PHYLOGENOMIC ANALYSIS OF THE RECENTLY RADIATED SAPSUCKER (*SPHYRAPICUS*) SUPERSPECIES COMPLEX

ASHLEY IRENE JOY CURTIS
BSc, Brigham Young University - Hawaii, 2011

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

MASTER OF SCIENCE

Department of Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

PHYLOGENOMIC ANALYSIS OF THE RECENTLY RADIATED SAPSUCKER (*SPHYRAPICUS*) SUPERSPECIES COMPLEX

ASHLEY IRENE JOY CURTIS

Date of Defence: December 14, 2016

Dr. Theresa Burg Supervisor	Associate Professor	PhD
Dr. Jim Thomas Thesis Examination Committee Member	Professor	PhD
Dr. Andrew Iwaniuk Thesis Examination Committee Member	Associate Professor	PhD
Dr. Sean Rogers External Examiner University of Calgary Calgary, Alberta	Associate Professor	PhD
Dr. Tony Russell Chair, Thesis Examination Committee	Assistant Professor	PhD

ABSTRACT

Sphyrapicus varius, S. ruber, and *S. nuchalis* are members of a closely related and recently radiated superspecies complex widely distributed throughout North America, with areas of range overlap and hybrid zones. I used genomic data to determine the evolutionary history of the three species. Three distinct genetic clusters corresponding to phenotypic species identification using both neutral and non-neutral genetic markers were identified along with possible candidate genes under selection. These genes may have played a role in adaptive evolution driving speciation. Additionally, I studied the genomic structure of a hybrid zone between *S. varius* and *S. nuchalis.* It was composed of advanced generation hybrids with lots of admixture within the zone, but no introgression of interspecific alleles outside of the contact zone. Within the zone backcrosses differed along a north to south gradient, where southern individuals were more skewed to *S. nuchalis* ancestry and northern individuals to *S. varius.*

ACKNOWLEDGEMENTS

I feel like completing a Masters becomes a bit of a community endeavor including all those special friends and family of the individual pursuing the degree. I know so many people who have encouraged me to push through the slumps and stick with it so that when they ask the inevitable "Are you done yet?" question, I can finally say "YES!!" As my wonderfully patient and supremely knowledgeable supervisor Theresa Burg can attest to, this has been a long journey full of ups and downs. I want to thank Theresa for pushing me to do something that is so new and challenging for me and for being a great sounding board and rational mind behind this whole study. She has truly become someone that I admire and I appreciate all the time and effort she has put into helping me feel like I've learned the skills, theories, and techniques that will help me to adapt and move on in my future career.

I have to thank my extremely patient family, without whom, I wouldn't have gotten anywhere. I'd still be sitting there feeling sorry for myself that "it doesn't work." Thanks to Nathan, Natasha, Alycia, and Allisan – the best siblings in the world – for coming in to the lab in snowstorms and on late nights to help me mash up bird tissue, pipette blood and solutions, and label tubes (their favourite jobs!). Thank you to my amazing Mom and Dad who were always asking what they could do to help, though really the only thing they could do was be there and be supportive – which they always were. They also listened to each amazing new "breakthrough" I had with the same level of excitement the 20th time they heard it as the first time. Thanks to my friends and family who understood and had patience with me and continued to invite me to outings in spite

of my numerous and frequent cancellations – your persistence kept me sane and human with the knowledge that you loved me in spite of my Masters.

I'd like to thank my lab mates who put up with me, kept me sane, and helped me to figure out the convoluted ins and outs of graduate work. Thanks to Rachael for her determination to make sure everyone relaxed with social outings but who also showed me how to work hard and smart when in the lab. Thanks to Colin for showing me the ropes during our summer field sampling – that was one of the best summers ever and I'll treasure it, the memories, and our friendship forever. I feel like you have to stay friends for life with someone who knows how bad you can smell after not showering for as long as we went without showers. I'd also like to thank those wonderful people from Project Feeder Watch who allowed us to sample in their yards and often gave us a place to sleep and a shower (bless their hearts).

