

Genomic Differentiation of Northern Goshawks in Coastal British Columbia

by

Kenneth K. Askelson

B.Sc., the University of Wyoming, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies
(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

August 2018

© Kenneth K. Askelson, 2018

Committee Page

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

Genomic Differentiation of Northern Goshawks in Coastal British Columbia

submitted by Kenneth K. Askelson in partial fulfillment of the requirements for

the degree of Master of Science

in Zoology

Examining Committee:

Darren Irwin

Supervisor

Dolph Schluter

Supervisory Committee Member

Sally Aitken

Supervisory Committee Member

Loren Rieseberg

Additional Examiner

Abstract

Understanding the process by which populations become genetically differentiated from one another has been a central goal of population genetics since its inception. With the loss of biodiversity across the globe, we lose information regarding how populations of organisms separate and become genetically distinct. An organism that exemplifies this issue is a subspecies of the Northern Goshawk (*Accipiter gentilis laingi*; hereafter simple “*laingi*”) which is classified as Threatened in coastal British Columbia under the Species at Risk Act (Canada) and the Endangered Species Act (USA). Using genotyping-by-sequencing (GBS) data across thousands of SNPs (single nucleotide polymorphisms), we investigate the genetic differentiation of this subspecies and infer the processes governing its distinctiveness. We find that Northern Goshawks on the archipelago of Haida Gwaii are distinct from other populations, clearly separating in principal component analyses, and have a wider distribution of F_{ST} and D_{XY} when compared to other populations. In other populations, we recovered weak differentiation ranging from coastal BC to Maine; these populations likely represent the other North American subspecies *Accipiter gentilis atricapillus* (hereafter, simply “*atricapillus*”). The second phase of our research was to clarify the range of the *laingi* subspecies, which previously had been under debate. By selecting *laingi*-informative single nucleotide polymorphisms (SNPs) from our sequencing data we were able to develop SNP genotyping assays that allowed for the inclusion of hundreds of additional low-quality samples. Using these assays, we find that *laingi* goshawks are largely restricted to Haida Gwaii. Additionally, we wanted to understand the processes driving differentiation in Haida Gwaii and gene flow between other populations and Haida Gwaii. We found that even though Haida Gwaii is a small population, strong selection is likely

shaping the genome. By jointly analyzing our GBS and genotyping data we find that gene flow between Haida Gwaii and other populations is likely low. This thesis contributes to knowledge of a Threatened bird of prey and more generally to how evolutionary distinctness evolves in geographically separated populations of organisms.

Lay Summary

In coastal British Columbia, there is a Threatened subspecies of Northern Goshawk (*Accipiter gentilis laingi*). While this subspecies has been described morphologically, little is known about the range and extent to which it is genetically distinct from the other North American subspecies, *A. g. atricapillus*. In this thesis we investigate the distinctness and genetic range of *laingi* by using a combination of genomic and genetic tools. We find that Northern Goshawks on the archipelago of Haida Gwaii are differentiated from those in other regions in North America. Additional findings of our work indicate that selection has likely shaped the genome of goshawks in Haida Gwaii and little gene flow occurs with other populations. This work contributes to the understanding of genetic differentiation of this Threatened bird of prey. By clarifying the range of *laingi*, we hope that this will aid in the management of Northern Goshawks in British Columbia.

Preface

In the thesis described below, I collected samples from museums and other contributors, extracted samples and prepared them for DNA sequencing and genotyping, and analyzed the data. Each of these steps was done in collaboration with my advisor Dr. Darren Irwin and Research Associate Dr. Armando Geraldés. Dr. Geraldés worked closely with me on extracting samples and preparing them for sequencing and genotyping, as well as analyzing the data. Dr. Geraldés co-wrote (with Dr. Irwin) the Genome B.C. grant that funded most of this research. In the writing below, I use the term “we” to acknowledge Dr. Geraldés and Dr. Irwin’s important role in this research.

Table of Contents

Abstract	iii
Lay Summary	v
Preface	vi
Table of Contents.....	vii
List of Tables	ix
List of Figures	x
List of Abbreviations.....	xii
Acknowledgements	xiii
CHAPTER 1: General Introduction	1
1.1 Importance of Research.....	6
1.2 Thesis Goals	8
CHAPTER 2: Genomic Differentiation of Goshawks in British Columbia	10
2.1 Introduction.....	10
2.2 Methods	13
2.2.1 Sampling Distribution	13
2.2.2 DNA Extraction, GBS Preparation, and Sequencing	13
2.2.3 GBS Data	14
2.2.4 Genotyping Assays	16
2.2.5 Dendrogram of Evolutionary Relationships.....	18
2.2.6 Principal Component Analysis	18
2.2.7 Inbreeding and Average Heterozygosity	18
2.2.8 Outlier Loci.....	19
2.2.9 Variation in Differentiation Across the Genome	20
2.2.10 Genotyping Assay Filtering and Analyses	20
2.2.11 Gene Flow	21
2.3 Results.....	22
2.3.1 Evolutionary Relationships.....	22
2.3.2 Genomic Differentiation	22
2.3.3 Range of <i>A. g. laingi</i>	23
2.3.4 Outlier Loci Detection	23
2.3.5 Variation in Differentiation Across the Genome	24

2.3.6 Gene Flow	25
2.4 Discussion.....	28
CHAPTER 3: General Conclusion	33
Bibliography	54

List of Tables

Table 1. All samples received as shown by populations. Columns “GBS” and “Genotyping” indicate how many samples were used in each methodology. For the populations in which we have GBS data we’ve included the average expected heterozygosity for that population and the standard error of the mean.	38
Table 2. F_{ST} , allele frequencies and Missing data for laingi specific genotyping assays. Set 1 indicates the original filtering parameters and set 2 indicates updated filtering (see methods). “Freq” indicates allele frequency and in these columns Taq indicates the allele frequency in the genotyping data. Δ “Freq” is the difference between frequencies found in GBS and genotyping data. NOGO3 is marked with an asterisk because this assay failed to amplify and was excluded from further analysis.....	38
Table 3. Nm between population pairs from the shared SNP analysis. Acronyms are as follows: Haida Gwaii (HG), Alexander Archipelago (AA), British Columbia (BC), Europe (EU), and North America (NA).....	39
Table 4. Flanking sequences of laingi specific genotyping assays.	39

List of Figures

Figure 1. Dorsal images of juvenile HG goshawk (above) and juvenile mainland goshawk (below). Photos are color corrected.	40
Figure 2. Ventral images of juvenile HG goshawk (above) and juvenile mainland goshawk (below). Photos are color corrected.	40
Figure 3. Total samples received of all goshawk samples in North America by population. Defining populations follows that of how GBS populations were defined. Acronyms explained are AK (Alaska), AA (Alexander Archipelago), AZ (Arizona), BC (British Columbia), CA (California), HG (Haida Gwaii), VI (Vancouver Island), WA (Washington).	41
Figure 4. Dendrogram of evolutionary relationships of Northern Goshawks and outgroups. N.A. are North American Goshawks, EU are European, and HG is Haida Gwaii.	41
Figure 5. PCA of 124 samples excluding outgroups using 6006 SNPs. See Table 1 for acronym definitions.	42
Figure 6. PCA using 6058 SNPs of 119 samples excluding European and outgroups. See Table 1 for acronym definitions.	43
Figure 7. PCA using 6257 SNPs using 107 samples excluding HG, European, and outgroup samples. See Table 1 for acronym definitions.	44
Figure 8. Distribution of laingi index across each population. AZ is Arizona, CA California, and YK is the Yukon. Each circle is an individual.	44
Figure 9. Distribution of samples used in TaqMan assays. Points are colored from blue to red indicating percent laingi with red being highly laingi. Range of laingi index does correspond from 0-1 as loci are not fixed between populations.	45
Figure 10. Inbreeding coefficients (F_{IS}) of 119 samples across GBS populations using 24,311 SNPs.	46
Figure 11. Distribution of F_{ST} among 9850 SNPs with 119 samples from HG, VI, AA, AK, WA, BC, and East fit to chi-squared. See Table 1 for acronym definitions.	46
Figure 12. A zoomed in view of the right tail of the distribution from figure 9.	47
Figure 13. F_{ST} of 9850 SNPs with 119 samples from HG, VI, AA, AK, WA, BC, and East. A total of 114 outlier loci at q-value (p-value corrected for a false discovery rate) < 0.01 highlighted. See Table 1 for acronym definitions.	47
Figure 14. Distribution of F_{ST} of 10217 SNPs with 107 samples excluding HG fit to chi-squared.	48
Figure 15. A zoomed in view of the right tail of the distribution from figure 12.	48
Figure 16. Loci F_{ST} against expected heterozygosity, HG excluded for 10217 SNPs with 107 samples excluding HG. A total of 7 outlier loci at q-value (p-value corrected for a false discovery rate) < 0.01 highlighted.	49
Figure 17. Mean π against D_{XY} for NW and East in 5000bp windows (each dot is a window). Dot color is scaled from blue to red to represent the average F_{ST} in that window.	49
Figure 18 Mean π against D_{XY} for NW and HG in 5000bp windows. Dot color is scaled from blue to red to represent the average F_{ST} in that window.	50
Figure 19. Mean π against D_{XY} for East and HG in 5000bp windows. Dot color is scaled from blue to red to represent the average F_{ST} in that window.	50

Figure 20. Rolling F_{ST} , D_{XY} , and π for three populations across genomic scaffold NW_010972720.1 (~11.2 million bp). The colors correspond to groups used in our genome scans and are as follows: Green is HG vs. NW, Orange is HG vs. East, and Purple is NW vs. East. For π colors correspond to populations used in genome scan analyses: blue is HG, red is NW, and yellow is East. The star represents the region to be suspected under selection. 51

Figure 21. Rolling F_{ST} , D_{XY} , and π for three populations across genomic scaffold NW_010973238.1 (~5.5 million bp). The colors correspond to groups used in our genome scans and are as follows: Green is HG vs. NW, Orange is HG vs. East, and Purple is NW vs. East. For π colors correspond to populations used in genome scan analyses: blue is HG, red is NW, and yellow is East. The star represents the windowed region that is suspected to be under selection. 52

Figure 22. Proportion of shared SNPs across varying minor allele frequencies between populations. Populations with historic isolation and no gene flow should share few SNPs across all minor allele frequencies (NA_EU). Populations that have recently diverged or have high levels gene flow should share a large majority of SNPs across all minor allele frequencies (BC_East). However, populations that have been historically isolated pressure should share less SNPs across minor allele frequencies. 52

Figure 23. Observed heterozygosity plotted against laingi index from the 10 loci in our genotyping assays. Each symbol corresponds to a sample. See Table 1 for population acronyms. Index value of one indicates samples that are highly laingi. Samples at the top of the triangle can be inferred as F1 hybrids, and samples in the middle of the triangle as F2 hybrids. Samples approaching one or zero are considered backcrosses until they reach zero or one. The exception in this case is that these loci are not fixed between populations so there are no samples with laingi index of one..... 53

List of Abbreviations

Genotyping-by-sequencing (GBS)
Single nucleotide polymorphism (SNP)
Alexander Archipelago (AA)
Arizona (AZ)
British Columbia (BC)
California (CA)
Europe (EU)
Haida Gwaii (HG)
North America (NA)
Northwest (NW)
Vancouver Island (VI)
Washington (WA)

Acknowledgements

I would like to thank our research funding contributors from Genome British Columbia, British Columbia Ministry of Forests, Lands and Natural Resource Operations, Coast Forest Products Association, and Western Forest Products. I would also like to thank all of our sample contributors who sent us samples to perform the research described below: American Museum of Natural History, Beaty Biodiversity Museum, Burke Museum, Field Museum of Natural History, New York State Museum, UCLA Dickey Collection of Birds and Mammals, the University of Alaska Museum, the University of Arizona Ornithological Collection, and the University of Michigan Zoological Collections. We also received a great deal of help from individuals in coordinating and receiving samples, for this effort I would like to thank Janice Anderson, Bryce Bancroft, Carita Bergman, Victoria Bowes, Christina Carrières, John Deal, David Dickson, Frank Doyle, Steve Gordon, Leanne Harris, Bill Harrower, Molly Hudson, Brent Matsuda, Sue McDonald, Erica McLaren, Ildiko Szabo, Sandy Talbot, Tania Tripp, Thijs Valkenburg, Alysha Van Deft, Martina Versteeg, Warren Warttig, Berry Wijdeven, Melanie Wilson, and Kevin Winker. For funding my graduate stipend and travel expenses I would like to thank the University of British Columbia Zoology Department, John R Grace Fellowship, David W. Strangway Fellowship, NSERC Discovery Grant RGPIN-2017-03919 (to D. Irwin), and the Werner and Hildegard Hesse Research Award in Ornithology. I was lucky enough to have both Dr. Sally Aitken and Dr. Dolph Schluter serve as my committee members for this thesis. Their insight and advice improved not only this thesis but my skills as a scientist and for that I thank you both. I would also like to thank my collaborator on this project Dr. Armando Geraldés for working with

me in every phase of this research project and providing a constant source of guidance. Finally, I would like to thank my advisor Dr. Darren Irwin who was an invaluable resource, who guided me through this research and helping me improve my skills as a scientist.

CHAPTER 1: General Introduction

Since Darwin, the study of evolution and speciation has increased the understanding of why there are so many species across the globe. Yet on a planet with an estimated 8.7 million species, only a small fraction has been studied (Mora et al. 2011). Currently, the International Ornithological Congress (IOC) has over 10,000 listed bird species (IOC World Bird List 2018). Cataloguing this diversity and characterizing the processes governing it remains a monumental task. This becomes especially pressing due to the current, exceptional rates of climatic change and species extinction (Ceballos et al. 2015).

Cataloging species diversity with genetic tools has become important in the case of scenarios in which the genetic information of species is permanently lost; in addition, understanding the genetics of populations can be critical to develop the most effective recovery strategies (Pimm et al. 2006; McCallum 2008; Hedrick and Fredrickson 2010). A little over a decade ago, conservation genetics was usually limited to methods that were costly or utilized a small set of genetic markers. These markers, such as mitochondrial DNA or microsatellites, cover an extremely small portion of the genome and may not reflect the full ancestral history of organisms (Funk and Omland 2003; Toews and Brelsford 2012). A solution to this has been integrating current genomic tools such as RADseq (restriction-site associated DNA) and GBS (genotyping-by-sequencing) into applied conservation; a field currently referred to as conservation genomics (Narum et al. 2013).

Broadly, these methods work by randomly sampling the genome by using a common restriction enzyme to cut DNA into thousands of short fragments (Elshire et al. 2011; Andrews et al. 2016). These fragments are then ligated to unique sample-identifying barcodes and a

common sequencing adapter (Elshire et al. 2011; Andrews et al. 2016). These short fragments are then pooled together and sequenced. RADseq and GBS provide a significant advantage in conservation projects on two major fronts: these tools provide hundreds of thousands of markers across the genome and they are cost effective by allowing many samples to be sequenced per lane in a sequencer (Elshire et al. 2011; Andrews et al. 2016). With these technologies, conservation projects can collect information over a much larger proportion of the genome.

The use of conservation genomics is playing an important role for species of conservation concern. For example, in Tasmanian Devils (*Sarcophilus harrisii*), next-generation sequencing was used to examine if low genetic diversity was a contributing factor to their decline from transmissible facial tumors (Miller et al. 2011). Researchers studying the endangered channel island fox (*Urocyon littoralis*) on the island of San Nicholas were able to show that there are vast genomic regions with low heterozygosity that have accumulated deleterious mutations (Robinson et al. 2016). Finally, Der Sarkissian et al. (2015) used whole genome sequencing to investigate the consequences of long term captivity and conservation of Przewalski's horses (*Equus ferus przewalskii*). They found that these long-term effects resulted in inbreeding and introgression from domestic horses (Der Sarkissian et al. 2015).

As these examples show, genomic data can shed light on the complicated situations typically involved with species of conservation concern. Beyond these applications, genomics can definitively separate cryptic species or uncover that species once thought to be deeply diverged have only small regions of genetic differentiation (Toews and Irwin 2008; Bradbury et al. 2014; Mason and Taylor 2015; Toews et al. 2016). As an example, Wagner et al. (2013) used

genomic data to clarify species boundaries in cichlids in Lake Victoria. They found that as they increased the number of genetic markers in their analyses, groups of cichlids formed clear monophyletic groups (Wagner et al. 2013).

Despite having unprecedented levels of genetic information, it can still be difficult to infer more about complex evolutionary histories. For example, what kind of processes have taken place to drive differentiation? Is it the result of local adaptation, sexual selection, genetic drift, or a combination of processes? The difficulties of parsing these out is especially true for species of conservation interest as these are often understudied species that lack high quality reference genomes. A reference genome is an important component to sequencing projects. It allows researchers to align their sequences to locations in the reference genome to determine where those sequences are located. It also allows researchers to find differences in nucleotides between the reference genome and sequence data.

Without a closely related reference genome, researchers are left with few options to proceed. They can either assemble and align their sequencing data "*de novo*", align to a transcriptome, or align to a distantly related reference genome (Catchen et al. 2013; Puritz et al. 2014). In a *de novo* assembly, the chromosomal location of genetic regions is unknown. By aligning to a distantly related genome, sequence data is lost when it aligns poorly to the reference genome. Most would consider that either choice poses a significant drawback when compared to using a high-quality reference genome. For species at risk of extinction, retaining as much data as possible is important for maximizing the amount of knowledge conservation managers can obtain about a species of concern. If researchers lose data due to technical issues, they may miss critically informative genetic information.

When considering these kinds of restrictions, it quickly becomes important that researchers focus not only on the patterns of differentiation, but the processes driving differentiation. This matters for conservation because, as an example, if conservation managers want to implement genetic rescue of populations by translocation of individuals from one population to another, understanding the evolutionary relationships between those populations can inform decision-making processes (Waller 2015; Whiteley et al. 2015). If the declining population differs genetically from the source population due to local adaptation, the new arrivals may be at a selective disadvantage and may not help recover the population. By implementing conservation through a genomic framework, situations like this can be circumvented.

In this thesis we use genomic analyses to inform the conservation of Northern Goshawks (*Accipiter gentilis*) (hereafter referred to as “goshawks”), a species that is understudied, difficult to study, and has long been of conservation interest (Reynolds 1983; Crocker-Bedford 1990; Woodbridge and Hargis 2006; Dickson et al. 2014). Goshawks are large diurnal birds of prey with a holarctic distribution (Brown and Amadon 1989). They are a sexually dimorphic, with females being larger than males on average (Storer 1966). Their wingspan is three to four feet and they weigh one to three and a half pounds (The Peregrine Fund 2018). Goshawks nest and breed between May and August and have clutch sizes ranging from two to four eggs (Lewis et al. 2004; Sauer et al. 2017). They are considered generalists, consuming a mix of avian and mammalian prey, but can have specialized diets based on regional food resources (Squires and Reynolds 1997; Miller et al. 2014). Goshawks typically reside in forest

habitats with a dense upper canopy and open understory with available flyways (Reynolds 1983).

