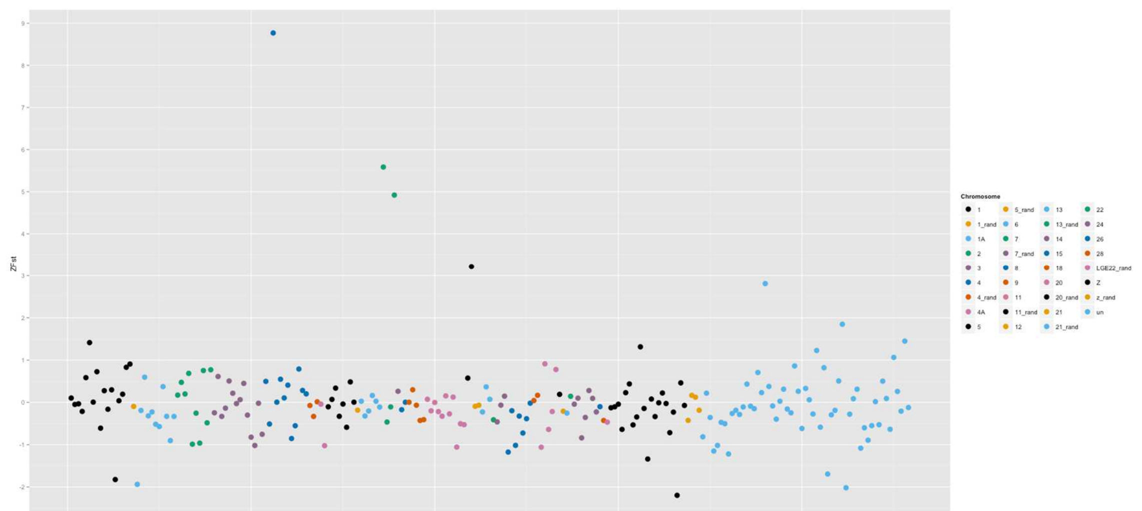


Figure 5-3. ZFst scores for 15 kb bins of the Laysan finch genome.



Population structure with outlier loci

Both DAPC and PCA showed that Laysan finch populations exhibit discernible structure when considering the 51 directionally selected loci (Figures 5-4 and 5-5). Using the twstats program in SmartPCA identified the first three principal components as significant. Together, they explained 45.8% of the variation in the data. Clustering and group assignment analysis showed that individuals from North and Southeast (and to a lesser extent, Grass) were more similar to each other than they were to Laysan (Figure 5-6).

Figure 5-4. Plot of the first two eigenvectors from DAPC for 51 directionally selected SNPs.

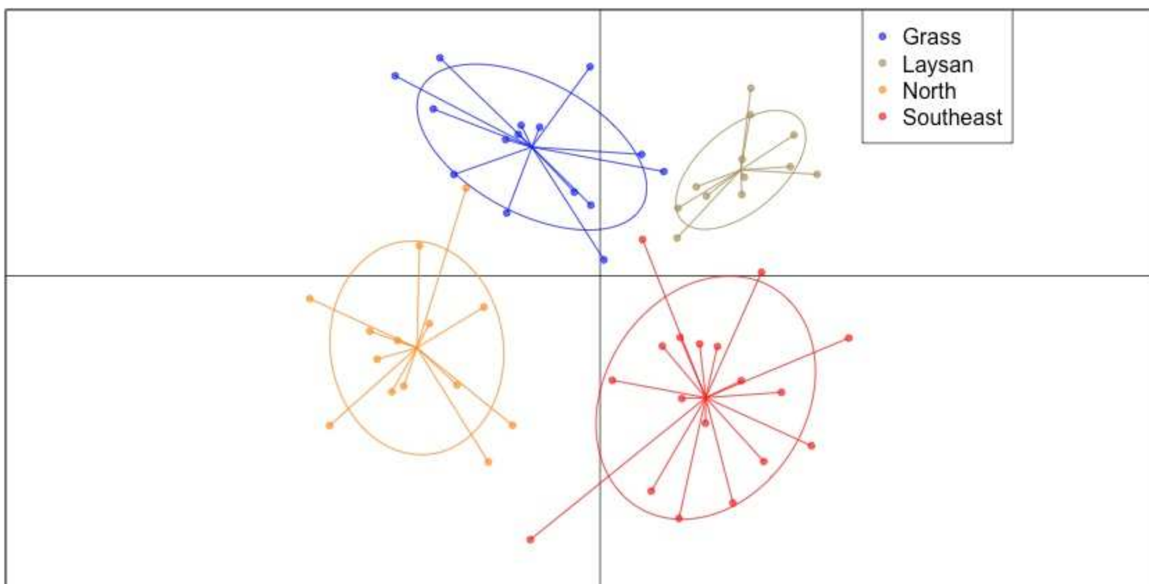


Figure 5-5. Clustering and group assignment based on 51 directionally detected SNPs.

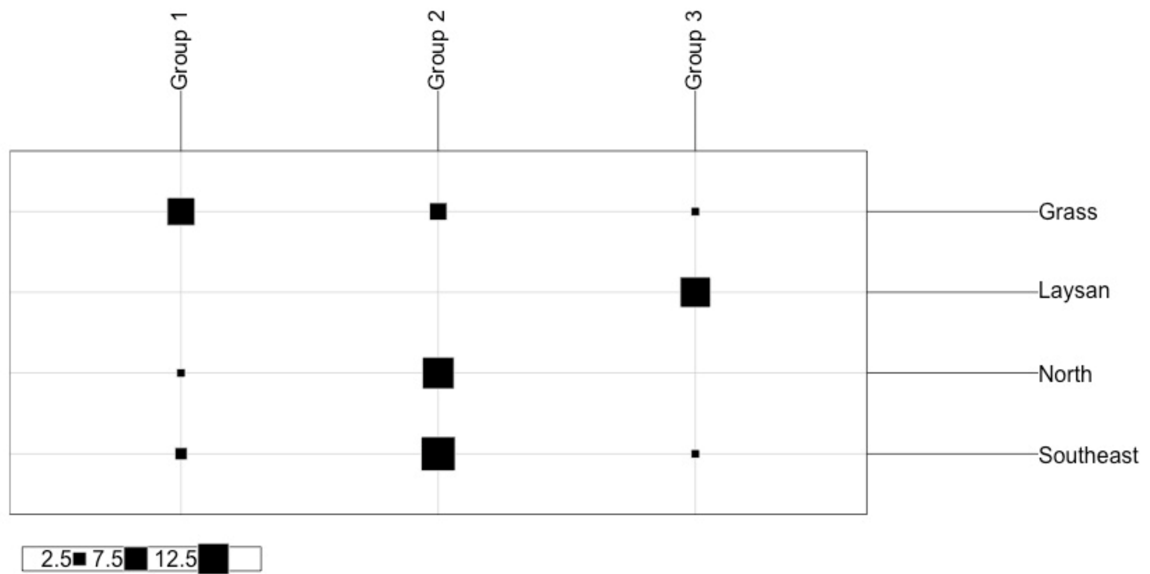
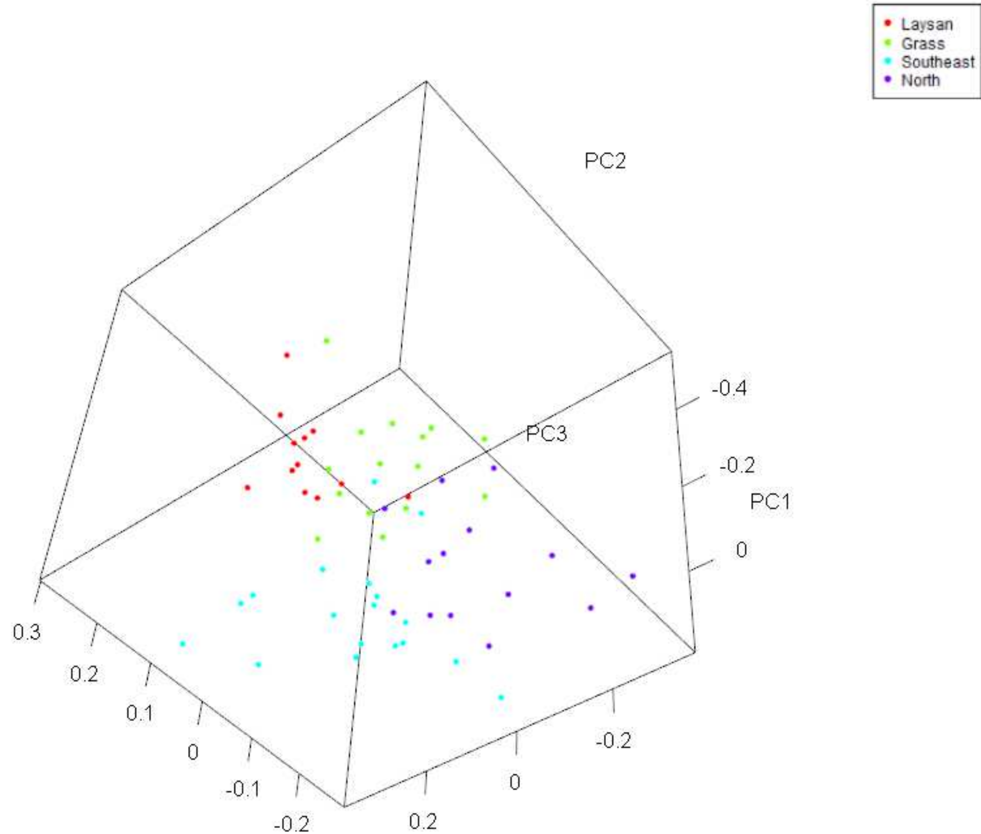


Figure 5-6. Individual loadings for the first three principal components from PCA using 51 directionally selected SNPs.



Functional significance

Results of blasting a 200 bp sequence centered on each of the 51 outlier loci are shown in Table 5-1. Some queries returned the same result or results and are listed in the same row of the table. Many sequences did not return any hits. It is interesting to note that several outlier loci on Un2 returned hits Although Un2 is composed of concatenated contigs whose position in the genome is unknown, the fact that Un2 contains outlier loci

which blast to predicted functional sequences suggests that these contigs are still valuable resources. No SNPs that we genotyped were located in or within 1,000 bp of any genes that had been identified in other studies of beak morphology.