None of my analyses could have been completed if not for the wonderful computer science people who assisted me along the way. A special thank you to Kevin Vadnais who wrote a script for something that was way over my head and out of my league. Thank you, too, to Sanjay for the wonderful birthday present of my computer script that does that one thing that I wanted it to do. They are really the only reason I got over a lot of those computer hurdles you face when you're a biologist trying to make it in a computer science world.

This research was funded by the Natural Sciences and Engineering Research Council (NSERC; TMB) and Alberta Innovates (TMB) and would not have been possible without it. Lastly, I'd like to thank the University of Lethbridge and the Biological Sciences community for helping and allowing me to complete this degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AB	Alberta
AFLP	amplified fragment length polymorphism
AFS	Allele Frequency Spectrum
AUC	area under the receiver operating characteristic curve
BC	British Columbia
BIC	Bayesian Information Criterion
BSC	Biological Species Concept
bp	base pairs
CA	California
CAB	central Alberta
CIS	Cordilleran Ice Sheet
COI	cytochrome oxidase I
DAPC	discriminant analysis of principal components
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ENM	ecological niche modelling
ESS	effective sample size

GBS	Genotyping by Sequencing
GBIF	Global Biodiversity Inventory Facility
GO	gene ontology
IGD	Institute for Genomic Diversity
ILS	incomplete lineage sorting
ka	thousand years ago
km	kilometre
km^2	square kilometre
LIG	last interglacial
LIS	Laurentide Ice Sheet
LGM	last glacial maximum
m	metre
M	molar (moles/litre)
Ma	million years ago
MAF	minor allele frequency
MCMC	Markov Chain Monte Carlo
mg	milligrams
mL	millilitre

mm^2	square millimetre
mtDNA	mitochondrial DNA
mxMAF	maximum minor allele frequency
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
ng	nanogram
NGS	next-generation sequencing
NaOAc	sodium acetate
NM	New Mexico
OR	Oregon
RAD	restriction-site associated DNA
RNA	ribonucleic acid
rpm	revolutions per minute
RRL	reduced representation library
SAB	southern Alberta
SDM	species distribution modeling
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism

UNEAK	universal network enabled analysis kit
UTexas	University of Texas
WA	Washington
^{o}C	degrees Celsius
^{o}N	degrees north of the equator
λ	lambda
µg	micrograms
µL	microlitre
1X V	one times volume
χ^2	Chi-square

CHAPTER ONE: General Introduction

Speciation is the process through which one species splits into two or more diverging lineages. Ecological and genetic processes such as climate change, natural selection, mate choice, and genetic drift all influence the evolution of a species. Evolution through ecological processes relies on environmental pressures acting directly on morphology and subsequently resulting in a change in genetic composition of a population. Speciation can occur through gradual changes in the genetic composition of populations (gradualism) or through bursts of rapid changes in environmental pressures or mutation (punctuated equilibrium). The continuous process of evolution along with recent species' radiations during the severe climate fluctuations of the Pleistocene Epoch have resulted in unclear relationships between recently diverged species (Rissler & Apodaca 2007; Shaffer & Thomson 2007). Reproductive isolation, monophyly, ecological divergence, or morphological separation between two diverging lineages are common criteria for species delimitation but often one method alone is not sufficient to resolve species' relationships (Bacon *et al.* 2012; Shaffer & Thomson 2007).

Studies on speciation use multiple evolutionary, molecular, and systematic techniques to understand complex relationships between recently diverged species. One such multidisciplinary field is phylogenomics. It utilizes population genetics, systematics, genomics, and computational methods and theories congruently (DeSalle & Rosenfeld 2013) to determine genetic variation, paleo- and current-geographic distribution, and gene function and adaptation. These separate components combined provide fine scale resolution to delineate species (Rissler & Apodaca 2007; Shaffer & Thomson 2007).

1.1 Phylogeography

Phylogeography is the study of the geographic variation in patterns of genetic structure (Avise *et al.* 1987). Studies involving phylogeography use both paleo- and current-distribution to predict how species' ranges have changed throughout history. The identification of current and historic areas of allopatry, sympatry, and parapatry and how distribution affects genetic structure assist phylogenomic analyses in understanding speciation. Most speciation events of modern species can be traced to climate changes during the Pliocene and Pleistocene Epochs (Avise & Walker 1998; Avise *et al.* 1998; Weir & Schluter 2004). The Earth's history has been marked by recurring glacial and interglacial cycles. These cycles contracted and expanded species ranges resulting in cyclical periods of sympatric, allopatric and parapatric distributions (Hewitt 1996, 2004).