In the application of conservation genomics to understand patterns of differentiation and the processes governing it, the goshawk represents an excellent candidate for study. The current consensus is that there is a Threatened subspecies of goshawk (*A. g. laingi*) in coastal British Columbia (BC), Alaska, and Washington, whereas the other North American subspecies, *A. g. atricapillus*, range covers much of the rest of North America (Squires and Reynolds 1997; COSEWIC 2000; Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008). However, the geographic boundaries between of the two subspecies is under much debate.

The coastal subspecies was first described by P.A. Taverner (1940) based on examination of museum specimens, in which he found that birds on the archipelago of Haida Gwaii had darker plumage than birds from most other places, and he designated the subspecies name *A. g. laingi* to describe these goshawks. He stated that there may be intermediate forms on Vancouver Island (Taverner 1940).

Additional morphological analyses showed that the goshawks on Haida Gwaii are on average slightly smaller than mainland goshawks (Johnson 1989). Currently, the range is described as all of coastal BC, Haida Gwaii, Vancouver Island, the Alexander Archipelago of Alaska, and the Olympic Peninsula of Washington (Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008). This currently described range is based on habitat modeling, radio telemetry and morphometrics (Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008). Despite this being considered the range of *laingi*, there has been little genetic evidence

to identify *laingi* (Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008; Parks Canada Agency 2017).

By examining a subspecies of goshawk on Haida Gwaii, our research joins a set of studies that has investigated genetic differentiation in organisms of Haida Gwaii. Many studies have focused on the role Haida Gwaii may have played in acting as a glacial refugium (Byun et al. 1997; Demboski et al. 1999; Topp and Winker 2008; Bull et al. 2010). In an assessment of several avian species, Topp and Winker (2008) used the cytochrome *b* region of the mitochondria to assess population structure in Steller's Jay (*Cyanocitta stelleri*), Northern Saw-whet owl (*Aegolius acadicus*), Pine Grosbeak (*Pinicola enucleator*), Hairy Woodpecker (*Leuconotopicus villosus*), and Chestnut-backed Chickadee (*Poecile rufescens*) on Haida Gwaii. For all of these species, except the Chestnut-backed Chickadee, there is population differentiation between Haida Gwaii and other populations (Topp and Winker 2008).

On Haida Gwaii, there is also a unique subspecies of Black Bear (*Ursus americanus carlottae*) (Osgood 1901). To investigate this subspecies, Byun et al. (1997) used Cytochrome *b* genetic data to examine population structuring of black bears in North America. They did not find structuring between Haida Gwaii and other populations but uncovered divergent coastal and interior clades (Byun et al. 1997). Byun et al. 1997 concluded that a region near Haida Gwaii likely played an important role as a glacial refugium for the coastal clade, but this finding has been debated (Demboski et al. 1999).

1.1 Importance of Research

The interior subspecies of North American goshawks (*A. g. atricapillus*) has long been a species of management concern for wildlife agencies (Reynolds 1983; Crocker-Bedford 1990;

Woodbridge and Hargis 2006; Dickson et al. 2014). Goshawk success has been used as an indicator for overall forest health, thus simplifying forest management objectives to implementing practices that result in increased fitness of goshawks (Goodell and Seager 2015). As previously stated, the *laingi* subspecies is currently listed as Threatened due to declining populations (COSEWIC 2000). In the original 2000 decision to list *laingi* as Threatened in Canada, the main factor listed was loss of quality forest habitat (COSEWIC 2000). However, other factors have emerged as important variables for affecting the success of *laingi* goshawks. On Haida Gwaii, Sooty Grouse (*Dendragapus fuliginosus*) seems to be an important component of goshawk diet (Doyle 2005). Sooty Grouse on Haida Gwaii have recently declined, likely the result of increasing competition for food with invasive deer (Doyle 2006). Recently, Haida Gwaii goshawks have suffered additional human-caused mortality because goshawks occasionally predate chickens (Parks Canada Agency 2017). Together, these factors affecting *laingi* goshawks present challenges for conservation managers.

Management of a species or subspecies is particularly difficult when the genetic distinctness and range is ambiguous. This ambiguity affects wildlife and forest managers alike. In BC, forestry represents an important economic sector. With a lack of genetic information, it can be difficult for forest managers to ensure that their activities are not affecting organisms designated for conservation. To assist in the management of this Threatened subspecies, we set out to clarify the genetic status and range of *laingi* in BC and across North America in general (COSEWIC 200).

1.2 Thesis Goals

In this thesis, we set out to answer key questions regarding the this Threatened bird of prey in BC (COSEWIC 200). The first was, is there spatial genetic structuring of goshawks in BC? To answer this question, we examined the amount of separation between sampling regions in a principal component analysis.

Given that we did find spatial genetic structuring, our second question of this thesis was to ask, what is the current genetic range of the *laingi* subspecies? As stated previously, the current range is ambiguous. To answer this, we developed 10 *laingi* specific genotyping assays to incorporate hundreds of low quality DNA samples to find where the highest densities of *laingi* specific genotypes are located. This is important because without a clear demarcation of *laingi*, conservation managers may be led astray in identifying where *laingi* is, reducing the efficacy of recovery planning.

The third question we asked was, if there are genetically identifiable populations in BC, particularly on Haida Gwaii (the range where *laingi* was originally described), what role does selection have in their genetic differentiation? To answer this, we used two different approaches: the first method is an F_{ST} outlier tests to see if we could uncover loci selected for local adaptation and the second method was examining outlier regions in sliding windows across the genome. Given that Haida Gwaii is a small and isolated population (Parks Canada Agency 2017), finding a detectable signal of selection in Haida Gwaii would suggest that selection must be very strong to produce a signal beyond the noise of drift.

Finally, as our last question we ask what are the rates of gene flow between populations of goshawks, specifically in relation to Haida Gwaii? We estimate the number of migrants (Nm)

per generation between pairs of populations using the average frequency of private alleles, a method developed by Barton and Slatkin (1986). In addition, to assess gene flow on a more qualitative scale, we examine shared single nucleotide polymorphisms in our sequencing data and use our genotyping assays to uncover any samples with high levels of heterozygosity. These individuals would be inferred to be highly admixed and would represent more contemporaneous gene flow. Describing gene flow between populations is important, particularly for Haida Gwaii, because if Haida Gwaii has limited gene flow, a small effective population size, and is in decline, this places it in greater peril of extirpation.

CHAPTER 2: Genomic Differentiation of Goshawks in British Columbia

2.1 Introduction

Characterizing the genetic structure of populations is central to the field of population genetics. By describing the extent and geographic pattern of population divergence and diversity, interesting and complex patterns have been uncovered involving seemingly cryptic or seemingly divergent species pairs (Toews and Irwin 2008; Bradbury et al. 2014; Mason and Taylor 2015; Toews et al. 2016). For species with conservation concerns, these complex patterns of genetic differentiation can provide valuable insights for management (Miller et al. 2011; Robinson et al. 2016; Der Sarkissian et al. 2015).

In coastal British Columbia (BC) there is a subspecies of Northern Goshawk that is listed as Threatened under Canada's Species at Risk Act (*Accipiter gentilis laingi*) (COSEWIC 2000). Currently, the range of the *laingi* subspecies is considered to cover a large portion of coastal BC, Vancouver Island (VI), the archipelago of Haida Gwaii (HG), Southeast Alaska and the Olympic Peninsula (Northern Goshawk (*Accipiter gentilis*) Recovery Team 2008). The other North American subspecies, *atricapillus*, is distributed in the interior of BC and across the rest of North America; this does not include a Mexican subspecies, *apache*, whose status is contentious (Van Rossem 1938; Boyce et al. 2006). When P.A. Taverner first described *laingi* on HG in 1940, he described it as slightly darker than *atricapillus* (Taverner 1940). These differences in plumage can be difficult to discern. As a qualitative example, in Figures 1 and 2 we provide photos of the differences between juvenile HG and juvenile mainland forms of goshawks. The HG form does appear to be slightly darker in certain plumage characteristics, like

breast feathers, but these differences are not very distinct. Because of this difficulty in visual identification, the distinctiveness and range of the *laingi* form is unclear.

Earlier analyses of genetic differentiation among goshawks have produced equivocal results with regard to *laingi* (Bayard De Volo et al. 2013; Sonsthagen et al. 2012). One of the larger assessments of goshawk genetic diversity, which used mitochondrial DNA haplotypes, did not sample from HG and did not directly try to determine if *laingi* was a distinct genetic population (Bayard De Volo et al. 2013). Other research has utilized microsatellite and mitochondrial markers to assess population structure in the Alexander Archipelago (AA), Alaska (AK), and coastal BC, but again this research did not directly test the genetic uniqueness of goshawks suspected of being *laingi* (Sonsthagen et al. 2012).

In this study we asked, whether there is a genetically identifiable population of Northern Goshawks in BC that would represent *laingi*? We are especially focused on individuals from HG, as HG represents the original range of *laingi* described by Taverner (1940). To assess genetic structure within goshawks we sequenced high quality tissue samples using a modified genotyping-by-sequencing (GBS) protocol (Elshire et al 2011; Alcaide et al. 2014). Using this data, we examined the genetic differentiation of *laingi* relative to *atricapillus* using a principal component analysis. Finding genetic differentiation would be surprising as goshawks on HG are morphologically quite similar (Taverner 1940; Johnson 1989).

Following our first question, we would like to know the exact range of genetically differentiated populations. We selected a panel of candidate *laingi* informative loci from our GBS data set to develop genotyping assays. This allowed for the incorporation of hundreds of

lower quality DNA samples because these genotyping assays only require a small amount of DNA. Using this genotyping information, we clarify the range of *laingi* in BC on a detailed scale.

If there is a genetically differentiated population we would also like to understand what patterns are driving divergence. On HG in particular, our null expectation would be that drift is driving divergence as HG is a small and likely isolated population (Parks Canada Agency 2017). So, if we found signals of selection in the genome, this selection would likely have to be quite strong to be effective on such a small population and to maintain this signal despite the obscuring effects of drift. By using an F_{ST} outlier analysis and genome scans that visualize the heterogenous landscape of the genome, we infer whether selection has played a strong role in differentiation.

Due to the conservation status of goshawks in BC, we also wanted to determine rates of gene flow between populations. This is important because for small populations like HG, if the rates of gene flow are low they are at a greater risk of extirpation (Parks Canada Agency 2017). Using our sequencing data, we assess the proportion of shared SNPs between populations and quantify the number of migrants per generation (Nm) based on private alleles (Barton and Slatkin 1986). By analyzing shared SNPs in our sequencing data and levels of heterozygosity in our genotyping data, we assess current gene flow on a qualitative basis.

By investigating genetic divergence, the processes driving differentiation, and the geographic extent of goshawk subspecies, this study contributes to both informing conservation decisions and the understanding of evolution in goshawks. Due to the current decline of this subspecies and the ambiguity of its range, our results will assist the management of this Threatened bird of prey (COSEWIC 2000; COSEWIC 2013; Parks Canada Agency 2017).

2.2 Methods

2.2.1 Sampling Distribution

We received over 450 tissue samples from across North America (NA) and Europe (EU) (Table 1 and Fig. 3). We received these samples from wildlife biologists, forest managers, and museums. They were comprised of blood, tissue, toepad, and feather specimens. Samples received from other countries were done so under the guidelines of CITES permits. We focused on collecting samples from the Haida Gwaii archipelago (HG) because this is the original range of *laingi* as described by Taverner (1940). We also focused on regions of coastal BC, Vancouver Island (VI), and Alexander Archipelago (AA), as these regions are likely to be areas connected by gene flow with HG. In our widespread sampling of Northern Goshawks (Fig. 3), we included samples from the eastern United States (“East” being east of the Rocky Mountains). We also included samples from Europe (EU) samples (*A. g. gentilis*) as a likely, distantly related, population of Northern Goshawks which serves as a baseline for any genetic differentiation between HG and other populations. We also refer to the “Northwest” as a group for genome scans, this refers to populations in the Pacific Northwest i.e. AK, BC, AA, VI, and Washington (WA).

2.2.2 DNA Extraction, GBS Preparation, and Sequencing

Toe-pad and tissue samples were extracted using a standard phenol-chloroform protocol. Feather samples were extracted by combining our phenol-chloroform protocol with a feather digestion step outlined by Bayard de Volo et al. (2008). For sequencing, we selected a subset of our samples that had the highest DNA quality. However, we did include several lower

quality samples, as they were from critical regions (i.e., HG and VI). Overall, we selected 160 samples to be sequenced in two 96-well plates. Several low-quality samples were sequenced multiple times across the two plates.

For the development of our GBS data set, we followed the methods outlined by Alcaide et al. (2014) using GBS primers A and B, with restriction enzyme PstI, and modifications to certain methods. The GBS-Primer A (as described in Alcaide et al. 2014) was:

5'- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

The GBS-Primer B (as described in Alcaide et al. 2014) was:

5'-CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT-3'

After the ligation step, the volume of DNA solution product for GBS fragment enrichment varied between 4-10 μ l. We used higher volumes when samples failed at lower volumes. In the GBS fragment enrichment step, we used 18 cycles of 98°C for 10 seconds, 65°C for 30 seconds and 72°C for 30 seconds, followed by an extension of 72°C for 5 minutes and then held at 4°C. Our last modification was the size selection of fragments. Alcaide et al. (2014) selected fragment sizes from 300-400 bp. In our reactions, we obtained a DNA band in this range in the gel electrophoresis indicating that one genetic region was preferentially amplified. To adjust for this, we size selected fragments in the 400-500 bp range. GBS libraries were sequenced at Genome Quebec with Illumina HiSeq 2000/2500 paired-end sequencing.

2.2.3 GBS Data

We de-multiplexed our reads using a custom perl script (described in Irwin et al. 2016). We then trimmed our reads using Trimmomatic-0.36 to remove low quality and short reads using these settings: TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30 (Bolger et al. 2014).

“TRAILING” indicates we removed the ends of reads if quality scores dropped below 3 and “SLIDINGWINDOW” means we used a sliding window of 4 base pairs and dropped reads if the quality scores in those windows dropped below 10 (Bolger et al. 2014). Finally, “MINLEN” means that if the total length of a read dropped below 30, the read was discarded (Bolger et al. 2014).

We then aligned our reads using BWA, Samtools, and Picard tools (Li and Durbin 2009; Li et al. 2009; Broad Institute 2018). We aligned our reads to the Bald Eagle (*Haliaeetus leucocephalus*) reference genome, which comprises 1,023 scaffolds unanchored to chromosomes (Warren et al. 2014; accession JPRR000000000.1). For most samples, greater than 90% of reads mapped; however, around 40% of these reads from each Northern Goshawk sample had low mapping quality (MAPQ score < 20) and were excluded in later analyses.

To call variant SNPs (single nucleotide polymorphisms) we used HaplotypeCaller in the GATK program to produce individual GVCF files (McKenna et al. 2010). We selected variants from the SNP calling and then merged those into a single VCF file using GenotypeGVCF again in the GATK program (McKenna et al. 2010). Our data file preparation for genome scans analyses, analyzing the heterogenous structure of the genome, was similar except in the variant selection stage with GenotypingGVCF we used the ‘allSites’ flag to retain invariant sites and the -L flag to select specific scaffolds (McKenna et al. 2010).

To filter our variant-only VCF file we used VCFtools and changed filtering parameters based on the analysis being performed (Danecek et al. 2011). For our analyses, we excluded samples that failed sequencing or had greater than 80% missing data across the total 2.8 million SNPs called. For each SNP we calculated observed heterozygosity across all samples and

excluded SNPs with observed heterozygosity greater than 0.6 (Irwin et al. 2016). SNPs that were not biallelic were excluded from further analysis.

To assess whether our data set contained closely related individuals, we calculated kinship coefficients using IBDKing in the snprerate package in R, for these data sets and found that 13 samples were closely related (Zheng et al. 2012; R Core Team 2017) We removed one individual from each of six pairs and in one case two individuals from a triplet of relatives, excluding a total of 7 samples from further analysis. After this initial filtering our variant data set contained 128 samples and approximately 2.8 million SNPs (Table 3).

The Bald Eagle reference genome has not been assembled into chromosomes; rather it is comprised of 1,023 scaffolds. Thus, we did not have location information for sex specific markers. We eliminated sex-specific markers by calculating pairwise F_{ST} between males and females in VCFtools using the “-weir-fst” flag (Danecek et al. 2011). We eliminated SNPs with male-female F_{ST} greater than 0.25 as putative sex specific markers. In total, 554 SNPs were removed.

2.2.4 Genotyping Assays

To clarify the range of the *laingi* subspecies, we selected 11 informative loci for which we developed custom genotyping assays. These were selected based on high F_{ST} calculated by VCFtools, genotyping quality scores, and read depth (Danecek et al. 2011). Originally, F_{ST} was calculated using a data set that had not removed related individuals and sex specific loci and used a pairwise comparison of 12 HG samples and 116 other NA samples. The filtering parameters removed SNPs with greater than 30% missing data, less than 10 genotyping quality and a minor allele frequency less than 0.5% (F_{ST} set 1, in Table 2). In our updated filtering (see

filtering for OutFlank analysis) we calculated F_{ST} for these loci with 12 HG samples and 107 other samples (F_{ST} set 2, in Table 2). As well, we selected a subset of 12 samples from other populations to calculate F_{ST} with HG to correct for any possible bias from unbalanced sampling (F_{ST} set 2, 12 subset, in Table 2). In each case, the F_{ST} of these loci were in the top 1.2% of all SNPs (Table 2).

To develop TaqMan genotyping assays (Applied Biosystems, Thermo Fisher Scientific), we used flanking sequences of the loci chosen (Table 4). We then used these assays to genotype hundreds of additional samples (excluding some outgroups and samples with no remaining DNA). We genotyped both the high-quality samples used in GBS and low-quality feather and toe pad samples excluded from sequencing. To prepare DNA samples we used 1 μ l of a 5 ng/ μ l solution of DNA and followed the steps outlined in the TaqMan Sample-to-SNP genotyping assay protocol (Applied Biosystems, Thermo Fisher Scientific). To amplify these assays, we used a Viia 7 Real Time qPCR machine by Applied Biosystems (Thermo Fisher Scientific).

