The Genomic Basis of Population and Adaptive Divergence in *Buteo* Sister Species Across Multiple Evolutionary and Geographic Scales

By

EMILY ABERNATHY

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Ecology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Josh Hull, Chair

_____

Rachael Bay

_____

Andrea Schreier

Committee in Charge

2021

ACKNOWLEDGEMENTS

There are many people and organizations whose support and assistance made this dissertation possible. I would first like to thank the Swainson's Hawk Technical Advisory Committee and the UC Davis Graduate Group in Ecology for funding this research. Next, I would like to thank my dissertation committee, including my advisor, Josh Hull, for providing much needed feedback on all of my writing and helping me stick to my goals. To Andrea Schreier, thank you for always taking the time to check in on me when I most needed it and for making sure I always had financial support. To Rachael Bay, thank you for always being the second opinion I needed to make my difficult bioinformatics decisions and for letting me talk out all of my genomics problems. There is absolutely no way this dissertation would have come together without your computational and mental support. Thank you to my collaborator Kristen Ruegg for providing help in my project design, and Patty Parker and Holly Ernest for allowing me access to samples that enabled this whole dissertation to come together. I would like to thank Christian Gruppi, Teia Schweizer, Alisha Goodbla, and Grace Auringer for helping me with my library preparation and for answering all of my questions as I fumbled my way through my lab work.

Next, I would also like to thank Allen Fish and all of the volunteers at the Golden Gate Raptor Observatory for helping me first discover my love for raptors and allowing me to continue to be a part of your organization for the last six years. Also, thank you for providing financial support and samples for the rodenticide project which provided me with my first opportunity to become a published scientist. To Chris Briggs, I can't thank you enough for taking me under your wing, making me into a confident raptor bander, and for inspiring me to go to grad school and get my PhD. Next, I need to thank the Hull Enchilada for being the most amazing lab and for supporting my genomics work even though you had no idea what I was talking about most of the time.

And lastly, I need to thank all of my family and friends who have continuously supported me throughout this experience. To the Balboa Haus, thank you for providing me a safe space to grow into the researcher I am today. To Eric, thank you for being the worst grad student so I didn't have to be, to Sarah, thank you for always agreeing to proofread my writing, to Amy, thank you for always being down for a dance party study break, and to Michael, thank you for being Lupine's emergency contact, allowing me to raise a puppy and still make progress on my research. To my family, thank you for being proud of me even in times when I haven't been and for allowing me a safe place to land and write in a global pandemic. And last but not least, to Lupine, thank you for providing me with a fluffy distraction whenever I needed to come up for air, and for giving me a sense of purpose when things got tough.

ABSTRACT

The genetic basis of adaptation and divergence has been at the forefront of evolutionary and ecological studies for many years. As environments change, it becomes increasingly important to understand the genetic basis and drivers of population and adaptive divergence to preserve local adaptation relevant to conservation, and to provide insight into the maintenance of species diversity. In this dissertation, I use whole-genome sequencing to answer questions related to adaptation on multiple scales from local adaptation to speciation in two closely related species of raptors. First, I characterized population divergence and local adaptation in a highly migratory species, the Swainson's hawk (*Buteo swainsoni*), and find a clear distinction between the populations to the east and west of the continental divide, a pattern not yet documented in the Swainson's hawk. I also identified patterns of genomic divergence between the slightly distinct Central Valley population and the rest of the populations in hopes of using this information to identify the genetic basis of multiple phenotypic traits. Next, I investigated the adaptive radiation of the Galapagos hawk (*Buteo swainsoni*), to assess patterns of diversity and diversification across islands as well as to investigate the genetic basis of both morphological and behavioral traits. The size distinct populations of the Galapagos hawk are significantly genetically divergent and have very low levels of genetic diversity. I also identified candidate loci and genes involved in the mating system of the Galapagos hawk as well as six morphometric traits and found that morphology differed significantly across islands independent of body size. Lastly, I assembled and annotated a high-quality draft genome for the Swainson's hawk to be able to look at fine scale patterns of divergence between the Swainson's hawk and the Galapagos hawk to better understand the genomic basis of allopatric speciation. The genome-wide patterns in both relative and absolute divergence support our hypothesis of founder speciation without secondary contact and represent an important step towards an understanding of the complex interactions between evolutionary processes that shape allopatric speciation at the genomic level. Overall, this work provides important insight into the evolutionary processes that shape genomic divergence in natural populations while also providing information imperative to the conservation of two, threatened species.

TABLE OF CONTENTS

CHAPTER 1:

Using whole-genome data to characterize population and adaptive divergence in a highly-migratory raptor species.

ABSTRACT

Understanding the genetic basis of local adaptation is critical to the study of adaptive evolution and ecological conservation. Despite this importance, information is lacking on genes underlying local adaptation in non-model organisms. The Swainson's hawk is a widespread, diurnal raptor that is considered threatened in parts of its range. While almost entirely panmictic across Western North America, one population of Swainson's hawks in the Central Valley of California shows distinct phenotypic differences from other populations, and previous studies have found this population to be slightly genetically distinct. Here, we use whole-genome sequencing to investigate fine-scale patterns of neutral and adaptive differentiation in the Swainson's hawk (*Buteo swainsoni*), while also identifying genomic areas of divergence unique to the Central Valley population. We found clear evidence of differentiation between populations of Swainson's hawks to the East and West of the Rocky Mountains, coinciding with the continental divide. This indicates that mountain ranges may be a barrier to gene flow in the Swainson's hawk, a finding consistent with previous studies on raptors. We also found the Central Valley population to be the most differentiated from the other populations, but the Owen's Valley population was also slightly distinct. Lastly, our analysis provided a number of candidate genes putatively under selection in the Swainson's hawk, including genes with known functions in feather formation, coloration, and migration. Our results provide important information for delineation of conservation units and conservation priorities in the Swainson's

hawk while also showing the ability of whole-genome sequencing to uncover fine-scale population structure not seen in microsatellite data.

INTRODUCTION

Local adaptation, selection resulting in a higher relative fitness in one's local habitat compared to other habitats (Williams 1966), is extremely important to species persistence and the maintenance of biodiversity. Local adaptation among populations has been shown to play a critical role in maintaining genetic variation (Felsenstein 1976, Hedrick et al. 1976), and in initiating the divergence of new species (Turelli et al. 2001). Adaptive variation among populations is a strong determinant of species long-term viability, extinction probability, and their ability to increase in population size (Hohenlohe et al. 2020). Patterns of local adaptation can also be used to assess the adaptive potential of a species, information that is particularly important in light of recent and future climate change (Ruegg et al. 2018).

The discovery of genetic regions involved in local adaptation is important to conservation as it can help predict the performance of a genotype in a new environment, which can inform management decisions for threatened and endangered wildlife (Funk et al. 2012). Because local adaptation and specialization can often end in speciation (Nosil 2012), identifying the genes underlying local adaptation may help identify the origins of species diversity and delineate the process of speciation (Nosil and Schluter 2011). Despite its importance, the genetic basis of local adaptation remains poorly understood, especially in non-model organisms where genomic resources are lacking. Advancements in genomic technology have allowed the identification of candidate regions under selection to become a critical part of modern studies in conservation genetics and adaptation.

To generate genome-wide data sets, many genomic studies of non-model organisms currently use reduced-representation methods such as RAD-seq, which sequences a small portion of the genome at a sufficient depth to make reliable inferences about individual genotypes. While very cost-effective and fast, by only sequencing part of the genome, these methods may miss key areas that are under selection (Lowry et al. 2017). Recent advances in technology combined with decreases in sequencing costs have now made whole genome sequencing a popular and viable approach in non-model organisms. Unlike reduced-representation methods, sequencing the entire genome, even at low coverage, can result in the detection of fine-scale signals of selection that otherwise would have been missed (Pespeni et al. 2012). Often, the traits undergoing selection are not known, so loci can be identified that show patterns of local adaptation through differentiation outlier methods which do not require any phenotypic information. These methods search for alleles that occur at a high frequency, as loci under selection are expected to be at higher frequencies in populations where they increase fitness, and at lower frequencies in populations where they decrease fitness. Therefore, loci involved in local adaptation will often show greater than average genetic differentiation among populations (Lewontin and Krakauer 1973, Beaumont 2005).

Here, we aim to study the genome-wide patterns of differentiation and selection in the widespread, migratory Swainson's hawk (*Buteo swainsoni*). The Swainson's hawk is a diurnal raptor species found throughout western North America, with a contemporary breeding range that covers much of the Great Basin, Great Plains, and southwestern deserts, with a geographically separate population located in the Central Valley of California (England et al. 1997). Swainson's hawks are highly migratory, with the vast majority of individuals overwintering in Argentina every year (Fuller et al. 1998). Once considered one of the most

abundant raptors in western North America, threats such as agricultural conversion and expansion of wind energy have caused considerable declines in Swainson's hawk populations over the last century. In the 1990's mass mortalities due to insecticide use occurred in Argentina, which is thought to have decreased the worldwide Swainson's hawk population by as much as 5% (Goldstein et al. 1999). In California, extensive loss of riparian habitat (up to 85% in some areas) (Katibah 1984) may be a main contributing factor in population declines across the state (Risenbrough 1989). Due to these declines the Swainson's hawk is listed as a species of concern in many states and has been listed as Threatened in the state of California since 1983.

Because of its conservation status, the Swainson's hawk has been closely monitored and the focus of many research efforts in California. A previous study using microsatellites found limited differentiation among populations but evidence of slight genetic distinctness between the Central Valley population and other breeding populations across the range (Hull et al. 2008a). Despite this limited evidence of genetic differentiation, the Central Valley population shows distinct differences in certain life-history traits. One of the most striking differences in Central Valley Swainson's hawks is the high frequency of dark-morph individuals compared to other populations. The Swainson's hawk has an almost continuous plumage variation from light to dark, with most individuals being classified in one of three categories – light, medium and dark morph (Palmer 1988). The Central Valley population is comprised of approximately 85-90% of dark morph individuals, compared to less than 40% in populations outside of California, and as little as 1% in some eastern populations (Wheeler 2003). In birds, there is evidence of differential fitness based on coloration (Meunier et al. 2011), although the exact mechanisms behind the maintenance of plumage polymorphism in the Swainson's hawk is still unknown (Briggs et al. 2011). In addition to plumage differentiation, the Central Valley population shows

differences in migratory behavior with a mean natal dispersal of less than 10 km (Estep 1989, Woodbridge et al. 1995) compared to other populations some of which have mean natal dispersal distances of more than 100 km (Houston and Schmutz 1995). Similarly, the Central Valley population has been found to overwinter in western Mexico and Central America while other Swainson's hawk populations migrate to central Argentina (Houston 1990, Wheeler 2003). The persistence of these traits within the Central Valley indicates their potential to improve the fitness of the Swainson's hawks in this population. This divergence in key life history traits may be the result of local adaptation to the Central Valley environment, because in the absence of divergent natural selection, genetic differentiation in adaptive traits is expected to be erased by gene flow. If this differentiation persists in populations connected by gene flow, as is suggested in the Swainson's hawk by earlier microsatellite data, then it is likely due to ongoing natural selection related to differences in environmental conditions in the local habitat (Kawecki and Ebert 2004). These geographically structured patterns in phenotype, combined with limited dispersal, provide support for divergence and local adaptation within the Central Valley population. Identifying genes that are differentiated within the Central Valley population may provide important information about the presence and amount of adaptive variation in this population and may provide candidate genes for the genetic basis of divergent phenotypic traits within the Swainson's hawk.

To describe the extent and mechanisms of neutral and selective differentiation in breeding populations of Swainson's hawks, we took two approaches. First, we used whole-genome data to characterize fine-scale patterns of population differentiation and genetic variation across the range of the Swainson's hawk. Second, we investigated population level divergence, specifically within the Central Valley population, to identify important areas of differentiation

and candidate genes for selection. The results from this study can be used to guide the delineation of conservation units while also providing insight into the genetic basis of divergence in a threatened raptor species.

METHODS

Sample Collection

Whole blood and feathers were collected from both adult and nestling Swainson's hawk between 2003 and 2005. A total of 88 samples were collected from 9 distinct geographic locations across the breeding range (Figure 1.1). Ten samples were included from all populations except the Alberta population where only 8 samples were included. Blood was drawn from the medial metatarsal vein and stored in 1.2 ml of Longmire's lysis buffer (100 mM Tris pH 8.0, 100 mM EDTA, 10 mM NaCl, 0.5% SDS) before being stored at –80° C in a laboratory facility. Feathers were plucked from the breast of the birds and kept cool and dry in paper envelopes. QIAGEN DNeasy kits (QIAGEN Inc.) were used to isolate and extract DNA from 25 µl of blood/buffer solution or feather calamus and the resulting DNA was stored in a -80°C freezer until further use.

Library Preparation and Sequencing

We evaluated DNA quantity using a Qubit Fluorometer (Invitrogen) and quality was evaluated with agarose gel electrophoresis. The ten samples with the highest quantity of DNA and the highest molecular weight with limited smearing were retained from each population for library preparation. The whole genome sequencing protocol was modified from the newest version of an adapted Illumina protocol, first described in Therkildsen

and Palumbi (2017). In short, we made 3 modifications to maximize efficiency of DNA fragmentation and recovery in the desired range for sequencing for use with low input and low quality samples. This method also works with high quality samples at low input. First, we doubled the ratio of tagmentation enzyme to DNA which results in increased fragmentation and therefore shorter average fragment lengths. We also increased the tagmentation incubation time from 5 minutes to 20 minutes to ensure that the tagmentation enzyme had enough time to interact with the DNA. Lastly, we decreased the elongation time during the indexing PCR from 3 minutes to 30 seconds to preferentially increase amplification of shorter fragments within the targeted range for sequencing. Final individual libraries were pooled by equal copies and size selected to retain fragments in the 300-520 basepair range. Final QC was performed with a Qubit quantification and Tapestation for fragment analysis. Multiple paired-end libraries were sequenced at 2-4x coverage on an Illumina HiSeq 4000 at Novogene Corporation Inc. in Sacramento, CA. Studies have demonstrated that sequencing many individuals at a coverage as low as 1 read per locus provides more information about population parameters compared to sampling schemes with lower numbers of individuals and higher coverage (Buerkle and Gompert 2013, Fumagalli 2013).

Dataset preparation

We assessed the quality of the paired-end raw reads with FastQC v. 0.11.9 (Andrews 2010) and summarized with program MultiQC v. 1.8 (Ewels et al. 2016). Next, PCR duplicates were removed with FASTUNIQ v. 1.1 (Xu et al. 2012) and overlapping read pairs were collapsed into single reads with Flash2 (Magoč and Salzberg 2011). Next, reads were aligned to

the Swainson's hawk draft genome assembly (Abernathy et al. in prep) using BWA v. 0.7.16 (Li 2013).

Next, polymorphic sites were identified in ANGSD v. 0.930 (Korneliussen et al. 2014) using the following parameters: -uniqueOnly 1 -skipTriallelic 1 -minMapQ 30 -minQ 30 -doHWE 1 –maxHetFreq 0.5 -minInd n/2. Polymorphic sites were then identified with –pval 1 e 10-6 and –maf 0.05 cutoffs, and genotype likelihoods were calculated for both the whole genome dataset and just for the polymorphic sites using the GATK model.

Characterizing genetic diversity and structure

Population differentiation in Swainson's hawks was calculated using both model and distance-based approaches. First, individual admixture proportions were calculated using the genotype likelihoods in NGSadmix (Skotte et al. 2013). As suggested, ten iterations were performed using "K" ancestral populations from 1-9. These results were visualized in R v. 3.6.3 (R Core Team 2021) using the package POPHELPER v2.3.1 (Francis 2017). Next, the most likely number of ancestral populations was calculated in CLUMPAK (Kopelman et al. 2015), a software created specifically to aid in the interpretation of admixture results, using the Evanno method (Evanno et al. 2005). A principal component analysis (PCA) was conducted using a covariance matrix created in the program PCAngsd v.0.98 (Meisner and Albrechtsen 2018). This program is made specifically for low-coverage data and uses an iterative procedure based on genotype likelihoods to estimate the covariance matrix. These results and all further results were visualized in R v. 3.6.3 (R Core Team 2021) using package ggplot2 v. 3.3.3 (Wickham 2016). To test for a spatial pattern in our data we performed the non-parametric Spearman's rank order correlation between the principal components and the geographic location

of each sample. Next, we estimated pairwise F_{ST} in ANGSD by calculating the two-dimensional site frequency spectrum for each population pair. These joint-spectra were then used as priors to calculate allele frequencies at each site in order to estimate F_{ST}.

Estimates of genetic variation were calculated based on allele frequencies in ANGSD (Korneliussen et al. 2014). Nucleotide diversity and Watterson's theta estimates were calculated using a sliding window size of 50 kb and a step size of 10kb and final estimates were corrected by dividing by the number of sites with data in each window. To test for differences in genetic diversity related to geographic location, we classified five populations as "Western" (Butte Valley, Central Valley, Owen's Valley, Arizona, and Idaho) and four as "Eastern" (Alberta, Saskatchewan, Colorado and Texas) as defined by the continental divide. We next performed Wilcoxon rank sum tests to determine if there was a significant difference in theta and nucleotide diversity estimates between the Eastern and Western populations.

Identifying genes underlying differentiation

We identified candidate genes under selection in the Central Valley population using two approaches. First, we used per-site pairwise F_{ST} values calculated in ANGSD to identify areas of high divergence. We classified candidate sites as those that were within the top 5% of all F_{ST} values across all pairwise comparisons with the Central Valley population. Our second method for identifying genomic areas putatively under selection, was a probabilistic approach measuring differences in allele frequencies, calculated as pFst implemented in vcflib (Garrison 2020). This method uses a likelihood ratio test based on the binomial distribution and outputs p-values based on the chi-squared distribution with one degree of freedom (Garrison 2020). Next, we identified divergent genes as those within 100 kb of our candidate sites for both approaches. We used this

cut-off as recent studies have shown that SNPs may affect distant genes up to 200 kb away (Brodie et al. 2016).

To identify gene functions, we used the gene list analysis function in Panther v. 16.0 (Mi et al. 2020). Our candidate gene lists were annotated against the chicken (*Gallus gallus*) GO Ontology database DOI: 10.5281. We performed the PANTHER Overrepresentation Test for both the GO biological process and GO molecular function annotation datasets on our candidate genes. Specifically, we used the Fisher's Exact test, controlling for False Discovery Rate, and using the chicken genome as a reference list.

RESULTS

Data Generation

A total of 284.3 GB of raw data was produced across all samples. The number of reads per sample ranged from 517,994 to 46,647,118 for an average of 11,801,146 reads per sample. After paralog removal and quality filtering a total of 5,311,268 polymorphic sites dispersed across 4,710 scaffolds were retained for further analysis.

Population Differentiation and Genetic Variation

We found little distinction between all 9 populations, but the principal component analysis showed clear differentiation between the Eastern and Western populations (Figure 1.2). There was some clustering evident in the Central Valley population and the Owen's Valley population, although they were not completely distinct and showed some overlap with the other populations (Figure 1.3). We also found a significant correlation between the second principal component and longitude (Spearman's rank, $p=2.618x\ 10^{-11}$). Additionally, the PCA of the

California populations showed a clear distinction between all three populations (Figure 1.4). The clustering analysis in NgsAdmix showed some distinctness of the Central Valley population consistent with the results of the principal component analysis and previous studies. Although the Central Valley does seem more differentiated than the other populations in this analysis, there is still substantial admixture between the Central Valley and the other populations as individuals in other locations were assigned to the same population as the Central Valley and vice versa. Otherwise, there was no clear pattern of population structure although the most likely K was found to be K=3 (Figure S1.1). The pairwise F_{ST} values ranged from 0.048 to 0.061 with the Central Valley being the most distinct with an average pairwise F_{ST} value of 0.055 (Table 1.1).

With regard to genetic variation, we found that the Owen's valley population had the least amount of genetic variation, and Saskatchewan had the highest for both Watterson's θ and nucleotide diversity (π). Watterson's θ values ranged from 0.00164 to 0.0041 and nucleotide diversity from 0.00151 to 0.00233 (Table 1.2). The Central Valley has a θ value of 0.00233 and a π value of 0.00197, both values near the median for both statistics. Neither Watterson's theta ($p=0.2857$) or nucleotide diversity ($p=0.7302$) was significantly different between eastern and western populations.

Candidate Genes under Selection

Our first approach for identifying regions putatively under selection in the Central Valley population revealed 833 SNPs within the top 5% of all F_{ST} values for all pairwise comparisons with the Central Valley population. These candidate loci were spread across 72 scaffolds and had F_{ST} values ranging from 0.16 to 0.85. We identified 650 genes from our annotated Swainson's

hawk genome that were within 100 kb of our high F_{ST} sites. Of these genes, 463 mapped to the Chicken genome during gene classification. The overrepresentation analysis identified 30 molecular function, and 18 biological process GO groups (Figure 1.5) that were significantly overrepresented in our list of candidate genes from this approach. The full list of candidate loci and their associated genes from this approach can be found in the supplementary materials (Table S1.3). Our pFst analysis resulted in 137 sites with significantly divergent allele frequencies in our Central Valley population compared to the other eight populations. These loci were spread across 51 scaffolds. We identified 138 named genes within 100 kb of these sites and our overrepresentation analysis identified 9 overrepresented GO groups related to molecular function (Table S1.4). A comparison of sites and genes from our two approaches revealed that no sites overlapped between the two sets of candidate sites, but 16 genes were identified as putatively under selection in both methods (Table 1.3).

DISCUSSION

Using whole-genome data we were able to uncover patterns of population differentiation at a much finer scale than previous studies, providing important information for the conservation of a threatened species. While previously thought to be panmictic outside of the Central Valley of California, the differentiation between Swainson's hawk populations to the east and west of the Rocky Mountains shows that this mountain range may act as a barrier to gene flow. This pattern is consistent with findings of genetic structure in other wide-ranging raptor species (Hull et al. 2008b, Machado et al. 2018), but was previously undetected in the Swainson's hawk. Interestingly, studies on the migratory patterns of the Swainson's hawk have tracked individuals

from western breeding grounds crossing the Rocky Mountains during migration (Kochert et al. 2011), suggesting that the Rocky Mountains are not necessarily acting as a physical barrier between populations. High breeding site fidelity has also been well-documented in the Swainson's hawk (Schmutz et al. 2006, Woodbridge 1991) which may be a reasonable explanation for the east-west differentiation, although it is interesting that this site fidelity has not led to more differentiation at the population scale.

Another important finding of this study is confirmation of some genetic distinctness in the Central Valley population as shown in our population differentiation analyses. The Central Valley population showed some clustering away from the other populations in the principal component analysis as well as many individuals in the Central Valley population assigning to a single population in the admixture analysis, although there was overlap with other populations in both analyses. These results combined with the highest average pairwise F_{ST} indicate some genetic distinction between the Central Valley and the other eight populations. Previous hypotheses suggested the distinctness to be a result of population declines in the 19th and 20th centuries associated with human settlement and subsequent habitat loss (Hull et al. 2008a, England et al. 1997). Our findings suggest that mountain ranges could be a barrier to gene flow within the Swainons's hawk, and the Central Valley population is separated from all other populations by the Sierra Nevada Mountain range. Over time, the geography of the Central Valley could have decreased gene flow between the Central Valley populations and other populations, causing the Central Valley population to become slightly distinct. Interestingly, we also found some genetic distinctness in the Owen's Valley population. We believe that the mountain ranges may also limit dispersal between the Owen's Valley population and the others, as the Owen's Valley population is situated between the Sierra Nevada and White Mountain

ranges. These findings support mountain ranges acting as a dispersal barrier leading to genetic divergence in the Swainson's hawk.

In addition to describing the neutral variation between populations, we also used two approaches to identify loci and genes possibly under divergent selection between the Central Valley and the other populations. The 16 genes that overlapped between our two methods have the best support for being putatively under selection, but as no loci overlapped between the two methods, we believe both candidate loci and gene lists to still be important to understanding divergence in the Swainson's hawk. Given the known phenotypic difference between the Central Valley and the other populations, there were a few genes that we hypothesized may be under selection. These genes included those involved in coloration (e.g. MC1R; Theron et al. 2001), and migration (e.g. ADCYAP1; Mueller et al. 2011). Two genes from our pFst candidate gene list have known functions in birds related to these phenotypic differences between the Central Valley and the other populations. NEK2 was previously identified as a candidate gene for migration in White-crowned sparrows (Jones et al. 2008), and ANK1 is a candidate gene for pigmentation (Poelstra et al. 2013). Additionally, a single gene, TBX5, from our overlapping list of 16 candidate genes has known functions in birds. TBX5 has been identified as one of the causative genes of ptilopody, or leg feathering, in pigeons (Domyan et al. 2016). An additional gene from our high F_{ST} candidate gene list, GREM1, also has known functions in feather formation. GREM1 modulates BMP signaling and is a key regulator of the barb-generative zone topology (Li et al. 2017). Further investigation is needed to form a link between all of our candidate genes and fitness in the Central Valley population and provide evidence for their involvement in local adaptation.

Our results provide important information for the conservation of the Swainson's hawk, particularly in California. We provide support for the classification of three independent populations in California as well as new information regarding the genetic health of these populations as indicated by levels of genetic variation. Studies have shown a strong link between genetic variation and fitness (Reed and Frankham 2003) making the preservation of genetic variation in populations of importance to scientists and conservationists alike. Accordingly, we advise monitoring of the Owen's Valley population, which shows the least amount of genetic variation of all populations. The lower values of genetic diversity in this population may be a result of genetic drift from small population size combined with the mountain ranges creating geographic isolation. Our list of outlier loci and genes also provide an important resource for conservation managers as new conservation strategies are being implemented that prioritize the preservation of both functionally important, locally adapted genes and range-wide genetic diversity (Charruau et al. 2011). In fact, Funk et al. (2012) have highlighted the importance of incorporating genome-wide information on functional loci when defining conservation units.

In summary, this study has uncovered fine-scale population differentiation within the Swainson's hawk that not only provides important information for the conservation of a threatened species across its range and especially within California, but also shows the utility of genome-wide data to the field of conservation genetics. Our whole-genome approach has enabled us to identify candidate genes under selection from across the genome, the first step in in identifying the genomic basis of potentially adaptive life-history traits in the Swainson's hawk. Understanding the genetic architecture of adaptive traits, especially those that provide successful establishment in a new environment, provides conservation managers with the information needed to predict how a species may respond to changing environments (Barrett and

Schluter 2008). Given the threatened status of the Swainson's hawk, genomic studies will need to continue to be conducted to identify the selective pressures and genes that are driving adaptation within the Swainson's hawk. Specifically, the demographic history of the Swainson's hawk should be studied to decouple the effects of neutral processes such as genetic drift from selection and local adaptation, and a clear connection needs to be established between candidate loci and fitness, to confirm these loci as the genetic basis of local adaptation in the Swainson's hawk.

LITERATURE CITED

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. Trends in ecology & evolution, 23(1), 38-44.

Beaumont, M. A. (2005). Adaptation and speciation: what can Fst tell us?. *Trends in ecology & evolution*, *20*(8), 435-440.

Briggs, C. W., Collopy, M. W., & Woodbridge, B. (2011). Plumage polymorphism and fitness in Swainson's hawks. *Journal of Evolutionary Biology*, *24*(10), 2258-2268.

Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be?. *Nucleic acids research*, *44*(13), 6046-6054.

Charruau, P., Fernandes, C., OROZCO-terWENGEL, P., Peters, J., Hunter, L., Ziaie, H., ... & Burger, P.A. (2011). Phylogeography, genetic structure and population divergence time of cheetahs in Africa and Asia: evidence for long-term geographic isolates. Molecular Ecology, 20(4), 706-724.