1.1.1 The Pleistocene Epoch

The Quaternary Period is composed of two epochs; the Pleistocene (from approximately 1.8 Ma to 12 ka) and the Holocene (12 ka to present). The Pleistocene was characterized by cycles of cold global climate as a result of glaciation, followed by warmer intervals of interglacial periods. This variation in climate can be attributed to the Earth's uneven orbit around the sun, resulting in waxing and waning of ice sheets at regular intervals of 100, 41, and 23 thousand years (Clark *et al.* 1999; Hewitt 1996). The process of divergence throughout these climatic oscillations was dynamic. As the climate changed, species distribution changed to more closely track the expansion and contraction of suitable habitat (Hewitt 1996; Stewart & Lister 2001). As climate warmed, a species'

possible range expanded with increased suitable habitat leading to decreased diversity in rapidly expanding leading edge populations. Conversely, global cooling resulted in a reduction of a species' range which often became fragmented due to lack of contiguous suitable habitat. These oscillating range expansions moved species' boundaries north and south with the warming interglacials and cooling glacials vacillating between allopatric and sympatric distributions. Sudden changes in climate led to extirpation of populations in the uninhabitable part of the range and resulted in loss of genetic variation (Hewitt 1996, 2004).

Recent mitochondrial DNA molecular clock analyses of avian and mammalian species have shown that divergence of sister species dates to the end of the Pliocene Epoch in the Tertiary Period – around 2 Ma (Avise & Walker 1998; Avise *et al.* 1998); however, the rate of speciation is not the same in all organisms. The large diversity of cichlid fish in Lake Victoria have a history of a much faster radiation, having diverged from their most recent common ancestor approximately 700 ka (Meyer *et al.* 1990). In comparison, a study on the coalescent time of mitochondrial DNA of boreal forest avian superspecies showed that divergence of these superspecies occurred mostly during the early Pleistocene due to the fragmentation of the boreal forest from advancing ice sheets (Weir & Schluter 2004).

1.1.2 The Last Glacial Maximum

The cyclical fluctuation of glacial and interglacial periods that characterized the Pleistocene Epoch of the Quaternary Period had the greatest effect on temperate species.

The Last Glacial Maximum (LGM) was the largest and most recent glacial period of the Pleistocene lasting from approximately 26.5-14 ka on the North American continent. The ice sheets reached their maximum at 26.5 ka, which corresponded with a lowered sea level of approximately 130 m below present (Clark *et al.* 2009; Clark & Mix 2002; Dyke & Prest 1987). North America was covered by two main ice sheets, the Laurentide Ice Sheet (LIS) and the Cordilleran Ice Sheet (CIS) (Figure 1.1). The LIS extended across North America from the east coast to the eastern edge of the Cordillera and from the Queen Elizabeth Islands in the north down to at least 40 ^{o}N in the mid-western United States and contributed to between 40-92 m of sea level change – more than any other ice sheet (Dyke *et al.* 2002; Fulton & Prest 1987; Hewitt 1996). The CIS extended along the west side of the Rocky Mountains from south and central Yukon, through British Columbia down into Washington, Idaho and part of Montana covering about 1.5 million km^2 (Margold *et al.* 2013) compared to the approximately 11 million km^2 covered by the LIS (Fulton & Prest 1987). With these large ice sheets covering such a large portion of North America, species ranges were at their smallest (Beatty & Provan 2011).