Searching for annotations in a 15 kb region centered on outlier SNPs returned several results. The regions around 1: 113,126,126 and 24: 1,502,006 contained predicted proteins. The region around 11:16,785,598 contained predicted proteins as well as *Gallus gallus* HYDIN (hydrocephalus inducing homolog), which is involved in cilia motility and zebra finch expressed sequence tags (ESTs) from brain tissue. The region around 3: 8,367,300 contained predicted genes; a small portion of the edge of this region overlapped human CAPN8 (calpain 8) (Hata et al. 2001), which is involved in calcium binding. Finally, although there were two directionally selected SNPs on chromosome Un, only the region around one of them (Un: 124,122,245) contained annotations: olfactory receptors were reported in this region for many species, including human OR8D2.

Table 5-1. Blast hits for outlier loci detected by Lositan.

SNP Sequence Location	Blast Hit	Comments
24:1,501,906-1,502,106 Un2:1,761,683-1,761,883 Un2:2363,672-2,363,872	*Fringilla coelebs* clone pGS-1 satellite sequence	
4:10,037,643-10,037,843	*Agelaius phoeniceus* cosmid Rwcos3, partial sequence	Songbird Genomics: Analysis of 45 kb Upstream of a Polymorphic MHC Class II Gene in Red-Winged Blackbirds (*Agelaius phoeniceus*) (Gasper et al. 2001)
7:8,683,855-8,684,055	*Taeniopygia guttata* chromosome UNK clone TGMCBa-50H12, complete sequence; *Fringilla coelebs* clone pGS-1 satellite sequence; PREDICTED: *Corvus brachyrhynchos* transmembrane protein 18 (TMEM18), mRNA	
Un2:1589364-1589564	*Taeniopygia guttata* chromosome UNK clone TGMCBa-50H12, complete sequence	
Un2:14,736,637-14,736,837 Un2:14,736,641-14,736,841 Un2:21,363,506-21,363,706 Un2:25,205,997-25,206,197	PREDICTED: *Geospiza fortis* olfactory receptor 14C36-like (LOC102044608), mRNA	These sequences had many blast hits from bird species similar to the one shown at left, all for olfactory receptors

Neither of the high ZF_{ST} regions on chromosome 7 contained any annotations or predicted annotations. The region on chromosome 4 contained a predicted ortholog for E74-Like Factor 2 (ELF2) which was first identified in humans (Oettgen et al. 1996) and has orthologs in many species, including chicken (*Gallus gallus*). This gene functions in

sequence-specific DNA binding and transcription regulation, and it is involved in pathways for immune response signaling.

Discussion

In this study, we identified 51 outlier loci which are candidate targets of directional selection. Our study of adaptive variation was motivated by the finding that Laysan finches on different islands show morphological differentiation and the knowledge that translocation to PHR involved an environmental change, likely incorporating different selective pressures than on Laysan. We wanted to find out how adaptation might be shaping molecular diversity and divergence in these populations. Although previous work had shown no Laysan finch population differentiation at thousands of neutral SNPs (previous chapter), when the 51 outliers identified in our study were used, differentiation became apparent. These directionally selected SNPs could represent a response to different selective pressures due to the variation between Laysan and the islets of Pearl and Hermes Reef (Conant 1988, McClung 2005). For an endangered species like the Laysan finch, this information could be critical for conservation decisions. Prior knowledge of locally adaptive variants increases the chance of sampling all important variation in case of translocation or collection for captive breeding. In the event that captive breeding becomes necessary for the Laysan finch, which is a consideration as sea levels rise, knowledge of which variants or haplotypes are advantageous in the wild habitat could also help managers combat the effects of captive selection, which has been demonstrated to have severe fitness consequences when

captive-bred individuals are eventually released into the wild (Araki et al. 2007, Montgomery et al. 2010, Lacy et al. 2013).

We were able to use resources from well-studied model organisms to help identify the nature of features which may be adaptively important in Laysan finches. It is important to note that these results must be interpreted cautiously. Many functional features are conserved between species and birds have been shown to have a high degree of synteny and homology (Stapley et al. 2008), but until actual functional genetics studies are carried out, we cannot be certain as to the nature of specific regions of the Laysan finch genome. The closest model organism with available annotations is the zebra finch, approximately 33.5 million years diverged from the Hawaiian honeycreepers (Jetz et al. 2012). We have used zebra finch annotations (and inferred annotations applied from other species to zebra finch) to describe what types of features may be under selective pressure in Laysan finches. Two regions, on chromosomes 4 and 7, were identified in both the individual outlier locus analysis and bin-wise differentiation. On chromosome four, our results indicate that there may be some selective pressure related to the immune system. An outlier locus on chromosome four blasted to a sequence identified in a study of the region upstream of MHC in red-winged blackbirds (Gasper et al. 2001). This outlier was within a region of elevated ZF_{ST} which also contained a transcription factor known to act in immune pathways. It is logical that immune system features would be targets of selective pressure; MHC in particular is well-studied and is known to have a high level of diversity due to selective pressures (Hess and Edwards 2002). In the case of the Laysan finch, it is reasonable to suggest that the differing environments on PHR and Laysan could have exerted different selective pressures on the immune system, resulting

in divergence related to immune genes (Cohen 2002, Ekblom et al. 2007, Bichet et al. 2015). Selective pressure exerted by pathogens has been suggested as major factor in local adaptation in humans (Fumagalli et al. 2011), and pathogens could also be playing a role in local adaptation of Laysan finches. Molecular diversity in this region on chromosome four may therefore be a good candidate for future studies. Again, we caution that the results should be interpreted conservatively and further study is required to draw firm conclusions about functional significance of specific regions.

Other annotations identified in this study are more difficult to put in context. Many of them appear to be general, including cilia mobility, calcium binding, and olfactory receptors. Interestingly, several of the outlier loci found in contigs with unknown genomic location (i.e., on sequence Un2) had blast hits to olfactory receptors, as did one outlier located on chromosome Un. Olfactory receptors are the largest multigene family in humans (Niimura and Nei 2003); their ubiquity may have made them difficult to place in zebra finch and also in the honeycreeper draft genome, explaining why outlier loci associated with these genes cannot be properly placed in chromosomes. There were also several high ZF_{ST} regions and outlier SNPs associated with non-specific predicted protein or gene annotations. Although we cannot speculate as to how any of these annotations might provide functional adaptive significance in Laysan finches, it is reassuring that genomic locations identified in our analysis are likely to have a function. The paucity of information about genes or regulatory elements in proximity to these regions of interest highlights the need for further study into the mechanisms of adaptation. The genomics revolution has made it relatively easy to identify large numbers of loci and pinpoint genomic regions of interest, but studies of this nature seem to have

outpaced the critical work into functional genetics that will provide an explanation for how adaptation is occurring. In the meantime, conservation practitioners can use the information available from studies like ours to start to identify regions of importance for local adaptation and begin to develop strategies to ensure that variation in these regions is preserved.

General Conclusion

Bottlenecks and founder events are important demographic events that can influence the genetic variation and adaptive potential of populations. These events are especially important for endangered species, where preserving genetic diversity and adaptive variation are priorities. We used both a laboratory insect model and molecular genetic study of *in situ* bottlenecked populations to investigate how bottlenecks interact with island demographic dynamics and translocation to a novel environment to affect genetic variation.

Our experimental study of serial bottlenecks in a novel environment using the red flour beetle showed that phenotypic and additive variance in a quantitative trait could be increased when bottlenecks occur in a novel environment. Additive variation is necessary for selection to be effective, so an increase in additive variation may result in increased adaptive potential. However, any adaptive advantage would depend on the shape of the fitness landscape in the new environment- essentially, whether any advantageous loci were contained within the increased range of variation- and the genetic architecture of the trait in question. These are factors that would need to be determined on a case-by-case basis, but our results do indicate that the environment in which a bottleneck occurs is certainly an important consideration in determining the outcome for adaptive potential and quantitative genetic variation.

In our study of the serially bottleneck Laysan finch, we found that island demographic dynamics can also influence the outcome of bottlenecks on genetic variation. The modern Laysan population did not have lower neutral genetic diversity

than museum samples, in contrast to our predictions. It is possible that the naturally occurring population fluctuations on Laysan have created a situation where there is a very high level of drift, effectively keeping the number of rare alleles low. This effect, in combination with the relatively large size of bottleneck survivors on Laysan and the rapid population growth following the bottleneck there, enabled the preservation of existing variation through the Laysan bottleneck. Likewise, the relatively large number of individuals that founded the PHR population on Southeast was able to include enough of the variation from Laysan to make the populations genetically indistinguishable.

An implication of our results is that new variants are unlikely to be maintained in Laysan finch populations, and any adaptation that occurs would likely come from standing variation. As for many endangered species, especially those found on islands, there is a high level of risk for Laysan finches due to climate change, rising sea levels, and the introduction of invasive competitors, predators, and disease. The founder events on PHR were accompanied by a change in environmental factors, and although we could not estimate changes in phenotypic and additive variation as we did in our experimental study, we attempted to quantify changes in adaptive variation by detecting signatures of selection in the genome using SNP markers.

Although Laysan finch populations showed remarkable genetic homogeneity when considering genome-wide SNP markers, we did find 51 outlier loci showing greater than expected differentiation. Two of these loci were also identified when using a bin-wise approach to reduce stochastic noise in the data. Quantifying adaptive variation in natural populations is an emerging area of research, and identification of loci or regions under selection in this case did not point to a gene or genes that might explain whether

(or how) adaptation has occurred in the PHR populations. However, results such as ours could be very useful for conservation managers trying to identify locally advantageous variants to target for captive breeding or other conservation actions.

Appendix A. Sample information

Table A-1. Modern Laysan finch samples.