Genotyping calls were determined with two individuals (K. Askelson, A. Gerald) independently plotting the results from the qPCR reactions and were then comparing them for concordance. Through this process we determined that one of our genotyping assays failed and locus *NOGO3* was excluded from further analysis. For certain feathers we received, we had multiple feathers per nest location. For 9 of these sets, we extracted multiple feathers to determine if we could genotype two individuals from one nest site. For all but one nest, the genotypes were the same and were treated as one sample.

2.2.5 Dendrogram of Evolutionary Relationships

To create an unrooted dendrogram based on matrix distances to represent evolutionary relationships we used the `snpGdsDiss`, `snpGdsHCluster`, and `snpGdsCutTree` functions in `snpRelate` (Zheng et al. 2012). The input `vcf` file for this was filtered by excluding SNPs with greater than 30% missing data and genotyping quality less than 10, but it was not filtered on minor allele frequency or linkage disequilibrium (LD) to allow the inclusion of fixed differences in the outgroups. After filtering there were 229,678 SNPs.

2.2.6 Principal Component Analysis

For our Principal Component Analyses (PCA) we used a custom script in R (R Core Team 2017; Irwin et al. 2016). Missing data were imputed using the “`svdImpute`” method in the “`pcaMethods`” command. Our PCA scores were centered but not scaled as all of the data used in the PCA’s are of the same unit (SNPs) and do not require scaling (Irwin et al. 2016).

We filtered out SNPs with greater than 30% missing data, genotyping quality less than 10, and a minor allele frequency less than 5%. For our PCAs, to retain only independent SNPs we pruned these data sets on LD using the “`r`” method (Zheng et al. 2012) which selects SNPs with an R coefficient of 0.2 or greater in windows of 500,000bp. We performed three PCAs with subsets of individuals: 1) 124 samples including EU samples (6006 SNPs), 2) 119 samples excluding EU (6058 SNPs), and 3) 107 samples excluding EU and HG (6257 SNPs).

2.2.7 Inbreeding and Average Heterozygosity

To examine the effects of inbreeding and drift in each population, we calculated average population level expected heterozygosity and inbreeding coefficients (F_{IS}) for each individual in our North American sample (Table 1 and Figure 10). In addition to using the same filtering

performed in the PCAs, we also filtered on a minor allele frequency of 0.84% (of 119 samples) to include rare polymorphisms while removing single alleles that may be a result of sequencing error. We filtered SNPs on LD in the same manner as our PCA analyses in order to retain only independent loci and to prevent biasing estimates of inbreeding and expected heterozygosity. After filtering, we retained 24,311 SNPs for calculating inbreeding and heterozygosity in each population. Calculations for inbreeding and average expected heterozygosity per population were made by using “-het” and “-freq” respectively in VCFtools (Danecek et al. 2011). Because two SNPs were missing from all HG samples and one SNP was missing from all VI samples, these SNPs were excluded in the calculations of inbreeding and heterozygosity for all populations. Estimates of inbreeding and heterozygosity were rounded to the 3rd decimal point and calculated from a set of 24,308 SNPs.

2.2.8 Outlier Loci

To analyze outlier loci associated with local adaptation we used the R package “OutFlank” (Whitlock and Lotterhos 2015; R Core Team 2017). Our input file was filtered using the same parameters as our PCA filtering with the exception of filtering on LD. We chose not to filter on LD to allow the inclusion of any fixed differences and because OutFlank allows for some amount of LD in SNPs (Whitlock and Lotterhos 2015). We performed this analysis with and without HG, after filtering, in the analysis with HG we used 9850 SNPs and in the analysis without HG we used 10217 SNPs.

In the analysis with HG before fitting the chi-squared distribution, we trimmed 6% of the upper and lower ends of the F_{ST} distribution. After trimming, we fit the distribution of F_{ST} values between SNPs calculated by OutFlank between AK, AA, East, WA, VI, BC, and HG to a Chi-

squared distribution. SNPs that were deemed to be outliers had q-values, correcting for false discovery rates, less than 0.01 and expected heterozygosity greater than 0.1. We also performed this analysis excluding HG which was identical to our approach with HG except we trimmed 10% of the upper and lower ends of the F_{ST} as it created a good chi-squared fit.

2.2.9 Variation in Differentiation Across the Genome

To assess genomic differentiation in a sliding window across a portion of the genome, we selected a subset of 12 unrelated individuals from HG; 12 from the “Northwest” (NW) comprising a random subset of AK, AA, BC, VI, and WA; and 12 Eastern samples. To calculate Weir and Cockerham (1984) pairwise F_{ST} , genetic distance (D_{XY}), and nucleotide diversity (π) in 5000 bp (sequenced base pairs) windows over 555 scaffolds we used a custom R script (Irwin et al. 2016). We excluded indels and removed SNPs with greater than 60% missing data and 0.6 heterozygosity. We also removed SNPs with mapping quality less than 20 and read depth less than 10 among all individuals in this analysis.

To identify regions under possible adaptive selection we plot windowed mean π against D_{XY} of pairwise groups (Irwin et al. 2016). For windows of potential interest for signs of selection, we plot the scaffolds from which they originate (scaffold NW_010972720.1 and NW_010973238.1) in detail using a custom R script (Irwin et al. 2016).

2.2.10 Genotyping Assay Filtering and Analyses

For our genotyping assays we allowed up to 30% missing data per individual. We then excluded outgroups, EU, and samples with unknown locations from analysis. After this filtering, we had 386 samples (Table 1). For our assays, we calculated observed heterozygosity and *laingi* index (proportion of *laingi* alleles) for each sample (Fig. 8 and Fig. 9).

2.2.11 Gene Flow

To infer gene flow in our GBS data based on our genotyping results, we made subsets of 28 samples to compare the proportion of shared SNPs between populations (Stryjewski and Sorenson 2017). We compared BC and Eastern (18753 SNPs), HG and BC (15625 SNPs), HG and AA (17954 SNPs), and NA and EU (23972 SNPs). Filtering parameters were the same as the filtering for the OutFlank analysis except we removed SNPs with a minor allele frequency lower than 3.57% (less than 2 minor alleles in 28 samples) to include rare polymorphisms but remove single SNPs that could be a result of sequencing error.

To calculate the proportion of shared SNPs we used a custom R script to calculate how many SNPs occurred in both populations for varying minor allele frequencies divided by the total number of SNPs at that minor allele frequency cutoff. These proportions are then plotted against minor allele frequencies.

This type of analysis has been performed and explained by Stryjewski and Sorenson (2017). By examining the proportion of shared SNPs across varying minor allele frequencies one can infer recent and historic rates of gene flow (Slatkin 1985; Stryjewski and Sorenson 2017). The reasoning behind this is that rare polymorphisms that occur as new mutations should be restricted to individual populations in the absence of gene flow (Kimura and Ohta 1973). However, if polymorphisms present at a higher frequency in a population are not shared with other populations this would reflect either historic isolation or strong selection.

To accompany this, we calculated Nm with these pairs of population using a method involving the frequency of private alleles described by Barton and Slatkin (1986). This estimation was done in the genepop package in R (Rousset 2008; R Core Team 2017).

2.3 Results

Sequencing of our GBS libraries produced approximately 0.5 billion reads. Our initially filtered VCF file contained ~2.8 million SNPs. In the filtered VCF file that contained invariant sites there were ~28 million sites.

2.3.1 Evolutionary Relationships

We investigated evolutionary relationships using 229,678 SNPs (see Methods for Dendrogram filtering). All HG, NA, and EU Northern Goshawks are the most closely related (Fig. 4). Northern Goshawks are closely related to the Cooper's hawk (*Accipiter cooperi*) followed by the Sharp Shinned hawk (*Accipiter striatus*). The two most distantly related species included in this analysis are the Bald Eagle (*Haliaeetus leucocephalus*) and Red-Tailed Hawk (*Buteo jamaicensis*) (Fig. 4).

2.3.2 Genomic Differentiation

For our PCA including the EU subspecies, we retained after filtering 6006 SNPs. EU goshawks separate along PC1 and HG goshawks along PC2 (Fig. 5). Following this we removed EU goshawks from analysis and in this PCA, with 6058 SNPs, HG goshawks separate along PC1 and the remaining North American goshawks from coastal and interior regions seem to have weak population structuring along PC2 (Fig. 6). To inspect the non-HG relationships further, we removed HG samples from analysis and performed another PCA with 6257 SNPs (Fig. 7). These samples still appear to have weak population structuring. We consider this structuring weak because the differentiation among these regions relative to EU and HG populations (see in Figure 5), is quite small, and there are no clear gaps within the PCA.

For each of our populations in our GBS data set, we plotted inbreeding coefficients (F_{IS}) for each sample per population. We did not find a discernable increase of inbreeding for any of our populations (Fig. 10). The most outbred sample, represented by a negative F_{IS} coefficient, is the HG sample which is the most similar to other North American populations in Figure 6 along PC1. We also represented average expected heterozygosity per population in Table 1. HG is the lowest of all of the populations but not substantially lower.

2.3.3 Range of *A. g. laingi*

In Figure 8, we plot the *laingi* index from genotyping assays (percent *laingi* specific alleles) for each sample across our sampling populations. HG is clearly separated from the rest of the populations with the exception of one sample from the interior of BC which has a *laingi* index of 0.5.

To visualize *laingi* alleles across the landscape, we plot our sampling distribution map and color samples using a color spectrum that indicates the proportion of *laingi* alleles in each sample (Fig. 9). We see that samples with the largest proportion of *laingi* alleles are found in HG with some admixed samples in VI, AA, and one outlier in the BC interior in that it has a much higher *laingi* index relative to other samples in the same geographic region.

2.3.4 Outlier Loci Detection

To investigate loci that may be under selection we used the F_{ST} outlier detection method OutFlank (Whitlock and Lotterhos 2015) to examine outliers across 9850 SNPs that passed filtering. We conducted this test with seven NA goshawk sampling regions: HG, AA, VI, BC, AK, WA, and Eastern populations. Using the OutFlank program we calculated F_{ST} across all loci among our various populations and fit this distribution to a chi-squared distribution; chi-

squared being the distribution of F_{ST} we expect in a pure drift model (Fig. 11 and Fig. 12) (Whitlock and Lotterhos 2015).

Using a q-value threshold = 0.01, which accounts for a false discovery rate, we detected 114 loci that deviated from the chi-squared distribution and are potential candidates for local adaptation in HG goshawks (Fig. 13). With our chosen q-value, only 1-2 outliers detected should be false positives. We contrasted this with the same test with HG excluded (10217 SNPs) (Fig. 14 and Fig. 15). In this test we find only 7 loci that appear as outliers at a q-value = 0.01 (Fig. 16). This meets our expectations based on the distributions of F_{ST} seen in Figures 11, 12, 14, and 15. In Figure 11, the distribution of F_{ST} with HG has a much larger skew in its distribution, with some F_{ST} values greater than 0.8. In contrast, Figure 14 shows that without HG the distribution of F_{ST} does not go beyond 0.5.

2.3.5 Variation in Differentiation Across the Genome

To further explore signs of local adaptation in HG, we plotted 5000 sequenced bp windows of mean within-group variation (π) against absolute genetic distance (D_{XY}) between Northwest (“NW” see methods) and the East, NW and HG, and HG and the East (Fig. 17,18,19). In these figures, the most noticeable difference is the distribution of π and D_{XY} between pairs of populations. In figure 17, without HG, there are almost no windows with high D_{XY} relative to π whereas in figure 18 and 19, with HG, we see an increase in windows with high D_{XY} relative to π . Two windows (indicated by blue arrows in Fig 18 and 19) in the plots with HG show some evidence of divergent selection with high F_{ST} and low π .

To explore these regions further we plot the scaffolds they originate from in detail (Fig. 20 and Fig. 21). Neither of these scaffolds are particularly large; scaffold NW_010972720.1 is

~11.2 million bp and scaffold NW_010973238.1 is ~5.5 million bp long. In scaffold NW_010972720.1, the region of interest has lower π in HG and has high F_{ST} between HG and both NW and East, but low F_{ST} between NW and East (Fig. 20). For scaffold NW_010973238.1, the case is similar, in this region of interest F_{ST} is high between HG and the NW but not East and π is reduced in HG (Fig. 21).

2.3.6 Gene Flow

To qualitatively assess gene flow we used two methods, the first using the proportion of shared SNPs across varying minor allele frequencies. Our genotyping assay data suggests that AA, the BC coast, and Vancouver Island have a slightly higher proportion of *laingi* alleles relative to other populations, making it a likely candidate for gene flow with HG or a recent source of colonization to HG. To investigate this, we calculated the proportion of shared SNPs across varying minor allele frequencies between NA and EU goshawks, HG and AA, HG and BC, and BC and Eastern samples (Fig. 22). We chose these pairs of populations for the following reasons: NA and EU were chosen to serve as a baseline in our comparisons as populations who have long been separated geographically and should share few SNPs across all minor allele frequencies. NA and EU have been diverged in long enough isolation that there should be a substantial increase in variation of allele frequencies between these populations. The BC and Eastern comparison was used because our PCAs indicate that these sampling regions have high gene flow and/or such recent common ancestry that they have not diverged. Therefore, these populations should share a large number of SNPs across the range of minor allele frequencies. HG and BC were chosen to compare against HG and AA because the BC population is closer to HG than the East but appears to be just as differentiated from HG, judging from our PCAs.

Finally, HG and AA were chosen because our genotyping assay suggested that AA has a higher proportion of *laingi* alleles making AA likely to have a higher relative degree of gene flow with HG. VI and the BC coast were not included in this analysis as their GBS sample sizes ($N = 6$ and $N = 5$) were too small.

Between NA and EU, few SNPs are shared across all low (<0.2) minor allele frequencies examined. For BC and East, we find that the proportion of shared SNPs is moderately low at low minor allele frequencies (<0.05) but quickly approaches 1 as minor allele frequency increases (>0.05). These results meet our expectations (described above), with NA and EU having been long separated and therefore sharing few SNPs while BC and East having very weak differentiation and are therefore sharing most SNPs. When we compare populations to HG, the proportion of shared SNPs is lower than the BC-East comparison and shows a shallower increase as minor allele frequency increases. This suggests historic isolation in HG driving the lower proportion of shared SNPs compared to the proportion of shared SNPs between NW and East. Rare polymorphisms are equivalently shared in our comparisons with NA populations only, making it difficult to infer whether there has been recent gene flow or recent divergence.

For these pairs of populations, we also estimated the number of migrants (Nm) between pairs of populations (Table 3) using a method that utilizes private alleles in populations by Barton and Slatkin (1986) through the program genepop in R (Rousset 2008; R Core Team 2017). We find that Nm between HG and other populations is estimated to be 0.74 migrants per generation or lower, whereas Nm between BC and East is 1.14 migrants per generation. Nm between NA and EU is estimated to be an order of magnitude smaller at 0.03 per generation. These values meet our expectations from our shared SNP analysis. NA and EU share few SNPs

across all minor allele frequencies, so we would expect very few migrants per generation. The BC and East share many SNPs, so we would expect these populations to have the highest Nm . Finally, because of the shallow increase in shared SNPs across minor allele frequency with HG we would expect migration to be reduced with HG. Overall, we can conclude from these shared SNPs that HG has had a period of historic isolation relative to other population pairs.

To qualitatively assess recent gene flow we used our genotyping assays to see if there are many highly admixed individuals found with high heterozygosity. This would reflect more contemporaneous gene flow within the last several hundred years (accounting for older museum samples). To do this we compared observed heterozygosity and the proportion of *laingi* genotypes in our assay data by creating a “triangular plots” using observed heterozygosity plotted against *laingi* index (Fig. 23). This suggests that in our data set we have two highly admixed individuals from HG (seen in the top portion of the “triangle” in Figure 23). However, we also have two individuals with intermediate *laingi* index values but low heterozygosity which is what would be expected from advanced hybrids. It is important to note that our assays are not based on fixed differences between HG and the other populations, this means that samples may have higher heterozygosity and observed *laingi* indexes by chance. Regardless, it is likely that the two individuals that have high amounts of heterozygosity are migrants that have a large genomic portion of admixture between HG and other populations. This is because even though these loci are polymorphic between populations, the likelihood an individual would be heterozygous at multiple sites, without migration, by chance is extremely low.

2.4 Discussion

The key focal question of this thesis was to determine if there is a genetically identifiable population of Northern Goshawks in BC that would correspond to the subspecies *A. g. laingi*. With our genomic data we have shown that Northern Goshawks on Haida Gwaii (HG) separate into a distinct cluster in a PCA, whereas the remainder of North American (NA) goshawks show little differentiation (Fig. 5, 6, and 7). If *laingi* is to be described as a genetically distinct cluster, then goshawks from HG should be considered *laingi* while goshawks from all North American sampling regions should be considered *atricapillus*.

The amount of genomic differentiation between HG and other sampling regions in our PCA is surprising given that they have little morphological differences and in some cases are separated by only ~70 km over water. In contrast, we recover only weak population structuring spanning 2,000 km from Vancouver to Maine and 2,000 km from Alaska to Vancouver. This pattern of differentiation in goshawks is particularly unusual because for many species, there are boundaries of geographic isolation that separate eastern and western species and subspecies across North America (Proudfoot et al. 2007; Barker et al. 2008; Carling and Brumfield 2008). However, these barriers to gene flow for other organisms seem to pose little restriction to gene flow in goshawks (Table 3).

To place the genomic distinctness of HG into context, we assessed the evolutionary relationships between NA and EU goshawks with outgroups (Fig. 4). Northern Goshawks in HG were nested within NA, and NA goshawks were closely related to EU goshawks. Northern Goshawks as a whole were closely related to the Cooper's Hawk, and Sharp-Shinned Hawk. The Bald Eagle and Red-Tailed hawk were equally the most distantly related to Northern Goshawks.

These relationships correspond to previous work done on the Accipitridae family using mtDNA (Wink and Sauer-Gür 2004). The nested nature of HG in this dendrogram indicates that HG likely diverged from NA and not from a possible colonization event from EU.

Another main objective of this study was to clarify the range of *laingi* goshawks. If genomic relationships are a basis for defining subspecies, then our results conclusively demonstrate that *A. g. laingi* is restricted to HG (Fig. 8 and 9). One notable exception to this is a bird with an intermediate proportion of *laingi* alleles recorded as being collected in Vernon, BC. Vernon is ~950 km from HG in the interior of BC in a biome (warm and arid) that is significantly different from coastal BC. This bird is likely the product of a long-distance dispersal event and/or a complex admixture history. This sample was a museum specimen from the 1940s, so it is possible that this bird may have been used in falconry and later released in a different location.