1.1.3 Allopatric Speciation in Glacial Refugia

Along with reduction of species' ranges due to the large ice sheets covering most of North America, the habitat closest to the ice sheets became a temporary tundra partially due to the Arctic air off the ice sheets (Fulton & Prest 1987). As a result of the expansion and contraction cycles of suitable habitat from throughout the Pleistocene, temperate species were forced into small, allopatric areas of suitable habitat called glacial refugia. Phylogeographic and geological studies have identified multiple refugia during

the LGM. The area of the United States located just south of the ice sheets and down into Mexico was composed of multiple areas of refugia (Beatty & Provan 2010; Fulton & Prest 1987; Hewitt 2004). The Pacific Coast, south of the CIS, was another refugia for many species (Beatty & Provan 2010; Gugger & Sugita 2010) while the North Atlantic continental shelf may have been a third refugia along with other small nunataks (sky islands) on mountain peaks (Stewart & Lister 2001).

Changes in allele frequencies between populations through genetic drift due to prolonged separation in isolated refugia had the potential for allopatric speciation. Exposure to different cohabitating species and environmental pressures while in separate refugia may also have assisted speciation via adaptive divergence through selection on different genetic traits (Hewitt 1996). These traits could have been morphological or behavioural and may have resulted in reproductive isolation and further genetic divergence between two glacially isolated populations (Hewitt 1996; Stewart & Lister 2001).

1.1.4 Sympatric Speciation through Colonization and Expansion

Following the recession of the ice sheets after the LGM, species in refugia expanded into newly available suitable habitat. Populations isolated in allopatry during the glacial period came into secondary contact. Rapid expansion of species' ranges during the warmer interglacial periods resulted in reduced genetic variation in smaller peripheral populations due to genetic drift and founder effects (Hewitt 2004). If the time spent in isolated refugia was not sufficient to result in reproductive isolation, but did result in new

or different alleles between populations, mixing of different alleles from previously separated populations during secondary contact would result in increased genetic variation (Stewart & Lister 2001). Conversely, if sufficient genetic variation for reproductive isolation between the populations had occurred while in allopatry, sympatric and/or parapatric speciation would reinforce genetic differences between the two populations resulting in decreased gene flow between the diverging populations and a continued trajectory towards speciation. One cycle of allopatry followed by sympatry or parapatry may not be sufficient for speciation to occur (Weir & Schluter 2004), however, the multiple cycles of allopatric speciation during glacial periods followed by parapatric or sympatric speciation during the warm interglacials could eventually lead to the splitting of the two lineages into separate species.

1.2 Species Concepts

One problem with studying speciation involves the fundamental definition of a species. If speciation occurs through gradual changes in allele frequencies of a population, is each new variant classed as a new species? The biological species concept (BSC) as defined by Mayr (1963) lists a species as a group of populations that are able to interbreed. If two populations are not able to interbreed, they are classed as different species, however this definition is only viable when looking at sympatric populations. It is difficult to test this definition on allopatric populations and it is impractical in taxonomic classifications of asexual organisms. Another definition of a species is the morphological species concept which designates a species depending on shared phenotypic characteristics among individuals. The shortcomings of this definition can

lead to over splitting of taxa, especially in cases where sexual dimorphism and complex life stages can result in phenotypically distinct males and females or larvae and adults. A third definition, the evolutionary species concept, classifies species as an interbreeding group of organisms that share a common ancestor within a monophyletic group. One problem is determining what molecular markers and how many to use to determine monophyletic lineages.

1.3 Population Genetics

Population genetics is the study of allele frequencies and how they change from one population or meta-population to the next. Phylogenomics utilizes population genetic methods to determine fine scale genetic variation between species. Traditionally, mitochondrial DNA (mtDNA) and nuclear markers, often microsatellites, have been used to study speciation and population genetic structure. The assumption of microsatellite and mtDNA markers to be selectively neutral makes them ideal for genetic studies that focus on neutral processes such as genetic drift and gene flow (DeSalle & Rosenfeld 2013; McCormack *et al.* 2013). There are, however, limitations in the use of these methods. MtDNA is uniparentally inherited and may provide a biased historical view of a species or population (Zhang & Hewitt 2003). Microsatellites, on the other hand, are inherited from both parents, but present unclear historical relationships between alleles while at the same time exhibiting a highly variable mutation rate within and across different organisms and loci (Zhang & Hewitt 2003). While the use of microsatellites and mtDNA in population genetics is now common, recent technological advances have placed an increased demand on using large numbers of loci. Despite the standard practice of using

up to 20 microsatellite loci in traditional genetics studies, this still represents only a small fraction of the genome in comparison to some methods of NGS (Angeloni *et al.* 2012).