Domyan, E. T., Kronenberg, Z., Infante, C. R., Vickrey, A. I., Stringham, S. A., Bruders, R., ... & Shapiro, M. D. (2016). Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. elife, 5, e12115.

England, A.S., M.J. Bechard, C.S. Houston 1997. Swainson's Hawk (*Buteo swainsoni*). In: Poole A., F. Gill (eds) The birds of North America, No. 265. The Academy of Natural Sciences, Philadelphia PA, and The American Ornithologists' Union, Washington, DC.

Estep, J.A. 1989. Biology, movements, and habitat relationships of the Swainson's hawk in the Central Valley of California, 1986-87. California Department of Fish and Game, Nongame Bird and Mammal Section Report.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, 14(8), 2611-2620.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016).

Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. Annu. Rev. Genet., 10, 253–280.

Fuller, M. R., Seegar, W. S., & Schueck, L. S. (1998). Routes and travel rates of migrating Peregrine Falcons Falco peregrinus and Swainson's Hawks Buteo swainsoni in the Western Hemisphere. Journal of avian biology, 433-440.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. Trends in ecology & evolution, 27(9), 489-496.

Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources, 17*(1), 27-32. DOI: 10.1111/1755-0998.12509

Goldstein, M.I., T.E. Lacher, Jr, B. Woodbridge, M.J. Bechard, S.B. Canavelli, M.E. Zaccagnini, G.P. Covv, E.J. Scollon, R. Tribolet, and M.J. Hooper. 1999. Monocrotophos-induced mass mortality of Swainson's hawks in Argentina, 995-96. *Ecotoxicology* 8:201-214.

Hedrick, P.W., Ginevan, M.E. & Ewing, E.P. (1976). Genetic polymorphism in heterogeneous environments. Annu. Rev. Ecol. Syst., 7, 1–32.

Houston, C.S. 1990. Saskatchewan Swainson's Hawks. *American Birds* 4:215-220.

Houston, C.S. and J.K. Schmutz. 1995. Declining reproduction among Swainson's Hawks in prairie Canada. *Journal of Raptor Research* 29:198-201.

Hull, J.M., R. Anderson, M. Bradbury, J.A. Estep, and H.B. Ernest. 2008a. Population structure and genetic diversity in Swainson's Hawks (*Buteo swainsoni*): implications for conservation. *Conservation genetics* 9:305-316.

Hull, J. M., Hull, A. C., Sacks, B. N., Smith, J. P., & Ernest, H. B. (2008b). Landscape characteristics influence morphological and genetic differentiation in a widespread raptor (Buteo jamaicensis). *Molecular Ecology, 17*(3), 810-824.

Jones, S., Pfister-Genskow, M., Cirelli, C., & Benca, R. M. (2008). Changes in brain gene expression during migration in the white-crowned sparrow. Brain research bulletin, 76(5), 536-544.

Katibah, E. F. (1984). A brief history of riparian forests in the Central Valley of California. In *California riparian systems* (pp. 23-29). University of California Press.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. Ecology letters, 7(12), 1225-1241.

Kochert, M. N., Fuller, M. R., Schueck, L. S., Bond, L., Bechard, M. J., Woodbridge, B., Holroyd, G.L., Martell, M.S. & Banasch, U. (2011). Migration patterns, use of stopover areas, and austral summer movements of Swainson's Hawks. *The Condor, 113*(1), 89-106.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources, 15*(5), 1179-1191.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics, 15*(1), 356.

Le Corre, V., & Kremer, A. (2003). Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics, 164*(3), 1205-1219.

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics, 74*(1), 175-195.

Li, A., Figueroa, S., Jiang, T. X., Wu, P., Widelitz, R., Nie, Q., & Chuong, C. M. (2017). Diverse feather shape evolution enabled by coupling anisotropic signalling modules with self-organizing branching programme. Nature communications, 8(1), 1-13.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation.

Machado, A. P., Clément, L., Uva, V., Goudet, J., & Roulin, A. (2018). The Rocky Mountains as a dispersal barrier between barn owl (Tyto alba) populations in North America. *Journal of Biogeography*, *45*(6), 1288-1300.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957-2963.

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719-731.

Meunier, J., Pinto, S. F., Burri, R., & Roulin, A. (2011). Eumelanin-based coloration and fitness parameters in birds: a meta-analysis. *Behavioral Ecology and Sociobiology*, *65*(4), 559-567.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L., Mushayamaha T., and Thomas, P.D. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API , Nucl. Acids Res. (2020) doi: 10.1093/nar/gkaa1106s.

Mueller, J. C., Pulido, F., & Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1719), 2848-2856.

Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, *26*(4), 160-167.

Palmer, R.S. 1988. Handbook of North American Birds, vols. 4 and 5. Diurnal raptors (Parts 1 and 2). Yale University Press, New Haven, CT

Pespeni, M. H., Garfield, D. A., Manier, M. K., & Palumbi, S. R. (2012). Genome-wide polymorphisms show unexpected targets of natural selection. Proceedings of the Royal Society B: Biological Sciences, 279(1732), 1412-1420.

Poelstra, J. W., Ellegren, H., & Wolf, J. B. W. (2013). An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. Heredity, 111(6), 467-473.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation biology*, *17*(1), 230-237.

Risenbrough R.W., Schlorff R.W., Bloom P.H. and E.E. Littrell. 1989. Investigations of the decline of Swainson's Hawk populations in California. *Journal of Raptor Research* 23:63–71

Ruegg, K., Bay, R. A., Anderson, E. C., Saracco, J. F., Harrigan, R. J., Whitfield, M., ... & Smith, T. B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecology Letters, 21(7), 1085-1096.

Schmutz, J. K., McLoughlin, P. D., & Houston, C. S. (2006). Demography of Swainson's Hawks breeding in western Canada. *The Journal of wildlife management*, *70*(5), 1455-1460.
Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693-702.
Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular ecology resources, 17(2), 194-208.
Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R., & Mundy, N. I. (2001). The molecular basis of an avian polymorphism in the wild: a point mutation in the melanocortin-1 receptor is perfectly associated with melanism in the Bananaquit (Coereba flaveola). *Curr Biol*, *11*, 550-7.
Turelli, M., Barton, N.H. & Coyne, J.A. (2001). Theory and speciation. Trends Ecol. Evol., 16, 330–343.
Wheeler, B.K. 2003. Raptors of Western North America. Princeton, N.J.: Princeton University Press.
Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.
Williams, G.C. (1966). Adaptation and Natural Selection. Princeton University Press, Princeton
Woodbridge B. 1991. Habitat selection by nesting Swainson’s Hawk: A hierarchical Approach. M.Sc. Thesis, Oregon State Univ., Corvallis, OR.
Woodbridge, B.K., K.K. Finley, and P.H. Bloom. 1995. Reproductive performance, age structure, and natal dispersal of Swainson’s Hawks in the Butte Valley, California. *Journal of Raptor Research* 29:187-192.
Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, G. & Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. *PloS one*, *7*(12), e52249.

FIGURES AND TABLES

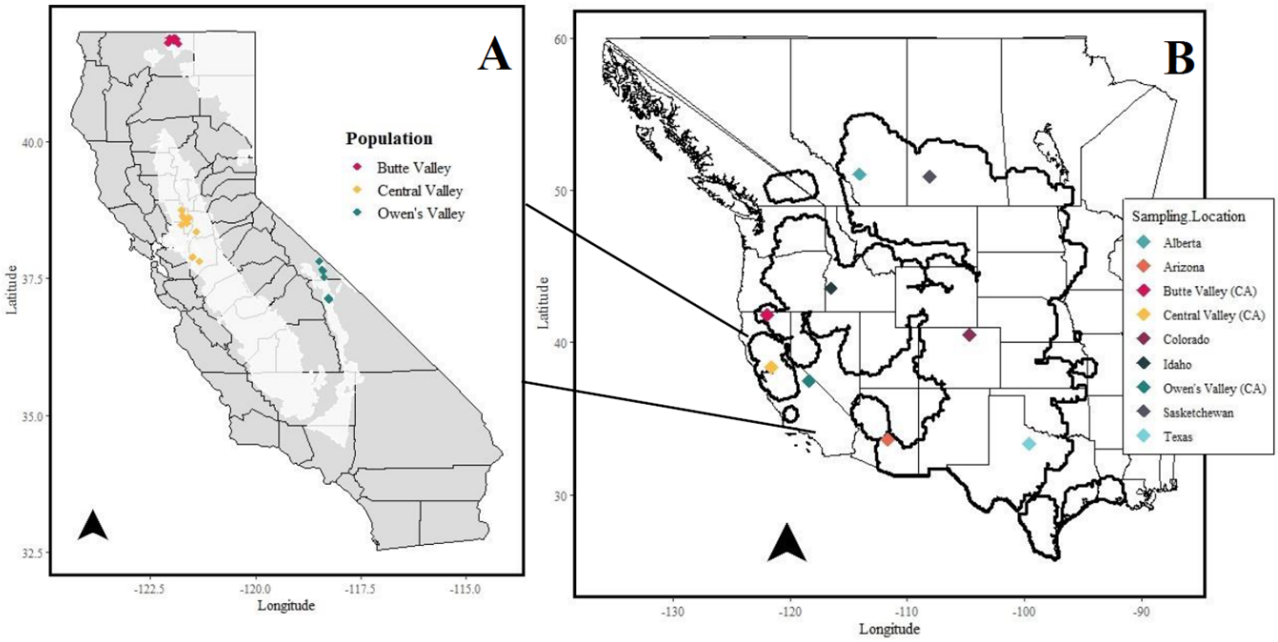


Figure 1.1 **A)** Map of California sampling locations. The shaded white area indicates the current range of the Swainson's hawk as determined by the Department of Fish and Wildlife and obtained through the California State Geoportal (www.gis.data.ca.gov). **B)** The 9 sampling locations of the Swainson's hawk populations. The thick black outline corresponds to breeding range determined by Breeding Bird Surveys from 2011-2015.

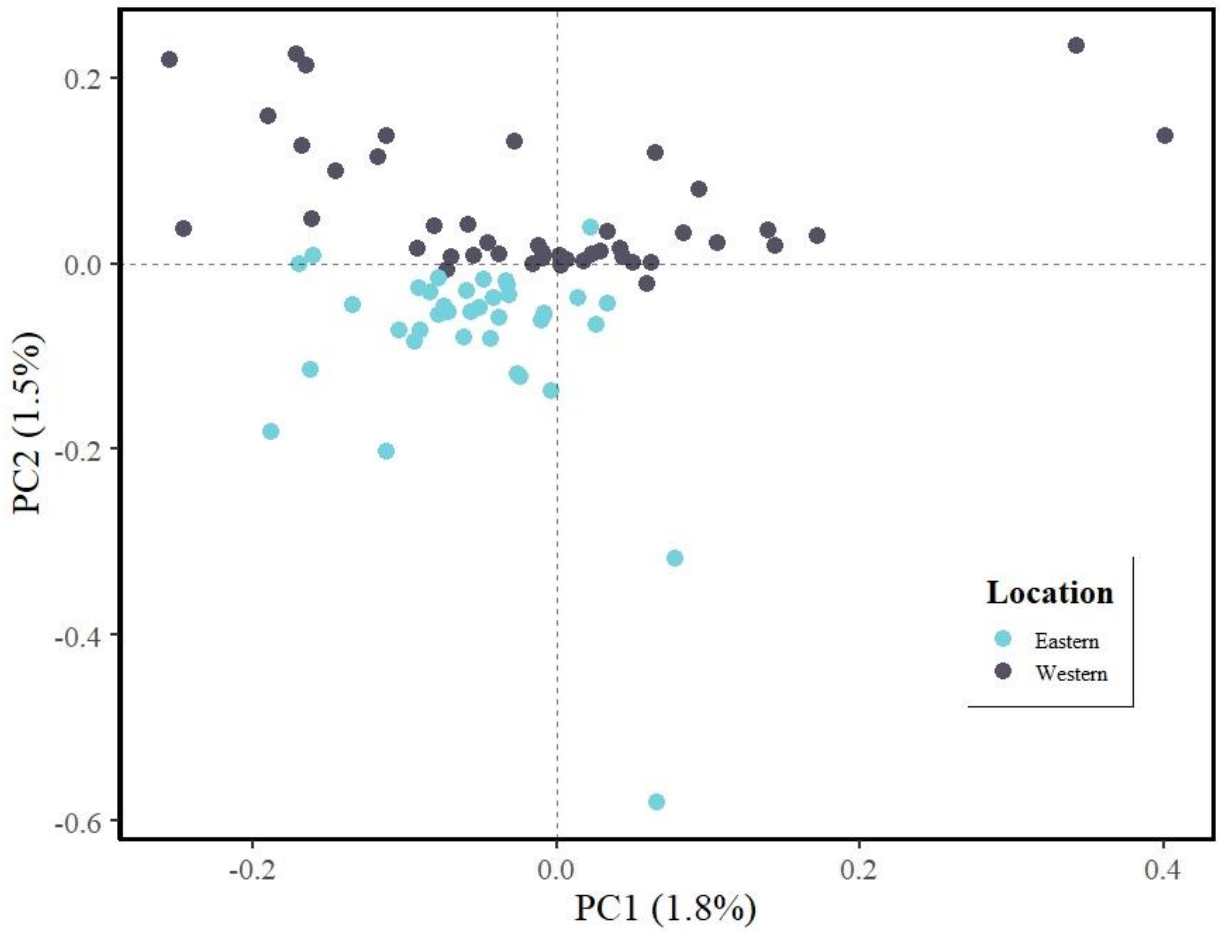


Figure 1.2. Principal Component Analysis of Eastern and Western Swainson's hawks as determined by the North American continental divide.

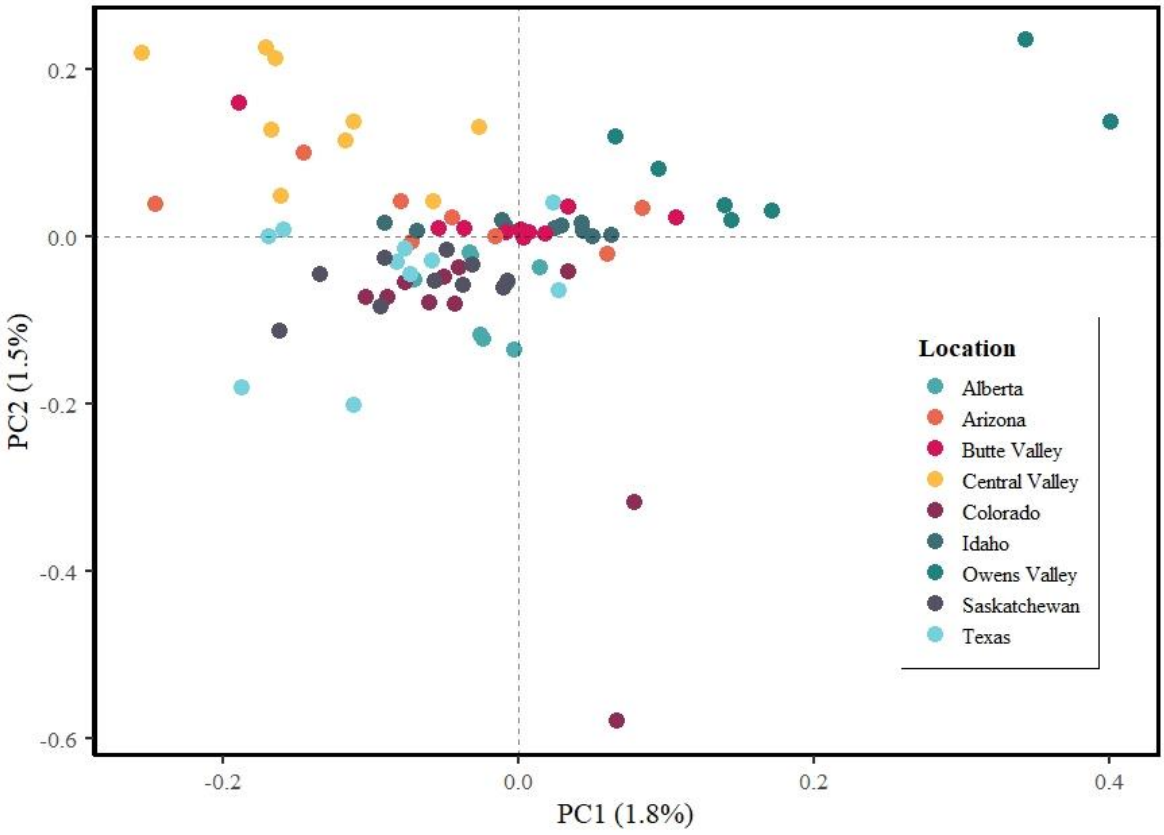


Figure 1.3. Principal Component Analysis of the nine Swainson's hawk populations.

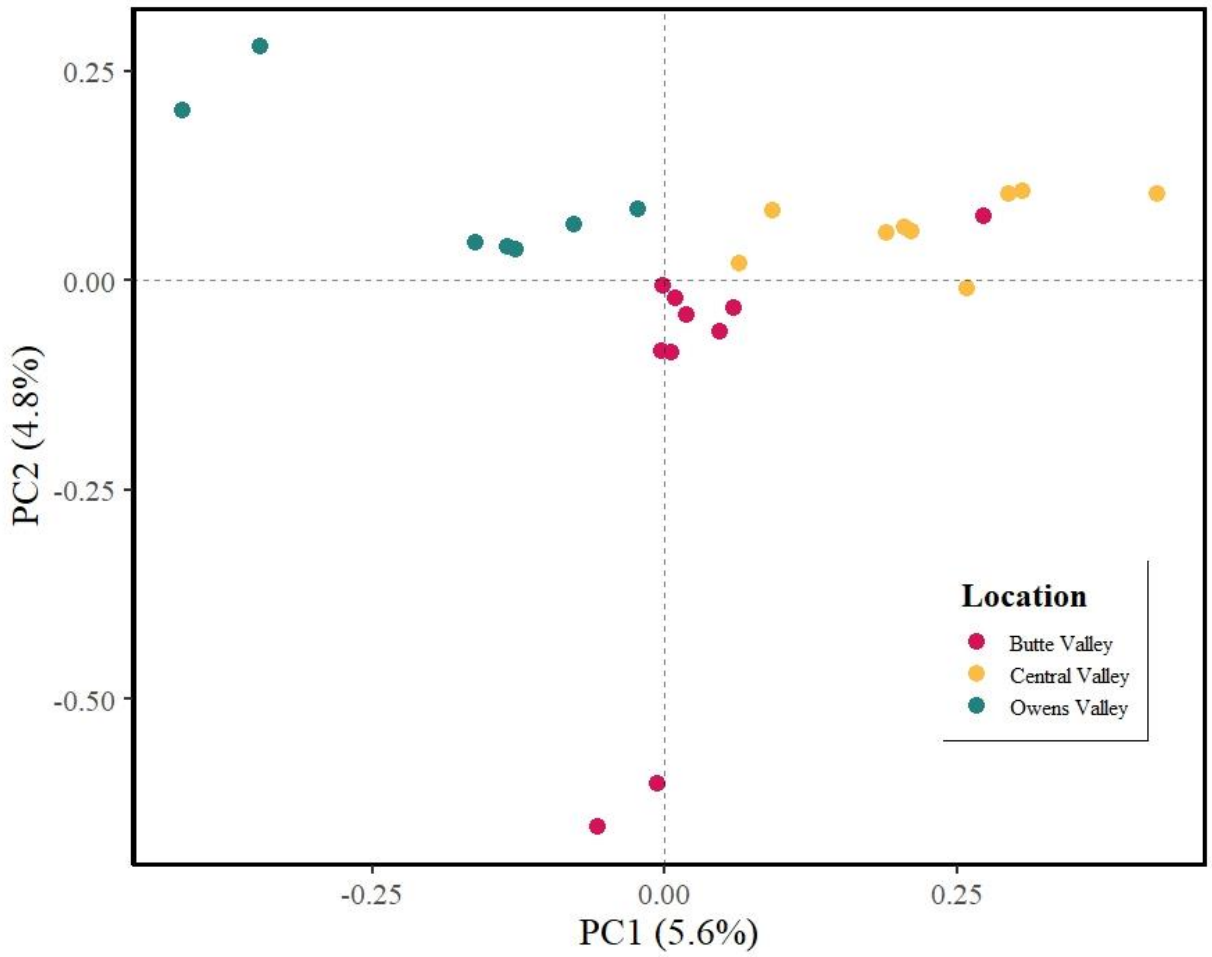


Figure 1.4. Principal Component Analysis of the California populations.

Table 1.1. Genome-wide FST values for all pairwise population comparisons.

	AB	**AZ**	**CE**	**CN**	**CO**	**CV**	**ID**	**SK**	**TX**
AB									
AZ	0.0565								
CE	0.0566	0.0521							
CN	0.0557	0.0533	0.0532						
CO	0.05	0.0502	0.0503	0.0496					
CV	0.0612	0.0529	0.0527	0.0546	0.0543				
ID	0.0534	0.0517	0.0509	0.0507	0.048	0.0548			
SK	0.054	0.0562	0.0556	0.0551	0.05	0.0601	0.0531		
TX	0.0522	0.0499	0.051	0.0508	0.0449	0.0542	0.049	0.0519	
Average	0.0549	0.0529	0.0528	0.0529	0.0497	0.0556	0.0515	0.0545	0.0505

Table 1.2. Genetic Diversity indices.

Population	Watterson's θ	Pairwise Theta (π)
Alberta	0.002159423	0.001594382
Arizona	0.002392244	0.001956844
Central Valley (CA)	0.002334536	0.001966433
Owen's Valley (CA)	0.001640728	0.001512081
Butte Valley (CA)	0.002285258	0.001652786
Colorado	0.002377951	0.001704432
Idaho	0.002322027	0.00167423
Saskatchewan	0.004114738	0.002322983
Texas	0.002405301	0.001704626

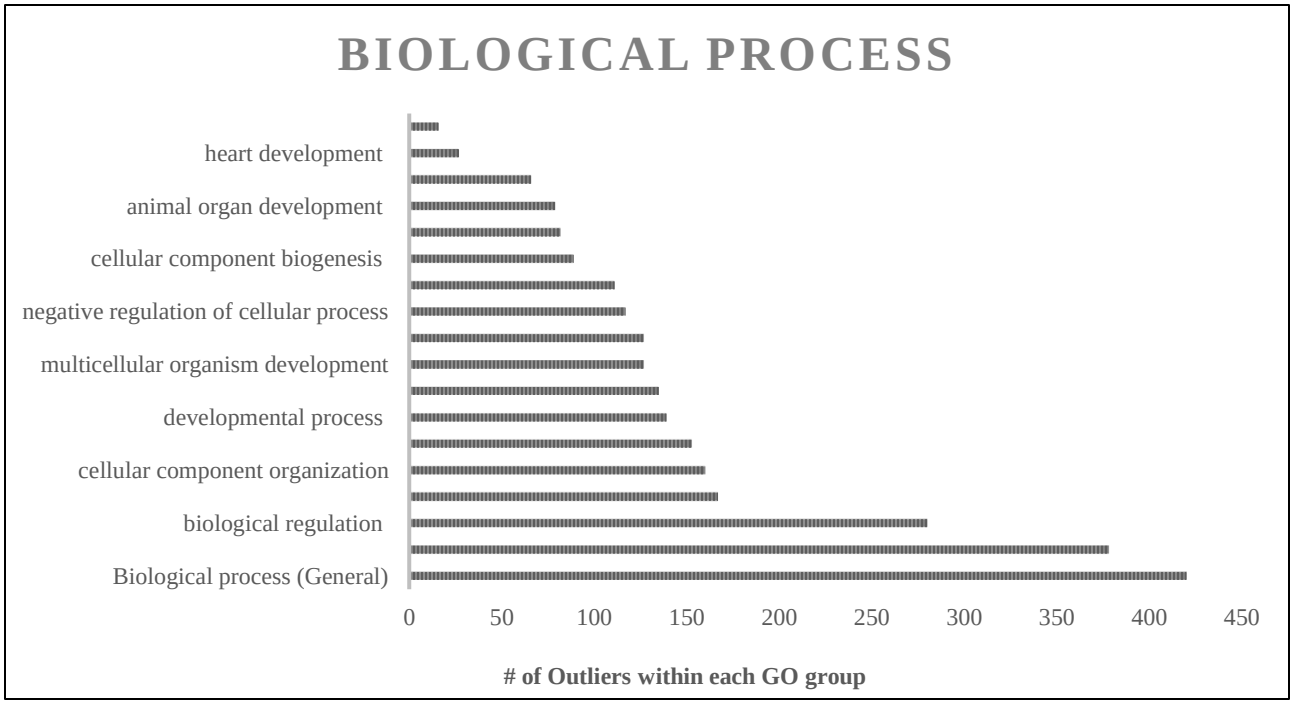


Figure 1.5. Biological process GO groups significantly overrepresented in our outlier dataset along with the number of outlier loci in each group.

Table 1.3: List of genes overlapping between our two candidate gene analyses.

Ensemble or UniProt ID	Gene Name
ENSGALG00000011816	C1QTNF1
ENSGALG00000012454	CYTH4
ENSGALG00000012446	ELFN2
ENSGALG00000033783	FER1L6
UniProtKB=P00368	GLUD1
ENSGALG00000042320	GTF2E1
ENSG00000122254	HS3ST2
ENSGALG00000005540	MICAL2
ENSGALG00000026119	MN1
ENSGALG00000043744	NDFIP1
ENSGALG00000012456	RAC2
ENSGALG00000006202	SCNN1B
ENSG00000187678	SPRY4
ENSG00000278195	SSTR3
ENSGALG00000008253	TBX5
ENSGALG00000006286	USP31

SUPPLEMENTARY FIGURES AND TABLES

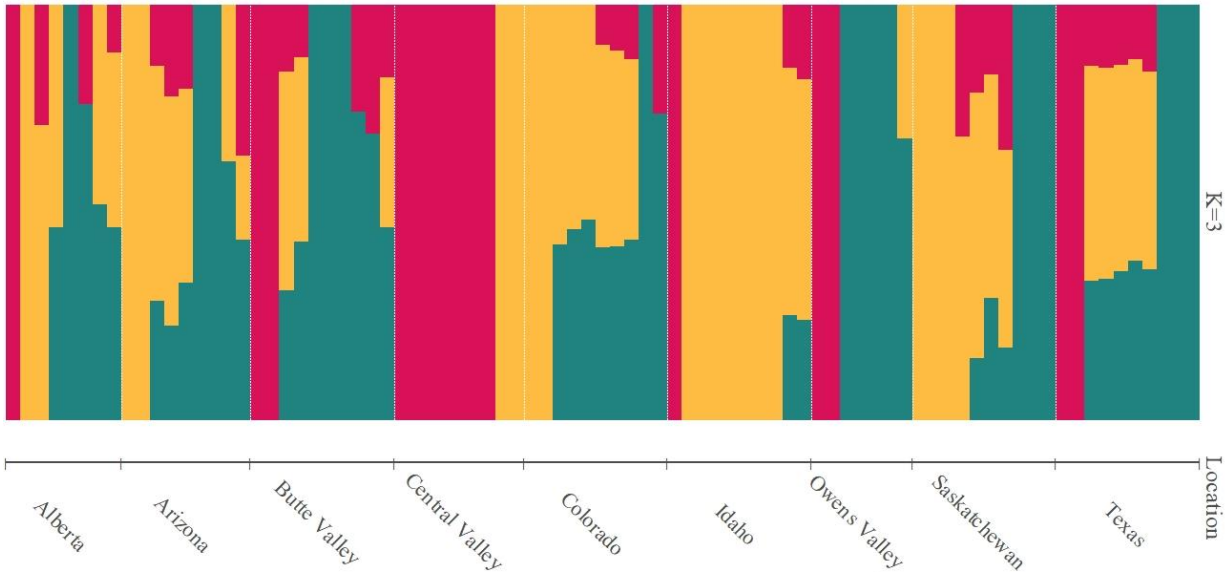


Figure S1.1: R Results of clustering analysis in NgsAdmix for K=3 across all 9 populations.