This clarified range description of *laingi* is a critical finding for the application of proper management for this subspecies. Previous work has designated the range of *laingi* as comprising HG, VI, AA, the BC coast, and the Olympic peninsula (Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008). Our results show that if *laingi* is defined as the genomic signature common in HG, the range of *laingi* is considerably smaller than previously described.

In this thesis, we also wanted to understand the processes governing differentiation and gene flow between populations. While we believe we have sufficient results to discuss these topics, they are more speculative than our conclusive results regarding the differentiation between HG and other NA populations or the range of the *laingi* genetic cluster. For the

remainder of this discussion, we discuss the processes driving differentiation and gene flow but issue a word of caution in the interpretation of these results.

HG is considered to be a small and isolated population (Parks Canada Agency 2017). Our null expectation is that genetic drift in this population would be strong and the impact of selection would be weak, given the strength of genetic drift. In a purely neutral model of a recently diverged population, we would expect our distribution of F_{ST} to approximate a chi-squared distribution with few if any high F_{ST} markers (Lewontin and Krakauer 1973). This neutral expectation is a poor fit when using the F_{ST} distribution with HG (Fig. 11 and 12). In this distribution we find a large skew with many high F_{ST} markers indicating 114 outlier loci (Fig. 13). When compared without HG, the F_{ST} distribution fits the neutral expectation quite well, only finding 7 outlier loci (Fig. 14, 15, 16). This provides evidence that selection has acted on some parts of the genome, however, this does not preclude the effects of drift. What we can infer is that for this signal of selection to persist in HG despite drift, selection is likely to be strong.

Following this we investigated signs of selection using genomic scans, made difficult by the fragmented nature of our reference genome. We find that some windows in our scan seem to show signs of selection with low π relative to D_{XY} , indicating selection has reduced standing genetic variation at these sites (Fig. 18 and 19). However, when we explore these scaffolds in detail, the windows of interest seem to be driven by a small set of SNPs and the reductions in π are subtle relative to the surrounding region (Fig. 20 and 21).

Other possible non-selection based processes are unlikely to explain the observed F_{ST} distribution with HG. Inbreeding could affect F_{ST} by reducing heterozygosity in HG, however, inbreeding is not supported in our data, as no population seems to exhibit signs of inbreeding

(Fig. 10). Background selection is unlikely to affect this distribution as recent research into this topic show that F_{ST} is robust against background selection (Matthey-Doret and Whitlock 2018). Most demographic effects are also unlikely to explain such a skew in F_{ST} distribution as OutFlank is robust to demographic effects, with the exception of allelic surfing due to range expansion (Klopfstein et al. 2005; Whitlock and Lotterhos 2015). However, allelic surfing remains an unsatisfactory explanation for the observed F_{ST} with HG. The first reason is that many variables control the success of a surfing event making it unlikely for any individual mutation to arise to high frequencies through range expansion (Klopfstein et al. 2005). Secondly, for this to be plausible with HG, HG would need to be the recipient of said range expansion. Large amounts of research have shown that for numerous taxa, HG is a glacial refugium and source of range expansion rather than a receiver of colonization (Shafer et al. 2010).

Our final question was to understand rates of gene flow between populations of goshawks, specifically in HG. Our estimation of gene flow using private alleles found that there is less than one migrant per generation to HG (Table 3). In our qualitative assessment of gene flow, the shared SNPs analysis indicated that HG has likely been historically isolated sharing a reduced number of SNPs with other populations (Fig. 22). As well, our genotyping assays only found two individuals that were highly admixed with *laingi* specific alleles on HG (Fig. 23). If these highly admixed individuals are a result of several generations of migration, then the estimation of less than one individual per generation to HG seems to be accurate.

Overall, we find that in BC there is a clearly genetically differentiated group of Northern Goshawks on HG. We also find, through genotyping hundreds of samples across a vast geographic range, that this differentiated group is restricted to HG. By analyzing the distribution

of F_{ST} we find compelling evidence for selection to be an important factor in effecting parts the genome of goshawks on HG. Finally, we find through multiple analyses that HG is likely to be quite isolated from gene flow, which places HG in greater peril of extinction. However, we stress that our results on selection and gene flow remain on the more speculative side, further work on these topics are needed to corroborate our findings.

CHAPTER 3: General Conclusion

The work in this thesis represents the integration of conservation genomics, evolutionary theory, and multiple genetic tools to better understand a Threatened subspecies (COSEWIC 200). We used next generation sequencing technology to generate SNPs from across the entire genome to determine genetic structuring. From this, we were able to utilize 10 informative markers to develop *laingi*-specific genotyping assays and incorporate hundreds of lower quality DNA samples. This strategy and our resulting data exemplify the importance of utilizing multiple data types and analyses to uncover complex patterns of genetic differentiation. Our results make a strong case for this type of study when working with a species for which high quality DNA samples may be limited, like Northern Goshawks.

The focal conclusion found in this thesis is that on the archipelago of Haida Gwaii, there is a genetically distinct population of Northern Goshawks. Goshawks residing in the remainder of continental North America appear largely unstructured genetically, despite having the Rocky Mountains as a potential barrier to gene flow.

This finding is important for a myriad of reasons interesting to the general public and conservation biologists. We are the first to document the genomic difference in Haida Gwaii goshawks. We also uncovered a system in which over an extremely large geographic area, an organism shows little genetic structuring but over a very narrow oceanic barrier we detect significant population structuring.

Another critical discovery in this study was that the range of *laingi* is largely restricted to Haida Gwaii. This may have significant implications to the management of this subspecies. The first is the new use of genotyping assays. Previously, conservation managers had no genetic tool

to differentiate *laingi* and *atricapillus* goshawks. Now that such a tool has been developed for the purposes of this research, conservation managers can use the assays we have developed to better inform their own projects. The second major implication of our study regards the legal status of *laingi* in BC. The current range for this Threatened subspecies is currently considered to encompass a large geographic area of coastal BC, Vancouver Island, the Olympic Peninsula, and Haida Gwaii (COSEWIC 200; Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008). If *laingi* is considered to correspond to the genetic cluster we have detected on Haida Gwaii, then it is likely *laingi* would be a candidate for Endangered status under criteria set by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) (COSEWIC 2015): the number of spatial locations *laingi* occurs is now one (just Haida Gwaii), and there are less than 250 individuals on Haida Gwaii (Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team 2008; COSEWIC 2015). From a potential change in status and distribution of *laingi*, conservation efforts may be more focused on Haida Gwaii rather than Vancouver Island and coastal BC. This may require changes to the current recovery projects taking place with the goal of increasing the number of *laingi* goshawks.

While the evidence from our genetic study shows a clear genetic cluster in HG that may be a result in at least part due to selection for local adaptation, our study has important limitations. For the remainder of this section I will discuss limitations to our study and future directions.

In our study we also found evidence of selection shaping the genome in the small and isolated population of HG; however, to directly test for the type of selection driving differentiation in Haida Gwaii would be extraordinarily difficult. Typically, most tests focusing

on local adaptation utilize reciprocal transplants. Doing so with goshawks would not be possible. Alternative methods would be to analyze this by proxy. Current evidence suggests that goshawks in Haida Gwaii are reliant on Sooty Grouse (Doyle 2006). However, across the range of goshawks, their diet can be dynamic indicating that having unique food sources may not be a good metric of adaptation to regional location (Miller et al. 2014). If the status of goshawks change, increased recovery on Haida Gwaii could benefit from the use of GPS tags on goshawks. From this, conservation managers could be able to more accurately estimate rates of immigration and then survival of those immigrants. While we did estimate geneflow with HG to be low using rates of Nm and other qualitative methods, a more direct estimate of contemporary migration is warranted to confirm our estimates. For example, if it was found that there are high immigration rates to HG, this could potentially lend support for local adaptation in Haida Gwaii because we would expect gene flow to obscure genetic differentiation and lessen the overall risk of extirpation of goshawks on Haida Gwaii.

Another limitation to this study is the use of a Bald Eagle reference genome (Warren et al. 2014; accession JPRR000000000.1). This poses drawbacks in understanding local adaptation, selection, and genomic differentiation in goshawks and in the population on HG. Through evolutionary time, many lasting genomic changes are likely to have taken place between goshawks and the Bald Eagle, such as mutations and rearrangements. If this is true, it is likely that we have missed a large portion of the goshawk genome (many reads in our GBS data set mapped, but a large portion mapped with low mapping quality). This poses an issue because with the lost data, we are likely missing out on genomic regions that are evolving the fastest:

underestimating our measures of genomic differentiation (this problem would affect all comparisons of goshawk populations).

Despite any drawbacks, the Bald Eagle genome remained extremely useful for our purposes. In our analyses with only variant sites we utilized 6,000 to 24,000 SNPs depending on the analysis, and in our genomic scans, which included invariant sites, we retained approximately 28,000,000 sites. It is extremely unlikely that with the inclusion of the reads lost to poor mapping quality that our primary conclusions would change. The inclusion of these lost reads would improve our genomic scans and give us a better understanding on the heterogenous nature of the genome in goshawks.

Moving forward, we hope to develop a high quality Northern Goshawk reference genome. Sequencing of high quality reference genomes has reached unprecedented accessibility with price reductions by companies scaling these technologies. The development and use of a reference genome would significantly improve this study. First, we would likely recover many of the reads lost by mapping to the Bald Eagle, including regions that are the fastest evolving between Haida Gwaii and mainland populations. Secondly, by developing a Northern Goshawk reference genome with current sequencing methods, we would likely be able to have genomic scaffolds anchored in chromosomal positions improving our precision on where genomic regions of differentiation occur. Additionally, with this genome we could have a better understanding of the genetic basis of outlier loci regions and the linkage relationships of those outliers.

Thanks to our contributors, our sample size is exceptional for a bird of prey that is notorious for being difficult to study. Despite this, we lack samples in areas that could be highly

informative. One region missing is the Rocky Mountains areas of Alberta, Montana, Idaho, Wyoming and Colorado. We also do not have genomic data from the American Southwest or Mexico. Sampling the Southwest and Mexico would be important as this region is home to a contentious goshawk subspecies *A. g. apache* (Van Rossem 1938; Boyce et al. 2006). Having GBS data for this region could reveal a complex evolutionary history of differentiation in a potentially third subspecies of Northern Goshawks in North America.

To conclude, our research has provided detailed insight in the genomic differentiation and range of a difficult to study bird of prey that is of conservation interest. We also demonstrate that this differentiation may be due, at least in part, to strong natural selection on Haida Gwaii which is seen in the strong skew in F_{ST} that contain outlier loci despite apparent gene flow. This provides insights on how species are maintained and separated through space and time. In a small and isolated population, the expectation would be that drift is playing a dominant role in any genetic differentiation obscuring signals from selection. In our data, we find that despite Haida Gwaii being a small and isolated archipelago population we recover signals of selection beyond the noise of drift.

We hope that with future use of a high-quality goshawk reference genome that we can learn more about the heterogenous landscape of the genome and how this population remains differentiated despite apparent gene flow. Lastly, this study represents a remarkable collaboration between academic researchers, conservation managers, forest managers, and private funding agencies. Our findings are a testament to this collaborative and cohesive process focused on scientific pursuits for the better understanding of a Threatened bird of prey (COSEWIC 200).

Table 1. All samples received as shown by populations. Columns “GBS” and “Genotyping” indicate how many samples were used in each methodology. For the populations in which we have GBS data we’ve included the average expected heterozygosity for that population and the standard error of the mean.

Populations	N received	N GBS	N Genotyping	expected Het	St. Error
Alexander Archipelago (AA)	20	17	20	0.065	0.00075
Alaska (AK)	23	26	23	0.067	0.00073
AK coast	20	NA	20	NA	NA
Arizona (AZ)	11	NA	9	NA	NA
British Columbia (BC)	106	16	103	0.065	0.00075
BC coast	89	NA	61	NA	NA
California (CA)	5	NA	4	NA	NA
East	41	29	40	0.072	0.00072
Haida Gwaii (HG)	22	12	14	0.054	0.00071
Vancouver Island (VI)	107	6	78	0.063	0.00082
Washington (WA)	11	13	11	0.065	0.00076
WA coast	3	NA	3	NA	NA
Outgroup	9	4	NA	NA	NA
Europe (EU)	9	5	NA	NA	NA
Unknown	3	NA	NA	NA	NA
Total	479	128	386		

Table 2. F_{ST} , allele frequencies and Missing data for *laingi* specific genotyping assays. Set 1 indicates the original filtering parameters and set 2 indicates updated filtering (see methods). “Freq” indicates allele frequency and in these columns Tag indicates the allele frequency in the genotyping data. Δ “Freq” is the difference between frequencies found in GBS and genotyping data. NOGO3 is marked with an asterisk because this assay failed to amplify and was excluded from further analysis.

Assay ID	Genome Location	FST Filter Set 1	FST Filter Set 2	FST Filter Set 2, 12 each	Freq HG	Freq NoHG	Freq HG Tag	Freq NoHG Tag	Δ Freq HG	Δ Freq NoHG	% Missing HG GBS	% Missing HG Tag
NOGO2	NW_010972611:746549	0.95	0.95	0.92	0.92	0.02	0.85	0.02	0.07	0.00	0.25	0.07
NOGO3*	NW_010972593:1636141	0.95	0.95	0.93	1.00	0.02	NA	NA	NA	NA	0.33	NA
NOGO4	NW_010973166:279399	0.94	0.94	0.95	0.94	0.03	0.88	0.03	0.06	0.00	0.58	0.07
NOGO6	NW_010972587:15841473	0.93	0.93	0.87	0.88	0.02	0.86	0.03	0.02	-0.01	0.50	0.00
NOGO7	NW_010973284:18188044	0.93	0.95	0.88	0.95	0.02	0.86	0.06	0.09	-0.03	0.42	0.00
NOGO9	NW_010973052:6718156	0.92	0.92	0.86	0.86	0.02	0.79	0.05	0.07	-0.02	0.33	0.00
NOGO12	NW_010972731:3090459	0.91	0.92	0.78	0.80	0.01	0.75	0.02	0.05	-0.01	0.17	0.00
NOGO16	NW_010972436:29246950	0.9	0.89	0.77	0.78	0.02	0.57	0.03	0.21	-0.01	0.25	0.00
NOGO21	NW_010972626:620268	0.88	0.87	0.70	0.71	0.02	0.69	0.01	0.02	0.01	0.42	0.07
NOGO28	NW_010972631:1004654	0.86	0.85	0.65	0.69	0.02	0.71	0.02	-0.03	0.00	0.25	0.00
NOGO29	NW_010972831:1904284	0.86	0.85	0.71	0.72	0.03	0.67	0.02	0.06	0.01	0.17	0.14
Top % SNPs		0.30	0.30	1.14								

Table 3. Nm between population pairs from the shared SNP analysis. Acronyms are as follows: Haida Gwaii (HG), Alexander Archipelago (AA), British Columbia (BC), Europe (EU), and North America (NA).

Population Pair	Nm
HG and AA	0.74
HG and BC	0.61
BC and East	1.14
EU and NA	0.03

Table 4. Flanking sequences of laingi specific genotyping assays.

Genotyping Loci	Flanking Sequences
NOG02	GACCTGCTGGAGCAGGAGGAGTTGGAGGCAGGAATCCTCAT GGTGAACAAGAAGGGC[G/A]AAGAGGAGCAAGCCAGTGAGTG GCTGGGGTCAACAT
NOG03	AAAGGAGAGATGCTCATAAGCATGCTGCTGACTGG CATTTCATG[C/T]CTCTGTAATCAAGGACAAGGTCAGGTGCTGGC CAACATGTCAGTTTGAAAAATATCACCAAGTCAGTGAAGCAA- CACTGCATTTACTTGCTCTATACCCTCATCTCATGGAATCAG
NOG04	TCAGTGGAAAAAGAGCAGCCTGTGAAGAGAAGGATATTTTA CTGCAGGACATCGGACCCAGTCACAGCTCTTGCCC- CAAAAAGCACTTGCCACTGCCTGGCCCCACCGGCCAGGGCTG GTCCAGGGCTGGCTGCATCCCCAGT[G/A]CCCGGACA GCGGCNNNNNNNNNNNNNCCCCAGCTGGGGGA- CACCTGGGGACGACTGTCCCCTGCTTCCC- CCCCGTGCATGACCCAGCAGCTGCCATCCCTCCGCTGTGCTTGCGGA GGGG
NOG06	AGGGGCTGCCACGCTGCAGCACAGCGGTGCTGTACTC[G/A]GTCC CACTCTGCTCTGTTTCAGGCTGGGCCACCCAAACCCTGCAACAGC- CGTACTTCTATGGATGCATTCATGTTTCCCACTCCTCCCTC
NOG07	GCTGCAGGGACATCACAAGTGTCCAGTGGC[C/A]ATGGGCACTACA TGCCTGAGCTTCCCTGCTGCCAGACGGATTTCTATGCCATGAAGT- CACTGCATG
NOG09	AATTGCAAGCTTATTGAAAACCATCAGCATATCCACCTGGAACCTG AAAGGTGAGGAAGAG[G/A]GCTTGGTGAACCCAGATAGCCTCAGCC TCATCAGCCCACGAGCAGCATGCT
NOG012	GCACAAAGCTATTTGGTAAAGGTATTTAATGGTGATTAC TGGGAAAATTCACAGCTTCCAGTGGTGTGCA- GATTCACAGTTTCACTG[C/T]GAGCAGCGTTGCTCTCCAG CACAGTCTGCACACCGCCTGCAACATCAA
NOG016	CAGGTATTTGCCACCTGCTGCATTTCTACCACAAGTGCATCTG CTCACAAGCAGCAGCGTGAACATGGGAAAAGC[G/A]ACACGCTGG GAGCAGAGGAGAAGACAGCAACACTCACCGCAC
NOG021	TGCAGGAAGAGCTGGAGCGCTTTCTGTCTCACTGGGGAGCAGTTCA CTCTGGTTAAATGGGCTCCATGCAGTCTGGCT GCTCAGCAGGAAGCATTTGATGCTGTCACGCTCTNNGT- GTGGCTCTGCTACTTCTGGGGGGCACAGAGAAAACCCTGA GGTTGAATCCTCATCTCCCTCCTCCTGTCTC
NOG028	GGCAC[C/T]GGGGCACCAGCCTGTGCTGCCA GCTCGTCTGGTGTG GTTTTTACGGGAATACCTGACCTAATTCCTCGGAGCTAAAATACCTGA CA[A/G]TAACAATTGAGCAACTGCCTCCTGGATGCAGGATGC GAAATAAGGATCAGT
NOG029	GCCCTGGGGAGGTGGCCCCACCCGTGTTGAAGACATTGATGTTGCCCC [G/A]TACATCCTGGTATGTGCAGAAAATCAAGTC



Figure 1. Dorsal images of juvenile HG goshawk (above) and juvenile mainland goshawk (below). Photos are color corrected.