Sample	Sample Type	Island	Status[1]	Total Illumina Reads	On-Target Reads[2]	Baits with Reads[3]	SNPs[4]
26574	Plasma	Grass	DNA too degraded				
84443	Plasma	Grass	DNA too degraded				
84444	Erythrocytes	Grass	Good coverage	296,380	21,457	6,991	10,66 5
84445	Plasma	Grass	DNA too degraded				
84445	Erythrocytes	Grass	Good coverage	614,351	40,674	11,439	11,24 9
84449	Erythrocytes	Grass	Good coverage	1,062,886	149,283	23,817	11,36 2
84450	Erythrocytes	Grass	Good coverage	351,455	20,589	5,661	10,58 3
84452	Erythrocytes	Grass	Good coverage	493,886	61,266	14,112	11,22 9
84126573	Plasma	Grass	Low coverage	210,928	14,803	4,807	
98184448	Erythrocytes	Grass	Good coverage	637,342	72,071	16,274	11,28 5
98184453	Blood	Grass	Library unsuccessful				
80013504 6	Plasma and erythrocytes	Grass	Good coverage	336,005	23,076	7,127	11,01 6
80013504	Plasma and	Grass	Good coverage	391,243	19,950	4,903	10,40

7	erythrocytes						6
800135048	Plasma and erythrocytes	Grass	Low coverage	182,623	7,457	949	
800135049	Plasma and erythrocytes	Grass	Low coverage	72,496	8,254	3,037	
800135050	Plasma and erythrocytes	Grass	Low coverage	257,182	16,841	5,911	
800135051	Plasma and erythrocytes	Grass	Low coverage	465,332	38,592	1,020	
800135052	Plasma and erythrocytes	Grass	Low coverage	10,475	856	409	
800135053	Plasma and erythrocytes	Grass	Low coverage	135,937	10,426	3,467	
800135054	Plasma and erythrocytes	Grass	Good coverage	34,199,645	insufficient memory		11,525
800135056	Plasma and erythrocytes	Grass	Low coverage	138,942	3,335	790	
800135057	Plasma and erythrocytes	Grass	Good coverage	1,127,072	173,359	26,442	11,515
800135058	Plasma and erythrocytes	Grass	Low coverage	40,256	3,073	1,500	
800135059	Plasma and erythrocytes	Grass	Good coverage	2,425,000	343,044	30,718	11,338
800135060	Plasma and erythrocytes	Grass	Low coverage	125,449	16,568	322	
804110329	Blood	Grass	Low coverage	364,046	11,748	1,139	
804110330	Blood	Grass	Good coverage	614,151	28,313	8,610	10,639
804110331	Blood	Grass	Low coverage	395,190	51,866	1,343	

804110333	Blood	Grass	Low coverage	33,011	3,596	259	
804110334	Blood	Grass	Good coverage	635,752	36,633	10,134	11,207
804110400	Blood	Grass	Low coverage	231,292	131,634	752	
804110403	Blood	Grass	Good coverage	1,036,692	119,589	22,274	11,083
841626573	Erythrocytes	Grass	Good coverage	296,909	20,597	6,312	10,838
800135227	Plasma and erythrocytes	Laysan	Good coverage	4,046,903	347,557	23,749	11,522
800135228	Plasma and erythrocytes	Laysan	Good coverage	1,008,405	98,488	19,182	11,403
800135230	Plasma and erythrocytes	Laysan	Low coverage	68,117	1,332	886	
800135231	Plasma and erythrocytes	Laysan	Low coverage	69,044	6,129	2,164	
800135233	Plasma and erythrocytes	Laysan	Good coverage	2,658,600	241,911	26,219	11,475
800135239	Plasma and erythrocytes	Laysan	Low coverage	65,552	8,759	4,622	
800135240	Plasma and erythrocytes	Laysan	Good coverage	1,016,178	92,537	18,358	10,925
800135802	Blood	Laysan	Good coverage	620,558	34,689	9,327	11,249
800135929	Blood	Laysan	Low coverage	247,688	15,263	4,727	
800135999	Blood	Laysan	Low coverage	502,746	56,210	1,488	
806185328	Blood	Laysan	Low coverage	62,705	4,257	2,008	

806185512	Blood	Laysan	Low coverage	53,850	4,203	1,708	
806185524	Blood	Laysan	Low coverage	92,979	15,428	6,117	
806185525	Blood	Laysan	Good coverage	287,440	31,868	9,209	10,712
806185546	Blood	Laysan	DNA too degraded				
806185552	Blood	Laysan	Good coverage	545,695	32,869	9,765	10,992
806185553	Blood	Laysan	Low coverage	28,621	5,046	2,612	
806185560	Blood	Laysan	Good coverage	14,922,668	insufficient memory		11,526
806185561	Blood	Laysan	Good coverage	402,896	29,989	7,518	10,955
806185565	Blood	Laysan	Low coverage	473,349	50,045	1,303	
806185566	Blood	Laysan	DNA too degraded				
806185567	Blood	Laysan	Low coverage	86,001	6,964	1,885	
806185568	Blood	Laysan	Good coverage	329,814	19,301	5,493	10,396
806185574	Blood	Laysan	Low coverage	1,241,580	66,387	2,945	
806185575	Blood	Laysan	Library unsuccessful				
806185576	Blood	Laysan	Low coverage	46	2	2	
806185577	Blood	Laysan	Low coverage	2,273	321	165	

806185578	Blood	Laysan	Low coverage	77,701	8,668	3,478	
806185579	Blood	Laysan	Good coverage	612,472	30,630	8,055	10,751
806185581	Blood	Laysan	Low coverage	161,584	1,837	952	
806185582	Blood	Laysan	Low coverage	51,306	4,644	1,301	
#2	Blood	Laysan	Low coverage	189,100	12,950	4,144	
#7	Blood	Laysan	Good coverage	814,290	117,513	19,328	11,336
83403	Erythrocytes	North	Good coverage	370,609	41,838	9,014	10,969
84345	Erythrocytes	North	DNA too degraded				
84346	Erythrocytes	North	Good coverage	16,267,439	insufficient memory		11,524
84347	Erythrocytes	North	Low coverage	18,810	341	101	
84348	Erythrocytes	North	Good coverage	5,464,336	insufficient memory		11,474
84350	Erythrocytes	North	Good coverage	310,060	24,328	8,046	10,866
84351	Erythrocytes	North	Good coverage	1,121,028	43,262	9,798	10,399
84404	Erythrocytes	North	Low coverage	131,500	1,910	335	
84405	Erythrocytes	North	Low coverage	539,649	60,568	1,119	
84406	Erythrocytes	North	Good coverage	1,579,366	210,752	25,837	11,468
84407	Erythrocytes	North	Good coverage	277,924	20,716	4,795	10,726
84408	Plasma and	North	Good coverage	335,206	209,722	29,079	10,90

	erythrocytes						6
84408	Erythrocytes	North	Good coverage	212,222	23,366	7,376	11,394
84409	Plasma and erythrocytes	North	Low coverage	94,449	12,451	4,547	
84410	Erythrocytes	North	Low coverage	484,586	8,896	2,115	
84411	Erythrocytes	North	Low coverage	192,287	14,638	5,081	
84126473	Plasma and erythrocytes	North	Low coverage	115,365	7,204	2,721	
84126487	Plasma and erythrocytes	North	Low coverage	114,231	13,284	5,736	
84126525	Erythrocytes	North	Good coverage	1,304,308	207,693	27,839	11,003
98184455	Erythrocytes	North	Good coverage	549,613	62,607	15,390	11,313
98184456	Erythrocytes	North	Low coverage	454,100	64,077	1,900	
98184457	Erythrocytes	North	Low coverage	70,230	4,505	1,590	
800135170	Plasma and erythrocytes	North	Library unsuccessful				
800135171	Plasma and erythrocytes	North	Low coverage	397,984	53,576	1,035	
800135172	Plasma and erythrocytes	North	Low coverage	160,795	11,846	4,269	
800135173	Plasma and erythrocytes	North	Good coverage	688,052	75,668	16,159	11,339
800135675	Blood	North	Good coverage	4,586,960	415,629	23,625	11,468
800135688	Blood	North	Good coverage	715,005	95,989	17,823	11,441
80013569	Blood	North	Low coverage	305,411	21,481	4,277	

6							
800135700	Blood	North	Low coverage	389,859	45,931	932	
804110346	Blood	North	Low coverage	226,409	14,333	3,592	
804110348	Blood	North	Low coverage	334,369	24,687	1,197	
804110353	Blood	North	Low coverage	184	5	5	
9818449	Erythrocytes	Southeast	Good coverage	1,448,922	151,096	23,208	11,493
97175819	Erythrocytes	Southeast	DNA too degraded				
97176172	Plasma and erythrocytes	Southeast	Good coverage	357,601	25,332	5,453	11,021
98184429	Plasma and erythrocytes	Southeast	DNA too degraded				
98184491	Erythrocytes	Southeast	Library unsuccessful				
98184492	Erythrocytes	Southeast	DNA too degraded				
98184493	Erythrocytes	Southeast	Good coverage	648,091	49,049	12,502	11,350
98184494	Erythrocytes	Southeast	Good coverage	236,507	30,514	10,722	10,511
98184497	Erythrocytes	Southeast	Low coverage	188,517	14,169	4,919	
98184498	Erythrocytes	Southeast	Low coverage	163,157	17,970	6,259	
98184500	Erythrocytes	Southeast	Good coverage	1,763,190	190,593	25,719	11,508

99167303	Erythrocytes	Southeast	Good coverage	27,253,040	insufficient memory		11,527
99167305	Erythrocytes	Southeast	Good coverage	233,770	17,909	6,175	10,568
99167308	Erythrocytes	Southeast	Low coverage	71,776	5,307	2,230	
99167309	Erythrocytes	Southeast	Good coverage	381,805	40,945	10,967	10,936
99167311	Erythrocytes	Southeast	Good coverage	475,175	32,216	5,924	11,165
99167312	Erythrocytes	Southeast	Low coverage	457,607	11,745	2,302	
99167314	Erythrocytes	Southeast	Low coverage	178,401	13,519	4,124	
99167321	Erythrocytes	Southeast	Good coverage	1,525,499	223,511	25,971	11,486
99167322	Erythrocytes	Southeast	Low coverage	132,741	11,428	4,334	
99167332	Erythrocytes	Southeast	Good coverage	449,849	28,503	8,546	11,070
99167334	Erythrocytes	Southeast	Good coverage	455,236	51,173	15,158	11,025
99167337	Erythrocytes	Southeast	Low coverage	115,746	10,611	3,972	
99167342	Erythrocytes	Southeast	Low coverage	839,417	33,562	1,727	
99167343	Erythrocytes	Southeast	Low coverage	99,575	14,407	5,104	
800135720	Blood	Southeast	Good coverage	1,109,455	111,652	18,605	11,404
804110322	Blood	Southeast	Good coverage	884,313	96,123	17,844	11,376

804110324	Blood	Southeast	Library unsuccessful				
804110326	Blood	Southeast	Good coverage	2,867,636	288,026	28,931	11,518
804110327	Blood	Southeast	Good coverage	1,137,972	220,679	27,075	11,221
804110338	Blood	Southeast	Low coverage	64,416	8,480	3,311	
804110342	Blood	Southeast	Good coverage	512,917	30,283	9,486	11,076
804110374	Blood	Southeast	DNA too degraded				
804110377	Blood	Southeast	Good coverage	1,554,039	85,879	16,312	11,494

1: Status of sample. 'DNA too degraded' - library prep was not attempted; 'Library unsuccessful' - too low quantity or quality for sequencing; 'Low coverage' - sequenced but not enough data to include sample in final data set; 'Good coverage' - sample included in final data set. 2: Indicates number of reads aligning to capture baits. This value could not be calcuated for samples with very large numbers of reads due to program limitations. 3: Number of baits with at least one read aligning. 4: How many of the 11,527 final filtered SNP set were genotyped in this sample.