Table S1.2. Biological process GO groups overrepresented in our high F_{ST} candidate genes for selection.

GO biological process	GO term	Observed # of Outliers	Expected # of Outliers	P-value
protein binding	GO:0005515	167	129.99	2.45E-04
organic cyclic compound binding	GO:0097159	159	119.87	7.00E-05
heterocyclic compound binding	GO:1901363	156	118.32	1.27E-04
small molecule binding	GO:0036094	96	54.38	5.79E-08
anion binding	GO:0043168	89	52.34	9.17E-07
nucleotide binding	GO:0000166	87	48.48	1.57E-07
nucleoside phosphate binding	GO:1901265	87	48.48	1.57E-07
carbohydrate derivative binding	GO:0097367	83	48.25	1.71E-06
purine nucleotide binding	GO:0017076	80	43.31	1.45E-07
ribonucleotide binding	GO:0032553	79	43.23	3.18E-07
purine ribonucleotide binding	GO:0032555	78	42.87	4.45E-07
purine ribonucleoside triphosphate binding	GO:0035639	75	41.57	1.11E-06
adenyl nucleotide binding	GO:0030554	67	34.69	4.93E-07
adenyl ribonucleotide binding	GO:0032559	66	34.45	7.40E-07
ATP binding	GO:0005524	64	33.39	1.29E-06
pyrophosphatase activity	GO:0016462	40	21.17	1.94E-04
hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	GO:0016818	40	21.23	1.99E-04
hydrolase activity, acting on acid anhydrides	GO:0016817	40	21.23	1.99E-04
protein kinase activity	GO:0004672	29	13.82	2.98E-04
passive transmembrane transporter activity	GO:0022803	26	10.87	8.14E-05
channel activity	GO:0015267	26	10.87	8.14E-05
cation channel activity	GO:0005261	24	7.82	3.04E-06
ion channel activity	GO:0005216	24	9.97	1.48E-04
metal ion transmembrane transporter activity	GO:0046873	23	9.16	1.54E-04

Table S1.3. Biological process GO groups overrepresented in our high F_{ST} candidate genes for selection.

GO biological process	GO term	Observed # of Outliers	Expected # of Outliers	P-value
Biological process (General)	GO:0008150	420	376.21	3.81E-08
cellular process	GO:0009987	378	321.82	5.55E-09
biological regulation	GO:0065007	280	232.88	1.73E-05
cellular component organization or biogenesis	GO:0071840	167	111.67	1.38E-08
cellular component organization	GO:0016043	160	107.27	4.36E-08
multicellular organismal process	GO:0032501	153	100.07	2.01E-08
developmental process	GO:0032502	139	88.32	2.39E-08
anatomical structure development	GO:0048856	135	81.85	2.79E-09
multicellular organism development	GO:0007275	127	73.85	8.71E-10
negative regulation of biological process	GO:0048519	127	88.45	1.65E-05
negative regulation of cellular process	GO:0048523	117	81.38	4.47E-05
system development	GO:0048731	111	67.12	1.20E-07
cellular component biogenesis	GO:0044085	89	45.79	2.74E-09
cellular component assembly	GO:0022607	82	40.87	3.11E-09
animal organ development	GO:0048513	79	47.76	1.43E-05
anatomical structure morphogenesis	GO:0009653	66	38.88	3.30E-05
heart development	GO:0007507	27	9.58	3.33E-06
cell junction assembly	GO:0034329	16	4.58	3.26E-05

Figure S1.4: Molecular Function GO groups overrepresented in our pFst candidate genes for selection.

GO Molecular Function	GO term	Observed # of Outliers	Expected # of Outliers	P-value
potassium channel activity	GO:0005267	7	0.78	2.57E-02
potassium ion transmembrane transporter activity	GO:0015079	7	1.01	4.22E-02
gated channel activity	GO:0022836	11	2.13	2.69E-02
cation channel activity	GO:0005261	11	2.26	1.49E-02
ion channel activity	GO:0005216	13	2.87	3.36E-02
metal ion transmembrane transporter activity	GO:0046873	11	2.64	5.14E-02
passive transmembrane transporter activity	GO:0022803	13	3.13	2.07E-02
channel activity	GO:0015267	13	3.13	1.66E-02
ion transmembrane transporter activity	GO:0015075	16	5.27	4.58E-02

CHAPTER 2:

Uncovering the genetic basis of adaptation and diversification in the Galápagos Hawk (*Buteo galapagoensis*)

ABSTRACT

Studying adaptive radiations is integral to our understanding of the mechanisms driving rapid speciation. On islands, studying the genetic basis of adaptation will help scientists answer many fundamental questions in evolutionary biology. The Galápagos hawk (*Buteo swainsoni*) is the only endemic diurnal raptor on the Galápagos archipelago and is thought to be undergoing active speciation across islands. Here, we use whole-genome sequencing to characterize fine-scale population structure and genetic diversity as well as investigate the prevalence and genetic basis of both morphological divergence and divergence in mating system in the Galápagos hawk across the archipelago. We found significant population differentiation among islands, indicating a minimum of six distinct population, as well as very low levels of genetic variation among all islands. Additionally, we found 3 loci associated with cooperative polyandry, which varies among islands, as well as numerous loci associated with 6 morphometric traits that vary significantly across islands independent of body size. These results provide insight into the mechanisms driving speciation in the Galápagos hawk across islands by providing a unique opportunity to compare divergence patterns in both behavioral and morphological traits. Lastly, these results also provide important information into the conservation of an ecologically important, yet vulnerable species.

INTRODUCTION

Island ecosystems have long been thought of as natural laboratories for evolutionary studies because of their discrete geographic boundaries and large diversity of flora and fauna (Emerson 2002). Because islands are separated by physical oceanic barriers gene flow is reduced between populations, often resulting in diversification and adaptation to the local environment (Emerson 2002). In fact, rapid phenotypic change is often observed between mainland and island taxa (Grant 1998). This evolution of phenotypic and ecological diversity within a rapidly multiplying lineage, typically after the colonization of a new environment, is known as an adaptive radiation (Schluter 2000). The most classic example of an adaptive radiation in nature is that of Darwin's finches, which evolved into 18 species in response to different ecological conditions and feeding habits (Grant and Grant 2008), after colonizing the Galápagos archipelago ~1.5 million years ago (Petren et al. 2005).

Genome-wide patterns of divergence within an adaptive radiation provide exciting opportunities for studying the genomic architecture of diversification. Studying the genetic basis of adaptation in natural populations may help scientists address fundamental questions such as the number of genes involved in adaptation, their distribution across the genome, and even whether adaptation occurs through the rapid fixation of new mutations or draws from standing genetic variation (Barrett and Schluter 2008). Adaptive radiations, especially those in island environments, can overcome some of the major limitations of identifying specific loci underpinning adaptation due to the presence of population replicates and high rates of convergent evolution (Berner and Salzburger 2015). Additionally, new technological advances that allow for sequencing of whole-genomes across many populations undergoing adaptive

radiations increase our ability to identify adaptive loci as genome-wide data can detect loci under selection that would have been missed with other, reduced-representation sequencing methods, even at very low-coverage (Pespeni et al. 2012).

While the study of adaptive radiations can provide important insight into the mechanisms and drivers of ecological diversity, recent radiations are particularly important in providing insights into evolutionary and adaptive processes because extinctions are minimal and phenotypic differences between species are small and therefore more interpretable in the earliest stages of divergence (Lack 1947). The Galápagos hawk (*Buteo galapagoensis*), which arrived on the Galápagos archipelago as recently as 126,000 years ago shows clear phenotypic differentiation among island populations, suggesting that it is in the earliest stages of divergence and is therefore an ideal species to study the genetic basis of adaptive evolutionary change (Bollmer et al. 2006). The Galápagos hawk is the only diurnal raptor found on the archipelago and is presently found on 8 islands, although it historically inhabited 11 (Bollmer et al. 2006). Among islands, the Galápagos hawk has been shown to have strong genetic differentiation as well as distinct differences in phenotypic and behavioral traits such as body morphology and mating system (Bollmer et al. 2003, 2005).

In some taxa, morphological evolution is faster in island populations compared to mainland populations over relatively short time-scales (Millien 2006). In fact, many studies use the rate of morphological divergence on islands as evidence for the strong role of selection in shaping divergence over the role of genetic drift (Sendell-Price et al. 2020). In birds, there is evidence for the 'island rule' where on islands large birds evolve towards smaller sizes, and small birds evolve towards larger sizes (Clegg and Owens 2002). Body size is seen as a particularly adaptive trait as it can influence many characteristics such as dispersal

potential, ecological interactions, and resource acquisition. The Galápagos hawk shows significant differences in body size across islands (Bollmer et al. 2003) although the genetic basis of this differentiation is unknown. Divergence in body morphology over such a short evolutionary time period could be an indication that the Galápagos hawk is in the earliest stages of speciation across islands as evolutionary theory predicts a coupling between rates of morphological change and speciation (Stanley 1975). Therefore, the identification of the loci underlying these adaptations may provide important information into the adaptive mechanisms behind speciation and the relative contributions of morphological divergence to the speciation process.

In addition to morphological differences, the Galápagos hawk also shows clear differentiation in mating system among islands. The Galápagos hawk exhibits cooperative polyandry, where one female mates with up to 8 unrelated males although group sizes are typically 2-3 males per mating group (Faaborg and Patterson 1981, Faaborg et al. 1995). Paternity is shared within broods and mating groups are territorial (Faaborg et al. 1995). Among islands, there are large differences in the degree of polyandry with more polyandrous islands averaging larger group sizes. For example, on Espanola pairs are completely monogamous, while group sizes on other islands average between 2.5-4.5 birds (Bollmer et al. 2003). These differences are of particular interest as group size has been shown to positively affect adult survivorship, indicating an adaptive benefit of this trait (Rivera-Parra et al. 2012). Studying the genetic basis of polyandry is particularly exciting and important because the majority of studies on adaptive radiations have focused only on morphological traits (Losos et al. 1998, Abzhanov et al. 2004), so the genetic basis of behavioral traits is largely unexplored. Also, behavioral traits, especially those related to mating, are a critical part of avian speciation, as they can often result

in premating incompatibilities and prezygotic isolation which is thought to be the major driving force of reproductive isolation in birds (Hinde 1959).

The Galápagos hawk is currently listed as vulnerable by the IUCN because of its small population size (roughly 400-500 individuals) and narrow geographic range (Birdlife International). As populations continue to decline, the Galápagos hawk will face many threats to its viability and genetic health such as inbreeding (Keller and Waller 2002). Inbreeding depression is known to have many negative effects on population health and reproduction, including increased rates of hatching failure in birds (Briskie and Mackintosh 2004), and there is evidence that inbreeding depression increases extinction risk in wild populations (O'Grady et al. 2006). Using genomics to measure fine-scale genetic variation is an integral part of the conservation process and the discovery of genetic regions involved in diversification can help predict the performance of a genotype in a new environment, information that can improve management decisions for threatened or declining species (Funk et al. 2012).

Here we aim to use the Galápagos hawk to investigate the genetic basis of adaptive evolutionary change in a diurnal, island endemic raptor by 1) using whole genome sequencing to characterize patterns of divergence and diversity between islands and 2) investigating the genes underlying morphology and mating system by establishing links between these traits and loci under selection. As the Galápagos hawk is thought to be in the early stages of divergence across islands, this is a unique opportunity to study the mechanisms behind rapid divergence in both morphologic and behavioral traits and compare their relative contributions to the process of speciation. And, as the Galápagos hawk is listed as vulnerable by the IUCN, these results will provide information integral to the prioritization and implementation of conservation actions.

METHODS

Sample Collection

Feather and whole blood samples were collected from 140 Galápagos hawks across 8 islands in the Galápagos archipelago from 1998-2003 (Figure 2.1). Hawks were captured with either bal-chatris or rope nooses on poles and standard morphometric measurements were taken. These measurements included wing chord, tail length, cranium length (from the posterior of the cranium to the tip of the mandible), culmen length, bill depth (the vertical distance from dorsal to ventral surface of mandible at anterior edge of cere) and hallux claw length. Additionally, two 50-µL samples of blood were drawn from the brachial vein through venipuncture. These samples were stored in 500 -µL of Longmire's buffer (Longmire et al. 1987) at ambient temperature. Genomic DNA was extracted using a modification of phenol/chloroform extraction (Sambrook et al. 1989) that included a final purification by dialysis against TNE_2. The extracted samples were then stored in a -80°C freezer until further use.

Library Preparation and Sequencing

We evaluated DNA quantity using a Qubit Fluorometer (Invitrogen) and quality was evaluated with agarose gel electrophoresis. The whole genome sequencing protocol was modified from the newest version of an adapted Illumina protocol, first described in Therkildsen and Palumbi (2017). In short, we made 3 modifications to maximize efficiency of DNA fragmentation and recovery in the desired range for sequencing for use with low input and low quality samples. This method also works with high quality samples at low input. First, we doubled the ratio of tagmentation enzyme to DNA which results in increased fragmentation and therefore shorter average fragment lengths. We also increased the tagmentation incubation time

from 5 minutes to 20 minutes to ensure that the tagmentation enzyme had sufficient time to interact with the DNA. Lastly, we decreased the elongation time during the indexing PCR from 3 minutes to 30 seconds to preferentially increase amplification of shorter fragments within the targeted range for sequencing. Final individual libraries were pooled by equal copies and size selected to retain fragments in the 300-520 base pair range. Final QC was performed with a Qubit quantification and Tapestation for fragment analysis. Multiple paired-end libraries were sequenced at 2-4x coverage on an Illumina HiSeq 4000 at Novogene Corporation Inc. in Sacramento, CA. Studies have demonstrated that sequencing many individuals at a coverage as low as 1 read per locus provides more information about population parameters compared to sampling schemes with lower numbers of individuals and higher coverage (Buerkle and Gompert 2013, Fumagalli et al. 2013).

Read Processing and Variant Calling

We assessed the quality of the paired-end raw reads with FastQC v. 0.11.9 (Andrews 2010) and summarized with program MultiQC v. 1.8 (Ewels et al. 2016). Next, PCR duplicates were removed with FASTUNIQ v. 1.1 (Xu et al. 2012) and overlapping read pairs were collapsed into single reads with Flash2 (Magoč and Salzberg 2011). Next, reads were aligned to the Swainson's hawk draft genome assembly (Abernathy et al. in prep) using BWA v. 0.7.16 (Li 2013).

We created two main datasets, one that included genome-wide data, and one that only included variants. For our variant dataset, we identified polymorphic sites in ANGSD v. 0.930 (Korneliussen et al. 2014) using the following parameters: -uniqueOnly 1 -skipTriallelic 1 -minMapQ 30 -minQ 30 -doHWE 1 –maxHetFreq 0.5 -minInd n/2. Polymorphic sites were then

identified with –pval 1 e 10-6 and –maf 0.05 cutoffs, and genotype likelihoods were calculated for both the whole genome dataset and just for the polymorphic sites using the GATK model.

Population Parameters

First, we calculated genetic variation using two measures, Watterson's theta and Pairwise theta. We used the genome-wide genotype likelihoods calculated in ANGSD to calculate the maximum likelihood estimate of the folded site allele frequency spectrum using the Swainson's hawk genome as both the reference and ancestral states. Next, the thetas were calculated in ANGSD from the site frequency spectrum using the formulas described in Korneliussen et al. (2014) across sliding windows of 50 kb with a step of 10kb. We then corrected the theta values by dividing the output by the number of sites per window and calculated the genome-wide value by averaging across all windows. To test the relationship between theta and island size, we performed the non-parametric Spearman's rank order correlation between both Watterson's theta and Pairwise theta, and island size in km^2 as reported in Bollmer et al. (2005). We also estimated genome-wide heterozygosity for each population using the EM algorithm implemented in realSFS within ANGSD to get heterozygosity values for each individual, and then individual heterozygosity was averaged within populations.

To identify population structure across islands, population differentiation was calculated using both model and distance-based approaches. First, individual admixture proportions were calculated using the genotype likelihoods in NGSadmix (Skotte et al. 2013). Ten iterations were performed using "K" ancestral populations from 1-8. These results were visualized in R v. 3.6.3 (R Core Team 2021) using the package POPHELPER v2.3.1 (Francis 2017) and as all iterations were extremely consistent, no further iterations were conducted. Next, the most likely number of

ancestral populations was calculated in CLUMPAK (Kopelman et al. 2015), a software created specifically to aid in the interpretation of admixture results, using the Evanno method (Evanno et al. 2005). A principal component analysis (PCA) was conducted using a covariance matrix created in program PCAngsd v.0.98 (Meisner and Albrechtsen 2018). This program is made specifically for low-depth data and uses an iterative procedure based on genotype likelihoods to estimate the covariance matrix. These results and all further results were visualized in R v. 3.6.3 (R Core Team 2021) using package ggplot2 v. 3.3.3 (Wickham 2016). Next, we estimated pairwise F_{ST} for all population pairs in ANGSD by calculating the two-dimensional site frequency spectrum for each population pair. These joint-spectrums were then used as priors to calculate allele frequencies at each site in order to estimate F_{ST}. Lastly, to better understand the phylogenetic relationships between islands we used the covariance matrix output from PCAngsd v.0.98 (Meisner and Albrechtsen 2018) to construct a neighbor-joining tree which we then visualized using the R package phytools v. 0.7-47(Revell 2012).

Morphological Trait Divergence

To investigate how our morphological traits varied by island, we first split our data by sex (males =95, females =45), as Galápagos hawks are known to exhibit reverse sexual size dimorphism. Only the males were included in the rest of the analysis as the sample size for females was too low on many of the islands. Previous studies have shown that the body size of the Galápagos hawk varies significantly by island (Bollmer et al. 2003). Because of this, we performed a size correction on our morphological traits to see if these traits vary independently from body size across islands. We used wing chord as our measure of body size as this trait was found to be the strongest predictor of overall body size in previous studies on the Galápagos

Hawk (Bollmer et al. 2003) and Merlins (*Falco columbarius*)(Warkentin et al. 2016). We performed the size corrections by regressing each of the five other trats against wing chord, and then using the residuals of this analysis as our new size corrected trait values. The size corrected trait values were used for all subsequent analyses for hallux, culmen, cranium, tail, and bill depth.

Next, we tested for significant differences in morphology across islands by performing one-way ANOVAs on our six morphometric traits. For each trait, outliers were removed that were above the 75^{th} or below the 25^{th} percentile of values by a factor of 1.5x the interquartile range. This removed between 2-4 samples for each trait. We also tested for pairwise differences in the means by performing Tukey's HSD test for all pairwise comparisons of islands. Next, we performed a principal component analysis on our morphological traits to visualize how traits clustered among islands.

To investigate the potential for a genetic basis of these morphological traits, we performed Spearman's rank correlation tests between the size corrected trait values, and the first seven principal components (PCs) from our PCA conducted on our genomic data. The first seven PC's were retained as they explained almost 100% of the genomic variation. We also used Spearman's rank correlation tests to investigate correlations between the first seven genomic PC's and the first two morphological PC's.

Next, we used genome-wide association tests implemented in ANGSD and described in Skotte et al. 2012 to find candidate loci associated with our six morphological traits as well as our first two morphological principal components. This test takes genotype uncertainty into account by using a generalized linear framework with posterior genotype probabilities to calculate a score statistic for the joint likelihood of the observed phenotypes. We removed

outliers for each trait as described above, and we included the first seven genomic principal components as covariates to correct for the large amount of population differentiation in our dataset (Price et al. 2006). We identified candidate loci as those with a significance value less than 5×10^{-4} to be conservative, as false-positives are a common issue in this type of analysis. Next, we identified genes that were within 200kb of our candidate loci. We used this cut-off as recent studies have shown that SNPs may affect or be linked to distant genes (Brodie et al. 2016). To identify gene functions, we used the gene list analysis function in Panther v. 16.0 (Mi et al. 2020). Our outlier gene list was annotated against the chicken (*Gallus gallus*) GO Ontology database DOI: 10.5281. We performed the PANTHER Overrepresentation Test for both the GO biological process and GO molecular function annotation datasets on our outlier loci. Specifically, we used the Fisher's Exact test, controlling for False Discovery Rate, and using the chicken genome as a reference list.

Genomic Basis of Polyandry

To identify loci associated with cooperative polyandry, we classified our island populations into two categories, those that are "less polyandrous" and those that are "more polyandrous" following Bollmer et al. (2005). Espanola, Santa Fe, Pinzon, and Fernandina are classified as "less polyandrous" as they typically have less than 2 males per polyandrous group, and Isabela, Marchena, Santiago, and Pinta, are classified as "more polyandrous" as they have on average greater than 2 males per polyandrous group (Bollmer et al. 2005). We were able to account for differences based on population structure as our most phylogenetically close islands, Fernandina and Isabela, and Pinzon and Santiago were both classified as having one "more polyandrous" and one "less polyandrous" island We used program pFst implemented

in vcflib (Garrison 2020) to detect significant differences in allele frequencies between all pairwise comparisons of "more polyandrous" versus "less polyandrous" populations. This program is a probabilistic approach that uses a likelihood ratio test based on the binomial distribution that outputs p-values based on the chi-squared distribution with one degree of freedom. Two sets of candidate loci with p-values greater than 0.05 were assembled, those that overlapped between all pairwise comparisons, and also those that only overlapped between our phylogenetically close pairwise comparisons. Next, we identified candidate genes as those that were within 200kb of our candidate loci.

RESULTS

Dataset

A total of 233G of raw data were produced for all 140 samples. The number of reads per sample ranged from 6,028,833 to 14,001,794 for an average of 9, 734,660 reads per sample and an average coverage of 2x. An average of 97.04% of reads from each sample mapped to the Swainson's hawk reference genome. After filtering out paralogs and low quality reads, the final genome-wide dataset consisted of 1,093,747,856 sites, and the final SNP dataset consisted of 797,067 sites.

Population Structure and Genetic Diversity

Our principal component analysis showed six clearly differentiated populations (Figure 2.2). Two sets of islands clustered together, Pinzon and Santiago, and Isabela and Fernandina, while the last four islands, Santa Fe, Pinta, Marchena, and Espanola, clustered individually. Our admixture analysis showed the same pattern. Although the most likely number of populations

calculated using the Evanno method was determined to be 5, based on individual Q values and what we know of the distribution of the islands and the life-history of the Galápagos hawk, we have determined that the most likely number of populations is 6, with the same groupings as determined above in the principal component analysis (Figure 2.3). The values for pairwise F_{ST} ranged from 0.042 to 0.594 with an overall archipelago value of 0.333. Espanola has the highest average pairwise F_{ST} value of 0.4729 and Isabela has the lowest with a value of 0.230. The two lowest pairwise values were between our two sets of islands that grouped together in our structure analyses, Isabela and Fernandina (0.0421) and Pinzon and Santiago (0.1205) (Table 2.1). Lastly, the neighbor joining tree confirmed that Pinzon and Santiago, and Isabela and Fernandina are each other's closest relatives, as they grouped as sister taxa (Figure 2.4).

For our genetic diversity estimates, we found that overall genetic diversity was very low (Table 2.2). The average θ value was 0.000379 and the average nucleotide diversity (π) was 0.000179 across all 8 islands. Pinzon had the lowest θ (0.00025216) and Santiago the highest (0.0004632), while Santa Fe had the lowest (0.0001394) and Isabela the highest (0.0002236) nucleotide diversity. The average heterozygosity for the archipelago was 0.221, with values ranging from 0.175 (Santa Fe) to 0.2628 (Isabela). We also found a significant positive relationship between island size and Watterson's θ (Spearman's rank, *p=0.03676*) and a non-significant but positive relationship between island size and nucleotide diversity (Spearman's rank, *p=0.05759*).

Morphological Differentiation

We found that independent of body size, tail, cranium, bill depth, and culmen differed significantly between islands, but hallux did not (Figure 2.5). Specifically, for tail length, Pinzon,

Isabela and Fernandina had longer tails than we would expect for their body size, while Marchena, Pinta, and Santiago had shorter tails than we would expect. Additionally, Fernandina had the largest cranium compared to body size, while Marchena had the smallest. Bill depth was similar across body size except for Pinzon, which had a much smaller bill depth relative to body size. Lastly, Marchena, Pinta, Pinzon, and Santa Fe had slightly smaller culmen measurements than expected for their body size. Wing chord also varied significantly between Islands with on average Espanola having the largest wing chord, and Marchena the smallest.

Our principal component analysis of the six morphometric traits did not cluster clearly by islands with the first principal component (PC1) explaining 26.13% of the total variance, and the second principal component (PC2) explaining 20.78% (Figure 2.6). Cranium and bill depth had the highest loading on PC1, while wing chord and tail loaded most strongly onto PC2 (Table 2.3). When compared to the principal components from our genomic PCA, we found a strong correlation between morphometric PC1 and genomic PC1 (Spearman's rank, $p=5.579x10^{-7}$). When comparing our morphometric phenotypes to the first seven genomic PC's, we found that tail and culmen length correlated most strongly with genomic PC1, cranium length correlated most strongly with genomic PC2, and bill depth correlated most strongly with genomic PC5 (Table 2.4).

The genome-wide association tests for the six morphometric traits and the first 2 morphometric principal components resulted in the identification of numerous candidate genes for each trait and the results have been visualized in manhattan plots (Figure 2.7, 2.8). Using a p-value cutoff of 5×10^{-4}, the number of significant sites and the number of genes within 200kb of these sites is summarized in Table 2.5. Cranium had the highest number of significant sites followed by wing, tail, culmen, bill, and lastly hallux. As expected, cranium also had the highest

number of associated genes, and hallux had the lowest. Two genes, EML6 and SPBN1 were significant for both tail and wing chord, and one gene, ANO2, was shared between cranium and wing chord. No other genes overlapped between the morphologic traits. Our overrepresentation tests resulted in four overrepresented biological process Gene Ontology (GO) terms for the cranium (Table 2.6), and no overrepresented GO terms for the other five morphometric traits.

The association tests for PC1 and PC2 resulted in 307 and 27 significant sites with the same p-value cutoff of $5x10^{-4}$ (Figure 2.8). For PC1, there was a total of 209 genes within 200kb of the 307 significant sites (Suppl. Table 2.1,2.2). Fifteen of these genes overlapped with the genes found for the six morphometric traits. GRM1 overlapped with tail, EIF42A with cranium, ARHGEF33 with bill, CDCP1, CDH7, CLEC3B, DUSP12, MAPK8IP3, METTL16, NAA50, SIDT1, SORCS3, SPICE1, TENT5C with culmen, and PCNX2 with wing chord. No genes overlapped between PC1 and hallux. For PC2, 38 candidate genes were identified within 200kb of the 27 significant sites. Two of these genes overlapped with the significant genes from the morphometric traits, ASXL3 with tail, and IMPG1 with wing chord. The statistical overrepresentation test did not result in any overrepresented GO groups for the genes associated with either PC1 or PC2.