Figure 2. Ventral images of juvenile HG goshawk (above) and juvenile mainland goshawk (below). Photos are color corrected.

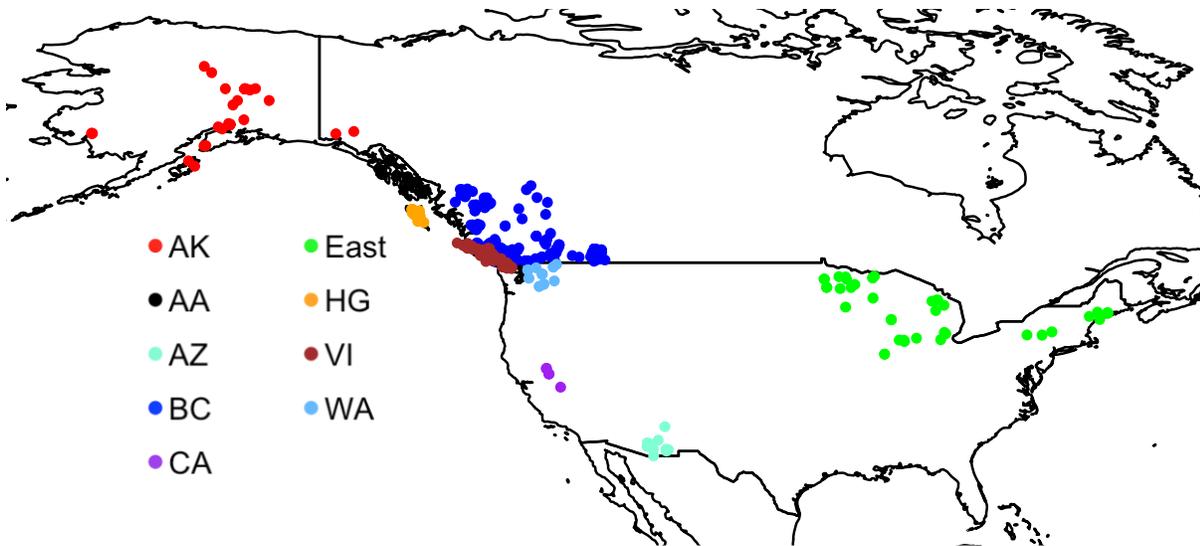


Figure 3. Total samples received of all goshawk samples in North America by population. Defining populations follows that of how GBS populations were defined. Acronyms explained are AK (Alaska), AA (Alexander Archipelago), AZ (Arizona), BC (British Columbia), CA (California), HG (Haida Gwaii), VI (Vancouver Island), WA (Washington). Each dot corresponds to an individual bird.

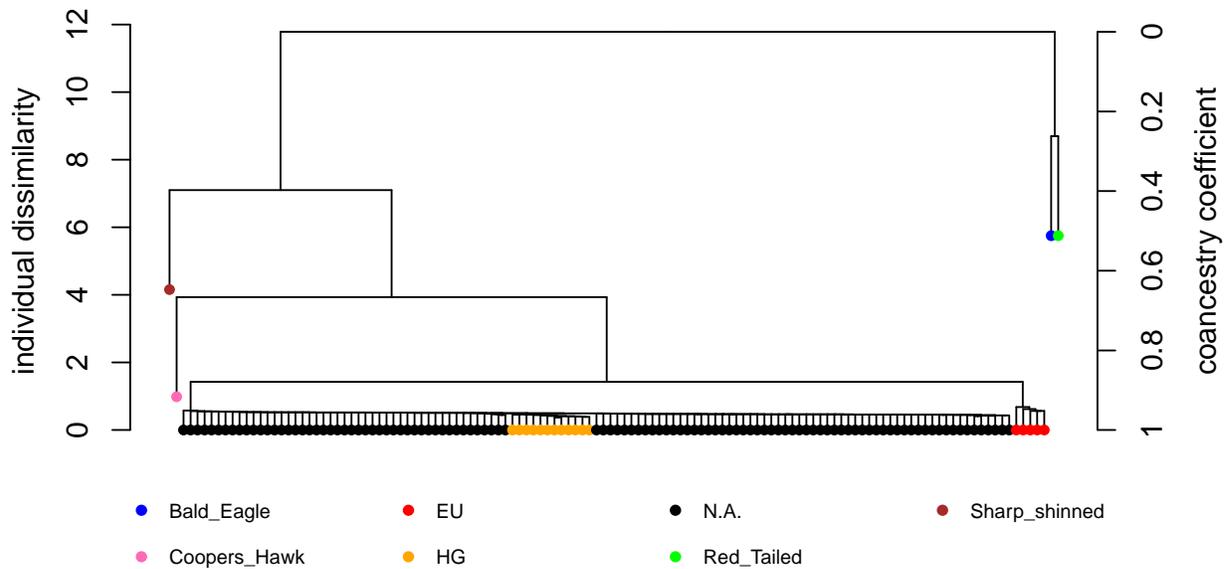


Figure 4. Dendrogram of evolutionary relationships of Northern Goshawks and outgroups. N.A. are North American Goshawks, EU are European, and HG is Haida Gwaii. Each dot corresponds to an individual bird.

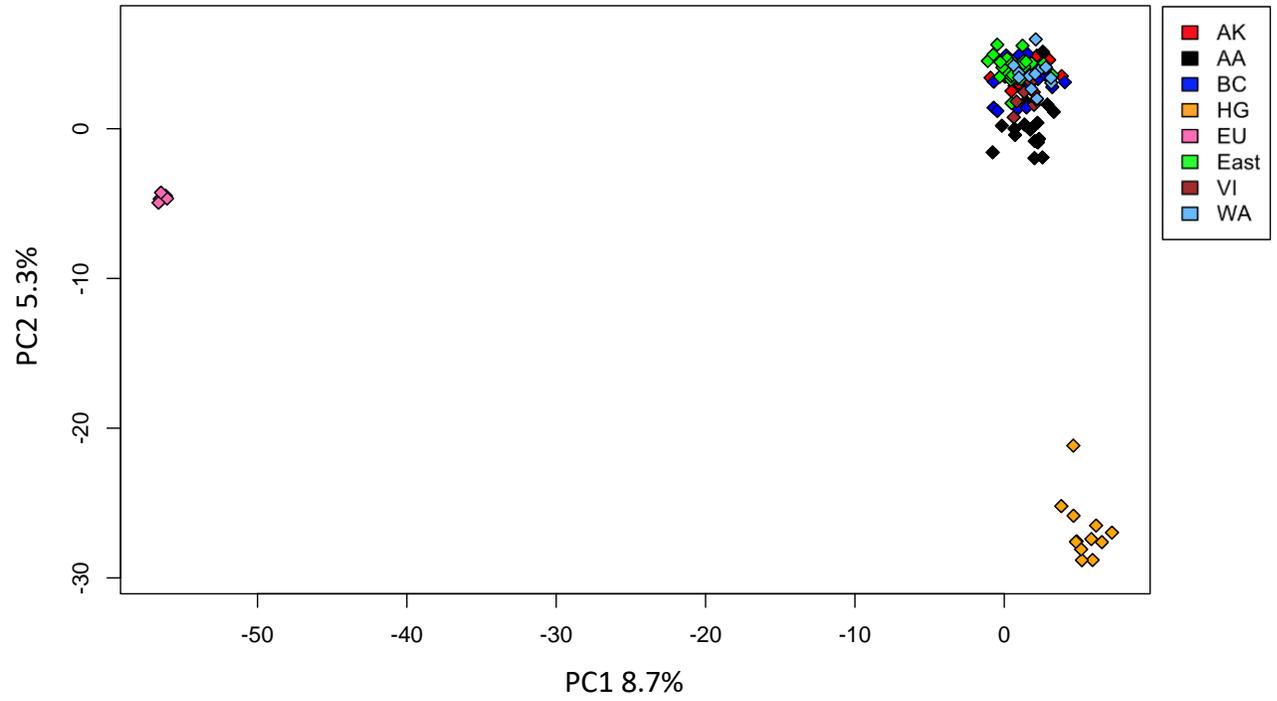


Figure 5. PCA of 124 samples excluding outgroups using 6006 SNPs. See Table 1 for acronym definitions. Each diamond corresponds to an individual bird.

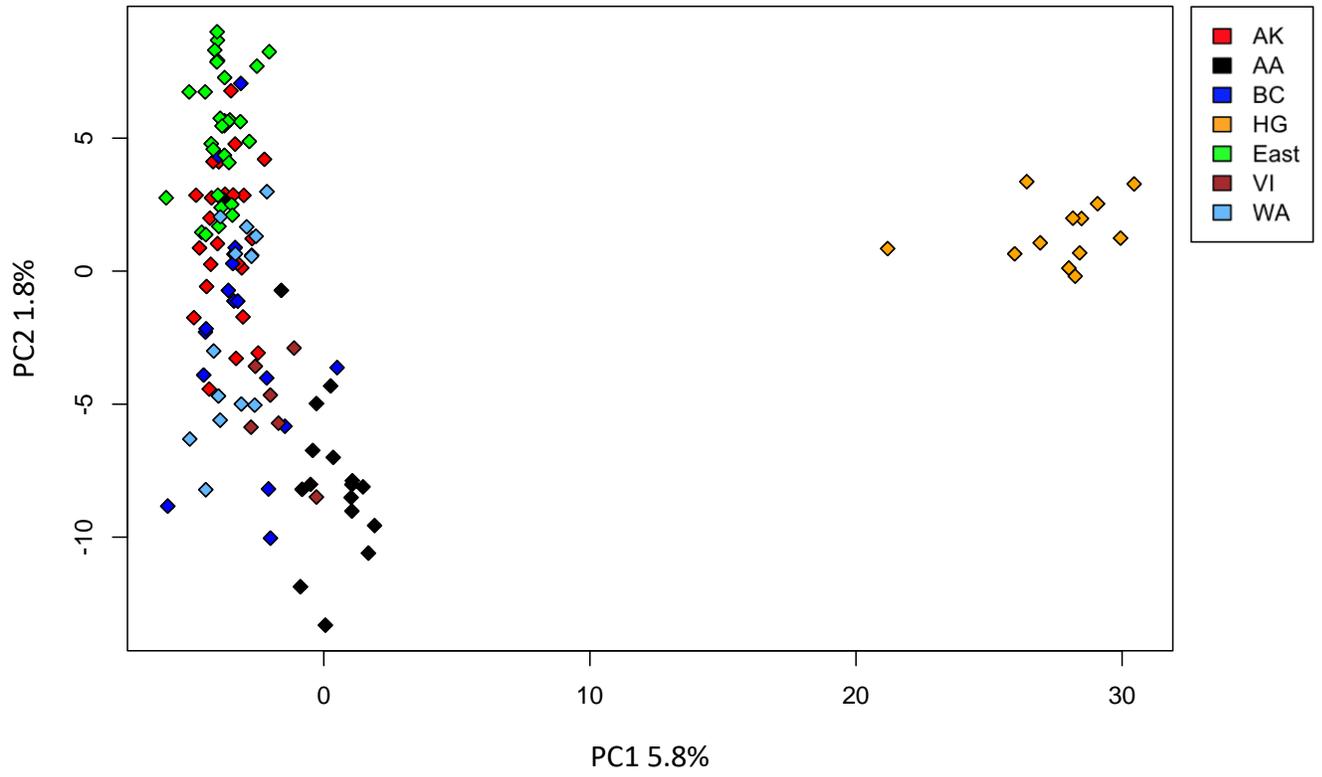


Figure 6. PCA using 6058 SNPs of 119 samples excluding European and outgroups. See Table 1 for acronym definitions. Each diamond corresponds to an individual bird.

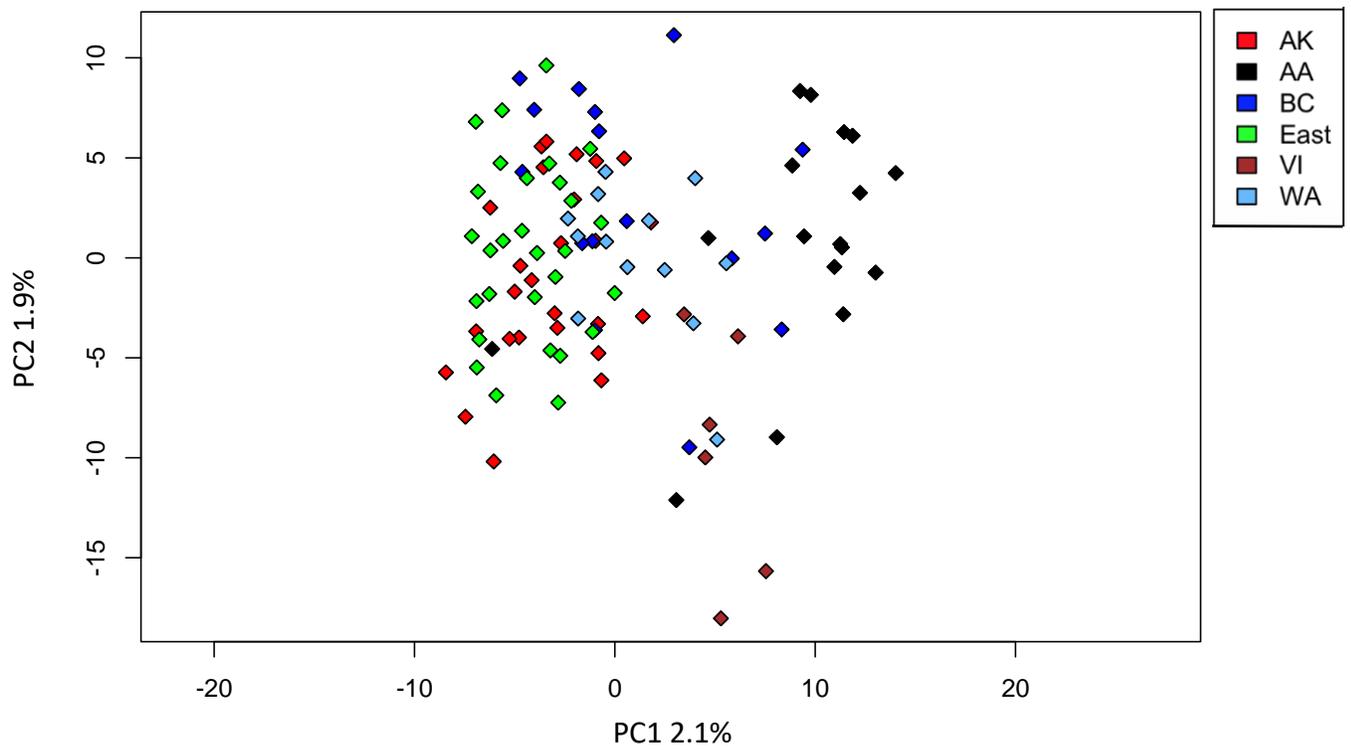


Figure 7. PCA using 6257 SNPs using 107 samples excluding HG, European, and outgroup samples. See Table 1 for acronym definitions. Each diamond corresponds to an individual bird.

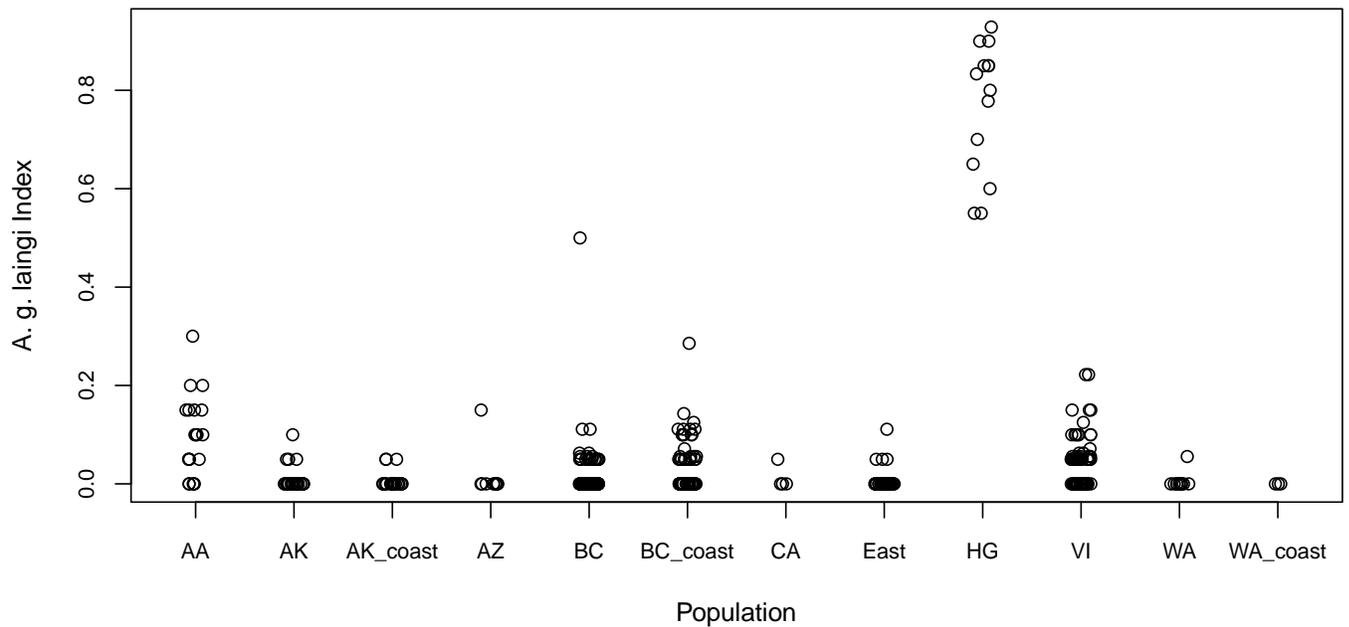


Figure 8. Distribution of *laingi* index across each population. AZ is Arizona, CA California, and YK is the Yukon. Each circle is an individual bird.