More recently, single nucleotide polymorphisms (SNPs) have been used in population genetic studies. Although SNPs contain relatively less variability than other markers, their reduced rates of substitution and homoplasy along with their biparental inheritance and increased frequency throughout the genome more than make up for it because you can increase the number of SNPs screened (Freeland *et al.* 2011; Shaffer & Thomson 2007). Recent advances in technology have resulted in increased ease and reduced cost of obtaining SNP data allowing for the application of SNP data to population genetic and speciation studies.

1.3.1 Next-generation Sequencing

The advent of multiparallel sequencing procedures in recent years has allowed for faster and cheaper sequencing of large numbers of loci from throughout the genome. This new technique is known as next-generation sequencing (NGS). NGS uses thousands of SNPs from multiple locations throughout the genome to uncover genetic variation. We can now go from studies using 5-20 microsatellite loci to studies using tens of thousands of SNP loci (Angeloni *et al.* 2011; Baird *et al.* 2008) for comparative or lower cost (Puritz *et al.* 2012; Shaffer & Thomson 2007). The portion of the genome represented may still be small, but the widespread distribution of SNPs throughout the genome provides more information than traditional mitochondrial and microsatellite loci and decreases potential bias of a few loci.

The location of traditional markers within neutral, non-coding regions of the genome may be beneficial for studying neutral processes – such as drift, dispersal, and habitat fragmentation – but the question arises as to whether they can satisfactorily measure non-neutral processes (Angeloni *et al.* 2012; Ouborg *et al.* 2010). Genetic drift through population bottlenecks and long distance dispersal are only a couple of examples of neutral processes that drive speciation through changes in allele frequencies in populations. Non-neutral processes, such as selection and adaptation of populations to varying environmental pressures, are other factors that affect speciation. As described previously, during the cyclical pattern of climate change throughout the last few millennia, both neutral (bottlenecks) and non-neutral (adaptation to local refugia) processes would have been important in species diverging into separate lineages. Non-neutral markers are located within functional portions of the genome (coding regions, splice sites, non-coding RNA) that may be subject to selection by environmental cues and are therefore involved in adaptive responses that may play a role in genetic divergence of different species (Angeloni *et al.* 2012). NGS screens the genome for large numbers of loci to identify candidate genes (genes under selection) that are subject to selection and thereby allowing us to study the effect of adaptive genes on individual or population fitness as a result of changing environment and local adaptation (Freeland *et al.* 2011; Ouborg *et al.* 2010; Shaffer & Thomson 2007; Zhang *et al.* 2011).

Another benefit that NGS provides genetic studies is the reduction in locus bias through the use of multiple unlinked loci. Reducing this bias is important for genetic studies because it allows for comparison of genetic variation not only specific to a population as a result of adaptation, but also neutral genetic variation as a result of gene

flow and evolution. NGS is also useful as a stepping-stone to developing markers for research on non-model species (Angeloni *et al.* 2012). The discovery and development of genetic markers – both neutral and non-neutral – at a high rate while simultaneously sequencing the genome through NGS reduces labour and costs traditionally involved in marker production (Ekblom & Galindo 2011). Finally, NGS also has the ability to sequence whole genomes of organisms faster, cheaper, and easier than traditional sequencing methods and as a result, the amount of genetic data available for non-model organisms has increased dramatically (Metzker 2010; Zhang *et al.* 2011).

1.3.2 Genotyping by sequencing

Genotyping by sequencing (GBS) is a form of NGS that utilizes restriction enzymes to create a reduced representation library (RRL). This RRL is made up of multiple restriction-site associated DNA (RAD) tags which provide high marker density throughout the genome (Davey *et al.* 2011; Elshire *et al.* 2011). The use of restriction enzymes to create the library allows for the use of this method with non-model organisms or organisms for which little or no previous genetic work has been done. GBS allows for large numbers of individuals to be sequenced simultaneously for population genetic studies and it can be used to determine fine scale resolution of closely related species or populations (Davey *et al.* 2011; Emerson *et al.* 2010).