Identification of Candidate Genes Associated with Polyandry

Our pFst analysis resulted in the detection of three significant ($P < 0.05$) loci that overlapped between all of our pairwise comparisons between more and less polyandrous islands (Table 2.7). These sites were all located on different scaffolds, but two of them mapped to the same chromosome. Seven genes were found within 200kb of these three loci, BARX2, TMEM45B, APLP2, PRDM10, CLPTIM1L, CHMP5, and PCBP3. When just comparing our

most phylogenetically close pairs of islands, Pinzon and Santiago, and Fernandina and Isabela, 150 significant loci overlapped between the four pairwise comparisons. We identified 148 genes within 200kb of these candidate loci. Our statistical overrepresentation analysis revealed 11 overrepresented GO groups related to molecular function, and 127 overrepresented GO groups related to biological processes (Table 2.8,2.9).

DISCUSSION

In this study we investigated the genetic basis of adaptive evolutionary change in the Galápagos hawk. We used whole-genome sequencing to find evidence of strong genetic differentiation between island populations as well as very low levels of genetic variation across the archipelago. Additionally, we investigated patterns of divergence in morphology by finding significant variation in morphometric traits across islands independent of overall body size. Lastly, we investigated the genetic basis of these morphometric traits as well as cooperative polyandry, by identifying loci and genes putatively under selection for these phenotypic differences.

We found strong evidence for 6 distinct population units across the eight occupied islands, findings that are consistent with previous studies on the Galápagos hawk using fewer markers (Bollmer et al. 2005, 2006). Our admixture analyses and pairwise F_{ST} comparisons show that Isabela and Fernandina can be classified as a single population, as they clustered together in all of our differentiation analyses and had the lowest pairwise F_{ST} value of 0.04. This is not surprising as Fernandina and Isabela are the closest islands geographically, being only about 4.5 km apart at their closest point. This indicates recent admixture between these islands possibly enabled by their geographic closeness. Pinzon and Santiago also clustered together in our

admixture analyses but had a slightly higher pairwise F_{ST} value of 0.12. While this is the second lowest pairwise F_{ST} value among our island pairs, this is still considered to be moderate differentiation (Hartl and Clark 1997). While we did not formally test for isolation by distance, our pairwise F_{ST} results are consistent with geographically close islands being more genetically similar to each other than islands that are geographically farther apart. This conclusion is also supported by the known aversion that soaring raptors have to crossing large bodies of water (Kerlinger 1985) which would result in minimal dispersal between islands.

The large divergence and high levels of F_{ST} between populations is not surprising due to the rapid rate at which founder effects can change the frequency of alleles in a population (Excoffier et al. 2009). F_{ST} is thought to increase sharply in the beginning of colonization (Austerlitz et al. 1997), before decreasing over time with the homogenizing effects of migration, but we know migration is very low in this system. Additionally, we hypothesize that much of this divergence is also due to genetic drift, as previous studies have found that genetic drift drives genome-wide divergence in many island ecosystems (Funk et al. 2016). Still, these pairwise F_{ST} values are very high, at levels above what is often used to delineate species. This suggests the need for further investigation into whether or not the different islands can be classified as distinct species or sub-species.

Neutral genetic variation has been well characterized in many endemic species across the Galápagos, although native Galápagos bird species show varying patterns in genetic variation at neutral loci (Bollmer and Nims 2018). Ours is one of few studies to look at patterns of genetic variation in a Galápagos endemic species on the whole-genome scale. The Galápagos hawk shows much lower levels of genome-wide genetic diversity than its closest mainland relative, the Swainson's hawk (Abernathy et al. in prep) a pattern that has also been found consistently in

other endemic island bird populations (Frankham 1997). The differences in genetic diversity between islands may provide insight into the colonization history of the archipelago as genetic diversity has been found to steadily decrease with concurrent colonization events due to re-occurring bottlenecks and founder events (Austerlitz et al. 1997). Alternatively, the levels of genetic diversity may also be driven by differences in island size as we found that genetic diversity does increase significantly in larger islands. This is also consistent with previous findings of a positive, significant relationship between nucleotide diversity and island size in the Galápagos hawk using minisatellite markers (Bollmer et al. 2005), and also in the Galápagos mockingbird (*Mimus* spp.) (Hoeck et al. 2010). Island populations are limited in size by the islands that they inhabit, and there is a known relationship between genetic variation and population size in wildlife species (Frankham 1996). Mating system is also known to have an effect on genetic diversity by influencing effective population size (Nunney 1993), but it has been previously found that in the Galápagos hawk, mating system does not have an effect on genetic diversity (Bollmer et al. 2005). Our understanding of the true effect of mating system on the evolutionary patterns of divergence in the Galápagos hawk can be informed by identifying specific genes associated with differences in mating system across the archipelago.

Our investigation into the genes underlying polyandry in the Galápagos hawk recovered three loci and seven genes that had significantly divergent allele frequencies between our more and less polyandrous islands. If these divergent allele frequencies were due to random genetic drift, it is unlikely that these same loci would significantly diverge in allele frequency independently of each other in the more versus less polyandrous islands. In addition, because our most closely related island pairs have different degrees of polyandry, we believe this pattern to be a result of convergent evolution for these similar loci across islands, although we also

recognize the possibility of common ancestry playing a role in the frequency of these alleles. Given that convergent evolution is common in adaptive radiations (Berner and Salzburger 2015) we find strong support for these loci playing a role in determining polyandrous group sizes in the Galápagos hawk. Although not many studies have looked at the genetic architecture of mating systems, early evidence suggests strong potential for loci of major effect (Lamichhaney et al. 2016, Tuttle et al. 2016), as has also been found in other behavioral traits such as migratory tendency (Bensch et al. 2002, Delmore et al. 2016). Our finding of only three possible loci associated with the degree of polyandry supports this hypothesis. While none of the genes associated with our candidate loci have been previously connected to mating systems, a number of them have been found as candidate genes under selection in various taxa. CLPTM1L has been found to be a candidate gene for apoptosis in chickens (Gu et al. 2020), BARX2 is known to be an important regulator of muscle growth and repair (Makarenkova and Meech 2012), APLP2 has been found as a candidate gene under selection in the seminal fluid of house sparrows (Rowe et al. 2020), and PRDM10 may play a role in sexual dichromatism in birds (Gazda et al. 2020). Many hypotheses have been proposed for why polyandry has evolved in general, such as a means of avoiding inbreeding (Cornell and Tregenza 2007) and defense against genomic incompatibility (Zeh and Zeh 1997). Because cooperative polyandry as a mating system is so rare, the exact selective mechanisms driving the evolution of cooperative polyandry are unknown, but the identification of genes associated with the degree of polyandry in the Galápagos hawk may provide insight into the biological mechanisms that underly this behavior.

While overall body size has been previously described as divergent across islands in the Galápagos hawk, we found evidence of significant differences in morphology independent of body size. Galápagos hawks are known to follow the ‘island rule’ (Clegg and Owens 2002) with

a documented increase in body size since arriving on the archipelago, often associated with competitive release and a decrease in predation (Lomolino 1985), but information on the evolution or morphological traits independent of body size was previously lacking. Of the morphological traits that varied significantly across islands, two were related to feather length (wing chord and tail) and the other three were skeletal features related to cranial morphology (cranium, culmen, and bill depth). A common trend on islands is the evolution of flightlessness in birds (Slikas et al. 2002). This trend is characterized by a reduction in flight muscles (Wright and Steadman 2012, Livezey 1992) and a reallocation of mass from the forelimbs to the hindlimbs (Gaspar et al. 2020). It has been hypothesized that longer wings and tails evolve in island birds to compensate for the evolution of smaller flight muscles (Wright and Steadman 2012), which may explain some of the differences in wing chord and tail length we found in the Galápagos hawk. While the Galápagos hawk still has the ability to fly, studies have shown that island birds tend to evolve on a trajectory toward flightlessness, even if they retain the ability to fly (Wright et al. 2016). We recognize that differences in feather characteristics across islands could also be due to feather wear and molt patterns, so more study is needed to find a correlation between changes in wing and tail length and an evolutionary trajectory towards flightlessness in the Galápagos hawk.

The significant differences in cranial features across islands independent of body size are of particular interest because size has previously been found to explain about 80% of cranial shape variation in raptors (Bright et al. 2016), and, birds that eat terrestrial vertebrates, like the Galápagos hawk, were recently found to have the slowest rates of cranial evolution of all avian dietary niches (Felice et al. 2019). We were also surprised to find that the three cranial morphology traits varied independently of each other across islands because cranial features are

known to covary strongly in many avian lineages (Klingenberg and Marugán-Lobón, 2013). One explanation for these morphological differences would be differences in diet and foraging habits across islands since cranial morphology is known to be correlated with the ecology and diet of many avian clades (Jonsson et al. 2012, Lovette et al. 2002). This is especially true in island radiations as evidenced by the evolution of different cranial features across different ecological niches in Hawaiian honeycreepers and Darwin's finches (Tokita et al. 2016). While the Galápagos hawk is the only diurnal raptor on the archipelago, and broadly occupies the same ecological niche on each island, little is known about the specific dietary preferences of the Galápagos hawk among islands. Therefore, the specific contributions of diet and behavior to the differential evolution of morphological traits across islands is unclear. Alternatively, because each island may have been colonized by very few individuals, it is possible that differences in cranial morphology are due mostly to founder effects and genetic drift, as morphological variation has recently been determined to be accounted for primarily by drift in some island populations (Sendell-Price et al. 2020).

Interestingly, while morphological traits are known to be adaptive in many systems, the strong correlation found between the morphometric principal components and the genomic principal components suggests a possible neutral evolutionary trajectory of these traits, most likely influenced by strong genetic drift following the colonization of each island. We also found significant correlations between morphometric traits and genetic variation, and we interpret this pattern as supporting a genetic basis of many of these traits. To better describe this genetic basis, our association analyses revealed candidate loci and genes significantly associated with the variation in all of our morphometric traits and in our first two principal components. Although we used a conservative significance threshold of 5×10^{-4}, we recognize that some of these

candidate loci may be false positives due to the large amount of population structure in our data, and how population stratification is known to increase the number of false-positives in association studies (Thomas and Witte 2002). The presence of multiple peaks in the association studies spread out across the genome, rather than a single site, suggests a polygenic nature to these traits. This means that much of the variation is likely explained by many sites with smaller effect sizes, a trend that has also been previously found for morphological traits in multiple passerine species (Duntsch et al. 2020, Silva et al. 2017). To better understand how these genes affect morphological traits in the Galápagos hawk further study needs to be conducted on the functionality of these traits and predictive modeling is needed to decrease the discovery of false-positives, a common issue in association studies.

Because of the small population sizes and very low genetic-variation, the Galápagos hawk is seen as a priority for conservation. The prioritization of conservation actions through the establishment of discrete conservation units is particularly important in subdivided populations (Toro and Caballero 2005). The high levels of differentiation we found among populations suggests the need for classifying each island as a distinct conservation unit. Additionally, measures of genome-wide genetic diversity are essential to the conservation process as genome-wide genetic diversity has recently been found to be the best approach to prevent the loss of adaptive potential and inbreeding depression (Kardos et al. 2021). The identification of loci involved in polyandry and morphology will also provide the opportunity to assess the diversity of these loci across populations in an attempt to measure adaptive potential among islands.

The study of adaptive radiations is integral to our understanding of the mechanisms that drive speciation and diversification (Schluter 2000). Few studies have investigated the roles of both behavioral and morphological divergence in driving species radiations. Here, we provide

knowledge into the genetic architecture of both mating system and morphology in the Galápagos hawk in hopes of better understanding the relative roles of these traits in propelling divergence among islands. Behavioral mechanisms that affect reproductive isolation can sometimes lead to divergence in morphological traits. We did not find any connection between divergence in mating behavior and morphology as no morphometric traits varied similarly in more versus less polyandrous islands. To truly understand the drivers of diversification and determine whether or not there is a connection between divergence in morphology and mating behavior, further ecological study is needed, particularly on the diet of the Galápagos hawk, and whether or not polyandrous group sizes are stable over time. These results provide a foundation for future investigations into the specific molecular mechanisms underlying adaptive traits within the Galápagos hawk. Additionally, these results will provide more knowledge about the genetic architecture of these traits, which will help us understand the true effect of genetic diversity loss on the adaptive potential of small populations. Lastly, the identification of putatively adaptive loci combined with our measures of genetic differentiation and diversity provide guidance for conservation practices of the threatened Galápagos hawk.

LITERATURE CITED

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, *305*(5689), 1462-1465.

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Austerlitz, F., Jung-Muller, B., Godelle, B., & Gouyon, P. H. (1997). Evolution of coalescence times, genetic diversity and structure during colonization. Theoretical population biology, 51(2), 148-164.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in ecology & evolution*, *23*(1), 38-44.

Bensch, S., Åkesson, S., & Irwin, D. E. (2002). The use of AFLP to find an informative SNP: genetic differences across a migratory divide in willow warblers. Molecular Ecology, 11(11), 2359-2366.

Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), 491-499.

BirdLife International. (2012). Buteo galapagoensis. The IUCN Red List of Threatened Species 2012: e.T22695909A37899125. http://dx.doi.org/10.2305/IUCN.UK.2012-1.RLTS.T22695909A37899125.en. Downloaded on 26 November 2016

Bollmer, J.L., T. Sanchez, M.D. Cannon, D. Sanchez, B. Cannon, J.C. Bednarz, T. De Vries, M. S. Struve, and P.G. Parker. 2003. Variation in morphology and mating system among island populations of Galápagos hawks. The Condor. 105: 428-428.

Bollmer, J. L., Whiteman, N. K., Cannon, M. D., Bednarz, J. C., Vries, T. D., & Parker, P. G. (2005). Population genetics of the Galápagos hawk (Buteo galapagoensis): genetic monomorphism within isolated populations. The Auk, 122(4), 1210-1224.Bright, J. A., Marugán-Lobón, J., Cobb, S. N., & Rayfield, E. J. (2016). The shapes of bird beaks are highly controlled by nondietary factors. *Proceedings of the National Academy of Sciences*, *113*(19), 5352-5357.

Bollmer, J. L., Kimball, R. T., Whiteman, N. K., Sarasola, J. H., & Parker, P. G. (2006). Phylogeography of the Galápagos hawk (Buteo galapagoensis): a recent arrival to the Galápagos Islands. Molecular phylogenetics and evolution, 39(1), 237-247.

Bollmer, J. L., & Nims, B. D. (2018). Genetic diversity in endemic galápagos birds: patterns and implications. In *Disease Ecology* (pp. 83-111). Springer, Cham.

Bright, J. A., Marugán-Lobón, J., Cobb, S. N., & Rayfield, E. J. (2016). The shapes of bird beaks are highly controlled by nondietary factors. Proceedings of the National Academy of Sciences, 113(19), 5352-5357.

Briskie, J. V., & Mackintosh, M. (2004). Hatching failure increases with severity of population bottlenecks in birds. Proceedings of the National Academy of Sciences, 101(2), 558-561.

Brodie, A., Azaria, J. R., & Ofran, Y. (2016). How far from the SNP may the causative genes be?. *Nucleic acids research*, *44*(13), 6046-6054.

Buerkle, A.C., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go?. Molecular ecology, 22(11), 3028-3035.

Clegg, S. M., & Owens, P. F. (2002). The 'island rule'in birds: medium body size and its ecological explanation. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1498), 1359-1365.

Cornell, S. J., & Tregenza, T. (2007). A new theory for the evolution of polyandry as a means of inbreeding avoidance. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1627), 2873-2879.

Delmore, K. E., Toews, D. P., Germain, R. R., Owens, G. L., & Irwin, D. E. (2016). The genetics of seasonal migration and plumage color. Current Biology, 26(16), 2167-2173.

Duntsch, L., Tomotani, B. M., de Villemereuil, P., Brekke, P., Lee, K. D., Ewen, J. G., & Santure, A. W. (2020). Polygenic basis for adaptive morphological variation in a threatened Aotearoa| New Zealand bird, the hihi (Notiomystis cincta). *Proceedings of the Royal Society B*, *287*(1933), 20200948.

Emerson, B. C. (2002). Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular ecology*, *11*(6), 951-966.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology, 14(8), 2611-2620.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016).

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481-501.

Faaborg, J., & Patterson, C. B. (1981). The characteristics and occurrence of cooperative polyandry. *Ibis*, *123*(4), 477-484.

Faaborg, J., P.G. Parker, L. DeLay, Tj de Vries, J.C. Bednarz, S. Maria Paz, J. Naranjo, and T.A. White. (1995). Confirmation of the cooperative polyandry in the Galápagos hawk (Buteo galapagoensis). Behav Ecol Sociobiol. 36: 83-90.

Felice, R. N., Tobias, J. A., Pigot, A. L., & Goswami, A. (2019). Dietary niche and the evolution of cranial morphology in birds. *Proceedings of the Royal Society B*, *286*(1897), 20182677.

Francis, R. M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27-32. DOI: 10.1111/1755-0998.12509

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. Conservation biology, 10(6), 1500-1508.

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

Fumagalli, M., Vieira, F. G., Korneliussen, T. S., Linderoth, T., Huerta-Sánchez, E., Albrechtsen, A., & Nielsen, R. (2013). Quantifying population genetic differentiation from next-generation sequencing data. *Genetics*, *195*(3), 979-992.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in ecology & evolution*, *27*(9), 489-496.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., Ghalambor, C.K., MAldonado, J.E., Rick, T.C., Day, M.D., Polato, N.R., Fitzpatrick, S.W., Coonan, T.J., Crooks, K.R., Dillon, A., Garcelon, D.K., King, J.L., Boser, C.L., Gould, N. & Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular ecology*, *25*(10), 2176-2194.

Garrison E. Vcflib, a simple C++ library for parsing and manipulating VCF files. 2020. https://github. com/vcflib/vcflib.

Gaspar, J., Gibb, G. C., & Trewick, S. A. (2020). Convergent morphological responses to loss of flight in rails (Aves: Rallidae). Ecology and Evolution, 10(13), 6186-6207.
Gazda, M. A., Araújo, P. M., Lopes, R. J., Toomey, M. B., Andrade, P., Afonso, S., Marques, C., Nunes, L., Pereira, P., Trigo, S., Hill, G.E., Corbo, J. & Carneiro, M. (2020). A genetic mechanism for sexual dichromatism in birds. Science, 368(6496), 1270-1274.
Grant, P.R. (1998). Patterns on island and microevolution. In: Grant PR, ed. Evolution in Islands. Oxford: Oxford University Press.
Grant, P.R., Grant, B.R. (2008). How and why species multiply: the radiation of Darwin's finches. Princeton University Press, Princeton, NJ.
Grant, P.R., Grant, B.R. (2010). Conspecific versus heterospecific gene
Gu, J., Liang, Q., Liu, C., & Li, S. (2020). Genomic analyses reveal adaptation to hot arid and harsh environments in native chickens of China. *Frontiers in genetics*, *11*.
Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). Principles of population genetics (Vol. 116). Sunderland, MA: Sinauer associates.
Hinde, R. A. (1959). Behaviour and speciation in birds and lower vertebrates. Biological Reviews, 34(1), 85-127.
Hoeck, P. E., Bollmer, J. L., Parker, P. G., & Keller, L. F. (2010). Differentiation with drift: a spatio-temporal genetic analysis of Galápagos mockingbird populations (Mimus spp.). Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1543), 1127-1138.
Jønsson, K. A., Fabre, P. H., Fritz, S. A., Etienne, R. S., Ricklefs, R. E., Jørgensen, T. B., Fjeldsa, J., Rahbek, C., Ericson, G.P., Woog, F., Pasquet, R. & Irestedt, M. (2012). Ecological and evolutionary determinants for the adaptive radiation of the Madagascan vangas. *Proceedings of the National Academy of Sciences*, *109*(17), 6620-6625.
Kardos, M., Armstrong, E., Fitzpatrick, S., Hauser, S., Hedrick, P., Miller, J., Tallmon, D.A., and W.C. Funk (2021). The crucial role of genome-wide genetic variation in conservation. [preprint] available at doi: https://doi.org/10.1101/2021.07.05.451163
Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in ecology & evolution*, *17*(5), 230-241.
Kerlinger, P. (1985). Water-crossing behavior of raptors during migration. *The Wilson Bulletin*, *97*(1), 109-113.
Klingenberg, C. P., & Marugán-Lobón, J. (2013). Evolutionary covariation in geometric morphometric data: analyzing integration, modularity, and allometry in a phylogenetic context. *Systematic biology*, *62*(4), 591-610.
Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*, *15*(5), 1179-1191.
Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, *15*(1), 356.
Lack, D. L. (1947). Darwin's finches. Cambridge University Press, Cambridge.
Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., Kerje, S.,Gustafson, U., Shi, C.,Chen, W.,Liang, X., Huang, L.,Want, J., Liang, E., Wu, Q., Lee, S.M., Xu, X., Hoglund, J., Liu, X. & Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nature genetics, 48(1), 84-88.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2

Livezey, B. C. (1992). Morphological corollaries and ecological implications of flightlessness in the kakapo (Psittaciformes: Strigops habroptilus). Journal of morphology, 213(1), 105-145.

Lomolino, M.V. (1985) Body size of mammals on islands: the island rule re-examined. American Naturalist, 125, 310–316

Longmire, J. L., Albright, K. L., Lewis, A. K., Meincke, L. J., & Hildebrand, C. E. (1987). A rapid and simple method for the isolation of high molecular weight cellular and chromosome-specific DNA in solution without the use of organic solvents. *Nucleic acids research*, *15*(2), 859.

Lovette, I. J., Bermingham, E., & Ricklefs, R. E. (2002). Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1486), 37-42.

Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., & Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, *279*(5359), 2115-2118.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957-2963.

Makarenkova, H. P., & Meech, R. (2012). Barx homeobox family in muscle development and regeneration. *International review of cell and molecular biology*, *297*, 117-173.

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719-731.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L., Mushayamaha T., and Thomas, P.D. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API , Nucl. Acids Res. (2020) doi: 10.1093/nar/gkaa1106s.

Millien, V. (2006). Morphological evolution is accelerated among island mammals. *PLoS Biol*, *4*(10), e321.

Nunney, L. (1993). The influence of mating system and overlapping generations on effective population size. Evolution, 47(5), 1329-1341.

O'Grady, J. J., Brook, B. W., Reed, D. H., Ballou, J. D., Tonkyn, D. W., & Frankham, R. (2006). Realistic levels of inbreeding depression strongly affect extinction risk in wild populations. *Biological conservation*, *133*(1), 42-51.

Pespeni, M. H., Garfield, D. A., Manier, M. K., & Palumbi, S. R. (2012). Genome-wide polymorphisms show unexpected targets of natural selection. Proceedings of the Royal Society B: Biological Sciences, 279(1732), 1412-1420.

Petren, K., Grant, P. R., Grant, B. R., & Keller, L. F. (2005). Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Molecular Ecology, 14(10), 2943-2957.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics, 38(8), 904-909.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3217-223. doi:10.1111/j.2041-210X.2011.00169.x
Rivera-Parra, J. L., Levenstein, K. M., Bednarz, J. C., Vargas, F. H., Carrion, V., & Parker, P. G. (2012). Implications of goat eradication on the survivorship of the Galapagos hawk. The Journal of Wildlife Management, 76(6), 1197-1204.
Rowe, M., Whittington, E., Borziak, K., Ravinet, M., Eroukhmanoff, F., Sætre, G. P., & Dorus, S. (2020). Molecular diversification of the seminal fluid proteome in a recently diverged passerine species pair. *Molecular biology and evolution*, *37*(2), 488-506.]
Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory
Schluter, D. (2000). The ecology of adaptive radiation. OUP Oxford.
Sendell-Price, A. T., Ruegg, K. C., & Clegg, S. M. (2020). Rapid morphological divergence following a human-mediated introduction: The role of drift and directional selection. Heredity, 124(4), 535-549.
Silva, C. N. S., McFarlane, S. E., Hagen, I. J., Rönnegård, L., Billing, A. M., Kvalnes, T., ... & Husby, A. (2017). Insights into the genetic architecture of morphological traits in two passerine bird species. Heredity, 119(3), 197-205.
Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2012). Association testing for next-generation sequencing data using score statistics. *Genetic epidemiology*, *36*(5), 430-437.
Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693-702.
Slikas, B., Olson, S. L., & Fleischer, R. C. (2002). Rapid, independent evolution of flightlessness in four species of Pacific Island rails (Rallidae): an analysis based on mitochondrial sequence data. Journal of Avian Biology, 33(1), 5-14.
Stanley, S. M. (1975). A theory of evolution above the species level. Proceedings of the National Academy of Sciences, 72(2), 646-650.
Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular ecology resources, 17(2), 194-208.
Thomas, D. C., & Witte, J. S. (2002). Point: population stratification: a problem for case-control studies of candidate-gene associations?. Cancer Epidemiology and Prevention Biomarkers, 11(6), 505-512.
Tokita, M., Yano, W., James, H. F., & Abzhanov, A. (2017). Cranial shape evolution in adaptive radiations of birds: comparative morphometrics of Darwin's finches and Hawaiian honeycreepers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150481.
Toro, M. A., & Caballero, A. (2005). Characterization and conservation of genetic diversity in subdivided populations. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1459), 1367-1378.
Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., Stager, M., Betuel, A., Cheviron, Z.A., Warren, W.C., Gonser, R.A. & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. Current Biology, 26(3), 344-350.
Warkentin, I. G., Espie, R. H., Lieske, D. J., & James, P. C. (2016).Variation in selection pressure acting on body size by age and sex in a reverse sexual size dimorphic raptor. Ibis, 158(3), 656-669.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wright, N. A., & Steadman, D. W. (2012). Insular avian adaptations on two Neotropical continental islands. Journal of biogeography, 39(10), 1891-1899.

Wright, N. A., Steadman, D. W., & Witt, C. C. (2016). Predictable evolution toward flightlessness in volant island birds. *Proceedings of the National Academy of Sciences*, *113*(17), 4765-4770.

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, G. & Chen, S. (2012). FastUniq: a fast de novo duplicates removal tool for paired short reads. *PloS one*, *7*(12), e52249.

Zeh, J. A., & Zeh, D. W. (1997). The evolution of polyandry II: post–copulatory defenses against genetic incompatibility. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1378), 69-75.