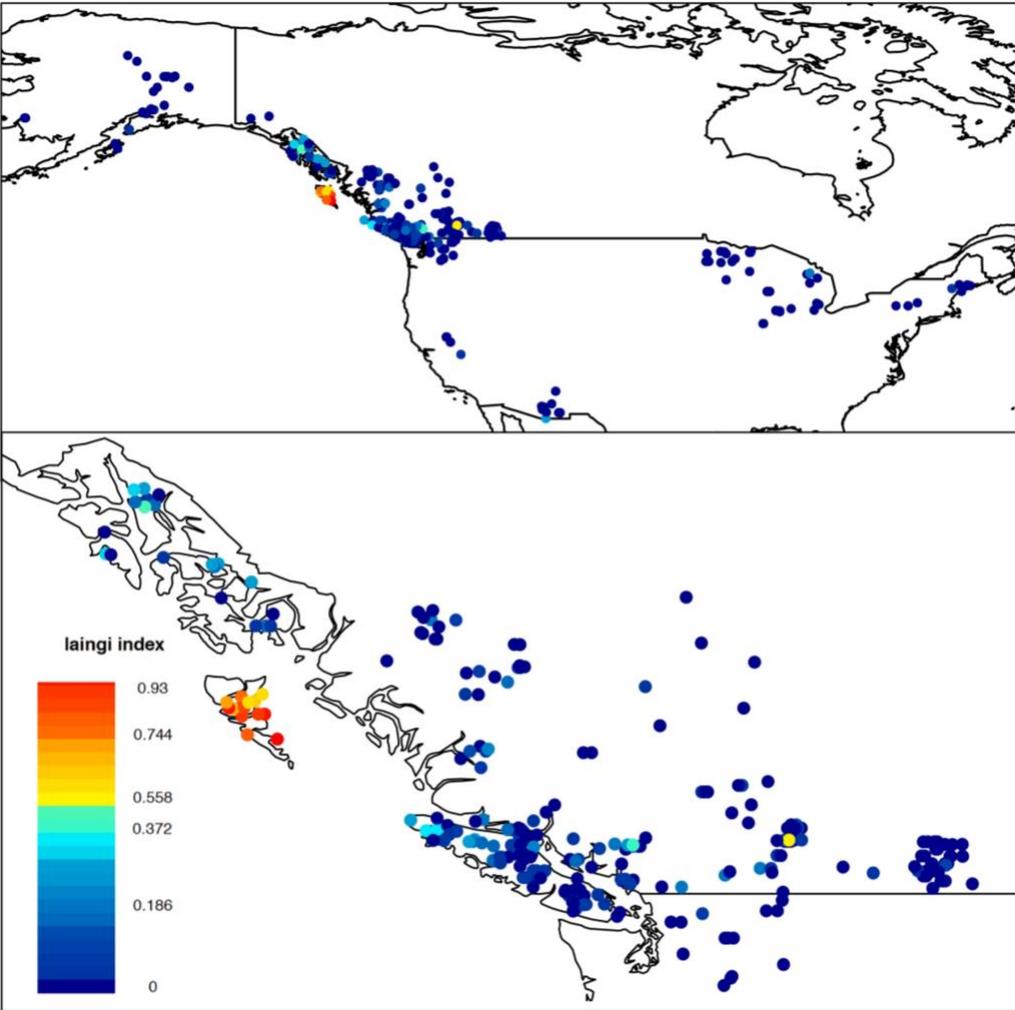


Figure 9. Distribution of samples used in TaqMan assays. Points are colored from blue to red indicating percent *laingi* with red being highly *laingi*. Range of *laingi* index does correspond from 0-1 as loci are not fixed between populations. Each dot corresponds to an individual bird.

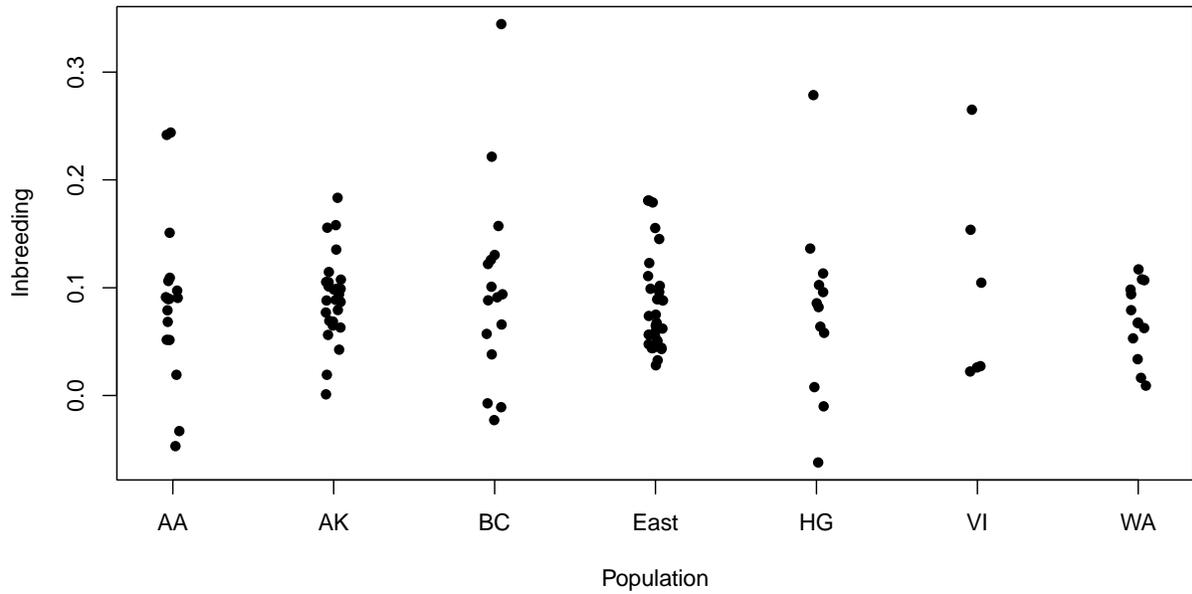


Figure 10. Inbreeding coefficients (F_{IS}) of 119 samples across GBS populations using 24,311 SNPs. Each circle corresponds to an individual bird.

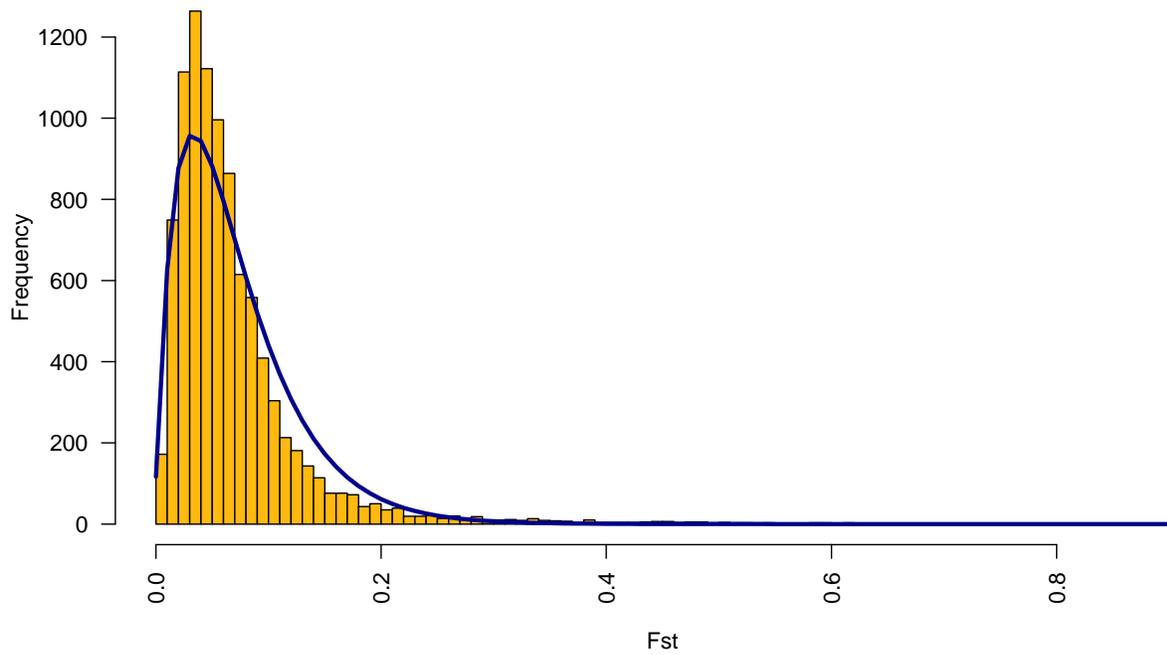


Figure 11. Distribution of F_{ST} among 9850 SNPs with 119 samples from HG, VI, AA, AK, WA, BC, and East fit to chi-squared. See Table 1 for acronym definitions.

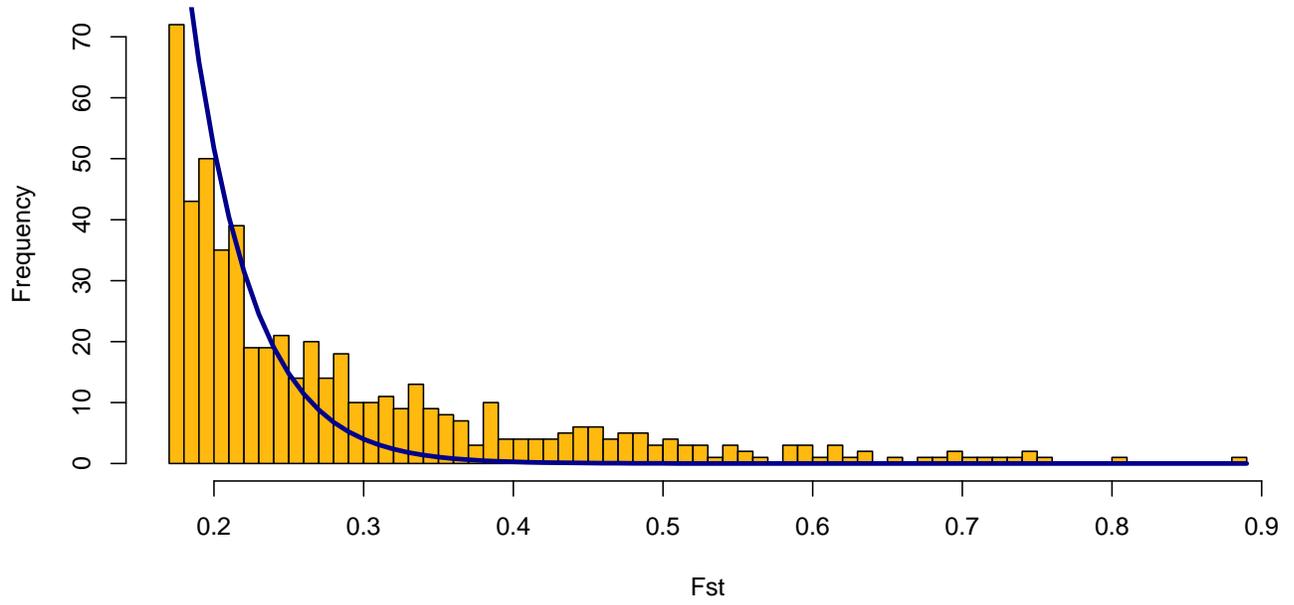


Figure 12. A zoomed in view of the right tail of the distribution from figure 9.

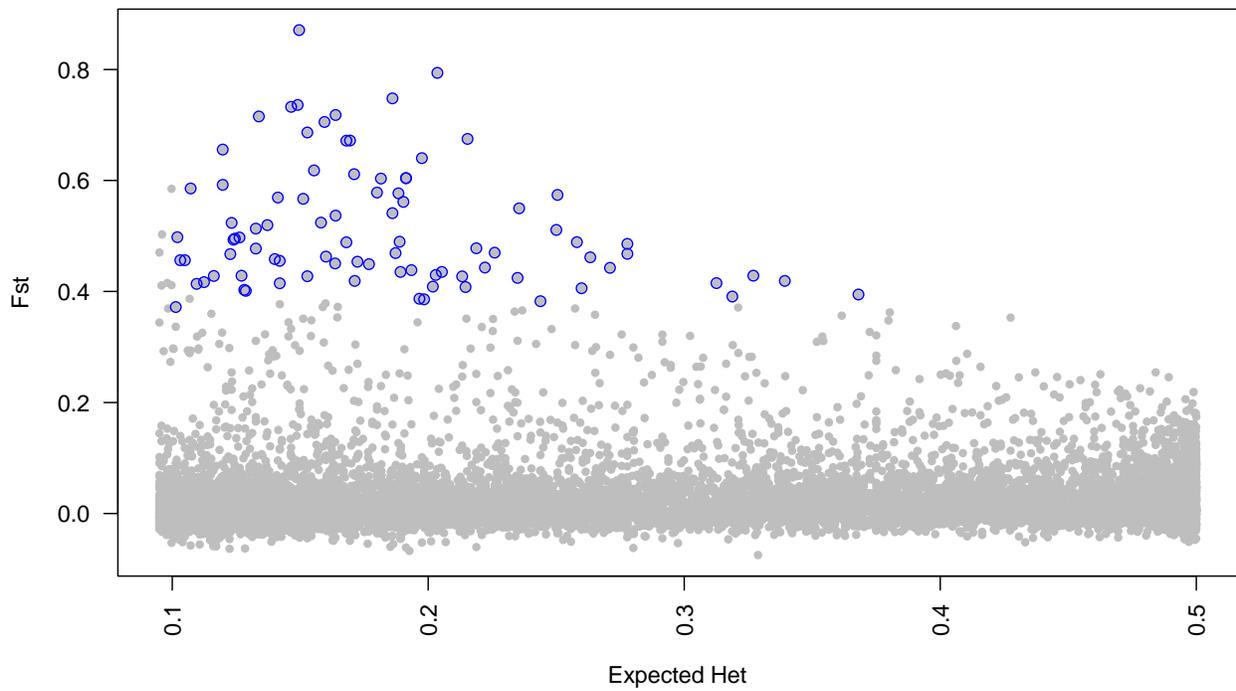


Figure 13. F_{ST} of 9850 SNPs with 119 samples from HG, VI, AA, AK, WA, BC, and East. A total of 114 outlier loci at q -value (p -value corrected for a false discovery rate) < 0.01 highlighted. See Table 1 for acronym definitions. Each circle is an individual SNP.

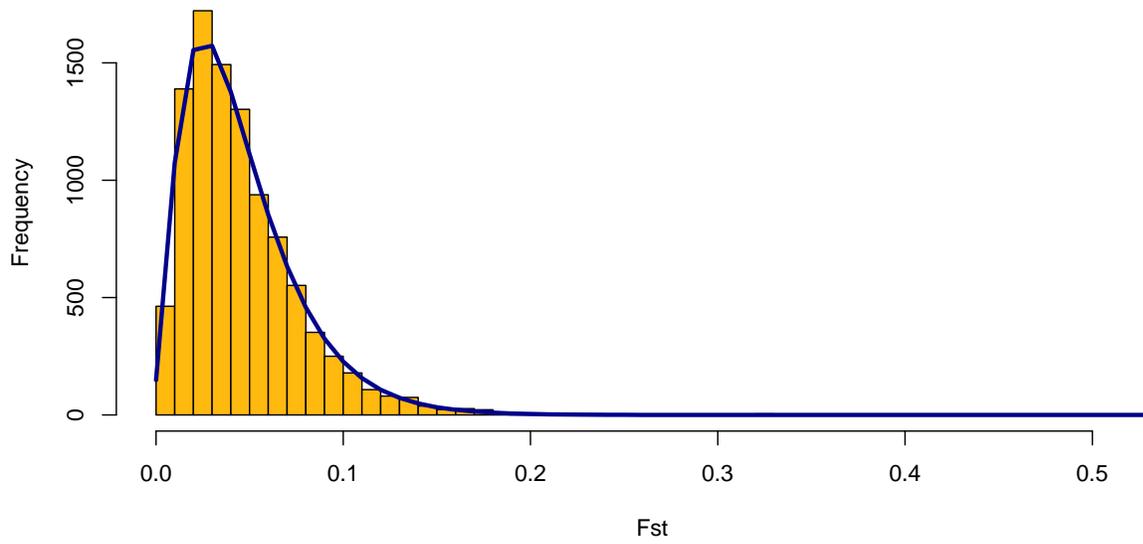


Figure 14. Distribution of F_{ST} of 10217 SNPs with 107 samples excluding HG fit to chi-squared.

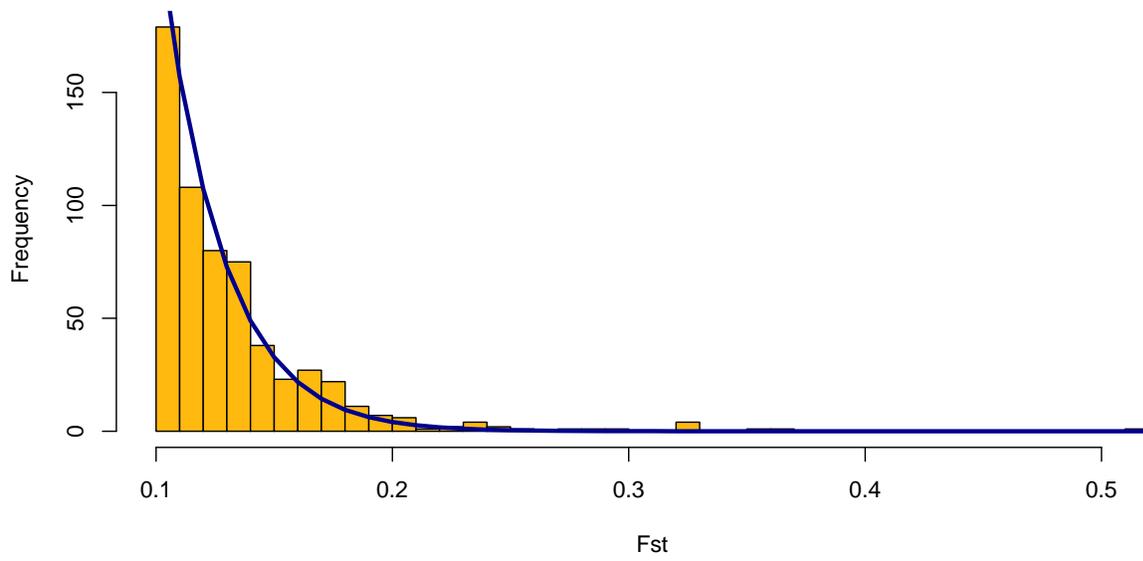


Figure 15. A zoomed in view of the right tail of the distribution from figure 12.