Table A-2. Museum Laysan finch samples.

Sample	Museum	Collection Date	Collector	Status[1]	Total Reads	On-Target Reads[2]	Baits with Reads[3]	SNPs[4]
AMNH 453658	American Museum of Natural History			Low coverage	6,546	1,027	637	
AMNH 453673	American Museum of Natural History			Library unsuccessful				
AMNH 788367	American Museum of Natural History	1911		Good coverage	349,232	29,839	9,708	11,052
CAS 83307	California Academy of Sciences	19 May 1902	Snyder, JO	Low coverage	12,047	161	44	
CAS 83308	California Academy of Sciences	19 May 1902	Fisher, WK	Low coverage	98,755	35,135	165	
CAS 83310	California Academy of Sciences	21 May 1902	Fisher, WK	Low coverage	69,546	7,111	161	
CAS 83311	California Academy of Sciences	23 May 1902	Fisher, WK	Low coverage	291,415	29,771	565	
CAS 83312	California Academy of Sciences	23 May 1902	Fisher, WK	Good coverage	33,552,892	insufficient memory		11,522
CAS 83313	California Academy of Sciences	23 May 1902	Fisher, WK	Low coverage	326,191	10,284	575	
CAS 83314	California Academy of Sciences	23 May 1902	Fisher, WK	Low coverage	93,588	3,634	494	

CAS 83315	California Academy of Sciences	23 May 1902	Fisher, WK	Good coverage	366,101	47,565	10,054	10,755
CAS 83316	California Academy of Sciences	23 May 1902	Fisher, WK	Low coverage	10,423	528	341	
CAS83305	California Academy of Sciences	17 May 1902	Snyder, JO	Library unsuccessful				
FMNH 188929	Field Museum of Natural History	1913		Good coverage	971,511	133,562	23,560	11,480
FMNH 188930	Field Museum of Natural History			Low coverage	19,837	1,628	795	
ROM 62812	Royal Ontario Museum	23Jun1891	Palmer	Good coverage	396,525	28,580	8,951	11,005
ROM 62814	Royal Ontario Museum	16Jun1891	Palmer	Good coverage	983,059	164,200	25,055	11,135
ROM 62815	Royal Ontario Museum			Low coverage	12,392	669	499	1,491
ROM 62816	Royal Ontario Museum	19Jun1891	Palmer	Good coverage	448,244	27,177	7,336	10,789
ROM 62820	Royal Ontario Museum	1Jan1913	Willet	Good coverage	379,740	23,819	7,229	10,974
ROM 62821	Royal Ontario Museum	2Jan1915 or 1913	Willet	Low coverage	138,609	6,705	2,052	7,497
UMMZ 121976	University of Michigan Museum of Zoology	Dec 25 1912	Bailey, Alfred	Low coverage	228,617	16,602	430	
UMMZ 121977	University of Michigan Museum of Zoology	Jan 2 1913	Bailey, Alfred	Library unsuccessful				

UMMZ 121978	University of Michigan Museum of Zoology	Jan 27 1913	Bailey, Alfred	Low coverage	79,004	5,370	1,476	
UMMZ 121979	University of Michigan Museum of Zoology	Feb 6 1913	Bailey, Alfred	Good coverage	1,553,987	202,252	25,468	11,479
UMMZ 121980	University of Michigan Museum of Zoology	Feb 6 1913	Bailey, Alfred	Good coverage	253,079	33,015	9,898	10,423
UMMZ 121981	University of Michigan Museum of Zoology	Feb 7 1913	Bailey, Alfred	Low coverage	24,546	3,711	1,584	
UMMZ 121982	University of Michigan Museum of Zoology	Feb 15 1913	Bailey, Alfred	Low coverage	5,269	197	139	
UMMZ 121983	University of Michigan Museum of Zoology	Feb 22 1913	Bailey, Alfred	Good coverage	989,296	100,820	17,895	11,476
UMMZ 121984	University of Michigan Museum of Zoology	Feb 24 1913	Bailey, Alfred	Low coverage	194,035	16,736	6,761	
UMMZ 121985	University of Michigan Museum of Zoology	Mar 1 1913		Low coverage	109,220	12,154	4,504	
UMMZ 121986	University of Michigan Museum of Zoology	Mar 1 1913		Good coverage	979,103	136,551	23,191	11,421
UMMZ 70838	University of Michigan Museum of Zoology	Jan 2 1913	Willet	Good coverage	1,368,294	184,842	26,583	11,478
USNM 189448	National Musuem of Natural History	8May1902		Library unsuccessful				

	(Smithsonian)						
USNM 189450	National Musuem of Natural History (Smithsonian)	19May1902		Library unsuccessful			
USNM 189454	National Musuem of Natural History (Smithsonian)	23May1902		Library unsuccessful			
USNM 189457	National Musuem of Natural History (Smithsonian)	21May1902		Low coverage	88,240	6,886	2,372
USNM 231200	National Musuem of Natural History (Smithsonian)	11May1911		Low coverage	75,530,934	insufficient memory	

1: Status of sample. 'DNA too degraded' - library prep was not attempted; 'Library unsuccessful' - too low quantity or quality for sequencing; 'Low coverage' - sequenced but not enough data to include sample in final data set; 'Good coverage' - sample included in final data set. 2: Indicates number of reads aligning to capture baits. This value could not be calcuated for samples with very large numbers of reads due to program limitations. 3: Number of baits with at least one read aligning. 4: How many of the 11,527 final filtered SNP set were genotyped in this sample.

References

Abzhanov, A., W. P. Kuo, C. Hartmann, B. R. Grant, P. R. Grant, and C. J. Tabin (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442:563–567.

Abzhanov, A., M. Protas, B. R. Grant, P. R. Grant, and C. J. Tabin (2004). Bmp4 and Morphological Variation of Beaks in Darwin's Finches.

Ackerman, M., W. Templin, J. Seeb, and L. Seeb (2013). Landscape heterogeneity and local adaptation define the spatial genetic structure of Pacific salmon in a pristine environment. Conservation Genetics 14:483–498. doi: 10.1007/s10592-012-0401-7

Albrechtsen, A., F. C. Nielsen, and R. Nielsen (2010). Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. Molecular Biology and Evolution 27:2534 –2547. doi: 10.1093/molbev/msq148

Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. Zoo Biology 5:181–190.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman (1990). Basic local alignment search tool. Journal of Molecular Biology 215:403–410. doi: 10.1016/S0022-2836(05)80360-2

Amadon, D. (1950). The Hawaiian honeycreepers (Aves, Drepaniidae). Bulletin of the AMNH 95:151–262.

Amos, W., and A. Balmford (2001). When does conservation genetics matter? Heredity 87:257–265.

Amos, W., and J. Harwood (1998). Factors affecting levels of genetic diversity in natural populations. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 353:177–186. doi: 10.1098/rstb.1998.0200

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart (2008). LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323. doi: 10.1186/1471-2105-9-323

Araki, H., B. Cooper, and M. S. Blouin (2007). Genetic Effects of Captive Breeding Cause a Rapid, Cumulative Fitness Decline in the Wild. Science 318:100–103. doi: 10.1126/science.1145621

Ardern, S. L., and D. M. Lambert (1997). Is the black robin in genetic peril? Molecular Ecology 6:21–28. doi: 10.1046/j.1365-294X.1997.00147.x

Aslam, M., J. Bastiaansen, M. Elferink, H.-J. Megens, R. Crooijmans, L. Blomberg, R. Fleischer, C. Van Tassell, T. Sonstegard, S. Schroeder, M. Groenen, and J. Long (2012). Whole genome SNP discovery and analysis of genetic diversity in Turkey (Meleagris gallopavo). BMC Genomics 13:391.

Atkinson, C., R. Dusek, K. Woods, and W. Iko (2000). Pathogenicity of avian malaria in experimentally-infected Hawaii Amakihi. J Wildl Dis 36:197–204.

Atkinson, C. T., and D. A. LaPointe (2009). Introduced Avian Diseases, Climate Change, and the Future of Hawaiian Honeycreepers. J. Avian Med. Surg. 23:53–63.

Atkinson, C. T., and M. D. Samuel (2010). Avian malaria Plasmodium relictum in native Hawaiian forest birds: epizootiology and demographic impacts on apapane Himatione sanguinea. Journal of Avian Biology 41:357–366. doi: 10.1111/j.1600-048X.2009.04915.x

Atkinson, C. T., K. L. Woods, R. J. Dusek, L. S. Sileo, and W. M. Iko (1995). Wildlife disease and conservation in Hawaii: Pathogenicity of avian malaria (Plasmodium relictum) in experimentally infected Iiwi (Vestiaria coccinea). Parasitology 111:S59–S69.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, et al. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In Current Protocols in Bioinformatics. John Wiley & Sons, Inc., pp. 11.10.1–11.10.33.