FIGURES AND TABLES

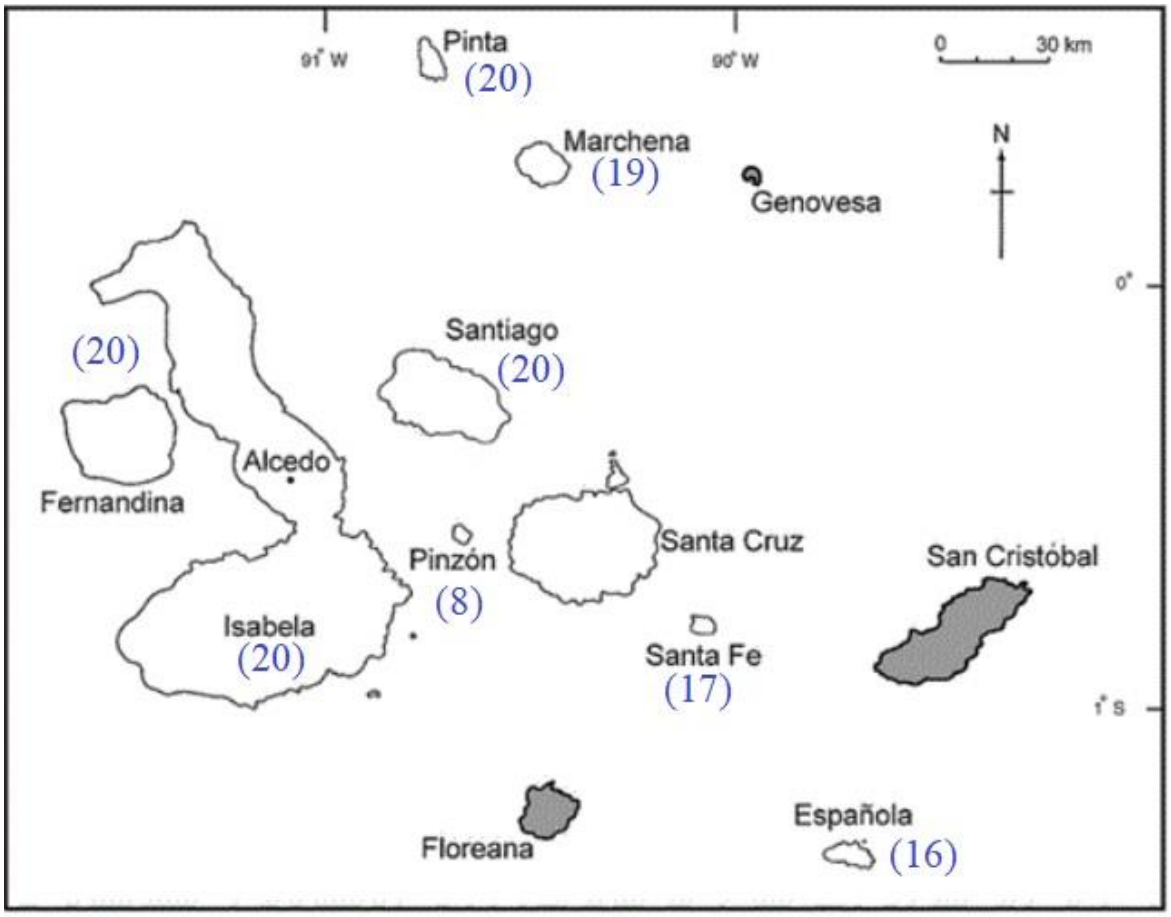


Figure 2.1. Map of the Galapagos archipelago adapted from Bollmer et al. 2003, with sample sizes in parentheses for the eight sampled islands.

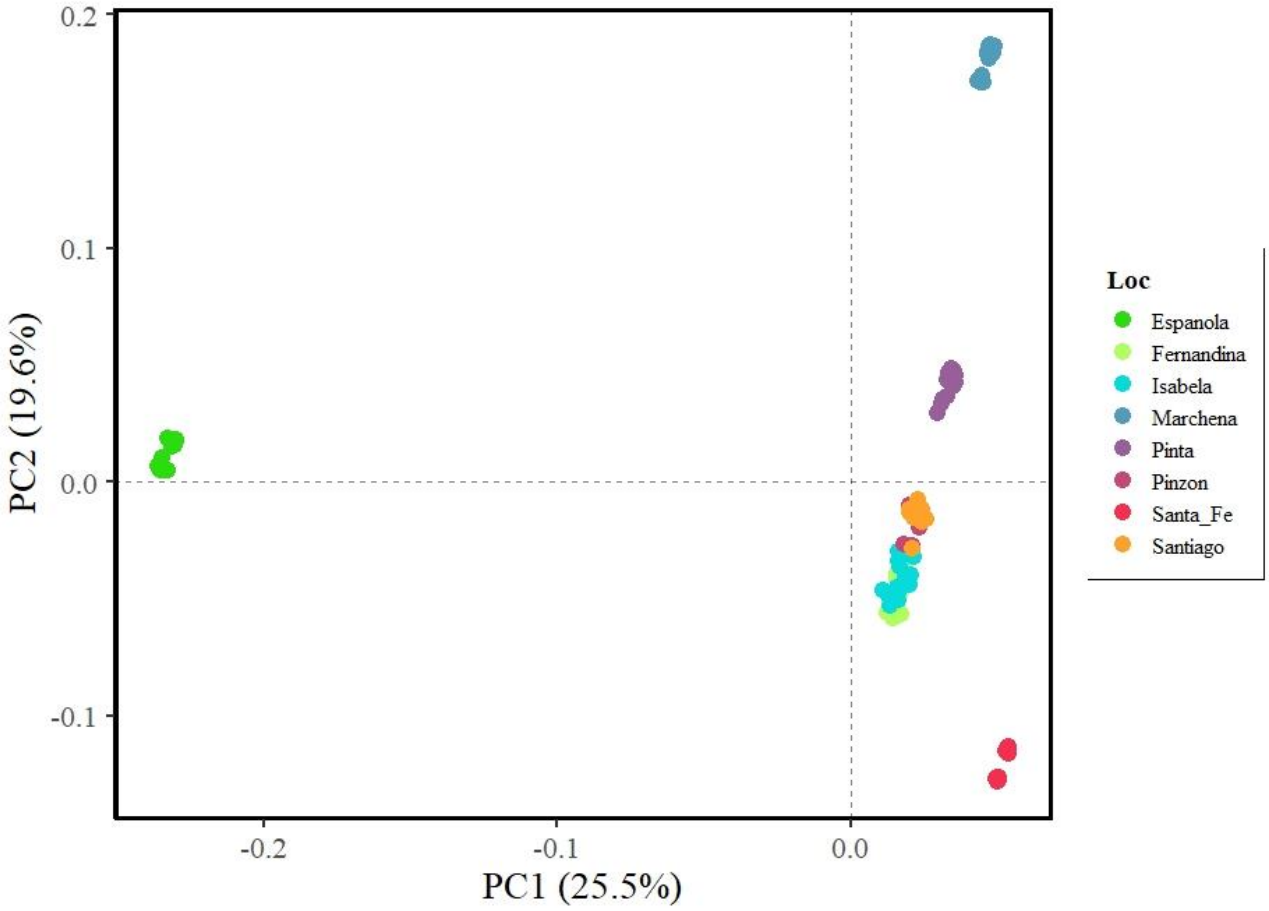


Figure 2.2. Principal components analysis of the SNP dataset for all 8 islands.

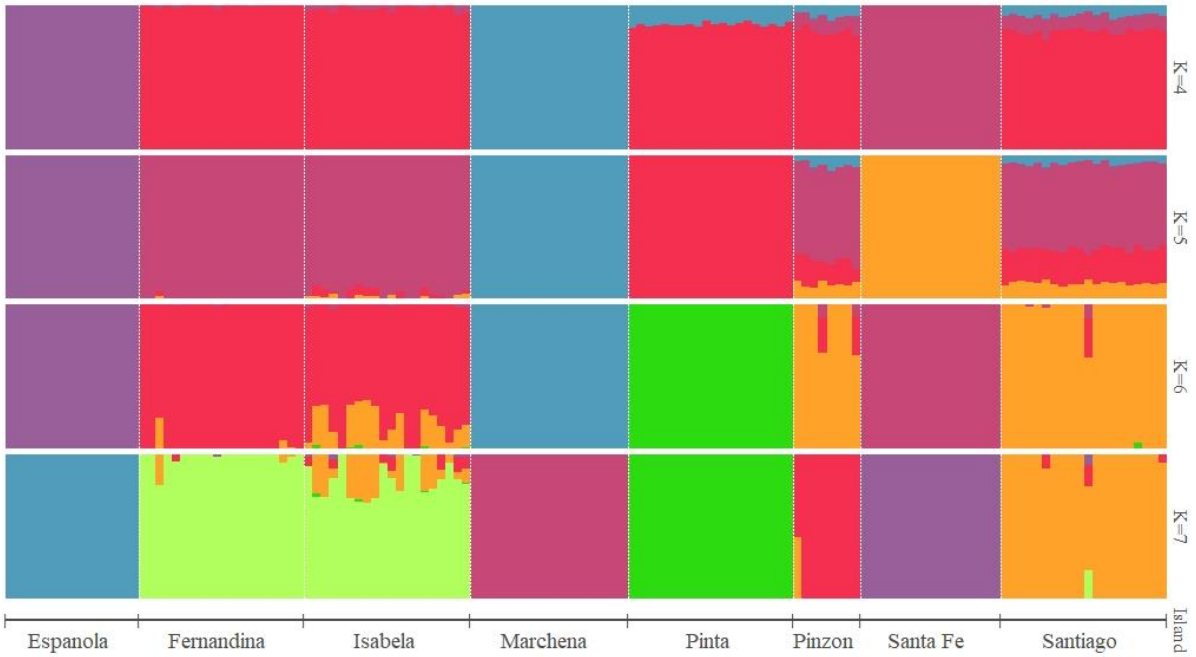


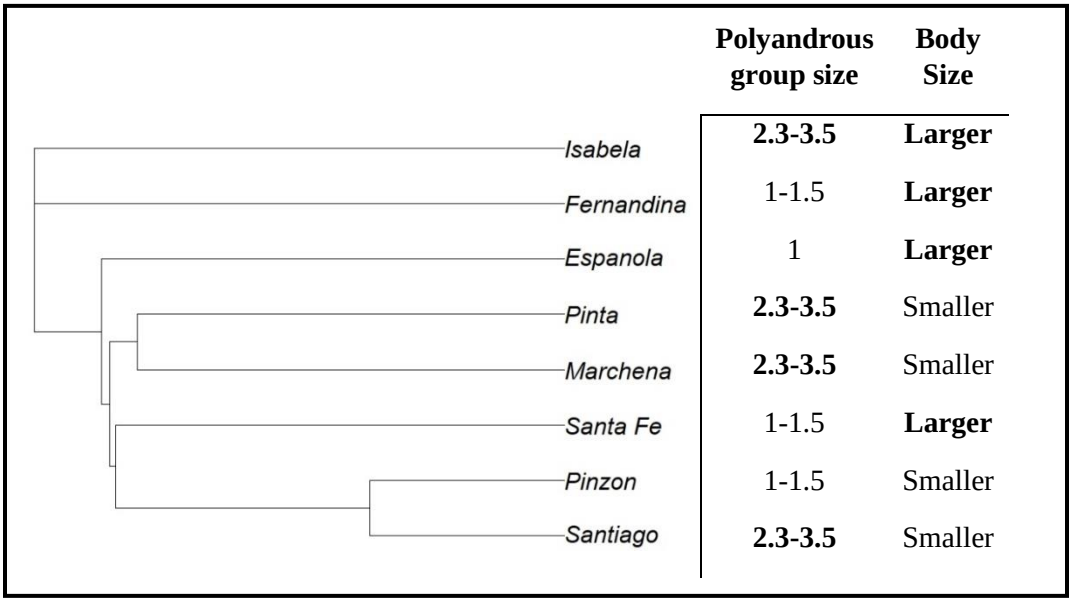
Figure 2.3. Admixture plots for all populations for K4 -K7

Table 2.1. Genome-wide F_{ST} for all pairwise Island comparisons

	Santa Fe	**Fernandina**	**Espanola**	**Isabela**	**Marchena**	**Pinzon**	**Pinta**
Santa Fe							
Fernandina	0.361689						
Espanola	0.594167	0.406239					
Isabela	0.331099	0.042094	0.382983				
Marchena	0.521719	0.356856	0.560531	0.328061			
Pinzon	0.413811	0.210918	0.477675	0.175155	0.397804		
Pinta	0.450334	0.261428	0.490743	0.230162	0.382281	0.296992	
Santiago	0.321188	0.1557	0.398364	0.120964	0.317247	0.120511	0.213008

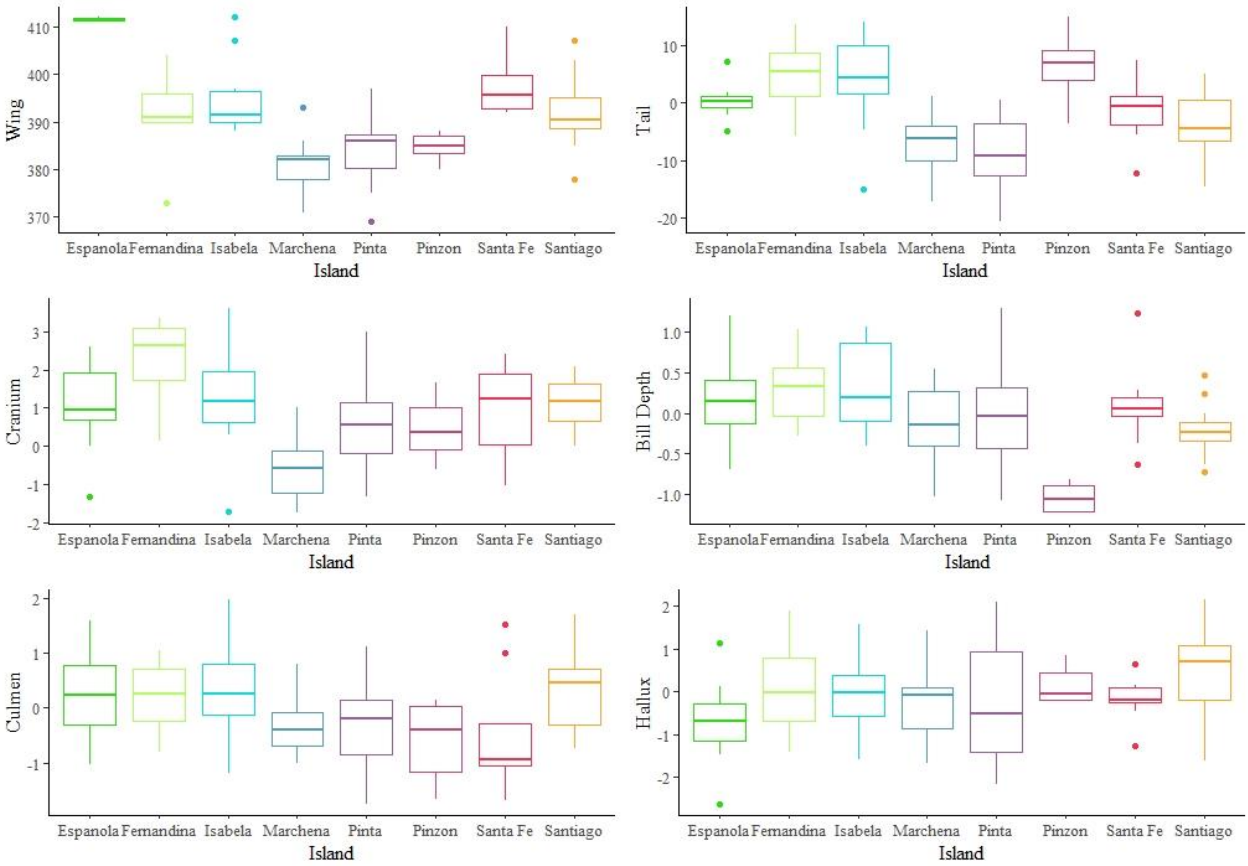
Table 2.2: Diversity and Island size estimates for all island populations

	Island Size	Watterson's θ	π	Avg. Heterozygosity
Santa Fe	24.8	0.000335857	0.000139362	0.1750412
Fernandina	647.6	0.000412272	0.000206703	0.26395
Espanola	61.1	0.000340097	0.000144597	0.190535
Isabela	4710.7	0.000444436	0.000223629	0.262835
Marchena	128.8	0.000360044	0.000152693	0.1809947
Pinzon	18.1	0.00025216	0.000157681	0.2368
Pinta	59.4	0.00042575	0.000185294	0.21081
Santiago	577.5	0.000463205	0.000219419	0.246315



	Polyandrous group size	Body Size
Isabela	**2.3-3.5**	**Larger**
Fernandina	1-1.5	**Larger**
Espanola	1	**Larger**
Pinta	**2.3-3.5**	Smaller
Marchena	**2.3-3.5**	Smaller
Santa Fe	1-1.5	**Larger**
Pinzon	1-1.5	Smaller
Santiago	**2.3-3.5**	Smaller

Figure 2.4. Phylogenetic relationships between the eight islands visualized in an unrooted neighbor-joining tree along with relevant life-history phenotypes.



	P-value
Tail	2.04×10^{-9} ***
Cranium	5.09×10^{-6} ***
Bill Depth	8.1×10^{-4} ***
Culmen	0.0172 *
Hallux	0.229
Wing	1.21×10^{-9} ***

Figure 2.5. Box plots showing differences in wing chord and size corrected morphometric features between islands.

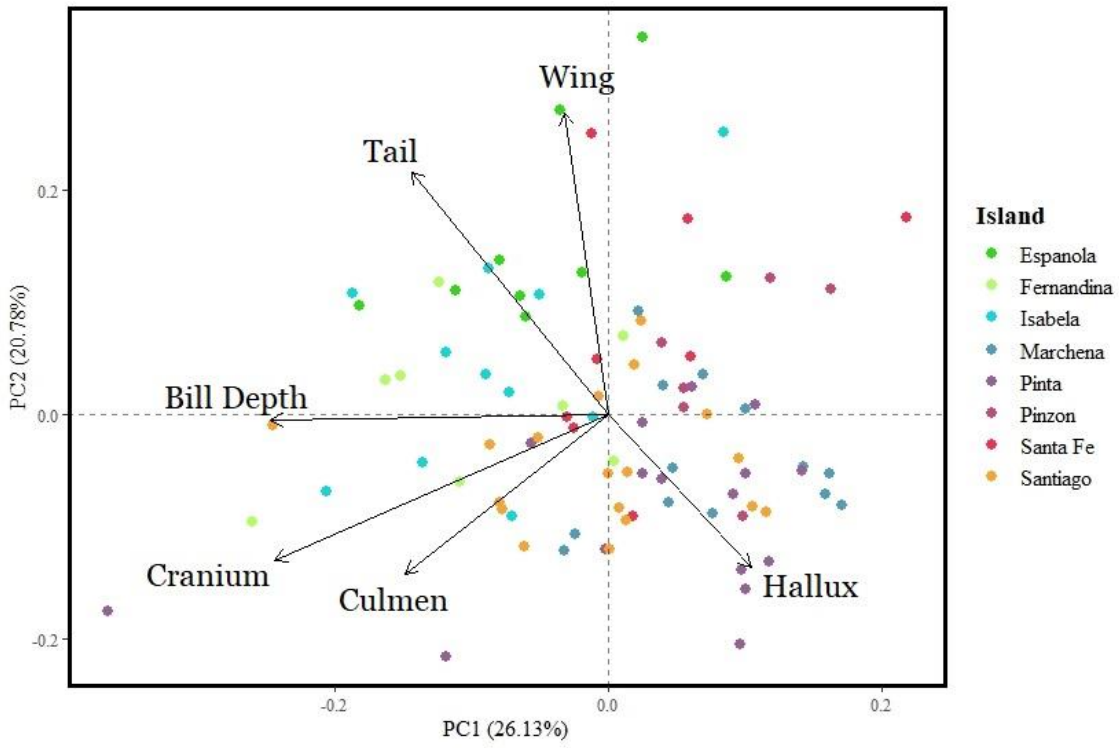


Figure 2.6. Principal Component Analysis of the six morphometric traits

Table 2.3. Morphological trait loadings for principal components 1 and 2.

	PC1	PC2
Wing	0.5775744	41.40218551
Hallux	6.1965015	10.56330296
Tail	11.8402199	26.86505792
Cranium	34.0008166	9.64569573
Culmen	2.6747863	11.51246062
Bill Depth	34.7101012	0.01129726

Table 2.4. Spearman's rank correlation rho's for the five morphometric traits and the first seven genomic principal components. Asterisks indicate a significant correlation with a p-value < 0.05.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**
Tail	0.578*****	0.463*****	-0.215*****	0.454*****	-0.102	-0.322*	-0.428*****
Cranium	-0.396*****	0.473*****	0.069	0.249*****	0.218*****	-0.329*	-0.13
Culmen	-0.268*****	0.074	0.164	0.226*****	0.025	0.222*	0.169
Bill Depth	-0.109	0.119	-0.025	0.018	0.374*****	-0.333*	-0.024
Hallux	-0.015	0.116	0.059	-0.147	-0.096	-0.051	-0.069

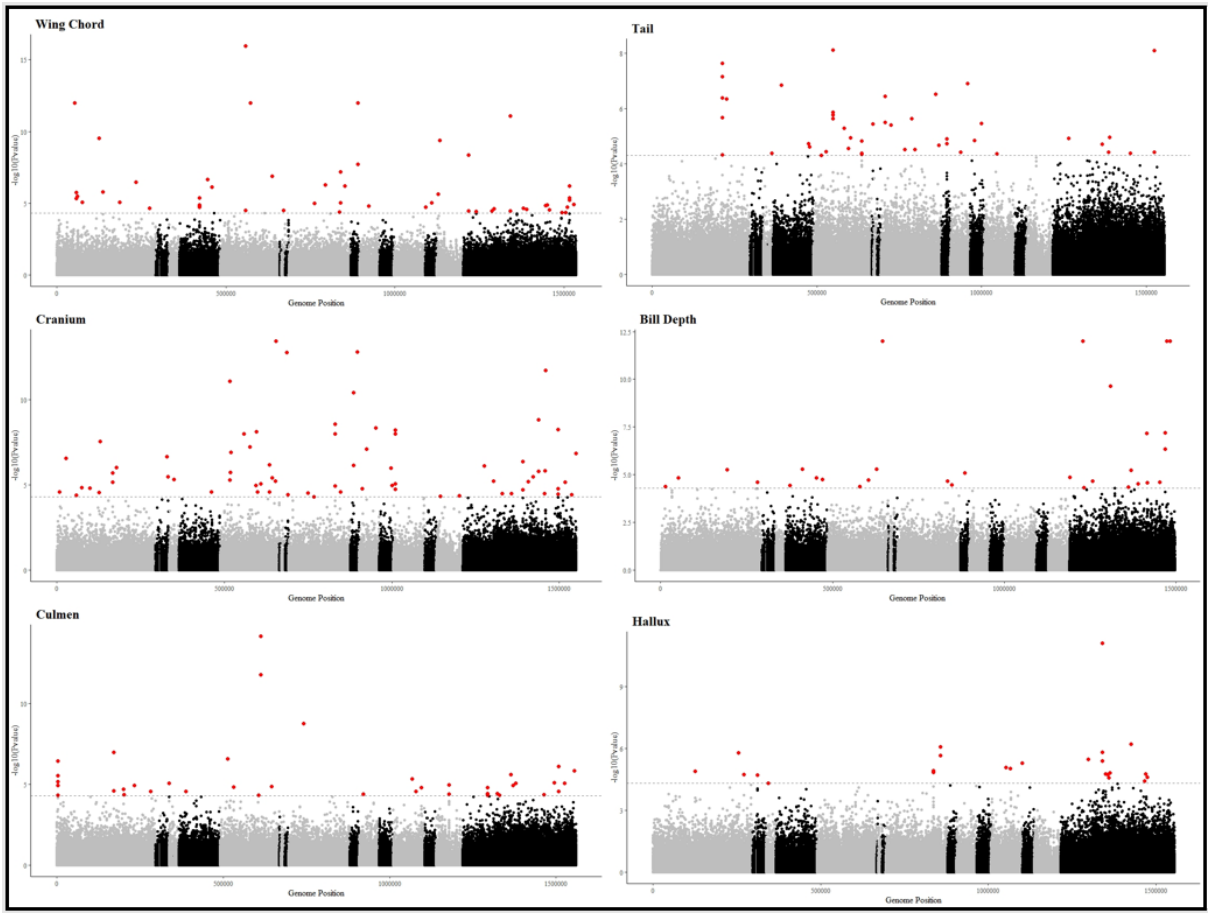


Figure 2.7. Manhattan plots of the six morphometric traits where the dashed horizontal line represents the significance threshold of 5×10^{-4} and the points in red are the significant, candidate loci.

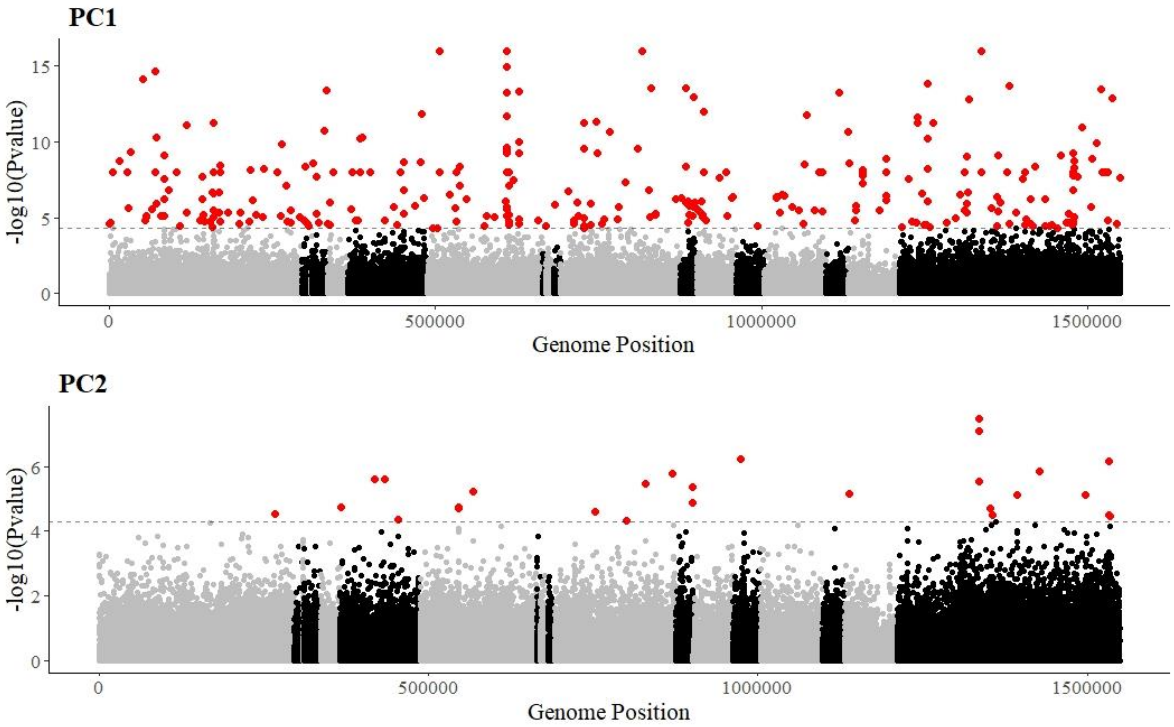


Figure 2.8. Manhattan plots of the first two morphometric PC's where the dashed horizontal line represents the significance threshold of 5 x 10^{-4} and the points in red are the significant, candidate loci.

Table 2.5. Summary of the results of the association analyses for the six morphometric traits of interest.