1.4 Study Species

The *Sphyrapicus* (sapsucker) genus is composed of four species: *Sphyrapicus thyroides, S. varius, S. ruber, and S. nuchalis*. This study focuses on the three most closely related species that comprise the *varius* superspecies complex: *S. ruber* (red-breasted), *S. nuchali*s (red-naped) and *S. varius* (yellow-bellied). A superspecies complex is defined as a monphyletic group of allopatric or parapatric species that have recently achieved species status as a result of becoming distinct, but were once classed as a single species (Amadon 1966). Previous studies on sapsuckers have focused on nuclear and mtDNA loci to determine genetic variation congenerically (Cicero & Johnson 1995; Johnson & Zink 1983; Seneviratne *et al.* 2012). Initial confusion about the taxonomy of these three species was a result of their recent divergence and low levels of genetic differentiation (Short & Morony 1970). Divergence in the *varius* complex dates back to the Pleistocene with *S. varius* diverging approximately 800 ka to 1.5 Ma and *S. nuchalis* and *S. ruber* diverging approximately 250 ka to 500 ka with the glacial cycles of the Pleistocene aiding in the completion of the speciation process (Weir & Schluter 2004). Phenotypically, *S. nuchalis* and *S. varius* sapsuckers look very similar while *S. ruber* are more phenotypically distinct compared to *S. nuchalis* and *S. varius*. Conversely, *S. varius* and *S. nuchalis* are more genetically distant from each other than *S. varius* to *S. ruber* (Cicero & Johnson 1995). Low levels of genetic differentiation have been found between *S. ruber* and *S. nuchalis* (Cicero & Johnson 1995).

All three species have distinct distribution but overlap at the edges of their ranges and hybridize (Figure 1.2). *S. varius* have a single hybrid zone with each of the other two species – one in central Alberta with *S. nuchalis* and one in northern BC with *S. ruber*

(Seneviratne *et al.* 2016; Seneviratne *et al.* 2012). In contrast, the hybrid zone between *S. ruber* and *S. nuchalis* sapsuckers is quite large, extending from southern BC to California (Johnson & Johnson 1985). *S. ruber* sapsuckers are not sexually dimorphic while *S. nuchalis* and *S. varius* are sexually dimorphic with the male of the species being the redder of the sexes (Figure 1.3) (Backhouse 2005). Sapsuckers make holes in trees for nest cavities and foraging for sap and insects. These nests and foraging sites are used in subsequent years by secondary cavity nesters, making sapsuckers a keynote species in forest ecosystems (Aitken & Martin 2007; Kalcounis & Brigham 1998; Rissler 1995).

Even though there was initially little evidence for genetic differentiation between *S. ruber* and *S. nuchalis*, their status as two separate species was solidified via the biological species concept when Johnson and Johnson (1985) determined that hybrids of the two were selected against, leading to a tension hybridization zone. Recent unpublished data (Pelton, Martin & Burg) have shown that the three species can be separated into three clades using the control region sequence of the mtDNA. The three clades are composed of one pure *S. nuchalis* clade, one mostly *S. varius* clade with some DNA from the other two species, and a third clade with a mix of the three species but composed of most *S. ruber* sapsuckers and half the *S. nuchalis*. The data, however, were not able to determine a pure *S. varius* or *S. ruber* clade and the pure *S. nuchalis* clade has no obvious geographic clustering. The ambiguity in the genetic differentiation between *S. ruber* and *S. nuchalis* seen in previous genetic studies using traditional markers is most likely the result of a recent radiation and is the rationale behind this study (Cicero & Johnson 1995; Johnson & Johnson 1985; Johnson & Zink 1983). NGS sampling of the entire genome for genetic variation along with phylogeographic inference of species

distribution and genomic analyses of adaptive speciation will contribute to understanding the processes resulting in the speciation of the sapsucker superspecies complex; if they underwent adaptive radiation and the possible mechanisms driving speciation.