Axelsson, E., A. Ratnakumar, M.-L. Arendt, K. Maqbool, M. T. Webster, M. Perloski, O. Liberg, J. M. Arnemo, A. Hedhammar, and K. Lindblad-Toh (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature 495:360–364. doi: 10.1038/nature11837

Backström, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer, A. Qvarnström, and H. Ellegren (2008). A Gene-Based Genetic Linkage Map of the Collared Flycatcher (Ficedula albicollis) Reveals Extensive Synteny and Gene-Order Conservation During 100 Million Years of Avian Evolution. Genetics 179:1479–1495. doi: 10.1534/genetics.108.088195

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE 3:e3376. doi: 10.1371/journal.pone.0003376

Ballou, J. D., and R. C. Lacy (1995). Identifying Genetically Important Individuals for Management of Genetic Variation in Pedigreed Populations. In Population Management for Survivial and Recovery: Analytical Methods and Strategies in Small Population Conservation. Columbia University Press, New York, pp. 76–111.

Banko, W. E., and P. C. Banko (2009). Historic Decline and Extinction. In Conservation Biology of Hawaiian Forest Birds. Yale University Press, pp. 25–58.

Barker, F. K., A. Cibois, P. Schikler, J. Feinstein, and J. Cracraft (2004). Phylogeny and diversification of the largest avian radiation. Proceedings of the National Academy of Sciences of the United States of America 101:11040–11045. doi: 10.1073/pnas.0401892101

Barrett, R. D. H., and D. Schluter (2008). Adaptation from standing genetic variation. Trends in Ecology & Evolution 23:38–44. doi: 10.1016/j.tree.2007.09.008

Barton, N. H., and J. Mallet (1996). Natural Selection and Random Genetic Drift as Causes of Evolution on Islands [and Discussion]. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 351:785–795. doi: 10.1098/rstb.1996.0073

Barton, N. H., and M. Turelli (1989). Evolutionary quantitative genetics: how little do we know? Annual Reviews of Genetics 23:337–370.

Barton, N. H., M. Turelli, and D. Houle (2004). Effects of genetic drift on variance components under a general model of epistasis. Evolution 58:2111–2132. doi: 10.1554/03-684

Bataillon, T., and M. Kirkpatrick (2000). Inbreeding depression due to mildly deleterious mutation in finite populations: size does matter. Genetical Research 75:75–81.

Beaumont, M. A., and D. J. Balding (2004). Identifying adaptive genetic divergence among populations from genome scans.

Beaumont, M. A., and R. A. Nichols (1996). Evaluating Loci for Use in the Genetic Analysis of Population Structure. Proceedings: Biological Sciences 263:1619–1626. doi: 10.2307/50648

Becker, W. (1984). Manual of quantitative genetics. 4th edition. Academic Enterprises, Pullman.

Beerli, P., and J. Felsenstein (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98:4563–4568. doi: 10.1073/pnas.081068098

Beerli, P., and M. Palczewski (2010). Unified Framework to Evaluate Panmixia and Migration Direction Among Multiple Sampling Locations. Genetics 185:313–326. doi: 10.1534/genetics.109.112532

Bichet, C., Y. Moodley, D. J. Penn, G. Sorci, and S. Garnier (2015). Genetic structure in insular and mainland populations of house sparrows (Passer domesticus) and their hemosporidian parasites. Ecology and Evolution 5:1639–1652. doi: 10.1002/ece3.1452

Blondel, J., P. C. Dias, P. Perret, M. Maistre, and M. M. Lambrechts (1999). Selection-Based Biodiversity at a Small Spatial Scale in a Low-Dispersing Insular Bird. Science 285:1399.

Blondel, J., D. W. Thomas, A. Charmantier, P. Perret, P. Bourgault, and M. M. Lambrechts (2006). A Thirty-Year Study of Phenotypic and Genetic Variation of Blue Tits in Mediterranean Habitat Mosaics. BioScience 56:661–673.

Bollmer, J. L., N. K. Whiteman, M. D. Cannon, J. C. Bednarz, T. De Vries, P. G. Parker, and K. Steenhof (2005). Population genetics of the Galapagos hawk (Buteo galapagoensis): genetic monomorphism within isolated populations.

Bouzat, J. (2010). Conservation genetics of population bottlenecks: the role of chance, selection, and history. Conservation Genetics 11:463–478.

Bowcock, A. M., J. R. Kidd, J. L. Mountain, J. M. Herbert, L. Carotenuto, K. K. Kidd, and L. Cavalli-Sforza (1991). Drift, Admixture, and Selection in Human Evolution: A Study with DNA Polymorphisms. Proceedings of the National Academy of Sciences of the United States of America 88:839–843.

Bradshaw, C. J. A., Y. Isagi, S. Kaneko, B. W. Brook, D. M. J. S. Bowman, and R. Frankham (2007). Low genetic diversity in the bottlenecked population of endangered non-native banteng in northern Australia. Molecular Ecology 16:2998–3008.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards (2003). The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution 18:249–256.

Bryant, E. H., S. A. McCommas, and L. M. Combs (1986). The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. Genetics 114:1191–1211.

Bryant, E. H., and L. M. Meffert (1988). Effect of an experimental bottleneck on morphological integration in the housefly. Evolution 42:698–707.

Bryant, E. H., and L. M. Meffert (1990). Multivariate phenotypic differentiation among bottleneck lines of the housefly. Evolution 44:660–668.

Bryant, E. H., and L. M. Meffert (1996). Morphometric differentiation in serially bottlenecked populations of housefly. Evolution 50:935–940.

Burt, D. W., C. Bruley, I. C. Dunn, C. T. Jones, A. Ramage, A. S. Law, D. R. Morrice, I. R. Paton, J. Smith, D. Windsor, A. Sazanov, et al. (1999). The dynamics of chromosome evolution in birds and mammals. Nature 402:411–413. doi: 10.1038/46555

van Buskirk, J., and Y. Willi (2006). The change in quantitative genetic variation with inbreeding. Evolution 60:2428–2434.

Callicrate, T., R. Dikow, J. Thomas, J. Mullikin, E. Jarvis, R. Fleischer, and N. C. S. Program (2014). Genomic resources for the endangered Hawaiian honeycreepers. BMC Genomics 15:1098.

Callicrate, T. E., R. C. Fleischer, and F. Siewerdt (2012). Exposure to a novel environment in conjunction with serial bottlenecks increases phenotypic and additive variation of a quantitative trait. Genomics and Quantitative Genetics 5:5–13.

Callicrate, T., R. C. Fleischer, J. Song, and S. Conant (no date). Serial bottlenecks in an endangered insular passerine, the Laysan finch. Manuscript in preparation.

Campos, P. F., T. Kristensen, L. Orlando, A. Sher, M. V. Kholodova, A. GÖTherstrÖM, M. Hofreiter, D. G. Drucker, P. Kosintsev, A. Tikhonov, G. F. Baryshnikov, et al. (2010). Ancient DNA sequences point to a large loss of mitochondrial genetic

diversity in the saiga antelope (Saiga tatarica) since the Pleistocene. Molecular Ecology 19:4863–4875. doi: 10.1111/j.1365-294X.2010.04826.x

Cardoso, M., M. Eldridge, M. Oakwood, B. Rankmore, W. Sherwin, and K. Firestone (2009). Effects of founder events on the genetic variation of translocated island populations: implications for conservation management of the northern quoll. Conservation Genetics 10:1719–1733.

Carson, H. L. (1990). Increased genetic variance after a population bottleneck. Trends in Ecology & Evolution 5:228–230.

Carson, H. L., and R. G. Wisotzkey (1989). Increase in genetic variance following a population bottleneck. The American Naturalist 134:668–673.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. G3: Genes, Genomes, Genetics 1:171 –182. doi: 10.1534/g3.111.000240

Chakraborty, R., and M. Nei (1977). Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. Evolution 31:347–356.

Charlesworth, B., M. Nordborg, and D. Charlesworth (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetics Research 70:155–174. doi: doi:null

Charpentier, M., J. M. Setchell, F. Prugnolle, L. A. Knapp, E. J. Wickings, P. Peignot, and M. Hossaert-McKey (2005). Genetic diversity and reproductive success in mandrills (Mandrillus sphinx). Proc Natl Acad Sci USA 102:16723–16728.

Cheverud, J. M., T. T. Vaughn, L. S. Pletscher, K. King-Ellison, J. Bailiff, E. Adams, C. Erickson, and A. Bonislawski (1999). Epistasis and the Evolution of Additive Genetic Variance in Populations That Pass Through a Bottleneck. Evolution 53:1009–1018.

Choiniere, A. D. (2008). Post-bottleneck inbreeding accumulation reduces fitness in laboratory populations of Tribolium castaneum under environmental stress.

Clegg, S. M., S. M. Degnan, J. Kikkawa, C. Moritz, A. Estoup, and I. P. F. Owens (2002a). Genetic Consequences of Sequential Founder Events by an Island-Colonizing Bird. Proceedings of the National Academy of Sciences of the United States of America 99:8127–8132.

Clegg, S. M., S. M. Degnan, C. Moritz, A. Estoup, J. Kikkawa, I. P. F. Owens, and S. Edwards (2002b). Microevolution in island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. Evolution 56:2090–2099. doi: 10.1554/0014-3820(2002)056[2090:miiftr]2.0.co;2

Cohen, S. (2002). Strong Positive Selection and Habitat-Specific Amino Acid Substitution Patterns in Mhc from an Estuarine Fish Under Intense Pollution Stress. Molecular Biology and Evolution 19:1870–1880.

Conant, S. (1988). Geographic variation in the Laysan Finch (Telespyza cantans). Evolutionary Ecology 2:270–282.

Crow, J. F., and M. Kimura (1970). An introduction to population genetics theory. Burgess Publishing, Minneapolis, MN.

Dalloul, R. A., J. A. Long, A. V. Zimin, L. Aslam, K. Beal, L. Ann Blomberg, P. Bouffard, D. W. Burt, O. Crasta, R. P. M. A. Crooijmans, K. Cooper, et al. (2010). Multi-Platform Next-Generation Sequencing of the Domestic Turkey (Meleagris gallopavo): Genome Assembly and Analysis. PLoS Biology 8:e1000475. doi: 10.1371/journal.pbio.1000475

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. Handsaker, G. Lunter, G. Marth, S. T. Sherry, G. McVean, et al. (2011). The Variant Call Format and VCFtools. Bioinformatics. doi: 10.1093/bioinformatics/btr330

Darwin, C. (1859). On the origin of species by means of natural selection, or the preservaiton of favoured races in the struggle for life. John Murray, London.