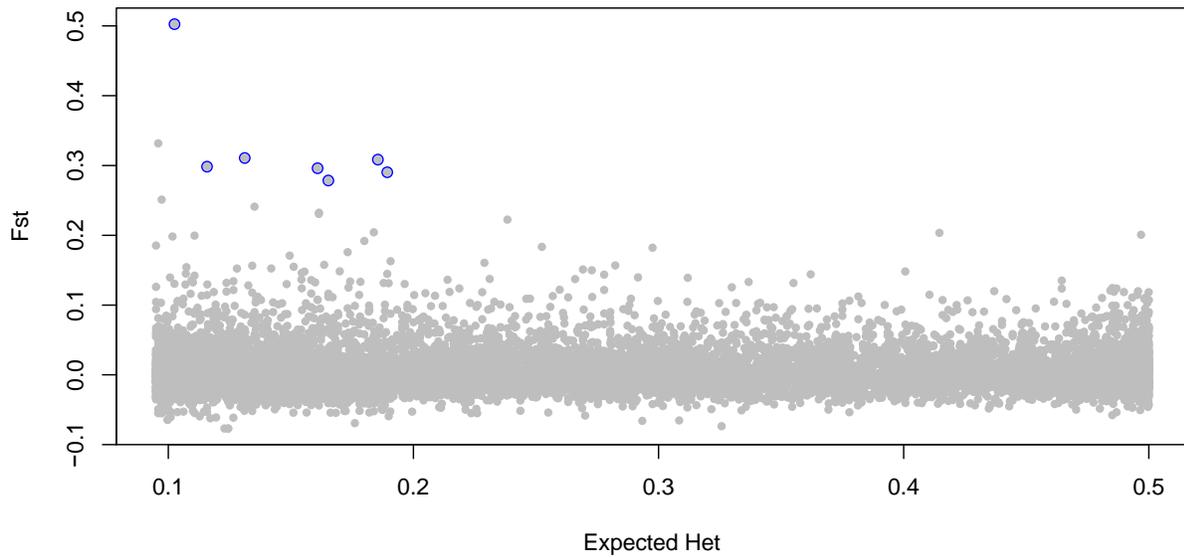


Figure 16. Loci F_{ST} against expected heterozygosity, HG excluded for 10217 SNPs with 107 samples excluding HG. A total of 7 outlier loci at q -value (p -value corrected for a false discovery rate) < 0.01 highlighted. Each circle is an individual SNP.

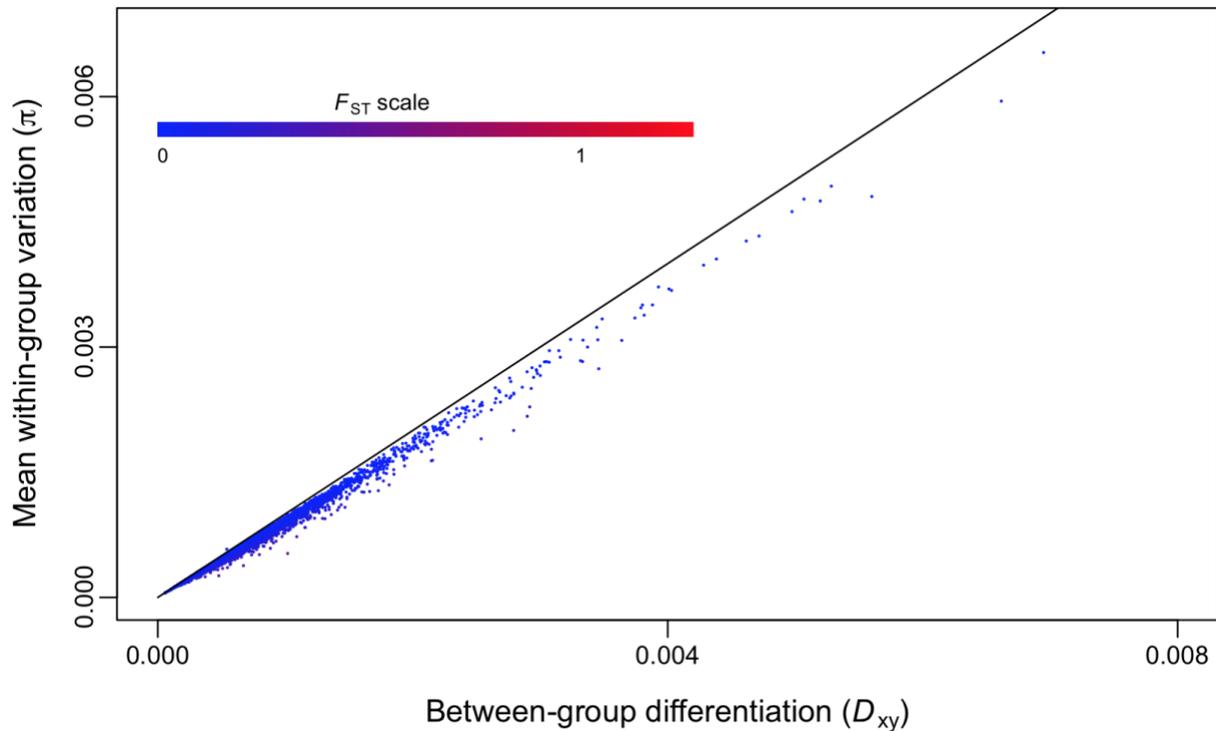


Figure 17. Mean π against D_{XY} for NW and East in 5000bp windows (each dot is a window). Dot color is scaled from blue to red to represent the average F_{ST} in that window.

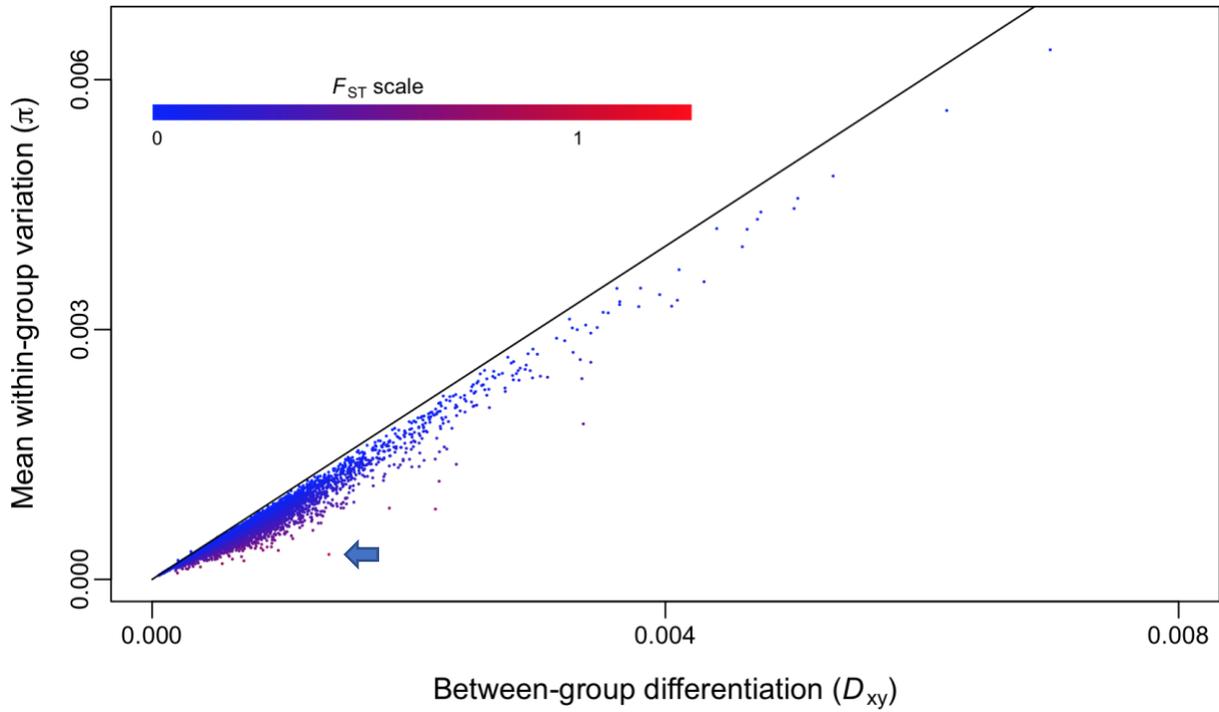


Figure 18 Mean π against D_{xy} for NW and HG in 5000bp windows (each dot is a window). Dot color is scaled from blue to red to represent the average F_{ST} in that window.

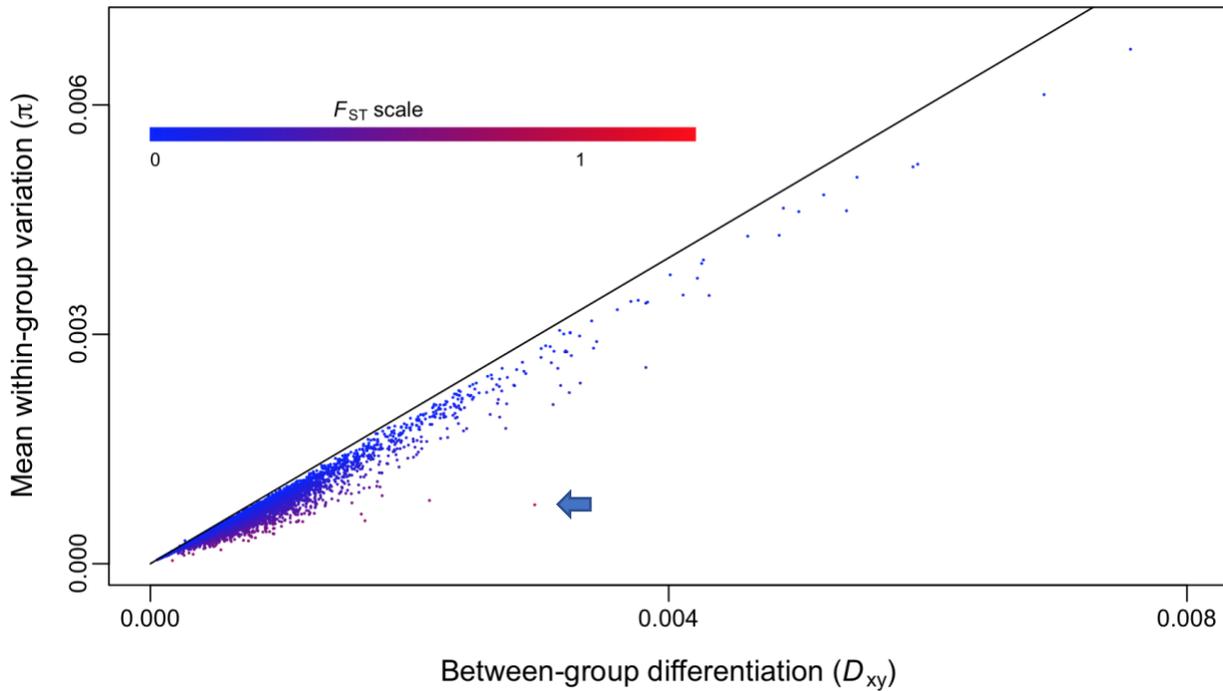


Figure 19. Mean π against D_{xy} for East and HG in 5000bp windows (each dot is a window). Dot color is scaled from blue to red to represent the average F_{ST} in that window.

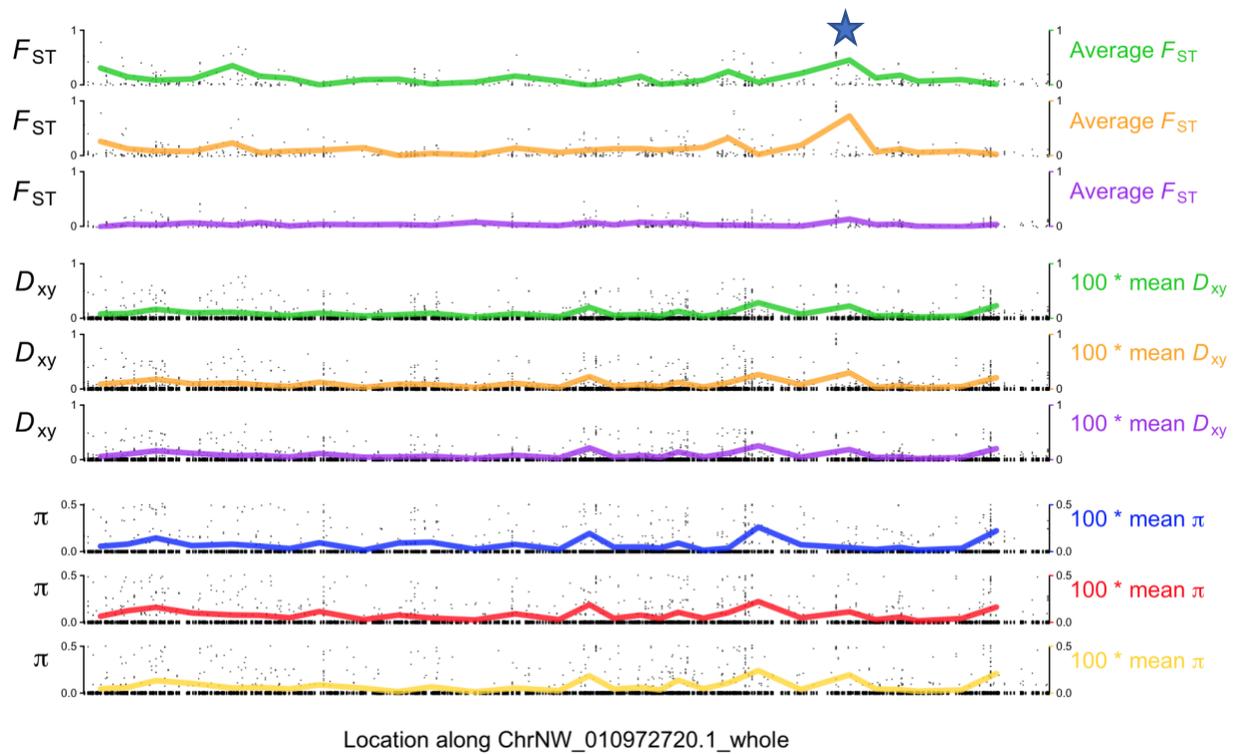


Figure 20. Rolling F_{ST} , D_{xy} , and π for three populations across genomic scaffold NW_010972720.1 (~11.2 million bp). The colors correspond to groups used in our genome scans and are as follows: Green is HG vs. NW, Orange is HG vs. East, and Purple is NW vs. East. For π colors correspond to populations used in genome scan analyses: blue is HG, red is NW, and yellow is East. The star represents the region to be suspected under selection. Each dot is an individual SNP.

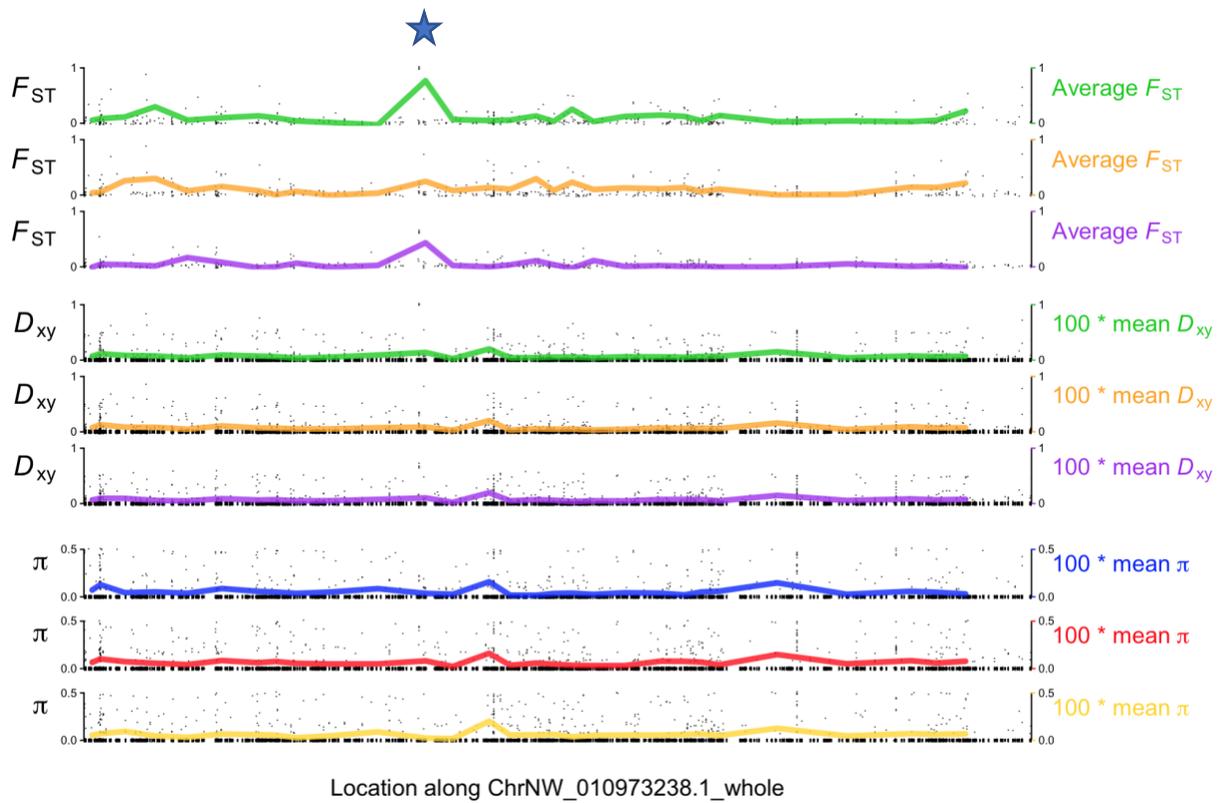


Figure 21. Rolling F_{ST} , D_{xy} , and π for three populations across genomic scaffold NW_010973238.1 (~5.5 million bp). The colors correspond to groups used in our genome scans and are as follows: Green is HG vs. NW, Orange is HG vs. East, and Purple is NW vs. East. For π colors correspond to populations used in genome scan analyses: blue is HG, red is NW, and yellow is East. The star represents the windowed region that is suspected to be under selection. Each dot is an individual SNP.