Morphometric Trait	Number of significant sites	Genes within 200kb of significant sites
Bill	31	ABCC4, ARHGEF33, BFSP1, CELF4, CHD3, ERG28, FLVCR2, IFT57, JDP2, KLC2, LAMB1, NEK9, NFS1, PCDH10, PCSK2, PLXNB1, POLR2A, PTK7, TGFB3, TMED10
Tail	45	ACP2, ADAMTS15, ADAMTS8, AMBRA1, APLP2, ARFGAP2, ARHGAP1, ASXL3, ATG13, CA2, DCLK3, DLGAP2, DLL1, EML6, GALNT10, GRIA1, GRM1, HARBI1, LRP4, MFAP2, MFAP3, NEUREXIN3, NRXN3, PACSIN3, PFKFB3, PRDM10, RAB33B, SAP30L, SPTBN1, ST14, TRIM67, ZBTB44
Culmen	42	BRI3BP, CDCP1, CDH7, CLEC3B, DUSP12, GALR2B, IRS2, MAPK8IP3, METTL16, MRPS34, NAA50, NCOR2, OTUB1, PLPP3, PRKG1, SCARB1, SIDT1, SLITRK5, SORCS3, SPICE1, TENT5C, USP42, UTP23
Cranium	71	ACTN1, ADCY9, ALG13, ANGPTL7, ANO2, ATXN1, BMP4, CADM3, CAPN3, CAPN6, CASZ1, CDCP2, CDIP1, CFAP20, DCAF5, DDHD1, DESI2, DIO1, DISP3, DUSP10, DUSP23, EIF4A2, ETNK2, EXD2, FAT4, FBXO2 FBXO6, GALNT9, GINS3, HMOX2, JPH2, KCND3, KCNK10, LDLRAD4, LNX2, LRRN2, MAD2L2, MASP2, MCM3, MGRN1, MOV10, MRPL37, MTARC2, MYBPC1, NDRG4, NEDD9, PAK3, PAQR8, PEX14, PFN3, PIK3C2B PLEKHA6, POU3F4, RASL11A, RIMS2, ROR2, SLC8A3, SMOC1, SOX13, SSBP3, SST2, TFAP2A, TMEM59, TRAP1, TRPC5, TSR3, TTC9, UBALD1, UBIAD1, WNT2B
Hallux	24	ARID5B, BCL11A, CARHSP1, DNM1, EIF3F, FZD4, KCTD12, KLF8, KYAT1, LRRC8A, MNX1, ODF2, PHYHD1, PTGES2, SORBS1, SPOUT1, SPTAN1, TMEM26, ZER1
Wing Chord	53	ANO2, CAPN2, CAPN8, DAAM1, EML6, GJA1, HTR1B, IMPG1, KCNA5, KIF6, L3HYPDH, MAN1C1, MTFR1L, PCNX2, RTN1A, RUNX3, SIX4, SIX6, SPTBN1, ST6GAL2, SYF2

Table 2.6. Statistically overrepresented GO groups for genes associated with Cranium size in the Galapagos Hawk

GO biological process complete	GO Term	Number of Genes	Expected Number of Genes	P-value	Corrected P-value
nuclear cell cycle DNA replication initiation	GO:1902315	3	0.03	7.27E-06	4.97E-02
cell cycle DNA replication initiation	GO:1902292	3	0.03	7.27E-06	3.31E-02
mitotic DNA replication initiation	GO:1902975	3	0.03	7.27E-06	2.48E-02
regulation of developmental process	GO:0050793	21	6.28	5.83E-07	7.98E-03

Table 2.7. Sites significantly associated with degree of polyandry in all pairwise comparisons

Chromosome	Scaffold	Position	Genes within 200kb
2	17	3361415	
2	23	17394076	CLPTM1L, CHMP5, PCBP3
3	214	18071529	BARX2, TMEM45B, APLP2, PRDM10

Table 2.8. List of statistically overrepresented GO categories related to biological process, for candidate genes associated with polyandry within 200kb of the overlapping significant sites between phylogenetically close islands.

GO biological process complete	GO Term	Number of Genes	Expected Number of Genes	P-value	Corrected P-value
negative regulation of cellular process	GO:0048523	61	27.89	6.92E-10	9.51E-06
multicellular organismal process	GO:0032501	69	34.23	7.08E-10	4.87E-06
negative regulation of biological process	GO:0048519	63	30.17	1.56E-09	7.14E-06
regulation of cellular process	GO:0050794	107	70.02	3.28E-09	1.13E-05
regulation of localization	GO:0032879	40	14.95	7.23E-09	1.99E-05
multicellular organism development	GO:0007275	55	25.33	1.07E-08	2.46E-05
anatomical structure development	GO:0048856	58	28.05	1.53E-08	3.01E-05
regulation of signaling	GO:0023051	46	19.4	1.64E-08	2.81E-05
developmental process	GO:0032502	61	30.25	1.91E-08	2.92E-05
biological regulation	GO:0065007	113	78.76	3.04E-08	4.18E-05
translational termination	GO:0006415	5	0.06	3.42E-08	4.27E-05
nervous system development	GO:0007399	34	12.06	3.80E-08	4.36E-05
cell-cell signaling	GO:0007267	21	4.96	3.96E-08	4.19E-05
regulation of cell communication	GO:0010646	45	19.31	5.86E-08	5.76E-05
regulation of biological process	GO:0050789	107	73.21	6.93E-08	6.35E-05
regulation of biological quality	GO:0065008	49	22.54	7.72E-08	6.64E-05

system development	GO:0048731	50	23.08	9.07E-08	7.34E-05
cellular process	GO:0009987	137	108.72	1.54E-07	1.17E-04
mitochondrial translational termination	GO:0070126	4	0.03	1.84E-07	1.34E-04
signaling	GO:0023052	58	29.96	2.00E-07	1.37E-04
localization	GO:0051179	63	34.44	3.15E-07	2.07E-04
mitochondrial translational elongation	GO:0070125	4	0.03	3.66E-07	2.29E-04
cell death	GO:0008219	17	3.78	3.75E-07	2.25E-04
cell communication	GO:0007154	58	30.57	4.36E-07	2.50E-04
apoptotic process	GO:0006915	16	3.43	5.32E-07	2.93E-04
regulation of developmental growth	GO:0048638	12	1.85	5.83E-07	3.09E-04
negative regulation of signaling	GO:0023057	24	7.63	8.20E-07	4.18E-04
programmed cell death	GO:0012501	16	3.62	1.06E-06	5.22E-04
regulation of cell growth	GO:0001558	12	1.99	1.21E-06	5.74E-04
regulation of transport	GO:0051049	26	8.97	1.22E-06	5.59E-04
protein-containing complex disassembly	GO:0032984	8	0.75	1.55E-06	6.87E-04
regulation of response to stimulus	GO:0048583	46	22.42	1.60E-06	6.89E-04
regulation of growth	GO:0040008	15	3.3	1.67E-06	6.95E-04
negative regulation of multicellular organismal process	GO:0051241	18	4.83	2.33E-06	9.42E-04
negative regulation of cell communication	GO:0010648	23	7.61	2.77E-06	1.09E-03
modulation of chemical synaptic transmission	GO:0050804	12	2.21	3.39E-06	1.30E-03
regulation of trans-synaptic signaling	GO:0099177	12	2.21	3.39E-06	1.26E-03
cellular response to stimulus	GO:0051716	65	38.42	3.69E-06	1.34E-03
cellular protein complex disassembly	GO:0043624	6	0.38	4.40E-06	1.55E-03
negative regulation of response to stimulus	GO:0048585	25	9.12	5.25E-06	1.81E-03

negative regulation of developmental growth	GO:0048640	7	0.64	5.91E-06	1.98E-03
animal organ development	GO:0048513	36	16.43	5.99E-06	1.96E-03
biological_process	GO:0008150	147	127.01	6.11E-06	1.91E-03
negative regulation of metabolic process	GO:0009892	37	17.08	7.35E-06	2.25E-03
regulation of ion transport	GO:0043269	14	3.27	7.38E-06	2.21E-03
cellular developmental process	GO:0048869	40	19.52	8.66E-06	2.54E-03
protein-containing complex subunit organization	GO:0043933	23	8.25	1.00E-05	2.87E-03
regulation of synaptic transmission, glutamatergic	GO:0051966	5	0.25	1.03E-05	2.89E-03
response to stimulus	GO:0050896	72	45.88	1.27E-05	3.49E-03
negative regulation of nervous system development	GO:0051961	7	0.72	1.29E-05	3.49E-03
negative regulation of signal transduction	GO:0009968	21	7.2	1.33E-05	3.51E-03
negative regulation of cellular metabolic process	GO:0031324	33	14.84	1.34E-05	3.48E-03
cell differentiation	GO:0030154	39	19.17	1.36E-05	3.46E-03
regulation of signal transduction	GO:0009966	36	17.03	1.46E-05	3.65E-03
regulation of multicellular organismal process	GO:0051239	32	14.09	1.53E-05	3.70E-03
positive regulation of biological process	GO:0048518	61	36.45	1.53E-05	3.76E-03
positive regulation of cellular process	GO:0048522	57	33.53	2.08E-05	4.94E-03
IRE1-mediated unfolded protein response	GO:0036498	3	0.03	2.16E-05	5.04E-03
signal transduction	GO:0007165	50	27.93	2.19E-05	5.02E-03
secretion	GO:0046903	12	2.68	2.20E-05	4.96E-03
cellular component disassembly	GO:0022411	9	1.46	2.27E-05	5.05E-03
negative regulation of macromolecule metabolic process	GO:0010605	34	15.94	2.60E-05	5.67E-03
skin epidermis development	GO:0098773	3	0.04	3.44E-05	7.39E-03

gland development	GO:0048732	9	1.56	3.82E-05	8.09E-03
regulation of dopamine secretion	GO:0014059	4	0.17	4.19E-05	8.73E-03
regulation of secretion by cell	GO:1903530	12	2.9	4.59E-05	9.43E-03
tissue development	GO:0009888	23	9.18	5.21E-05	1.05E-02
cell surface receptor signaling pathway	GO:0007166	27	11.85	5.80E-05	1.16E-02
response to chemical	GO:0042221	39	20.15	5.85E-05	1.15E-02
negative regulation of nitrogen compound metabolic process	GO:0051172	30	13.97	6.35E-05	1.23E-02
anatomical structure morphogenesis	GO:0009653	29	13.2	6.52E-05	1.24E-02
negative regulation of locomotion	GO:0040013	9	1.69	6.94E-05	1.31E-02
regulation of body fluid levels	GO:0050878	8	1.33	7.67E-05	1.43E-02
negative regulation of axonogenesis	GO:0050771	5	0.4	7.72E-05	1.42E-02
aging	GO:0007568	6	0.66	7.73E-05	1.40E-02
positive regulation of growth	GO:0045927	8	1.33	8.01E-05	1.43E-02
negative regulation of cell growth	GO:0030308	7	0.99	8.29E-05	1.46E-02
regulation of secretion	GO:0051046	12	3.1	8.75E-05	1.52E-02
response to organic substance	GO:0010033	29	13.54	9.06E-05	1.56E-02
platelet degranulation	GO:0002576	3	0.07	9.94E-05	1.69E-02
negative regulation of axon extension involved in axon guidance	GO:0048843	4	0.22	1.08E-04	1.79E-02
regulation of catecholamine secretion	GO:0050433	4	0.22	1.08E-04	1.77E-02
negative regulation of neurogenesis	GO:0050768	6	0.71	1.08E-04	1.80E-02
adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	GO:0007193	5	0.44	1.11E-04	1.80E-02
skin development	GO:0043588	6	0.72	1.15E-04	1.83E-02
secretion by cell	GO:0032940	10	2.25	1.16E-04	1.83E-02
regeneration	GO:0031099	5	0.44	1.21E-04	1.87E-02

response to stress	GO:0006950	33	16.54	1.21E-04	1.89E-02
animal organ regeneration	GO:0031100	3	0.08	1.32E-04	2.01E-02
endothelium development	GO:0003158	5	0.47	1.56E-04	2.33E-02
regulated exocytosis	GO:0045055	6	0.76	1.56E-04	2.35E-02
regulation of axon extension involved in axon guidance	GO:0048841	4	0.24	1.60E-04	2.37E-02
cellular amide metabolic process	GO:0043603	15	5	1.81E-04	2.65E-02
regulation of anatomical structure morphogenesis	GO:0022603	15	5	1.81E-04	2.63E-02
neural crest cell migration	GO:0001755	5	0.49	1.82E-04	2.61E-02
positive regulation of metabolic process	GO:0009893	40	22.31	2.05E-04	2.90E-02
transport	GO:0006810	44	25.59	2.06E-04	2.89E-02
regulation of cellular metabolic process	GO:0031323	58	37.07	2.06E-04	2.86E-02
cellular response to chemical stimulus	GO:0070887	29	14.12	2.11E-04	2.90E-02
negative regulation of adenylate cyclase activity	GO:0007194	3	0.1	2.15E-04	2.93E-02
regulation of transmembrane transport	GO:0034762	11	2.92	2.16E-04	2.91E-02
regulation of neurogenesis	GO:0050767	9	1.98	2.17E-04	2.90E-02
regulation of cellular catabolic process	GO:0031329	14	4.52	2.18E-04	2.89E-02
translational elongation	GO:0006414	4	0.27	2.28E-04	2.99E-02
stem cell differentiation	GO:0048863	7	1.18	2.37E-04	3.07E-02
cellular response to stress	GO:0033554	22	9.34	2.39E-04	3.07E-02
regulation of metabolic process	GO:0019222	62	40.97	2.45E-04	3.12E-02
regulation of molecular function	GO:0065009	35	18.44	2.50E-04	3.16E-02
establishment of localization	GO:0051234	45	26.64	2.59E-04	3.24E-02
peptide metabolic process	GO:0006518	12	3.51	2.61E-04	3.24E-02
regulation of nervous system development	GO:0051960	10	2.5	2.66E-04	3.26E-02
negative regulation of cyclase activity	GO:0031280	3	0.1	2.67E-04	3.25E-02

ameboidal-type cell migration	GO:0001667	7	1.21	2.80E-04	3.38E-02
regulation of proteolysis	GO:0030162	13	4.12	3.09E-04	3.69E-02
export from cell	GO:0140352	10	2.56	3.11E-04	3.69E-02
semaphorin-plexin signaling pathway	GO:0071526	4	0.31	3.49E-04	4.11E-02
negative regulation of axon extension	GO:0030517	4	0.31	3.49E-04	4.07E-02
Notch signaling pathway	GO:0007219	5	0.58	3.73E-04	4.31E-02
negative regulation of lyase activity	GO:0051350	3	0.12	3.94E-04	4.52E-02
exocytosis	GO:0006887	7	1.29	4.03E-04	4.58E-02
negative regulation of cell development	GO:0010721	6	0.92	4.07E-04	4.59E-02
glutamate receptor signaling pathway	GO:0007215	4	0.32	4.24E-04	4.71E-02
response to unfolded protein	GO:0006986	5	0.59	4.24E-04	4.74E-02
negative regulation of growth	GO:0045926	7	1.31	4.35E-04	4.79E-02
regulation of cell growth involved in cardiac muscle cell development	GO:0061050	2	0.02	4.41E-04	4.81E-02
positive regulation of developmental growth	GO:0048639	6	0.93	4.48E-04	4.85E-02
response to oxygen-containing compound	GO:1901700	17	6.69	4.53E-04	4.86E-02

Table 2.9. List of statistically overrepresented GO categories related to molecular function, for candidate genes associated with polyandry within 200kb of the overlapping significant sites between phylogenetically close islands

GO molecular function complete	GO Term	Number of Genes	Expected Number of Genes	P-value	Corrected P-value
Notch binding	GO:0005112	4	0.17	4.99E-05	4.13E-02
semaphorin receptor binding	GO:0030215	4	0.21	9.36E-05	4.31E-02
delayed rectifier potassium channel activity	GO:0005251	4	0.24	1.41E-04	4.86E-02
voltage-gated potassium channel activity	GO:0005249	6	0.68	8.85E-05	5.23E-02
potassium channel activity	GO:0005267	7	0.92	5.66E-05	3.91E-02
voltage-gated ion channel activity	GO:0005244	8	1.36	9.12E-05	4.72E-02
voltage-gated channel activity	GO:0022832	8	1.37	9.51E-05	3.94E-02
protein binding	GO:0005515	130	45.05	8.72E-45	3.61E-41
binding	GO:0005488	144	86.84	8.99E-24	1.86E-20
ion binding	GO:0043167	60	38.16	1.14E-04	4.29E-02
molecular function	GO:0003674	153	116.68	1.41E-15	1.95E-12

Table S2.1. List of genes significantly associated with morphometric PC1

Gene	
TENT5C	Terminal nucleotidyltransferase 5C
LGI3	LRRCT domain-containing protein
SPICE1	Coiled-coil domain-containing protein 52
FSTL4	Uncharacterized protein
SCHIP1	SCHIP-1 domain-containing protein
EPHA3	Ephrin type-A receptor 3
NEIL1	FPG_CAT domain-containing protein
ZHX3	Uncharacterized protein
BMP2K	Protein kinase domain-containing protein
RASD2	Uncharacterized protein
XYLT2	Protein xylosyltransferase
PBDC1	Polysacc_synt_4 domain-containing protein
TRPV6	ANK_REP_REGION domain-containing protein
IQSEC3	SEC7 domain-containing protein
MTCH1	Uncharacterized protein
STXBP5	Uncharacterized protein
NAA50	N-acetyltransferase domain-containing protein
METTL16	RNA N6-adenosine-methyltransferase METTL16
ABCA5	Uncharacterized protein
ATP12A	Sodium/potassium-transporting ATPase subunit alpha
SORCS3	PKD domain-containing protein
PI16	SCP domain-containing protein
TRIB1	Protein kinase domain-containing protein
SCO1	Uncharacterized protein
FKBP14	Peptidylprolyl isomerase
GRIN2C	Uncharacterized protein
KCNJ12	ATP-sensitive inward rectifier potassium channel 12
MAPK13	Mitogen-activated protein kinase
PLK2	Serine/threonine-protein kinase PLK
CDKN1B	Cyclin-dependent kinase inhibitor 1B
SIDT1	Uncharacterized protein
CCDC78	DUF4472 domain-containing protein
SIGMAR1	Sigma non-opioid intracellular receptor 1
NR3C2	Mineralocorticoid receptor
DUSP12	Protein-tyrosine-phosphatase
ASS1	Argininosuccinate synthase
EPHB5	Ephrin type-B receptor 5
GPR39	G_PROTEIN_RECEP_F1_2 domain-containing protein
CPLX3	Uncharacterized protein

CDH7	Cadherin-7
COL1A2	Fibrillar collagen NC1 domain-containing protein
CLEC3B	C-type lectin domain-containing protein
NME3	Nucleoside diphosphate kinase
SNAP29	t-SNARE coiled-coil homology domain-containing protein
CASQ2	Calsequestrin-2
ARID3A	Uncharacterized protein
RHOT2	Mitochondrial Rho GTPase
EEA1	Uncharacterized protein
NAT9	N-acetyltransferase domain-containing protein
NCS1	Neuronal calcium sensor 1
HTR2A	5-hydroxytryptamine receptor 2A
GAS7	Uncharacterized protein
PIM1	Serine/threonine-protein kinase
PDE6H	Rhodopsin-sensitive cGMP 3',5'-cyclic phosphodiesterase subunit gamma
SH3RF3	RING-type E3 ubiquitin transferase
ULK3	Serine/threonine-protein kinase ULK3
ARMC7	Uncharacterized protein
REEP4	Receptor expression-enhancing protein
VANGL1	Vang-like protein
XPO7	Importin N-terminal domain-containing protein
WASHC1	WASH complex subunit 1
B3GAT1	Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase
SIN3A	HDAC_interact domain-containing protein
ATXN1L	AXH domain-containing protein
SEMA3D	Semaphorin-3D
PRDM1	PR/SET domain 1
KPNA2	Importin subunit alpha
CBLN2	C1q domain-containing protein
CELF2	CUGBP Elav-like family member 2
GNA13	Uncharacterized protein
MID1	E3 ubiquitin-protein ligase Midline-1
PGRMC2	Cytochrome b5 heme-binding domain-containing protein
CDCP1	Uncharacterized protein
COMMD4	COMM domain-containing protein
MAPK8IP3	Uncharacterized protein
ZMAT3	Uncharacterized protein
SMARCB1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily B member 1
USP20	Ubiquitin carboxyl-terminal hydrolase
B4GALT3	Uncharacterized protein
WBP11	Uncharacterized protein
CCNA2	Cyclin-A2
TEN1	Teneurin-1

MRPL58	RF_PROK_I domain-containing protein
POU4F2	POU domain protein
SH3PXD2B	Uncharacterized protein
TEN1	Uncharacterized protein
GRAMD1B	VASt domain-containing protein
METTL16	U6 small nuclear RNA (adenine-(43)-N(6))-methyltransferase
BHLHE22	Class E basic helix-loop-helix protein 22
CDH7	Cadherin-7
TOP1	DNA topoisomerase I
CDR2L	Uncharacterized protein
DHX38	RNA helicase
APOLD1	Uncharacterized protein
SPSB3	SPRY domain-containing SOCS box protein 3
NFATC2	RHD domain-containing protein
MIS18A	Mis18 domain-containing protein
TEAD3	Transcriptional enhancer factor TEF-5
IGDCC3	Uncharacterized protein
GDAP1	Uncharacterized protein
SOX11	Transcription factor SOX-11
NTMT1	Uncharacterized protein
AMOTL1	Angiomotin_C domain-containing protein
NEURL1B	Uncharacterized protein
LHFPL5	LHFPL tetraspan subfamily member 5 protein
VSNL1	Visinin-like protein 1
SLC6A1	Transporter
HID1	Uncharacterized protein
GPR83	G_PROTEIN_RECEP_F1_2 domain-containing protein
JADE1	Uncharacterized protein
USH1G	ANK_REP_REGION domain-containing protein
NR2F2	COUP transcription factor 2
TOR1A	Torsin
EIF4A2	RNA helicase
SNX33	Sorting nexin
KSR2	Uncharacterized protein
KCNS3	BTB domain-containing protein
PLCG1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma
RASL12	Uncharacterized protein
CACNA1B	Voltage-dependent N-type calcium channel subunit alpha-1B
CELF2	CUGBP Elav-like family member 2
DUSP1	Dual specificity protein phosphatase
NUBP2	Cytosolic Fe-S cluster assembly factor NUBP2
PAF1	RNA polymerase II-associated factor 1 homolog
PARP16	Poly [ADP-ribose] polymerase
SRP68	Signal recognition particle subunit SRP68

CPSF4	Cleavage and polyadenylation specificity factor subunit 4
NRBP1	Protein kinase domain-containing protein
ZFHX4	Zinc finger homeobox protein 4
BHLHE22	Class B basic helix-loop-helix protein 5
MYH1B	Myosin-1B
BICD1	Uncharacterized protein
CNTN4	Uncharacterized protein
RHOT2	Mitochondrial Rho GTPase 2
WIPI1	Uncharacterized protein
PDCD7	Uncharacterized protein
WNT7B	Protein Wnt-7b
MAP2K6	Protein kinase domain-containing protein
FIGNL1	AAA domain-containing protein
DERL2	Derlin
ASB6	Uncharacterized protein
PTPN9	Uncharacterized protein
TMEM39A	Uncharacterized protein
NUP85	Nuclear pore complex protein
MMP11	ZnMc domain-containing protein
PHLPP2	PPM-type phosphatase domain-containing protein
ASS1	Argininosuccinate synthase
MCM5	DNA helicase
DDX5	DEAD box protein 5
ARHGEF33	DH domain-containing protein
ATP9A	Phospholipid-transporting ATPase
RTN4IP1	PKS_ER domain-containing protein
CASQ2	Calsequestrin
FBXL16	Uncharacterized protein
HPS3	Uncharacterized protein
AMDHD2	N-acetylglucosamine-6-phosphate deacetylase
B4GALT3	Uncharacterized protein
AQP3	Aquaglyceroporin-3
DCTN3	Uncharacterized protein
EPAS1	Endothelial PAS domain protein 1
DCTN2	Dynactin subunit 2
EXOC7	Exocyst complex component 7
GLUD1	Glutamate dehydrogenase 1, mitochondrial
BRPF3	Uncharacterized protein
BUD31	Protein BUD31 homolog
XPO7	Exportin-7
PCNX2	Pecanex_C domain-containing protein
CCNA2	Cyclin-A2
AP1G1	AP-1 complex subunit gamma
NHSL1	Uncharacterized protein

KCTD2	BTB domain-containing protein
PRRX2	Paired mesoderm homeobox protein 2
ARPC1B	Actin-related protein 2/3 complex subunit
NOL6	Nucleolar protein 6
FAM20A	Fam20C domain-containing protein
NOS1	Nitric oxide synthase
CDK3	Protein kinase domain-containing protein
DDX11	Helicase ATP-binding domain-containing protein
MAPK14	Mitogen-activated protein kinase
FADS6	FA_desaturase domain-containing protein
RGS9	Uncharacterized protein
STK10	Serine/threonine-protein kinase 10
COL1A2	Collagen alpha-2(I) chain
EHMT1	Uncharacterized protein
MSGN1	Mesogenin-1
MAP2K3	Protein kinase domain-containing protein
MAN2C1	Alpha-mannosidase
MED29	Intersex-like protein
TEAD3	Transcriptional enhancer factor TEF-5
EPHA3	Receptor protein-tyrosine kinase
EIF4A2	Eukaryotic initiation factor 4A-II
CHD1L	Uncharacterized protein
FNDC4	Fibronectin type-III domain-containing protein
GRM1	G_PROTEIN_RECEP_F3_4 domain-containing protein
OTOP2	Uncharacterized protein
UBTD2	Ubiquitin-like domain-containing protein
AGBL3	
ARHGAP31	
ATP5MF	
BMP1	
CHCHD2	
CHD8	
CHST4	
FGD2	
FKBP5	
GALR2A	
IGH1A	
IRX1A	
KARS1	
KBTBD13	
KCNK1	
MYH13	
MYH16	
NT5C	

NUDT6
PITPNC1
PIWIL2
PPP3CC
RNF40
RYR1
SCAMP5
SHISA9
TOR1B
TSSK2

Table S2.2. List of genes significantly associated with morphometric PC2

Gene	
RAD23A	UV excision repair protein RAD23
PKD2	Uncharacterized protein
PTK2B	Non-specific protein-tyrosine kinase
CC2D1A	Uncharacterized protein
NCOA1	Nuclear receptor coactivator
HS3ST5	Sulfotransferase
DNAJC27	DnaJ homolog subfamily C member 27
MICAL2	F-actin monooxygenase
WDR83OS	Protein Asterix
MYO6	Unconventional myosin-6
ADCY3	Adenylate cyclase type 3
CACNA1A	Voltage-dependent P/Q-type calcium channel subunit alpha-1A
CHRNA2	Neuronal acetylcholine receptor subunit alpha-2
MAGI1	Uncharacterized protein
CCDC25	Coiled-coil domain-containing protein 25
PTRHD1	Aminoacyl-tRNA hydrolase
IMPG1	Interphotoreceptor matrix proteoglycan 1
ABCG2	ATP-binding cassette sub-family G member 2
EFR3B	Uncharacterized protein
ESCO2	Uncharacterized protein
SCARA5	Scavenger receptor class A member 5
KCNK3	Potassium channel subfamily K member
LRIG1	Uncharacterized protein
CNN1	Calponin-1
NUDT9	Nudix hydrolase domain-containing protein
ASXL3	Uncharacterized protein
ELOF1	Transcription elongation factor 1 homolog
MYO6	Unconventional myosin-VI
CHRNA2	Neuronal acetylcholine receptor subunit alpha-2

BEST2
COA8
ELAVL3
GET3
MAN2B1
MAST1
NACC1
SCARA3
STMN4
TRMT1
WDR83

CHAPTER 3:

Characterizing genomic patterns of founder speciation in a recent arrival to the Galapagos, the Galapagos Hawk (*Buteo galapagoensis*)

ABSTRACT

Although allopatric speciation has formed the basis of foundational speciation theory, the specific role of founder events in allopatric speciation in nature is largely unknown. Founder events may be important drivers of speciation as they can cause a large shift in allele frequencies due to strong genetic drift. The Galapagos hawk (*Buteo galapagoensis*) is a recent arrival to the Galapagos archipelago, having diverged from its most recent common ancestor, the Swainson's hawk (*Buteo swainsoni*) just over 100,000 years ago. The Galapagos hawk is an ideal system to study allopatric founder speciation as it is thought to have diverged from a single founding event without any secondary contact. Additionally, island ecosystems like the Galapagos archipelago provide an excellent opportunity to study speciation in nature as they often promote ecological adaptation within discrete geographic boundaries. Here, we produce a high-quality genome assembly for the Swainson's hawk and characterize the genomic patterns of divergence underlying the speciation of the Galapagos hawk by resequencing individuals of both species across their distributions. We find that the genomic patterns of divergence between the Swainson's hawk and the Galapagos hawk are consistent with strong genetic drift following founder events, with no evidence of secondary contact. We also identify large genomic 'islands' of selection that may hold the key to understanding the rapid divergence of the two species. This study represents an important step towards an understanding of the complex interactions between evolutionary processes that shape allopatric speciation at the genomic level.