Davey, J. W., T. Cezard, P. Fuentes-Utrilla, C. Eland, K. Gharbi, and M. L. Blaxter (2013). Special features of RAD Sequencing data: implications for genotyping. Molecular Ecology 22:3151–3164. doi: 10.1111/mec.12084

Demont, M., W. U. Blanckenhorn, D. J. Hosken, and T. W. J. Garner (2008). Molecular and quantitative genetic differentiation across Europe in yellow dung flies. Journal of Evolutionary Biology 21:1492–1503.

DePristo, M., E. Banks, R. Poplin, K. Garimella, J. Maguire, C. Hartl, A. Philippakis, G. del Angel, M. Rivas, M. Hanna, A. McKenna, et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Reviews Genetics 43:491–8.

Dill, H. R., and W. A. Bryan (1912). Report of an expedition to Laysan Island in 1911.

Dinerstein, E., and G. F. McCracken (1990). Endangered Greater One-Horned Rhinoceros Carry High Levels of Genetic Variation. Conservation Biology 4:417–422.

Dudaniec, R. Y., B. E. Schlotfeldt, T. Bertozzi, S. C. Donnellan, and S. Kleindorfer (2011). Genetic and morphological divergence in island and mainland birds: Informing conservation priorities. Biological Conservation 144:2902–2912. doi: 10.1016/j.biocon.2011.08.007

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

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. doi: 10.1093/bioinformatics/btq461

Ekblom, R., S. A. SÆther, Pä. Jacobsson, P. Fiske, T. Sahlman, M. Grahn, J. A. KÅLÅS, and J. Höglund (2007). Spatial pattern of MHC class II variation in the great snipe (Gallinago media). Molecular Ecology 16:1439–1451. doi: 10.1111/j.1365-294X.2007.03281.x

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, S. Uebbing, and J. B. W. Wolf (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756–760. doi: 10.1038/nature11584

Ely, C. A., and R. B. Clapp (1973a). The natural history of Laysan Island, Northwestern Hawaiian Islands.

Ely, C. A., and R. B. Clapp (1973b). The natural history of Laysan Island, Northwestern Hawaiian Islands. Atoll Research Bulletin.

Excoffier, L., and H. E. Lischer (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Faircloth, B. C., and T. C. Glenn (2012). Not All Sequence Tags Are Created Equal: Designing and Validating Sequence Identification Tags Robust to Indels. PLoS ONE 7:e42543. doi: 10.1371/journal.pone.0042543

Falconer, D. S., and T. F. C. Mackay (1996). Introduction to quantitative genetics. Longman, New York.

Federico, C., C. Cantarella, C. Scavo, S. Saccone, B. Bed'Hom, and G. Bernardi (2005). Avian genomes: different karyotypes but a similar distribution of the GC-richest chromosome regions at interphase. Chromosome Research 13:785–793. doi: 10.1007/s10577-005-1012-7

Fisher, R. A. (1958). The genetical theory of natural selection, 2nd edn. Dover, New York.

Fleischer, R. C., S. Conant, and M. P. Morin (1991). Genetic variation in native and translocated populations of the Laysan finch (Telespiza cantans). Heredity 66:125–130.

Fleischer, R. C., C. E. McIntosh, and C. L. Tarr (1998). Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K–Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. Molecular Ecology 7:533–545. doi: 10.1046/j.1365-294x.1998.00364.x

Fleischer, R. C., S. L. Olson, H. F. James, and A. C. Cooper (2000). Identification of the Extinct Hawaiian Eagle (Haliaeetus) by mtDNA Sequence Analysis. The Auk 117:1051. doi: 10.1642/0004-8038(2000)117[1051:IOTEHE]2.0.CO;2

Fleischer, R. C., B. Silkas, J. Beadell, C. Atkins, C. E. McIntosh, and S. Conant (2007). Genetic variability and taxonomic status of the Nihoa and Laysan millerbirds. The Condor 109:954–962. doi: 10.1650/0010-5422(2007)109[954:GVATSO]2.0.CO;2

Foll, M., and O. Gaggiotti (2008). A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180:977–993. doi: 10.1534/genetics.108.092221

Ford, M. J. (2002). Selection in Captivity during Supportive Breeding May Reduce Fitness in the Wild. Conservation Biology 16:815–825. doi: 10.2307/3061228

Foster, J. T., B. L. Woodworth, L. E. Eggert, P. J. Hart, D. Palmer, D. C. Duffy, and R. C. Fleischer (2007). Genetic structure and evolved malaria resistance in Hawaiian honeycreepers. Molecular Ecology 16:4738–4746. doi: 10.1111/j.1365-294X.2007.03550.x

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

Frankham, R. (1998). Inbreeding and Extinction: Island Populations. Conservation Biology 12:665–675. doi: 10.2307/2387248

Frankham, R. (1999). Quantitative genetics in conservation biology. Genetical Research 74:237–244.

Frankham, R. (2005). Genetics and extinction. Biological Conservation 126:131–140.

Frankham, R., J. D. Ballou, and D. A. Briscoe (2002). Introduction to Conservation Genetics. Cambridge University Press, Cambridge.

Franklin, A. D., and F. Siewerdt (2011). Post-bottleneck inbreeding accumulation reduces fitness and adaptive potential in populations of Tribolium castaneum under environmental stress. Genomics and Quantitative Genetics 2:19–30.

Franklin, I. R. (1980). Evolutionary change in small populations. In Conservation Biology: an Evolutionary-ecological Perspective (M. E. Soule and B. A. Wilcox, Editors). Sinauer, Sunderland, pp. 135–150.

Fumagalli, M., M. Sironi, U. Pozzoli, A. Ferrer-Admettla, L. Pattini, and R. Nielsen (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLoS Genetics 7:e1002355. doi: 10.1371/journal.pgen.1002355

Ganapathy, G., J. T. Howard, J. M. Ward, J. Li, B. Li, Y. Li, Y. Xiong, Y. Zhang, S. Zhou, D. C. Schwartz, M. Schatz, et al. (2014). High-coverage sequencing and annotated assemblies of the budgerigar genome. GigaScience 3:11. doi: 10.1186/2047-217X-3-11

Gasper, J. S., T. Shiina, H. Inoko, and S. V. Edwards (2001). Songbird genomics: analysis of 45 kb upstream of a polymorphic Mhc class II gene in red-winged blackbirds (Agelaius phoeniceus). Genomics 75:26–34. doi: 10.1006/geno.2001.6596

Genome Sequencing Center, W. U. S. of M. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432. doi: 10.1038/nature03154

Goodnight, C. J. (1987). On the effect of founder events on epistatic genetic variance. Evolution 41:80–91.

Goodnight, C. J. (1988). Epistasis and the effect of founder events on the additive genetic variance. Evolution 42:441–454.

Grant, P. R. (1994). Population variation and hybridization: Comparison of finches from two archipelagos. Evolutionary Ecology 8:598–617. doi: 10.1007/bf01237844

Grant, P. R., and B. R. Grant (2002). Unpredictable Evolution in a 30-Year Study of Darwin's Finches. Science 296:707–711.

Griffin, D. K., L. B. W. Robertson, H. G. Tempest, and B. M. Skinner (2007). The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenetic and Genome Research 117:64–77.

Groombridge, J. J., D. A. Dawson, T. Burke, R. Prys-Jones, M. de L. Brooke, and N. Shah (2009). Evaluating the demographic history of the Seychelles kestrel (Falco araea): Genetic evidence for recovery from a population bottleneck following minimal conservation management. Biological Conservation 142:2250–2257.

Groombridge, J. J., C. G. Jones, M. W. Bruford, and R. A. Nichols (2000). Conservation biology: `Ghost' alleles of the Mauritius kestrel. Nature 403:616–616.

Hackett, S. J., R. T. Kimball, S. Reddy, R. C. K. Bowie, E. L. Braun, M. J. Braun, J. L. Chojnowski, W. A. Cox, K.-L. Han, J. Harshman, C. J. Huddleston, et al. (2008).

A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320:1763 –1768. doi: 10.1126/science.1157704

Hartl, G. B., and P. Hell (1994). Maintenance of high levels of allelic variation in spite of a severe bottleneck in population size: the brown bear (Ursus arctos) in the Western Carpathians. Biodiversity and Conservation 3:546–554.

Hata, S., K. Nishi, T. Kawamoto, H. J. Lee, H. Kawahara, T. Maeda, Y. Shintani, H. Sorimachi, and K. Suzuki (2001). Both the conserved and the unique gene structure of stomach-specific calpains reveal processes of calpain gene evolution. Journal of Molecular Evolution 53:191–203. doi: 10.1007/s002390010209

Hawkins, M. T. R., C. A. Hofman, T. Callicrate, M. M. McDonough, M. T. N. Tsuchiya, E. E. Gutiérrez, K. M. Helgen, and J. E. Maldonado (2015). In-Solution Hybridization for Mammalian Mitogenome enrichment: Pros, Cons, and Challenges Associated with Multiplexing Degraded DNA. Molecular Ecology Resources:n/a–n/a. doi: 10.1111/1755-0998.12448

Hedrick, P. W. (1999). Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. Evolution 53:313–318. doi: 10.2307/2640768

Hedrick, P. W. (2001). Conservation genetics: where are we now? Trends in Ecology & Evolution 16:629–636.

Hedrick, P. W. (2004). Recent developments in conservation genetics. Forest Ecology and Management 197:3–19.

Hedrick, P. W. (2005). Genetics of Populations. Third Edition. Jones and Bartlett Publishers, Sudbury, MA.

Hedrick, P. W., and S. T. Kalinowski (2000). Inbreeding depression in conservation biology. Annual Review of Ecology Evolution and Systematics 31:139–162.

Hess, C. M., and S. V. Edwards (2002). The Evolution of the Major Histocompatibility Complex in Birds: Scaling up and taking a genomic approach to the major histo compatibilty complex (MHC) of birds reveals surprising departures from generalities found in mammals in both large-scale structure and the mechanisms shaping the evolution of the MHC. BioScience 52:423–431. doi: 10.1641/0006-3568(2002)052[0423:TEOTMH]2.0.CO;2

Hess, J. E., N. R. Campbell, D. A. Close, M. F. Docker, and S. R. Narum (2013). Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. Molecular Ecology 22:2898–2916. doi: 10.1111/mec.12150

Hill, W. G. (1972). Estimation of genetic change. I. General theory and design of control populations. Animal Breeding Abstracts 40:1–15.