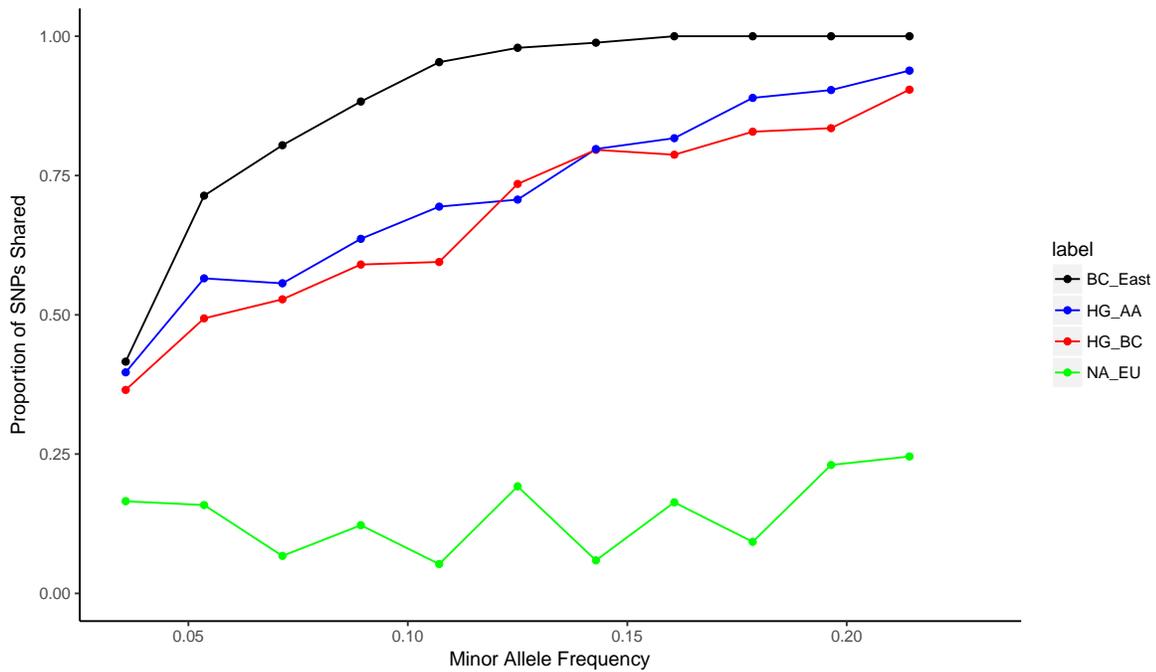


Figure 22. Proportion of shared SNPs across varying minor allele frequencies between populations. Populations with historic isolation and no gene flow should share few SNPs across all minor allele frequencies (NA_EU). Populations that have recently

diverged or have high levels gene flow should share a large majority of SNPs across all minor allele frequencies (BC_East). However, populations that have been historically isolated pressure should share less SNPs across minor allele frequencies.

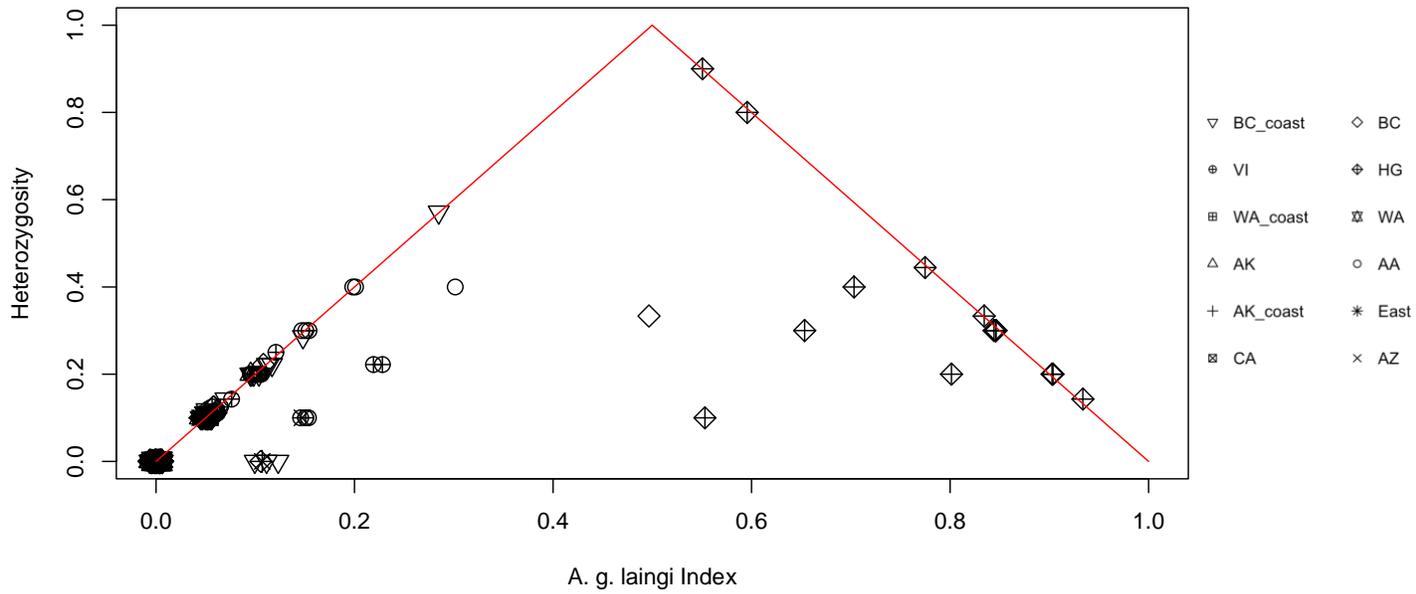


Figure 23. Observed heterozygosity plotted against *laingi* index from the 10 loci in our genotyping assays. Each symbol corresponds to a sample. See Table 1 for population acronyms. Index value of one indicates samples that are highly *laingi*. Samples at the top of the triangle can be inferred as F1 hybrids, and samples in the middle of the triangle as F2 hybrids. Samples approaching one or zero are considered backcrosses until they reach zero or one. The exception in this case is that these loci are not fixed between populations so there are no samples with *laingi* index of one.

Bibliography

- Alcaide, M., E.S. Scordato, T.D. Price, and D.E. Irwin. 2014. Genomic divergence in a ring species complex. *Nature* 511: 83.
- Andrews, K.R., J.M. Good, M.R. Miller, G. Luikart, and P.A. Hohenlohe. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* 17: 81.
- Barker, F.K., A.J. Vandergon, and S.M. Lanyon. 2008. Assessment of species limits among yellow-breasted meadowlarks (*Sturnella* spp.) using mitochondrial and sex-linked markers. *The Auk* 125: 869-879.
- Barton, N.H. and M.A. Slatkin. 1986. A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56: 409.
- Bayard De Volo, S., R.T. Reynolds, M.R. Douglas, and M.F. Antolin. 2008. An improved extraction method to increase DNA yield from molted feathers. *The Condor* 110: 762-766.
- Bayard De Volo, S., R.T. Reynolds, S.A. Sonsthagen, S.L. Talbot, and M.F. Antolin. 2013. Phylogeography, postglacial gene flow, and population history of North American northern goshawks (*Accipiter gentilis*). *The Auk* 130: 342-354.
- Bradbury, I.R., S. Bowman, T. Borza, P.V. Snelgrove, J.A. Hutchings, P.R. Berg, N. Rodríguez-Ezpeleta, J. Lighten, D.E. Ruzzante, C. Taggart, and P. Bentzen. 2014. Long distance linkage disequilibrium and limited hybridization suggest cryptic speciation in Atlantic cod. *PloS One* 9: e106380.

- Bolger, A.M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30: 2114-2120.
- Boyce, D.A., R.T. Reynolds, and R.T. Graham. 2006. Goshawk status and management: what do we know, what have we done, where are we going? *Studies in Avian Biology* 31: 312.
- Broad Institute. 2018. Picard Tools. <http://broadinstitute.github.io/picard/>
- Brown, L., and D. Amadon. 1989. Eagles, hawks and falcons of the world. Volume 2. McGraw-Hill Book Company, New York, NY.
- Bull, R.D., A. McCracken, A.J. Gaston, T.P. Birt, and V.L. Friesen. 2010. Evidence of recent population differentiation in Orange-crowned Warblers (*Vermivora celata*) in Haida Gwaii. *The Auk* 127: 23-34.
- Byun S., B.F. Koop, and T.E. Reimchen. 1997. North American black bear mtDNA phylogeography: implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* 51: 1647-1653.
- Carling, M.D., and R.T. Brumfield. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in Passerina buntings. *Genetics* 178: 363-377.
- Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores, and W.A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22: 3124-3140.
- Ceballos, G., P.R. Ehrlich, A.D. Barnosky, A. García, R.M. Pringle, and T.M. Palmer. 2015. Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* 1: e1400253.

- Crocker-Bedford, D.C. 1990. Goshawk reproduction and forest management. Wildlife Society Bulletin 18: 262-269.
- COSEWIC. 2000. COSEWIC assessment and update status report on the Northern Goshawk *laingi* subspecies *Accipiter gentilis laingi* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. www.sararegistry.gc.ca/status/status_e.cfm
- COSEWIC. 2013. COSEWIC assessment and status report on the Northern Goshawk *Accipiter gentilis laingi* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. www.registrelep-sararegistry.gc.ca/default_e.cfm
- COSEWIC. 2015. COSEWIC wildlife species assessment: quantitative criteria and guidelines. Canada Environment and Natural Resources. <https://www.canada.ca/en/environment-climate-change/services/committee-status-endangered-wildlife/wildlife-species-assessment-process-categories-guidelines/quantitative-criteria.html>
- Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, and G. McVean. 2011. The variant call format and VCFtools. Bioinformatics 27: 2156-2158.
- Demboski, J.R., K.D. Stone, and J.A. Cook. 1999. Further perspectives on the Haida Gwaii glacial refugium. Evolution 53: 2008-2012.
- Der Sarkissian, C., L. Ermini, M. Schubert, M.A. Yang, P. Librado, M. Fumagalli, H. Jónsson, G.K. Bar-Gal, A. Albrechtsen, F.G. Vieira, and B. Petersen. 2015. Evolutionary genomics and conservation of the endangered Przewalski's horse. Current Biology 25: 2577-2583.
- Dickson, B.G., T.D. Sisk, S.E. Sesnie, R.T. Reynolds, S.S. Rosenstock, C.D. Vojta, M.F. Ingraldi, and J.M. Rundall. 2014. Integrating single-species management and landscape conservation

- using regional habitat occurrence models: the northern goshawk in the Southwest, USA. *Landscape Ecology* 29: 803-815.
- Doyle, F. I. 2005. Ground-truthing Northern Goshawk and Marbled Murrelet Habitat Suitability Models and conducting Goshawk Nesting Surveys in Gwaii Haanas National Park Reserve and Haida Heritage Site: Year 2. Parks Canada. Haida Gwaii.
- Doyle, F. I. 2006. Blue Grouse Breeding Habitat on Haida Gwaii. Wildlife Dynamics Consulting. http://www.env.gov.bc.ca/wildlife/wsi/reports/4425_WSI_4425_RPT.PDF
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379.
- Funk, D.J., and K.E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34: 397-423.
- Goodell, J. M., and S.T. Seager. 2015. The Northern Goshawk on the Southern Blue Mountains and Malheur National Forest: A Technical Review of its Status, Ecology and Management. The Nature Conservancy: 1-93. <https://ir.library.oregonstate.edu/concern/defaults/47429970d>
- Hedrick, P.W., and R. Fredrickson. 2010. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics* 11: 615-626.
- IOC World Bird List, 2018 Master list v8.1. IOC World Bird List. <http://www.worldbirdnames.org/ioc-lists/master-list-2/>

- Irwin, D.E., M. Alcaide, K.E. Delmore, J.H. Irwin, and G.L. Owens. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology* 25: 4488-4507.
- Johnson, D.R. 1989. Body size of Northern Goshawks on coastal islands of British Columbia. *The Wilson Bulletin* 101: 637-639.
- Kimura, M., and T. Ohta. 1973. The age of a neutral mutant persisting in a finite population. *Genetics* 75: 199-212.
- Klopfstein, S., M. Currat, and L. Excoffier. 2005. The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* 23: 482-90.
- Lewis, S.B., M.R. Fuller, and K. Titus, 2004 A comparison of 3 methods for assessing raptor diet during the breeding season. *Wildlife Society Bulletin* 32: 373-385.
- Lewontin, R.C., and J. Krakauer. 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74: 175-195.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 25: 1754-60.
- Li, H., Handsaker, B., A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078-2079.
- Mason, N.A., and S.A. Taylor. 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology* 24: 3009-3025.

- Matthey-Doret, R., and M.C. Whitlock. 2018. Background selection and the statistics of population differentiation: consequences for detecting local adaptation. *bioRxiv* 1: 326256.
- McCallum, H. 2008. Tasmanian devil facial tumour disease: lessons for conservation biology. *Trends in Ecology & Evolution* 23: 631-637.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297-1303.
- Miller, R.A., J.D. Carlisle, and M.J. Bechard. 2014. Effects of prey abundance on breeding season diet of Northern Goshawks (*Accipiter gentilis*) within an unusual prey landscape. *Journal of Raptor Research* 48: 1-2.
- Miller, W., V.M. Hayes, A. Ratan, D.C. Petersen, N.E. Wittekindt, J. Miller, B. Walenz, J. Knight, J. Qi, F. Zhao, and Q. Wang. 2011. Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences* 108: 12348-12353.
- Mora, C., D.P. Tittensor, S. Adl, A.G. Simpson, and B. Worm. 2011. How many species are there on Earth and in the ocean? *PLoS Biology* 9: e1001127.
- Narum, S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22: 2841-2847.
- Northern Goshawk (*Accipiter gentilis laingi*) Recovery Team, 2008 Recovery strategy for the Northern Goshawk, *laingi* subspecies (*Accipiter gentilis laingi*) in British Columbia.

Prepared for the B.C. Ministry of Environment, Victoria, BC.

http://www.env.gov.bc.ca/wld/documents/recovery/rcvrystrat/northern_goshawk_rcvry_strat_200508.pdf

Osgood, W. H. 1901. Natural history of the Queen Charlotte Islands, British Columbia and natural history of the Cook Inlet region, Alaska. *North American Fauna* 21: 1-87

Parks Canada Agency. 2017. Recovery Strategy for the Northern Goshawk (*laingi*) subspecies (*Accipiter gentilis laingi*) in Canada [Proposed]. Species at Risk Act Recovery Strategy Series. Parks Canada Agency. http://www.sararegistry.gc.ca/virtual_sara/files/plans/Rs-NorthernGoshawkProposed-v00-2017Nov08-Eng.pdf

Pimm, S.L., L. Dollar, and O.L. Bass. 2006. The genetic rescue of the Florida panther. *Animal Conservation* 9: 115-122.

Proudfoot, G.A., F.R. Gehlbach, and R.L. Honeycutt. 2007. Mitochondrial DNA variation and phylogeography of the Eastern and Western screech-owls. *The Condor* 109: 617-628.

Puritz, J.B., C.M. Hollenbeck, and J.R. Gold, 2014 dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* 2: e431.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Reynolds, R.T. 1983. Management of western coniferous forest habitat for nesting accipiter hawks. US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station.

- Robinson, J.A., D. Ortega-Del Vecchyo, Z. Fan, B.Y. Kim, C.D. Marsden, K.E. Lohmueller, and R.K. Wayne. 2016. Genomic flatlining in the endangered island fox. *Current Biology* 26: 1183-1189.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8: 103-106.
- Sauer, J.R., D.K. Niven, J.E. Hines, D.J. Ziolkowski, K.L. Pardieck, J.E. Fallon, and W.A. Link. 2017. The North American breeding bird survey, results and analysis 1966-2015 (Version 2.07.2017). USGS Patuxent Wildlife Research Center.
- Shafer, A., C.I. Cullingham, S.D. Cote, and D.W. Coltman. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* 19: 4589-4621.
- Slatkin, M. 1985. Rare alleles as indicators of gene flow. *Evolution* 39: 53-65.
- Sonsthagen, S.A., E.L. McClaren, F.I. Doyle, K. Titus, G.K. Sage, R.E. Wilson, J.R. Gust, and S.L. Talbot. 2012. Identification of metapopulation dynamics among northern goshawks of the Alexander Archipelago, Alaska, and coastal British Columbia. *Conservation Genetics* 13: 1045-1057.
- Squires, J.R., and R.T. Reynolds. 1997. Northern Goshawk (*Accipiter gentilis*). Poole A and Gill FB, editors. The birds of North America, Number 298. Academy of Natural Science, Philadelphia, Pennsylvania, and American Ornithologists' Union, Washington. 1-32.
- Storer, R.W. 1966. Sexual dimorphism and food habits in three North American accipiters. *The Auk* 83: 423-436.

- Stryjewski, K.F., and M.D. Sorenson. 2017. Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution* 1: 1912.
- Taverner, P.A. 1940. Variation in the American goshawk. *The Condor* 42: 157-160.
- The Peregrine Fund. 2018. Northern Goshawk. www.peregrinefund.org/explore-raptors-species/Northern_Goshawk.03/05/2018.
- Toews, D.P., and D.E. Irwin. 2008. Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. *Molecular Ecology* 17: 2691-2705.
- Toews, D.P., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* 21: 3907-3930.
- Toews, D.P., and S.A. Taylor, R. Vallender, A. Brelsford, B.G. Butcher, P.W. Messer, and I.J. Lovette. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology* 26: 2313-2318.
- Topp, C.M. and K. Winker. 2008. Genetic patterns of differentiation among five landbird species from the Queen Charlotte Islands, British Columbia. *The Auk* 125: 461-472.
- Van Rossem, A.J. 1938. A Mexican race of the goshawk (*Accipiter gentiles*; Linnaeus). *Proceedings of the Biological Society of Washington* 51: 99-100.
- Wagner, C.E., I. Keller, S. Wittwer, O.M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology* 22: 787-798.
- Waller, D.M. 2015. Genetic rescue: a safe or risky bet? *Molecular Ecology* 24: 2595-2597.

- Warren, W., R. Agarwala, S. Shiryayev, and R.K. Wilson. 2014. The Genome Institute, Washington University School of Medicine, 4444 Forest Park, St. Louis, MO 63108, USA.
- Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Whiteley, A.R., S.W. Fitzpatrick, W.C. Funk, and D.A. Tallmon. 2015. Genetic rescue to the rescue. *Trends in Ecology & Evolution* 30: 42-49.
- Whitlock, M.C., and K.E. Lotterhos. 2015. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of FST. *The American Naturalist* 186: 24-36.
- Wink, M., and H. Sauer-Gürth. 2004. Phylogenetic relationships in diurnal raptors based on nucleotide sequences of mitochondrial and nuclear marker genes. *Raptors Worldwide*: 483-498.
- Woodbridge, B., and C.D. Hargis. 2006. Northern goshawk inventory and monitoring technical guide. General Technical Report. WO-71. Washington, DC: U.S. Department of Agriculture, Forest Service: 1-80.
https://www.fs.fed.us/rm/pubs_series/wo/wo_gtr071.pdf
- Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie, and B.S. Weir. 2012. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* 28: 3326-3328