INTRODUCTION

One of the main goals of evolutionary biology is to understand the patterns and processes that contribute to speciation. Recently, there has been considerable effort to document patterns of genome-wide divergence in closely related species, as it has become widely accepted that different genomic regions may convey varying information about the process of speciation (Nosil and Feder 2012). Although much has been learned about the speciation process in model organisms, little is known about the identity, effect size, or genomic distribution of loci involved in speciation in natural populations. Advancements in sequencing technology now allow for the low-cost creation of high-quality reference genomes which enable a more accurate characterization of divergent genomic architecture in non-model organisms.

Most commonly, scientists characterize the genomic underpinnings of speciation through the identification of genomic 'islands of divergence' (Feder et al. 2012) and barrier loci, specific loci that are known to contribute to reproductive isolation (Nosil and Schluter 2011). The specific type of speciation is thought to have a strong impact on the genomic architecture of divergence as sympatric and allopatric speciation models are known to evolve reproductive barriers at different rates (Seehausen et al. 2014). Most studies on speciation in avian systems involve taxa that are either currently are or have historically hybridized or taxa in which reproductive barriers are not yet complete (Ellegren 2012, Poelstra et al. 2014, Parchman 2013). While these studies are very informative, to truly understand the processes shaping genomes of closely related organisms, there needs to be a comparison of genomic patterns between sympatric and allopatric species (Feder et al. 2012, Cruicksank and Hahn 2014).

Founder speciation also called peripatric speciation, was first proposed by Ernst Mayr in 1954. Founder speciation a type of allopatric speciation in which a new species is formed from an isolated, smaller, peripheral population (Mayr 1954). In this model an organism undergoing long-distance dispersal experiences immediate reproductive isolation through geographic barriers, which drives speciation (Mayr 1963). Founder events have the potential to greatly reduce the genetic diversity within the new population which often leads to strong genetic drift resulting in a large change in the frequency of alleles. Although the exact mechanisms that result in speciation after a founder event are considered controversial (Templeton 2008), speciation associated with founder events, although rare, could play an important role in our understanding of evolution.

In part, the controversy surrounding founder speciation is based on the uncertainty around whether these events actually occur outside of existing theoretical models (Coyne and Orr 2004). The Galapagos hawk is a recent arrival to the Galapagos archipelago that is thought to still be undergoing active speciation (Bollmer et al. 2006). The Galapagos is thought to have split from its most recent ancestor, the Swainson's hawk (*Buteo swainsoni*) only about 126,000 years ago (Bollmer et al. 2006) when a small group of Swainson's hawk were possibly blown off course on their southernly migration to South America. While previous studies indicated a sister species relationship between the two taxa, recent studies have found that the Swainson's hawk is actually paraphyletic with respect to the Galapagos hawk (Hull et al. 2008). As the Galapagos hawk is thought to be the result of a single founding event with little to no secondary contact (Reising et al. 2003) and the two species have non-overlapping geographic ranges, this is a classic example of founder speciation in nature. Therefore, studying the speciation of the

Galapagos hawk will provide a rare opportunity to investigate the genomic mechanisms driving this type of speciation in natural populations.

The Galapagos hawk shows strong phenotypic divergence from the Swainson's hawk in three main areas: morphology, dispersal, and mating system. The Swainson's hawk is known for its long-distance migration from North America to Argentina and back every year (Fuller et al. 1998). The Galapagos hawk no longer migrates, and dispersal between islands is rare. Next, while the Swainson's hawk is monogamous, the Galapagos hawk exhibits varying degrees of cooperative polyandry across islands (Faaborg and Patterson 1981). Cooperative polyandry is where one female mates with multiple, unrelated males and all males provide paternal care (Faaborg et al. 1995). In the Galapagos hawk, polyandrous group sizes can vary but are stable over time, and the groups are territorial (Faaborg 1986). Most notably, the Swainson's hawk and the Galapagos hawk differ in their plumage coloration, and body size. The Swainson's hawk displays plumage polymorphism with light, dark, and intermediate color variations (Palmer et al. 1988). In contrast, the Galapagos hawk is entirely dark brown in color, although juveniles are lighter than adults. As for body size, the Galapagos hawk has evolved to be significantly larger than the Swainson's hawk (Hull et. al. 2008), and the Galapagos hawk has wider wings where the Swainson's hawk has narrower wings, a potential adaptation to its long-distance migration (Kerlinger 1989). These differences demonstrate large divergence in phenotypic traits between the two species over an evolutionarily-short time period, indicating the potential role of adaptive divergence in the differentiation of these species.

In this study, we combine a newly sequenced and annotated reference genome for the Swainson's hawk with individual re-sequencing data from 16 populations of Swainson's hawks and Galapagos hawks to investigate patterns of genome-wide divergence associated with founder

speciation. The creation of a high-quality draft genome for the Swainson's hawk enables the more accurate characterization of specific genomic regions and candidate genes associated with the speciation process (Fuentes-Pardo and Ruzzante 2017, Brandies et al. 2019). This genome will be the first high-quality annotated genome for the genus *Buteo* and will provide a key resource for studies of raptors. Because raptors have such diverse life histories and have populations spanning the entire globe, raptors can be used to study many important questions about adaptation and natural selection – questions that can now be more readily answered with new genomic resources. Our investigation into the divergent genomic landscape between the Galapagos hawk and the Swainson's hawk will also provide important insight into the relative contributions of adaptive divergence and genetic drift to the process of allopatric speciation.

METHODS

Swainson's Hawk Genome Assembly and Annotation

2 ml of whole blood was collected from the medial metatarsal vein of a female Swainson's hawk at the California Raptor Center in Davis, California. DNA was isolated from the whole blood using the Gentra PuregeneBlood Kit (Qiagen, Cat no =158445). Seven microliters of fresh blood preserved in EDTA was used as input for extraction with 900µl of lysis buffer. DNA was isolated following manufacturers' guidelines. DNA was quantified with a Qubit fluorometric assay and quality was assessed with nanodrop 260/280 and 260/230 values. gDNA fragment length was assessed using the Sage HLS pippin pulse system. Genomic DNA was adjusted to a concentration of 1.00 ng/µl and loaded on a Chromium Genome Chip. Whole genome sequencing libraries were prepared using Chromium Genome Library & Gel Bead Kit v.2 (10X Genomics, cat. 120258) and Chromium controller according to manufacturer's

instructions with one modification. Briefly, gDNA was combined with Master Mix, Genome Gel Beads, and partitioning oil to create Gel Bead-in-Emulsions (GEMs) on a Chromium Genome Chip. The GEMs' isothermally amplified barcoded DNA fragments were recovered for Illumina library construction. The post-GEM DNA was quantified prior to sequencing using a Bioanalyzer 2100 with an Agilent High sensitivity DNA kit (Agilent, cat. 5067-4626). Prior to Illumina library construction, the GEM amplification product was sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA) to approximately 375 bp (50 seconds at peak power = 175, duty factor = 10, and cycle/burst = 200). Then, the sheared GEMs were converted to a sequencing library following the 10X standard operating procedure. The library was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche) and sequenced on partial lane of NovaSeq6000 (Illumina, San Diego, CA) with paired-end 150 bp reads. The raw reads were assembled in Supernova v.2.1.1 (Weisenfeld et al. 2017), a program designed for the assembly of large diploid genomes from 10X Chromium linked-reads. To remove potential cross-contamination due to index-hopping from the 10X Genomics Chromium Genome kit, scaffolds were removed that had a median number of reads per barcode less than 10. Next, assembly completeness was measured in QUAST (Mikheenko et al. 2018) and in BUSCO v.5.1.2 using the aves dataset (Seppey et al. 2019). Next, the assembly was soft-masked in RepeatMasker (Smit et al. 2013) to avoid the prediction of false positive genes in low complexity and repetitive regions. Due to a lack of high-quality genomes within the genus Buteo, we used the BRAKER2 pipeline which is known for its high gene prediction accuracy in the absence of the annotation of a very closely related species (Bruna et al. 2021, Hoff et al. 2019, Hoff et al. 2016, Stanke et al. 2008, Stanke et al. 2006b). ProtHint was then used to generate hints using the Vertebrata OrthoDB protein database (Kriventseva et al. 2019) as reference protein sequences (Bruna et al. 2020).

These hints were then used for training in AUGUSTUS (Stanke et al. 2006) through the BRAKER2 pipeline. For functional annotation we conducted orthology assignment in eggNOG-mapper (Huerta-Cepas et al. 2019) and gathered family information from Interproscan 5.48-83.0 (Jones et al. 2014). These gene annotations were then run through Funannotate v. 1.8.1 (Palmer 2016) to create a set of final, combined annotations.

Next, we ordered our scaffolds by aligning the Swainson's hawk genome to the zebra finch (*Taeniopygia guttata*) genome. We used the optimized version of Satsuma, Satsuma2, to align our scaffolds to chromosomes based on a cross-correlation approach that uses a match scoring scheme to eliminate false hits (Grabherr et al. 2010). The zebra finch genome was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/assembly/GCF_008822105.2). Scaffolds were ordered according to their best alignment, and unplaced scaffolds were still included in further analyses but grouped separately.

Re-sequencing Dataset Preparation

Whole-genome data from 83 Swainson's hawks and 140 Galapagos hawks were generated according to the methods outlined in the previous chapters. Briefly, 83 Swainson's hawk were sampled from nine populations across the breeding range (see Figure 1.1), and 140 Galapagos hawks were sampled from eight islands across the Galapagos archipelago (see Figure 2.1). The sequences were aligned to our assembled Swainson's hawk reference genome using BWA v. 0.7.16 (Li 2013) and indexed bam files were created in Samtools v 1.10 (Li et al. 2009). Reads were filtered from the set of combined bam files in ANGSD with the following parameters: -uniqueOnly 1 -skipTriallelic 1 -minMapQ 30 -minQ 30 -doHWE 1 –maxHetFreq 0.5 -minInd n/2. Polymorphic sites were then identified with –pval 1 e 10-6 and –maf 0.05

cutoffs, and genotype likelihoods were calculated for both the whole genome dataset and just for the polymorphic sites usingthe GATK model.

Genome-wide Patterns of Differentiation

Population Genetics Statistics

To visualize the genome-wide variation between the Swainson's hawk and the Galapagos hawk a principal component analysis (PCA) was conducted using a covariance matrix created in program PCAngsd v.0.98 (Meisner and Albrechtsen 2018). This program is made specifically for low-depth data and uses an iterative procedure based on genotype likelihoods to estimate the covariance matrix. These results and all further results were visualized in R v. 3.6.3 (R Core Team 2021) using package ggplot2 v. 3.3.3 (Wickham 2016).

We next calculated nucleotide diversity and Tajima's D for all sites within scaffolds that contained an 'island' of divergence (see "Characterizing Islands of Divergence" below) larger than ten windows long. For each species, we used the genome-wide genotype likelihoods calculated in ANGSD to calculate the maximum likelihood estimate of the folded site allele frequency spectrum using the Swainson's hawk genome as both the reference and ancestral states. Next, nucleotide diversity and Tajima's D were calculated across the scaffolds of interest in ANGSD from the site frequency spectrum using the formulas described in Korneliussen et al. (2014) across sliding windows of 50 kb with a step of 10kb. We then corrected the theta values by dividing the output by the number of sites per window.

To analyze the change in allele frequencies between the species, we first calculated the number of rare alleles that were lost in the Galapagos hawk. We classified rare alleles as those

that have either a minor allele frequency greater than 0.9 or less than 0.1 in the Swainsons's hawk and a minor allele frequency less than 0.001 in the Galapagos hawk. Next, we calculated how many of those rare alleles have become the major allele in the Galapagos hawk. Lastly, we calculated how many genomic sites were fixed for different alleles between species, using a minor allele frequency of either greater than 0.999 or less than 0.001.

Characterizing Islands of Divergence

Next, to identify areas of genomic divergence, we estimated both relative and absolute divergence across the genome. Relative divergence, F_{ST}, was calculated in ANGSD for our polymorphic sites by first estimating the two-dimensional site frequency spectrum, and then using this spectrum as a prior for allele frequency calculations at each site. Then, F_{ST} was estimated across 50 kb sliding windows with a 10 kb step. Windows of high divergence were those that consisted of more than 100 sites and were within the top 1% of F_{ST} values across the genome. We then identified genomic 'islands' of divergence as groups of adjacent highly divergent windows, up to a maximum of four windows apart. Next, we identified annotated genes from the Swainson's hawk genome assembly that were within 100kb of our high F_{ST} windows. To investigate the function of these genes we used the gene list analysis function in Panther v. 16.0 (Mi et al. 2020) to annotate our gene list against the chicken (*Gallus gallus*) GO Ontology database DOI: 10.5281. We performed the PANTHER Overrepresentation Test for both the GO biological process and GO molecular function annotation datasets on our candidate genes. Specifically, we used the Fisher's Exact test, controlling for False Discovery Rate, and using the chicken genome as a reference list.

d_{XY} was calculated on the minor allele frequencies for our polymorphic sites using calcDxy.R (https://github.com/mfumagalli/ngsPopGen/blob/master/scripts/calcDxy.R). d_{XY} is an absolute measure of divergence as it is calculated as the average number of pairwise differences between sequences from two populations, excluding comparisons between sequences within populations. We calculated allele frequencies for both species separately in ANGSD using the –sites filter to specify the sites from our genomic dataset to ensure that we included sites that were fixed in either species. Next, we averaged the outputted d_{XY} values across the same 50 kb sliding windows that were used for the F_{ST} analysis above.

RESULTS

Genome Assembly and Annotation

A total of 74 Gb of data were generated encompassing 474,761,688 reads. The average coverage was 51.49 reads per site, with a GC content of 42.04%. The assembly was 1.17 Gb long which is consistent with the size of previous avian genomes. The assembly consists of 8,100 total scaffolds and the scaffold N50 is 20.8 Mb with the longest scaffold being 56.85 Mb long. Other summary statistics can be found in Table 3.1. In the annotation process, a total of 34,370 protein-coding genes were identified of which 5,922 genes were annotated with known protein names. Our BUSCO analysis found that 7,797 (93.5%) out of 8,338 highly conserved proteins within the Aves lineage were present in our genome assembly (Figure 3.1). Of the complete BUSCOs, 7,746 (92.9%) were single-copy and 51(0.6%) were duplicated. Of the remaining BUSCOs, 57 (0.7%) were fragmented and 484(5.8%) were missing.

Data Generation

A total of 284.3 Gb and 233 Gb of raw data was produced in the resequencing analysis of the Swainson's hawk and the Galapagos hawk, respectively. Additional statistics on these data can be found in the previous chapters. For our two genomic datasets, after quality filtering, our whole-genome dataset across both species consisted of 1,093,747,856 sites. Next, from this dataset we extracted only the variant sites, resulting in a final SNP dataset of 3,284,387 sites on 7,283 different scaffolds.

Characterizing Genomic Divergence

Population Genetics Statistics

The principal component analysis shows clear genetic differentiation between the Swainson's hawk and the Galapagos hawk (Figure 3.2). The first principal component explains 99.2% of the variation between the two species. It is also visually evident that there is more differentiation among the Galapagos hawk samples than the Swainson's hawk samples, as the Galapagos hawk samples are spread vertically across the extent y-axis, while all of the Swainson's hawk samples are clustered in a single area.

We found that there were 802,481 rare alleles in the Swainson's hawk, of which 5,102 (0.64%) were completely lost in the Galapagos hawk. Also, 304,854 (38%) rare alleles in the Swainson's hawk became the major allele in the Galapagos hawk, and 226,083 (28.17%) of those rare alleles are now fixed in the Galapagos hawk. Lastly, genome-wide, 45,022 sites are fixed for different alleles in the Swainson's hawk and the Galapagos hawk, which is 13.9% of all polymorphic sites between the two species.

Characterizing Islands of Divergence

To identify genomic areas of high divergence between the Swainson's hawk and the Galapagos hawk we classified the genome into 71,018 50kb sliding windows that each consisted of more than 100 polymorphic sites. The average genome-wide F_{ST} was 0.657 across all windows with more than 100 sites (Figure 3.3), and the density plot of the F_{ST} distribution has an obvious right-skew (Figure 3.4). Our windows of high F_{ST} included 711 windows across 517 scaffolds with an average F_{ST} of 0.87. Our analysis of the genes within and near these high F_{ST} windows revealed 219 genes of interest, although there were no statistically overrepresented GO groups for biological process or molecular function in comparison to the chicken genome. Once adjacent high F_{ST} windows were combined, we identified 162 'islands' of divergence with an average 'island' size of just under 90 kb (3.88 windows). Our main islands of interest were those that consisted of 10 or more adjacent windows, which revealed four 'islands' on three scaffolds that were retained for further analyses and visualization (Figures 3.5-3.7). The two largest 'islands' were located on scaffold 14 and were only separated by 30kb so we joined these into one, large, island consisting of 37 high F_{ST} windows across 560kb (Figure 3.5). Across the three scaffolds with large 'islands' of divergence, Tajima's D was much lower in the Swainson's hawk compared to the Galapagos hawk, although all values were consistently below 0 (Figure 3.8).

DISCUSSION

In this study we produced a high-quality genome assembly for the Swainson's hawk and characterized the genomic patterns of divergence underlying the speciation of the Galapagos hawk by resequencing individuals of both species across their distributions. Overall, we found

evidence of significant genomic differentiation between the Swainson's hawk and the Galapagos hawk characterized by genome-wide heterogenous peaks in measures of both relative and absolute divergence. The distribution of F_{ST} values as well as the patterns of genomic divergence support our hypothesis of a recent divergence with no active gene flow or secondary contact suggestive of true founder speciation.

Our principal component analysis (PCA) showed clear distinction between the two species, as expected. The PCA also showed that the Galapagos Hawk populations were spread out along the second principal component while the Swainson's hawk samples were clustered in a single group. This divergence in Galapagos hawk samples is most likely due to the limited gene flow among island populations combined with stochastic changes in allele frequencies from the founding of each island population. The divergence between the two species is further demonstrated by an average genome-wide F_{ST} of 0.657. This value is high compared to other allopatric avian species, but it is not unexpected in an island ecosystem. For example, Burri et al. (2015) found a maximum genome-wide F_{ST} of 0.303 in allopatric sister species of *Ficedula* flycatchers, but Funk et al. (2016) found an average genome-wide F_{ST} of 0.630 between mainland gray foxes (*Urocyon cinereoargenteus)* and six isolated island fox (*Urocyon littoralis*) populations. These results show the large amount of genomic divergence that has accumulated between the Swainson's hawk and the Galapagos hawk in a relatively short evolutionary time-scale.

Unlike speciation with gene flow models that predict few genomic areas of high divergence separated by mostly homogenous areas of low divergence (Turner et al. 2005), allopatric speciation is predicted to cause high divergence in many genomic regions. This is due to allopatric populations evolving independently in the face of genetic drift and selection without

the homogenizing effects of gene flow (Gavrilets 2004). We found an extremely heterogenous pattern of high differentiation in both relative and absolute measures of divergence, indicative of allopatric speciation. Further evidence supporting this hypothesis is the right-skewed density plot of genome-wide F_{ST} values. Nosil et al. (2012) found that in parapatric populations, the F_{ST} distributions tend to be L-shaped, but in allopatric populations the distribution is right skewed and the extent of the skew increases with geographic distance between populations. The large right-skew in the F_{ST} distribution between the Swainons's hawk and the Galapagos hawk is indicative of no secondary contact or gene flow occurring between these two species, which provides further evidence for strict founder speciation in this system. Theory also predicts that in the early stages of speciation, genetic drift will typically affect the entire genome whereas natural selection will only affect the specific region of the genome that dictate the adaptive phenotype(s) (Beaumont and Nichols 1996). Additionally, as shown in Chapters 1 and 2, there has been a large reduction in the genome-wide genetic diversity of the Galapagos hawk (avg. genome-wide diversity of 1.8×10^{-4}) compared to the Swainson's hawk (avg. genome-wide diversity of 1.8×10^{-3}). This pattern of large divergence across the entire genome combined with the loss of genetic diversity, indicates an important driver of the divergence between the two species may be the random nature of genetic drift from founder effects as each island was subsequently colonized by only few individuals.

Our identification of genes associated with high F_{ST} windows and 'islands' of differentiation provide insight into the genomic divergence between the Swainson's hawk and the Galapagos hawk. In our 'islands' we see a pattern of increased d_{XY} and F_{ST} and decreased nucleotide diversity compared to the rest of the scaffold. Many recent studies have linked patterns of high F_{ST} and low nucleotide diversity with linked selection in an ancestor before the lineage split

(Cruickshank and Hahn 2014, Burri et al. 2015). Specifically, in allopatric speciation, linked selection would likely either have no effect on d_{XY} or cause d_{XY} to decrease (Nachman and Payseur 2012). Our results are not consistent with this scenario as we found that d_{XY} increases along with F_{ST} in these regions. As d_{XY} is not affected by variation in present levels of polymorphism, high levels of d_{XY} are not indicative of linked selection. These patterns of co-occurring regions of high absolute (d_{XY}) and relative (F_{ST}) divergence are typically attributed to regions resistant to introgression (Cruickshank and Hahn 2014) in speciation with gene-flow models, but here, where we find no evidence of gene-flow, it is more likely that these areas developed due to strong genetic drift followed by local adaptation to differing environmental conditions. While these areas show the most divergence between the two species, we cannot confirm that any of these genomic regions are directly involved in the evolution of reproductive isolation as they could have evolved due to random stochastic events associated with founder speciation. Interestingly, within our high F_{ST} windows we identified two genes that have been previously recognized as candidate genes under selection in nocturnal raptors compared to their diurnal counterparts (Cho et al. 2019). RPE65 is involved in light detection and RDH8 is involved in retinol metabolism (Cho et al. 2019). These two genes, as well as the other 217 genes displaying particularly high divergence between the Swainson's hawk and the Galapagos hawk will provide an important resource for finding the genetic mechanisms driving pre- or post-zygotic reproductive isolation between these species.

Over the years, studies on island archipelagos have made important contributions to the development of the allopatric model of speciation. Although still controversial, island taxa also provide important opportunities to better understand the presence and validity of founder speciation (Mayr 1954). Our results provide evidence for the Galapagos hawk evolving through

true founder speciation, characterized by a loss of genetic variation in the initial founding event. The results shown here are consistent with strong genetic drift following demographic expansion from small founding groups, with large divergent 'islands' possibly experiencing reduced gene flow, although it is extremely difficult to decouple the evolutionary processes that could be driving this genomic divergence. To provide more insight into the role of divergent selection in defining these genomic 'islands' of divergence, future research should include a comprehensive study of recombination rate variation across the genome (Wolf and Ellegren 2017) because areas of low recombination can sometimes show divergence similar to that of divergent selection (Burri et al. 2015). Additionally, as several studies in natural populations have associated a variety of structural variations such as chromosomal inversions (Kirubakaran et al. 2016, Lamichhaney et al. 2015) with phenotypic divergence, a future step in this system is to incorporate chromosome-level data into the Swainson's hawk genome assembly. Overall, this study represents an important step towards an understanding of the complex interactions between evolutionary processes that shape allopatric speciation at the genomic level, and future studies on the divergence of the Galapagos hawk will provide a unique opportunity to study a rare example of founder speciation in a natural population.

LITERATURE CITED

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London. Series B: Biological Sciences, 263(1377), 1619-1626.

Bollmer, J. L., Kimball, R. T., Whiteman, N. K., Sarasola, J. H., & Parker, P. G. (2006). Phylogeography of the Galápagos hawk (Buteo galapagoensis): a recent arrival to the Galápagos Islands. Molecular phylogenetics and evolution, 39(1), 237-247.

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. Genes, 10(11), 846.

Bruna, T., Lomsadze, A., & Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics and Bioinformatics, 2(2), lqaa026.

Bruna, T., Hoff, K.J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. NAR Genomics and Bioinformatics 3(1):lqaa108, doi: 10.1093/nargab/lqaa108.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoi, L., Bures, S., Garamszegi, L.Z., Hogner, S., Moreno, J., Qvarnstron, A., Ruzic, M., Saether, S., Saetre, G., Torok, J. & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome research, 25(11), 1656-1665.

Cho, Y. S., Jun, J. H., Kim, J. A., Kim, H. M., Chung, O., Kang, S. G., Park, J.Y., Kim, H.J., Kim, S., Kim, H.J., Jang, J.H., Na, K.J., Kim, J., Park, S.G., Lee, H.Y., Manica, A., Mindell, D.P., Fuchs, J., Edwards, J.S., Weber, J.A., Witt, C.C., Yeo, J.H., Kim, S. & Bhak, J. (2019). Raptor genomes reveal evolutionary signatures of predatory and nocturnal lifestyles. Genome biology, 20(1), 1-11.

Coyne, J. A., & Orr, H. A. (2004). Speciation (Vol. 37). Sunderland, MA: Sinauer Associates.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular ecology, 23(13), 3133-3157.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Kunster, A., Makinen, H., Nadachowska-Bryzyska, Qvarnstrom, A., Uebbing, S. & Wolf, J. B. (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature, 491(7426), 756-760.

Faaborg, J., & Patterson, C. B. (1981). The characteristics and occurrence of cooperative polyandry. Ibis, 123(4), 477-484.

Faaborg, J. (1986). Reproductive success and survivorship of the Galapagos Hawk Buteo galapagoensis: potential costs and benefits of cooperative polyandry. Ibis, 128(3), 337-347.

Faaborg, J., Parker, P. G., DeLay, L., De Vries, T. J., Bednarz, J. C., Paz, S. M., Naranjo, J. & Waite, T. A. (1995). Confirmation of cooperative polyandry in the Galapagos hawk (Buteo galapagoensis). Behavioral Ecology and Sociobiology, 36(2), 83-90.

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends in genetics, 28(7), 342-350.

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular ecology, 26(20), 5369-5406.

Fuller, M. R., Seegar, W. S., & Schueck, L. S. (1998). Routes and travel rates of migrating Peregrine Falcons Falco peregrinus and Swainson's Hawks Buteo swainsoni in the Western Hemisphere. Journal of avian biology, 433-440.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., Ghalambor, C.K., MAldonado, J.E., Rick, T.C., Day, M.D., Polato, N.R., Fitzpatrick, S.W., Coonan, T.J., Crooks, K.R., Dillon, A., Garcelon, D.K., King, J.L., Boser, C.L., Gould, N. & Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular ecology*, *25*(10), 2176-2194.

Gavrilets, S. (2004). Fitness landscapes and the origin of species (MPB-41). Princeton University Press.

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics, 26(9), 1145-1151.

Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M. and Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics, 32(5):767-769.

Hoff, K.J., Lomsadze, A., Borodovsky, M. and Stanke, M. (2019). Whole-Genome Annotation with BRAKER. Methods Mol Biol. 1962:65-95, doi: 10.1007/978-1-4939-9173-0_5.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Von Mering, C. & Bork, P. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res, 47: D309–D314. doi: 10.1093/nar/gky1085

Hull, J. M., Savage, W. K., Bollmer, J. L., Kimball, R. T., Parker, P. G., Whiteman, N. K., & Ernest, H. B. (2008). On the origin of the Galapagos hawk: an examination of phenotypic differentiation and mitochondrial paraphyly. Biological Journal of the linnean Society, 95(4), 779-789.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWIlliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S., Lopez, R. & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics, 30(9), 1236-1240.