Hofreiter, M. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research 29:4793–4799. doi: 10.1093/nar/29.23.4793

James, H., and S. Olson (1991). Descriptions of thirty-two new species of birds from the Hawaiian Islands: Part II. Passeriformes. Ornithological Monographs 46:1–88.

Jarvi, S., C. Atkinson, and R. Fleischer (2001). Immunogenetics and resistance to avian malaria in Hawaiian honeycreepers (Drepanidinae). Studies in Avian Biology 22:254–263.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers (2012). The global diversity of birds in space and time. Nature 491:444–448. doi: 10.1038/nature11631

Johnson, N. K., J. A. Marten, and C. J. Ralph (1989). Genetic Evidence for the Origin and Relationships of Hawaiian Honeycreepers (Aves: Fringillidae). The Condor 91:379–396. doi: 10.2307/1368317

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

Jones, C. G., and D. V. Merton (2012). A tale of two islands: the rescue and recovery of endemic birds in New Zealand and Mauritius. In Reintroduction biology: integrating science and management. Wiley-Blackwell, Oxford, pp. 33–72.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61. doi: 10.1038/nature10944

Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, P. Olason, and H. Ellegren (2014). A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology 23:4035–4058. doi: 10.1111/mec.12810

Keller, L. F., K. J. Jeffery, P. Arcese, M. A. Beaumont, W. M. Hochachka, J. N. M. Smith, and M. W. Bruford (2001). Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. Proceedings of the Royal Society of London. Series B: Biological Sciences 268:1387–1394. doi: 10.1098/rspb.2001.1607

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

Kimura, M., and T. Ohta (1969). The Average Number of Generations until Extinction of an Individual Mutant Gene in a Finite Population. Genetics 63:701–709.

Kircher, M. (2012). Analysis of High-Throughput Ancient DNA Sequencing Data. In Ancient DNA (B. Shapiro and M. Hofreiter, Editors). Humana Press, Totowa, NJ, pp. 197–228.

Kirkpatrick, M. (1982). Quantum Evolution and Punctuated Equilibria in Continuous Genetic Characters. The American Naturalist 119:833–848.

Kleinjan, D. A., and V. van Heyningen (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. American Journal of Human Genetics 76:8–32. doi: 10.1086/426833

Lachance, J., and S. A. Tishkoff (2013). SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. BioEssays 35:780–786. doi: 10.1002/bies.201300014

Lacy, R. C., G. Alaks, and A. Walsh (2013). Evolution of Peromyscus leucopus Mice in Response to a Captive Environment. PLoS ONE 8:1–1.

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

Lamichhaney, S., J. Berglund, M. S. Almen, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerova, C.-J. Rubin, C. Wang, N. Zamani, B. R. Grant, et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution 30:314–334.

Lande, R. (1980). Genetic Variation and Phenotypic Evolution During Allopatric Speciation. The American Naturalist 116:463–479.

Lande, R. (1988). Genetics and Demography in Biological Conservation. Science 241:1455–1460.

Leberg, P. L. (1992). Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. Evolution 46:477–494.

de León, L. F., E. Bermingham, J. Podos, and A. P. Hendry (2010). Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365:1041–1052. doi: 10.1098/rstb.2009.0314

Lerner, H., and R. Fleischer (2010). Prospects for the Use of Next-Generation Sequencing Methods in Ornithology. The Auk 127:4–15. doi: doi:10.1525/auk.2010.127.1.4

Lerner, H., M. Meyer, M. Hofreiter, and R. Fleischer (2011). Multilocus resolution of the phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology 21:1838–1844.

Lewontin, R. C., and J. Krakauer (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74:175–195.

Li, H., and R. Durbin (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25:2078–2079. doi: 10.1093/bioinformatics/btp352

Lischer, H. E. L., and L. Excoffier (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. doi: 10.1093/bioinformatics/btr642

Lopez-Fanjul, C., and A. Villaverde (1989). Inbreeding Increases Genetic Variance for Viability in Drosophila melanogaster. Evolution 43:1800–1804.

Luikart, G., P. R. England, D. Tallmon, S. Jordan, and P. Taberlet (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics 4:981–994.

Lynch, M. (1988). Design and Analysis of Experiments on Random Drift and Inbreeding Depression. Genetics 120:791–807.

Maruyama, T., and P. A. Fuerst (1984). Population bottlenecks and nonequilibrium models in population genetics. I. Allele numbers when populations evolve from zero variability. Genetics 108:745–763.

Maruyama, T., and P. A. Fuerst (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111:675–689.

Mayr, E. (1966). Animal Species and Evolution. The Belknap Press of Harvard University Press, Cambridge, Massachusetts.

McClung, A. (2005). A population viability analysis of the Laysan finch (Telespiza cantans).

McKenna, A., E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. DePristo (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20:1297–303.

McQueen, H. A., J. Fantes, S. H. Cross, V. H. Clark, A. L. Archibald, and A. P. Bird (1996). CpG islands of chicken are concentrated on microchromosomes. Nat Genet 12:321–324. doi: 10.1038/ng0396-321

Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat Rev Genet 8:845–856.

Mitton, J. B. (1997). Selection in Natural Populations. Oxford University Press, Oxford.

Montgomery, M. E., L. M. Woodworth, P. R. England, D. A. Briscoe, and R. Frankham (2010). Widespread selective sweeps affecting microsatellites in Drosophila populations adapting to captivity: Implications for captive breeding programs. Biological Conservation 143:1842–1849. doi: 10.1016/j.biocon.2010.01.022

Morgulis, A., G. Coulouris, Y. Raytselis, T. L. Madden, R. Agarwala, and A. A. Schäffer (2008). Database indexing for production MegaBLAST searches. Bioinformatics (Oxford, England) 24:1757–1764. doi: 10.1093/bioinformatics/btn322

Morin, M. P., and S. Conant (1994). Variables influencing population estimates of an endangered passerine. Biological Conservation 67:73–84.

Morin, P., G. Luikart, R. Wayne, and the S. workshop group (2004). SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution 19:208–216.

Mounce, H. L., C. Raisin, D. L. Leonard, H. Wickenden, K. J. Swinnerton, and J. J. Groombridge (2014). Spatial genetic architecture of the critically-endangered Maui Parrotbill (Pseudonestor xanthophrys): management considerations for reintroduction strategies. Conservation Genetics. doi: 10.1007/s10592-014-0641-9

Mullikin, J. C., and Z. Ning (2003). The Phusion Assembler. Genome Research 13:81–90. doi: 10.1101/gr.731003

Mundy, N. I., P. Unitt, and D. S. Woodruff (1997). Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. Auk 114:126–129.

Munter, W. H. (1915). Report of destruction of bird life on Laysan Island.

Narum, S. R., and J. E. Hess (2011). Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources 11:184–194. doi: 10.1111/j.1755-0998.2011.02987.x

Nei, M., T. Maruyama, and R. Chakraborty (1975). The bottleneck effect and genetic variability in populations. Evolution 29:1–10.

Nichols, R. A., M. W. Bruford, and J. J. Groombridge (2001). Sustaining genetic variation in a small population: evidence from the Mauritius kestrel. Molecular Ecology 10:593–602. doi: 10.1046/j.1365-294x.2001.01204.x

Niimura, Y., and M. Nei (2003). Evolution of olfactory receptor genes in the human genome. Proceedings of the National Academy of Sciences 100:12235–12240. doi: 10.1073/pnas.1635157100

Nikolajewa, S. (2005). Common patterns in type II restriction enzyme binding sites. Nucleic Acids Research 33:2726–2733. doi: 10.1093/nar/gki575

Nosil, P., D. J. Funk, and D. Oritz-Barrientos (2009). Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375–402.

Oettgen, P., Y. Akbarali, J. Boltax, J. Best, C. Kunsch, and T. A. Libermann (1996). Characterization of NERF, a novel transcription factor related to the Ets factor ELF-1. Molecular and Cellular Biology 16:5091–5106.

Oleksyk, T. K., M. W. Smith, and S. J. O'Brien (2010). Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B: Biological Sciences 365:185–205. doi: 10.1098/rstb.2009.0219

Olson, S. (1996). History and ornithological journals of the Tanager expedition of 1923 to the Northwestern Hawaiian Islands, Johnston and Wake Island.

Orr, N., W. Back, J. Gu, P. Leegwater, P. Govindarajan, J. Conroy, B. Ducro, J. A. M. Van Arendonk, D. E. MacHugh, S. Ennis, E. W. Hill, and P. A. J. Brama (2010). Genome-wide SNP association–based localization of a dwarfism gene in Friesian dwarf horses. Animal Genetics 41:2–7. doi: 10.1111/j.1365-2052.2010.02091.x

Paten, B., J. Herrero, K. Beal, S. Fitzgerald, and E. Birney (2008). Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. Genome Research 18:1814–1828.

Patterson, N., A. L. Price, and D. Reich (2006). Population structure and eigenanalysis. PLoS Genetics 2:e190.

Paxinos, E. E., H. F. James, S. L. Olson, J. D. Ballou, J. A. Leonard, and R. C. Fleischer (2002). Prehistoric Decline of Genetic Diversity in the Nene. Science 296:1827. doi: 10.1126/science.296.5574.1827

Pimm, S. L., H. L. Jones, and J. Diamond (1988). On the Risk of Extinction. The American Naturalist 132:757–785. doi: 10.2307/2462261

Pimm, S. L., and A. Redfearn (1988). The variability of population densities. Nature 334:613–614. doi: 10.1038/334613a0

Pratt, T. K. (2009). Origins and evolution. In Conservation Biology of Hawaiian Forest Birds. Yale University Press, New Haven, CT, pp. 3–24.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. doi: 10.1038/ng1847

Price, J. P., and D. A. Clague (2002). How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. Proceedings of the Royal Society of London. Series B: Biological Sciences 269:2429–2435. doi: 10.1098/rspb.2002.2175

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81:559–575.

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

Raikow, R. (1977). The origin and evolution of the Hawaiian honeycreepers (Drepanididae). Living Bird 15:95–117.