Kerlinger, P. (1989). Flight strategies of migrating hawks. University of Chicago Press.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., De Rosa, M.C., Richino, B., Johansen, T., Ottera, H., Sonesson, A., Lien, S. & Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. Molecular ecology, 25(10), 2130-2143.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, *15*(1), 356.

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research, 47(D1), D807-D811.

Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., Kerje, S., Gustafson, U., Shi, C., Chen, W., Liang, X., Huang, L., Want, J., Liang, E., Wu, Q., Lee, S.M., Xu, X., Hoglund, J., Liu, X. & Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). Nature genetics, 48(1), 84-88.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078-2079.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2

Mayr, E. (1954). Change of genetic environment and evolution. In J.S. Huxley, A.C. Hardy, and E.B. Ford, eds., *Evolution as a Process*. London: G. Allen &c Unwin, pp. 157-180.

Mayr, E. (1963). Animal species and evolution. Harvard University Press.

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719-731.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L., Mushayamaha T., and Thomas, P.D. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API , Nucl. Acids Res. (2020) doi: 10.1093/nar/gkaa1106s.

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A. (2018). VErsatile genome assembly evalution with QUAST-LG. Bioinformatics, 34(13): i142-i150. doi: 10.1093/bioinformatics/bty266

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1587), 409-421.

Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. Trends in ecology & evolution, 26(4), 160-167.

Nosil, P. and Feder, J.L. (2012) Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B: Biol. Sci. 367, 332–342

Nosil, P., & Hohenlohe, P. A. (2012). Dimensionality of sexual isolation during reinforcement and ecological speciation in Timema cristinae stick insects. Evolutionary Ecology Research, 14(4), 467-485.

Palmer, R.S. 1988. Handbook of North American Birds, vols. 4 and 5. Diurnal raptors (Parts 1 and 2). Yale University Press, New Haven, CT

Palmer, J.M. (2016) Funannotate: Pipeline for Genome Annotation https://funannotate.readthedocs.io/en/latest/

Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., Zhang, G., Jarvis, E.D., Schlinger, B.A & Buerkle, C. A. (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. Molecular ecology, 22(12), 3304-3317.

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Wikelski, M. & Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science, 344(6190), 1410-1414.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riesing, M. J., Kruckenhauser, L., Gamauf, A., & Haring, E. (2003). Molecular phylogeny of the genus Buteo (Aves: Accipitridae) based on mitochondrial marker sequences. *Molecular phylogenetics and evolution*, *27*(2), 328-342.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C.L., Saetre, G., Bank, C., Brannstrom, A., Brelsford, A., Clarkson, C.S., Eroukhmanoff, F., Feder, J.L., Fischer, M.C., Foote, A.D., Franchini, P., Jiggins, C.D., Jones, F.C., Lindholm, A.K., Lucek, K., Maan, M.E., Marques, D.A., Martin, S.H., Matthews, B., Meir, J.I., Most, M., Nachman, M.W., Nonaja, E., Rennison, D.J., Schwarzer, J., Watson, E.T., Westram, A.J. & Widmer, A. (2014). Genomics and the origin of species. Nature Reviews Genetics, 15(3), 176-192.

Seppey, M., Manni, M., Zdobnov E.M. (2019) BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M. (eds) Gene Prediction. Methods in Molecular Biology, vol 1962. Humana, New York, NY. 2019 doi.org/10.1007/978-1-4939-9173-0_14. PMID:31020564

Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006a). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids research, 34(suppl_2), W435-W439.

Stanke. M., Schöffmann, O., Morgenstern, B. and Waack, S. (2006b). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7, 62.

Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics, doi: 10.1093/bioinformatics/btn013.

Templeton, A. R. (2008). The reality and importance of founder speciation in evolution. Bioessays, 30(5), 470-479.

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in Anopheles gambiae. PLoS biology, 3(9), e285.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome research*, *27*(5), 757-767

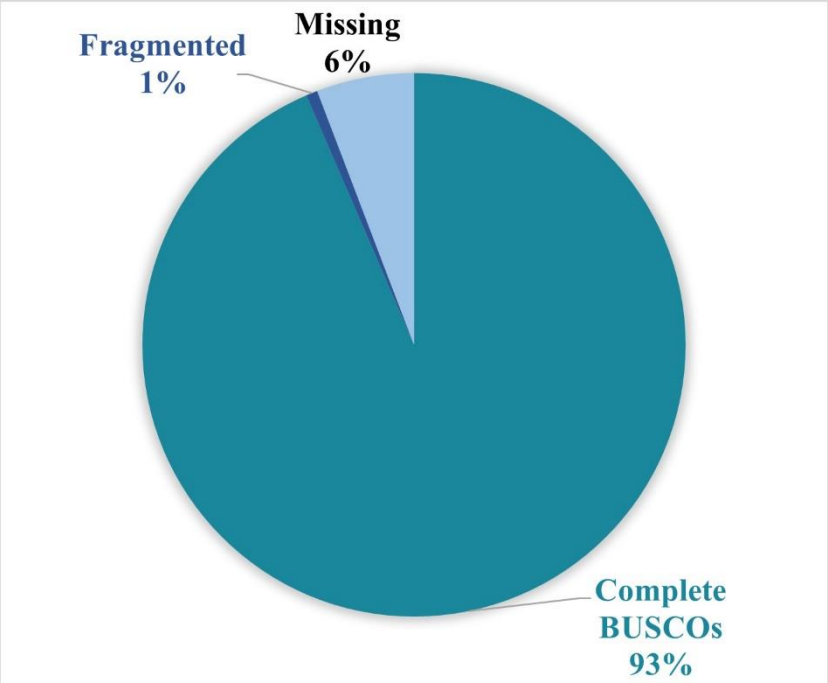
Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics, 18(2), 87-100.

FIGURES AND TABLES

Table 3.1. Summary of genome assembly and annotation results

Genome Assembly Statistics	
Raw Reads	474,761,688
Avg. Sequence Length	151 bp
Raw Coverage	51.49x
Effective Coverage	37.42x
GC Content	42.04%
Assembly Size	1.176 Gb
Estimated Genome Size	1.39 Gb
Total # of scaffolds	8,100
Scaffolds > 10kb	1,163
N50	20.8 Mb
N75	11.49 Mb
N's per 100kb	1968.72



C:93.5%[S:92.9%,D:0.6%],F:0.7%,M:5.8%,n:8338

Figure 3.1. BUSCO results based on the aves_odb10 dataset

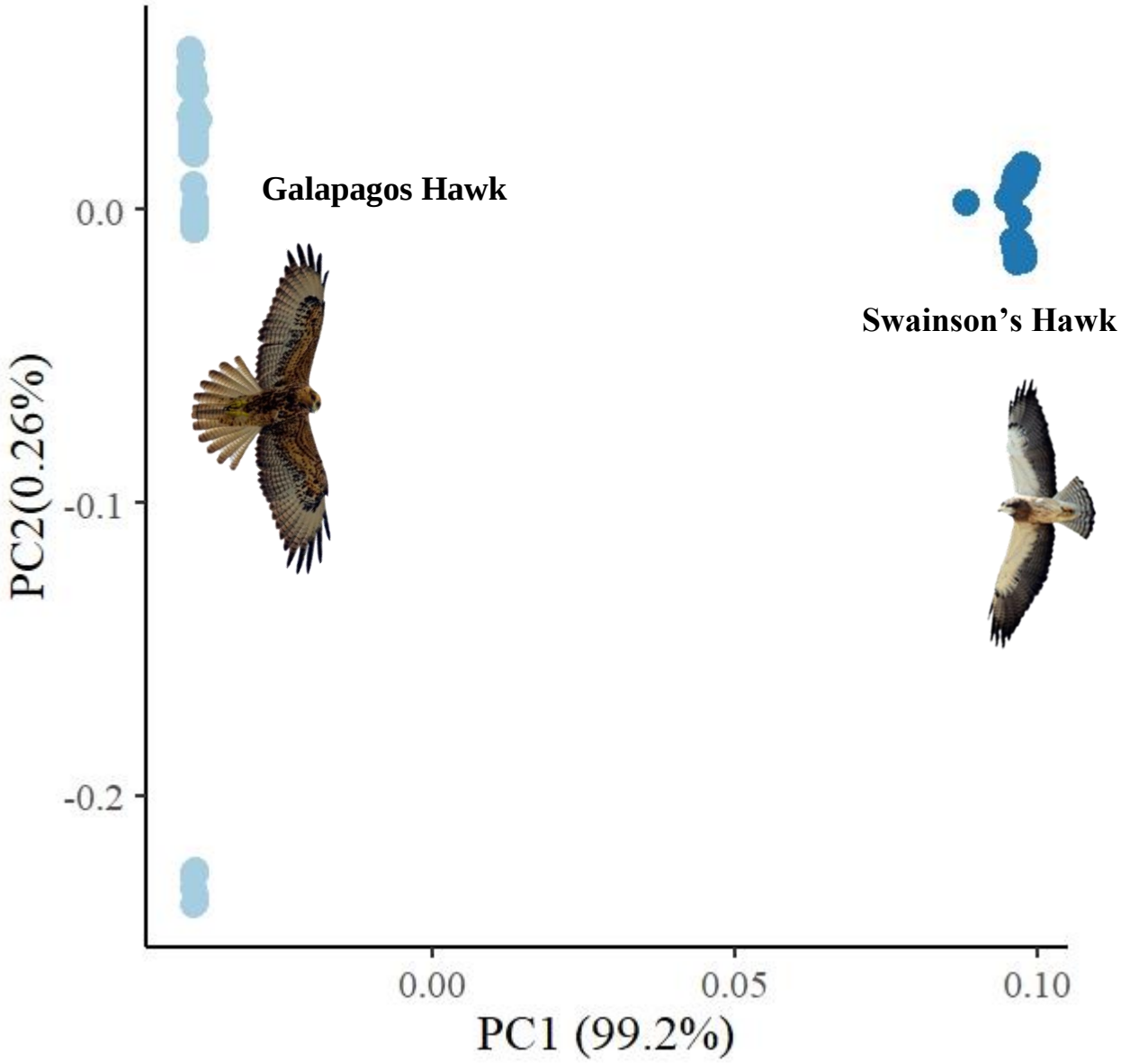


Figure 3.2. Principal component analysis showing the degree of differentiation between the Swainson's hawk and the Galapagos Hawk.

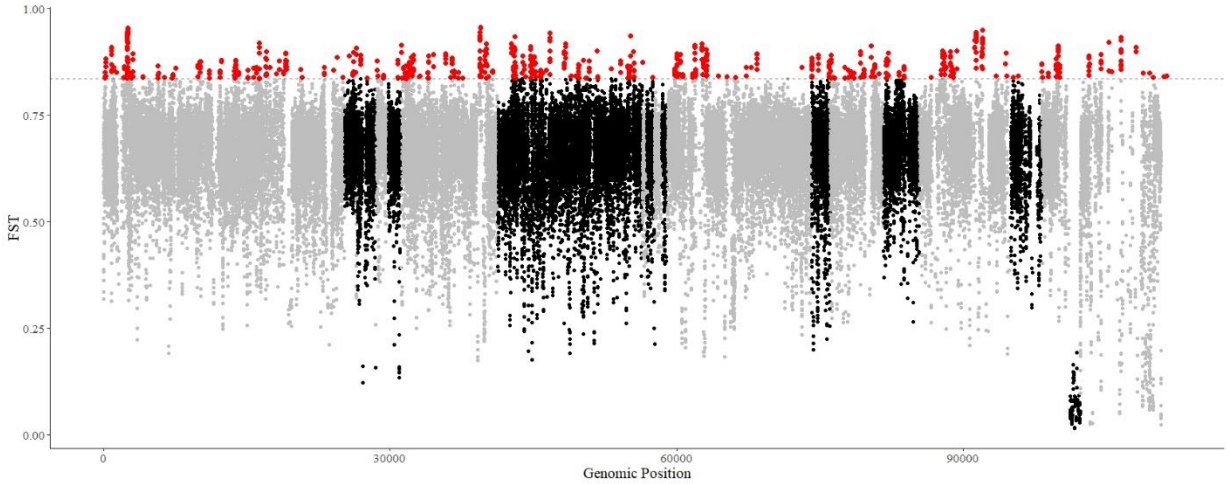


Figure 3.3. Genome-wide distribution of pairwise F_{ST} values in 50kb sliding windows. Red points indicate windows of high F_{ST} within the top 1% of FST values. The horizontal dotted line corresponds with the 99% high F_{ST} cutoff of 0.836

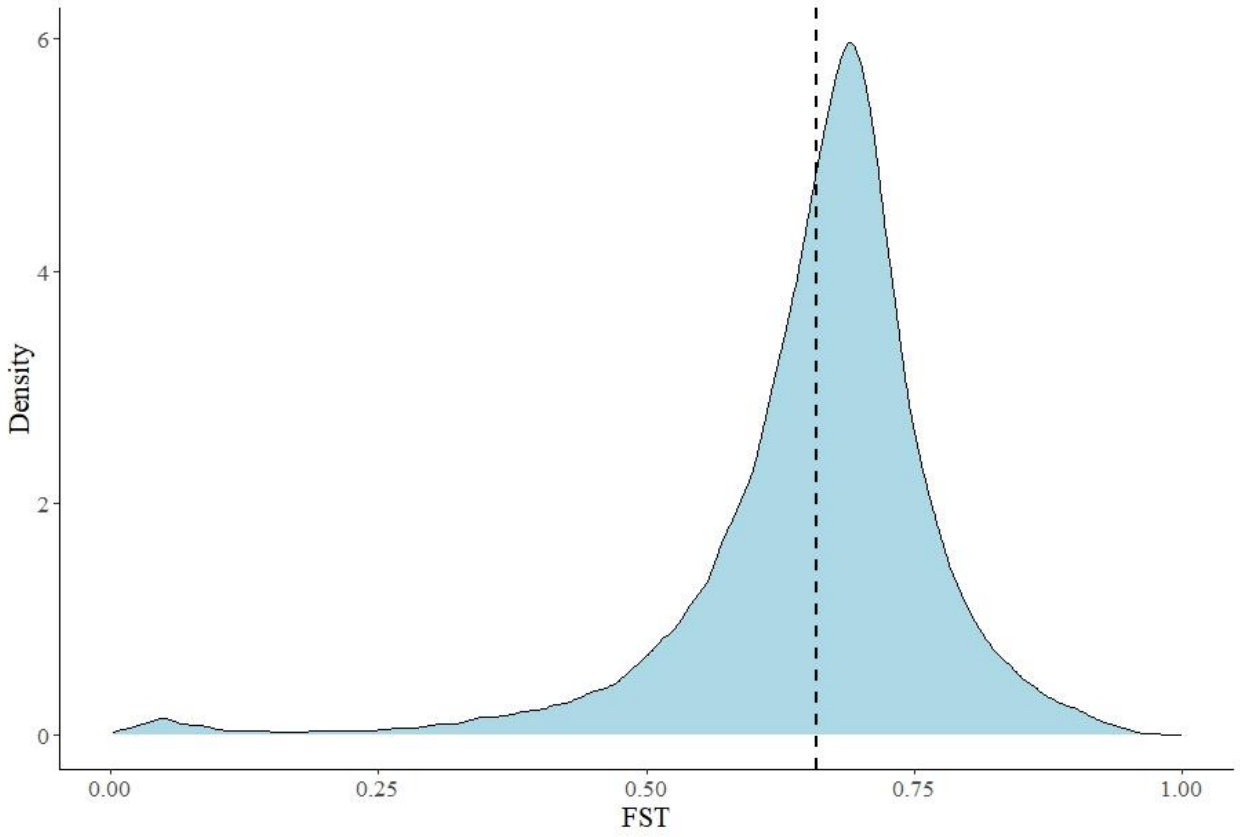


Figure 3.4. Density plot of genome-wide F_{ST} in 50 kb overlapping sliding windows, with a 10 kb step, dotted line corresponds to the mean genome-wide FST of 0.657

Scaffold	Region	Length(kb)	Windows	F_{ST}	Tajima's D SWHA	Tajima's D GAHA
14	24980000 - 25540000	560	37	0.893	-2.48	-0.799

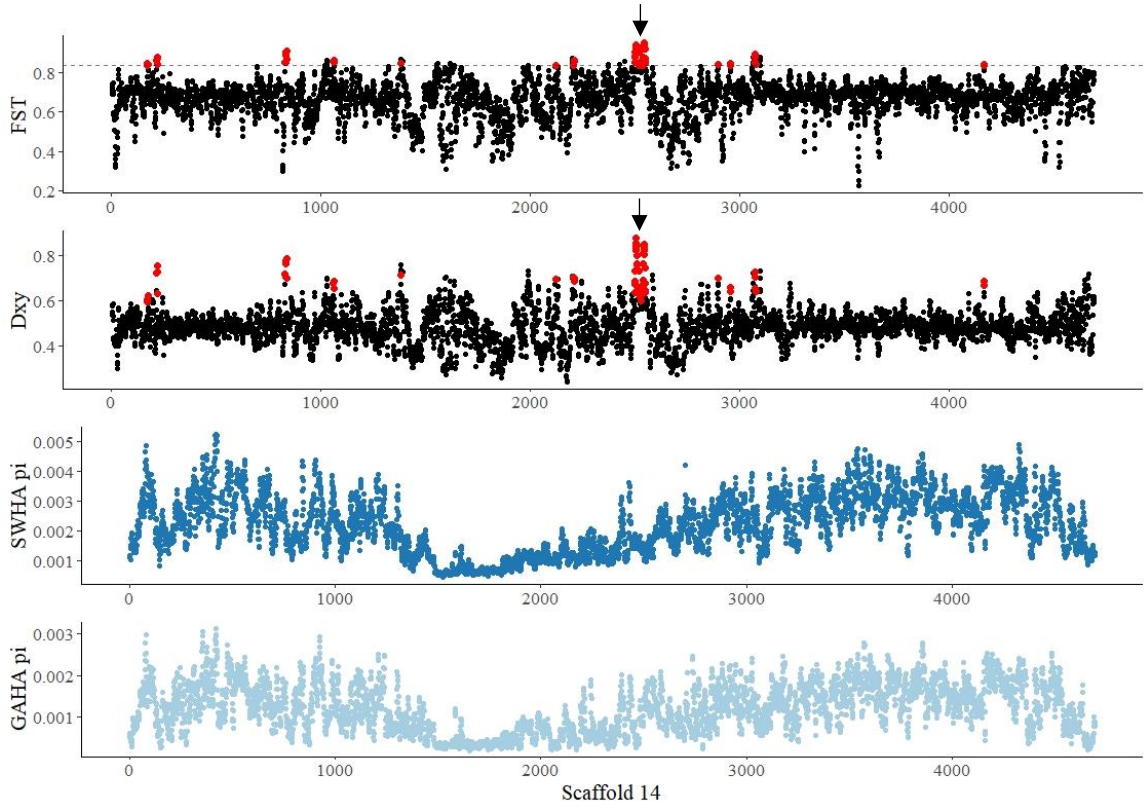


Figure 3.5. Visualization and summary statistics describing 'island' of divergence on Scaffold 14. The 'island' region is identified by the black arrow.

Scaffold	Region	Length(kb)	Windows	F_{ST}	Tajima's D SWHA	Tajima's D GAHA
28	17320000-	160	12	0.888	-2.50	-0.997

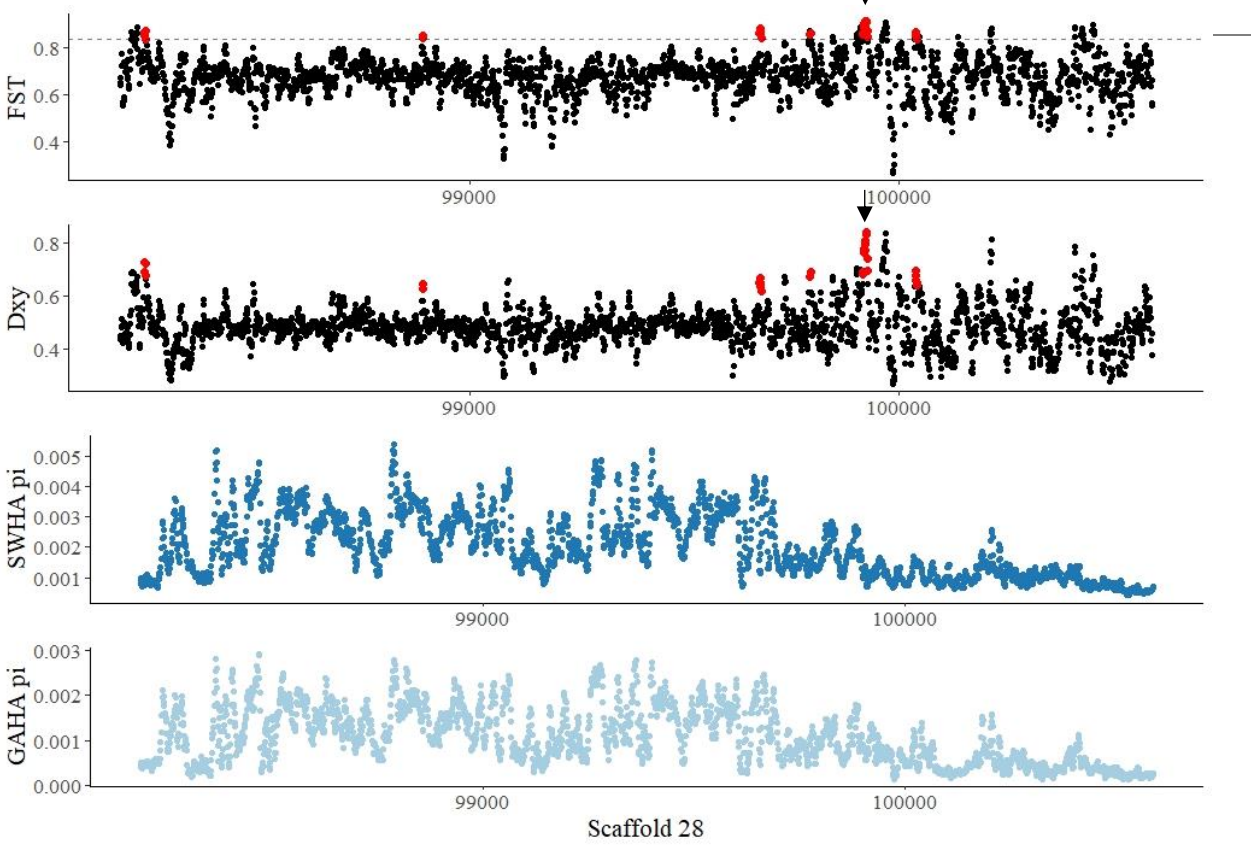


Figure 3.6. Visualization and summary statistics describing 'island' of divergence on Scaffold 28. The 'island' region is identified by the black arrow.

Scaffold	Region	Length(kb)	Windows	F_{ST}	Tajima's D SWHA	Tajima's D GAHA
92	3640000-3790000	150	11	0.938	-2.597	-.022

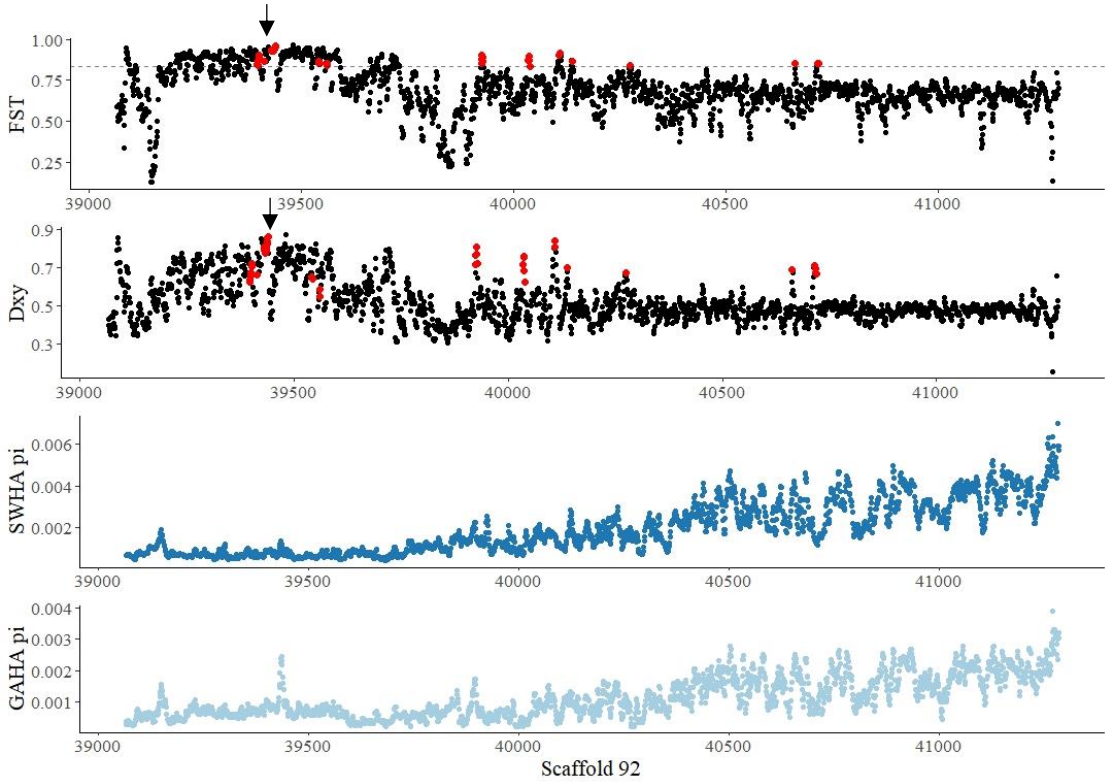


Figure 3.7. Visualization and summary statistics describing 'island' of divergence on Scaffold 92. The 'island' region is identified by the black arrow.

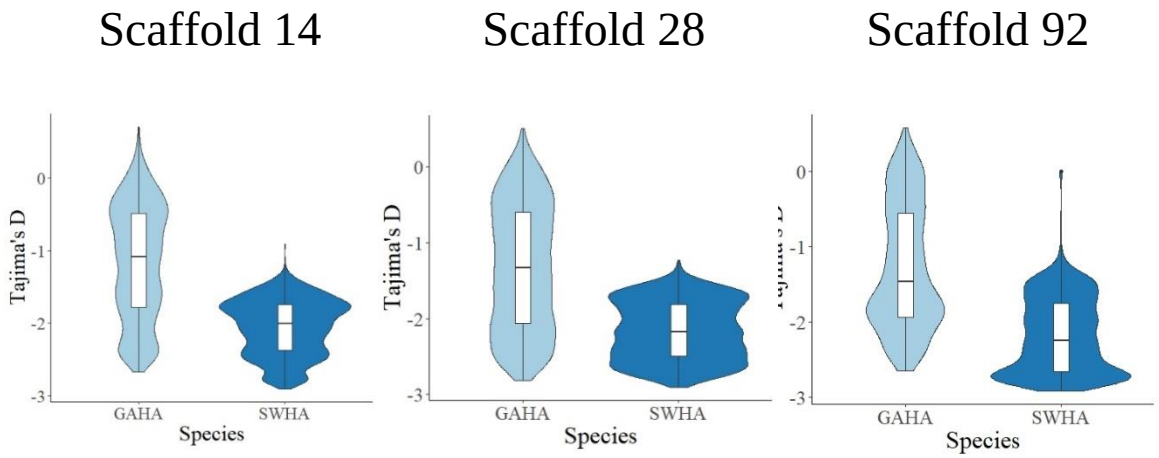


Figure 3.8. Comparison of Tajima's D distributions between the Swainson's hawk (SWHA) and the Galapagos Hawk (GAHA) for the three scaffolds with large divergent 'islands'.