Raj, A., M. Stephens, and J. K. Pritchard (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics 197:573–589. doi: 10.1534/genetics.114.164350

Rands, C. M., A. Darling, M. Fujita, L. Kong, M. T. Webster, C. Clabaut, R. D. Emes, A. Heger, S. Meader, M. B. Hawkins, M. B. Eisen, et al. (2013). Insights into the evolution of Darwin's finches from comparative analysis of the Geospiza magnirostris genome sequence. BMC Genomics 14:95. doi: 10.1186/1471-2164-14-95

Reding, D., L. Freed, R. Cann, and R. Fleischer (2010). Spatial and temporal patterns of genetic diversity in an endangered Hawaiian honeycreeper, the Hawaii Akepa (Loxops coccineus coccineus). Conservation Genetics 11:225–240.

Reding, D. M., J. T. Foster, H. F. James, H. D. Pratt, and R. C. Fleischer (2009). Convergent evolution of "creepers" in the Hawaiian honeycreeper radiation. Biology Letters 5:221–224. doi: 10.1098/rsbl.2008.0589

Reed, D. H., E. H. Lowe, D. A. Briscoe, and R. Frankham (2003). Fitness and Adaptation in a Novel Environment: Effect of Inbreeding, Prior Environment, and Lineage. Evolution 57:1822–1828.

Richards, L. P., and W. J. Bock (1973). Functional Anatomy and Adaptive Evolution of the Feeding Apparatus in the Hawaiian Honeycreeper Genus Loxops (Drepanididae). Ornithological Monographs:1–173. doi: 10.2307/40166695

van Riper, C. I., S. G. van Riper, M. L. Goff, and M. Laird (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. Ecological Monographs 56:327–344.

van Riper, C. I., and J. Scott (2001). Limiting factors affecting Hawaiian native birds. Studies in Avian Biology 22:221–233.

Rubin, C.-J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallbook, et al. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464:587–591. doi: 10.1038/nature08832

Saccheri, I. J., R. A. Nichols, and P. M. Brakefield (2001). Effects of bottlenecks on quantitative genetic variation in the butterfly Bicyclus anynana. Genetics Research 77:167–181.

Savidge, J. A. (1987). Extinction of an Island Forest Avifauna by an Introduced Snake. Ecology 68:660–668. doi: 10.2307/1938471

Seddon, N., W. Amos, R. A. Mulder, and J. A. Tobias (2004). Male heterozygosity predicts territory size, song structure and reproductive success in a cooperatively breeding bird. Proceedings of the Royal Society B: Biological Sciences 271:1823–1829.

Shafique, M., M. Ahmad, and M. A. Chaudry (2006). Feeding preference and development of Tribolium castaneum (Herbst.) in wheat products. Pakistan Journal of Zoology 38:27–31.

Shepherd, L. D., and D. M. Lambert (2008). Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. Mol. Ecol. 17:2174–2184. doi: 10.1111/j.1365-294X.2008.03749.x

Sincock, J., and E. Kridler (1977). The extinct and endangered endemic birds of the Northwestern Hawaiian Islands.

Slikas, B., I. B. Jones, S. R. Derrickson, and R. C. Fleischer (2000). Phylogenetic relationships of Micronesian white-eyes based on mitochondrial sequence data. The Auk 117:355–365. doi: 10.1642/0004-8038(2000)117[0355:PROMWE]2.0.CO;2

Smith, J., C. K. Bruley, I. R. Paton, I. Dunn, C. T. Jones, D. Windsor, D. R. Morrice, A. S. Law, J. Masabanda, A. Sazanov, D. Waddington, et al. (2000). Differences in gene density on chicken macrochromosomes and microchromosomes. Animal Genetics 31:96–103. doi: 10.1046/j.1365-2052.2000.00565.x

Smith, T. B. (1993). Disruptive selection and the genetic basis of bill size polymorphism in the African finch Pyrenestes. Nature 363:618–620.

Soulé, M., M. Gilpin, W. Conway, and T. Foose (1986). The millenium ark: How long a voyage, how many staterooms, how many passengers? Zoo biology 5:101–113.

Spencer, C. C., J. E. Neigel, and P. L. Leberg (2000). Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. Molecular Ecology 9:1517–1528.

Spielman, D., B. Brook, D. Briscoe, and R. Frankham (2004). Does Inbreeding and Loss of Genetic Diversity Decrease Disease Resistance? Conservation Genetics 5:439–448. doi: 10.1023/B:COGE.0000041030.76598.cd

Sprague, J. C. (2004). V. encelioides monitoring and associated research at Pearl & Hermes Reef, Northwestern Hawaiian Islands (May to August, 2003).

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

Stapley, J., T. R. Birkhead, T. Burke, and J. Slate (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. Genome Research 20:496–502. doi: 10.1101/gr.102095.109

Steadman, D. W. (2006). Extinction and biogeography of tropical pacific birds. University of Chicago, Chicago.

Stockwell, C. A., A. P. Hendry, and M. T. Kinnison (2003). Contemporary evolution meets conservation biology. Trends in Ecology & Evolution 18:94–101.

Swindell, W. R., and J. Bouzat (2005). Modeling the adaptive potential of isolated populations: experimental simulations using Drosophila. Evolution 59:2159–2169.

Tarr, C. L., J. D. Ballou, M. P. Morin, and S. Conant (2000). Microsatellite variation in simulated and natural founder populations of the Laysan finch (Telespiza cantans). Conservation Genetics 1:135–146.

Tarr, C. L., S. Conant, and R. C. Fleischer (1998). Founder events and variation at microsatellite loci in an insular passerine bird, the Laysan finch (Telespiza cantans). Molecular Ecology 7:719–731.

Tarr, C. L., and R. C. Fleischer (1993). Mitochondrial-DNA Variation and Evolutionary Relationships in the Amakihi Complex. The Auk 110:825–831. doi: 10.2307/4088636

Teer, J. K., L. L. Bonnycastle, P. S. Chines, N. F. Hansen, N. Aoyama, A. J. Swift, H. O. Abaan, T. J. Albert, E. H. Margulies, E. D. Green, F. S. Collins, et al. (2010). Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome research 20:1420–1431. doi: 10.1101/gr.106716.110

Templeton, A. R., and B. Read (1984). Factors eliminating inbreeding depression in a captive herd of Specke's gazelle. Zoo Biology 3:177–199.

Thomas, C. D., A. Cameron, R. E. Green, M. Bakkenes, L. J. Beaumont, Y. C. Collingham, B. F. N. Erasmus, M. F. de Siqueira, A. Grainger, L. Hannah, L. Hughes, et al. (2004). Extinction risk from climate change. Nature 427:145–148. doi: 10.1038/nature02121

USFWS, P. I. F. and W. O. (2008). Laysan finch (honeycreeper) (Telespiza cantans) 5-Year Review: Summary and Evaluation.

Vucetich, J. A., T. A. Waite, and L. Nunney (1997). Fluctuating Population Size and the Ratio of Effective to Census Population Size. Evolution 51:2017–2021. doi: 10.2307/2411022

Wade, M. J., S. M. Shuster, and L. Stevens (1996). Inbreeding: Its Effect on Response to Selection for Pupal Weight and the Heritable Variance in Fitness in the Flour Beetle, Tribolium castaneum. Evolution 50:723–733.

Wallace, A. R. (1880). Island life: or, the phenomena and causes of insular faunas and floras, including a revision and attempted solution of the problem of geological climates. Macmillan and Co., London.

Warner, R. E. (1959). Completion report- Midway and Laysan Islands bird studies.

Warner, R. E. (1968). The Role of Introduced Diseases in the Extinction of the Endemic Hawaiian Avifauna. The Condor 70:101–120. doi: 10.2307/1365954

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Kunstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, et al. (2010). The genome of a songbird. Nature 464:757–762.

Weir, B. S., and C. C. Cockerham (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358–1370.

Wetmore, A. (1925). Bird life among lava rock and coral sand. National Geographic Magazine 48:77–108.

White, N. D. G., C. J. Demianyk, and P. G. Fields (2000). Effects of red versus white wheat bran on rate of growth and feeding of some stored-product beetles. Canadian Journal of Plant Science 80:661–663.

Whitlock, M. C. (1995). Variance-Induced Peak Shifts. Evolution 49:252–259.

Wigginton, J. E., D. J. Cutler, and G. R. Abecasis (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. American Journal of Human Genetics 76:887–893.

Willis, J. H., and H. A. Orr (1993). Increased Heritable Variation Following Population Bottlenecks: The Role of Dominance. Evolution 47:949–957.

Willi, Y., J. Van Buskirk, and A. A. Hoffmann (2006). Limits to the Adaptive Potential of Small Populations. Annual Review of Ecology, Evolution, and Systematics 37:433–458. doi: doi:10.1146/annurev.ecolsys.37.091305.110145

Wisely, S. M., S. W. Buskirk, M. A. Fleming, D. B. McDonald, and E. A. Ostrander (2002). Genetic Diversity and Fitness in Black-Footed Ferrets Before and During a Bottleneck. J Hered 93:231–237. doi: 10.1093/jhered/93.4.231

Woodside, D. H., and R. J. Kramer (1961). A report on a survey trip to the Hawaiian Islands National Wildlife Refuge, March 1961.

Woodworth, B. L., C. T. Atkinson, D. A. LaPointe, P. J. Hart, C. S. Spiegel, E. J. Tweed, C. Henneman, J. LeBrun, T. Denette, R. DeMots, K. L. Kozar, et al. (2005). Host population persistence in the face of introduced vector-borne diseases: Hawaii amakihi and avian malaria. Proceedings of the National Academy of Sciences of the United States of America 102:1531–1536. doi: 10.1073/pnas.0409454102

Woodworth, L., M. Montgomery, D. Briscoe, and R. Frankham (2002). Rapid genetic deterioration in captive populations: Causes and conservation implications. Conservation Genetics 3:277–288. doi: 10.1023/A:1019954801089

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

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the 6th International Congress of Genetics 1:356–366.

Wright, S. (1951). The genetical structure of populations. Annals of Eugenics 15:323–354.

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller (2000). A greedy algorithm for aligning DNA sequences. Journal of computational biology : a journal of computational molecular cell biology 7:203–214. doi: 10.1089/10665270050081478

Zwickl, D. (2006). Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.