1.5 Thesis Aims

The goal of this thesis is to provide a phylogenomic analysis of speciation in sapsuckers. The first aim is to determine if NGS provides increased resolution into the degree of genetic differentiation within the sapsucker *varius* superspecies complex, particularly between *S. ruber* and *S. nuchalis*. Another aim is to determine the locations of refugial populations and possible contact zones for continued gene flow between the three species of sapsuckers during the LGM and the last interglacial period (LIG) and to compare this data with coalescent modelling data to determine the demographic histories of the species. The final aim of this thesis is to identify the genomic structure of the central Alberta hybrid zone between *S. nuchalis* and *S. varius* sapsuckers. Presently, the *varius* complex is unable to be broken into smaller monophyletic subunits containing single species using traditional molecular markers (Burg unpublished data; Cicero & Johnson 1995; Johnson & Zink 1983) and the classification of *S. ruber* and *S. nuchalis* as individual species is limited to reduced fitness of F_1 hybrids at one site in California and distinct morphology (Johnson & Johnson 1985). Also, little information is available on their distribution and potential contact during the LGM and LIG and how isolation and migration may have shaped their evolution.

This study incorporates the multidimensional requirements needed for a phylogenomic study in the form of genomic marker analyses for neutral and non-neutral genetic diversity, ecological niche modelling (ENM), and coalescent modelling. I used SNP markers collected from throughout the genome by GBS to determine genetic structure using population genetics techniques (Eisen & Fraser 2003; Elshire *et al.* 2011; Shaffer & Thomson 2007). The genome wide marker analysis will allow for increased resolution of genetic variation between *S. ruber* and *S. nuchalis* sapsuckers. Ecological niche modelling is used to determine present distributions of the three species and predict their distributions during both the LGM and the LIG. This allows me to determine if the species went through episodes of allopatric and sympatric or parapatric speciation to assist in determining if the species were isolated long enough to become fully diverged. The coalescent modelling is used to determine how speciation occurred under different models based on current genetic structure and to further expound on the ENM results. This method will assist me in determining if the genetic admixture between *S. ruber* and *S. nuchalis* is a result of incomplete lineage sorting due to recent divergence or a result of recent admixture of two fully diverged species following secondary contact.

1.6 Predictions

Due to the large number of SNPs produced from throughout the genome by GBS (Davey *et al.* 2011; Emerson *et al.* 2010; McCormack *et al.* 2013), I predict that I will find more genetic differentiation between *S. ruber* and *S. nuchalis* and that GBS will separate them into two genetically distinct, monophyletic species. I also predict that the ENM will show an overlapping refugial population between *S. nuchalis* and *S. ruber* in

the southern Rocky Mountains while the *S. varius* will be completely isolated in the east. The coalescent model of incomplete lineage sorting as a result of continued gene flow in refugia between *S. nuchalis* and *S. ruber* sapsuckers will be the most likely scenario driving speciation between *S. ruber* and *S. nuchalis*.

The second component is the study of the contact zone in Alberta between *S. nuchalis* and *S. varius*. I predict that hybrids between *S. nuchalis* and *S. varius* will show a higher ancestry to *S. nuchalis* due to the preferential selection of female *S. nuchalis* for males with more red (Johnson & Johnson 1985). This skew in ancestry will correspond to alleles under selection leading to understanding of fitness of the different parental forms.

1.7 Thesis Organization

This thesis is organized into four chapters. The first provides background into the process of speciation and the neutral (bottlenecks through range fragmentation during glaciations) and non-neutral (selection to different environmental pressures in refugia) processes that drive it, along with providing information on the genetic and systematic methods used to study speciation. Chapter 2 uses analyses of GBS data and spatial and coalescent modelling to determine the genetic variation within the sapsucker complex and the most likely model of speciation that explains current genome structure. Chapter 3 is a study of the central Alberta hybrid zone to determine genomic composition of admixture between *S. nuchalis* and *S. varius*. It uses both neutral and non-neutral markers to determine ancestry and fitness of hybrid individuals. The final chapter summarizes the results found in Chapters 2 and 3 focusing on the overall method of speciation and

genomic composition of the sapsucker *varius* superspecies complex. I also provide potential future work suggestions.

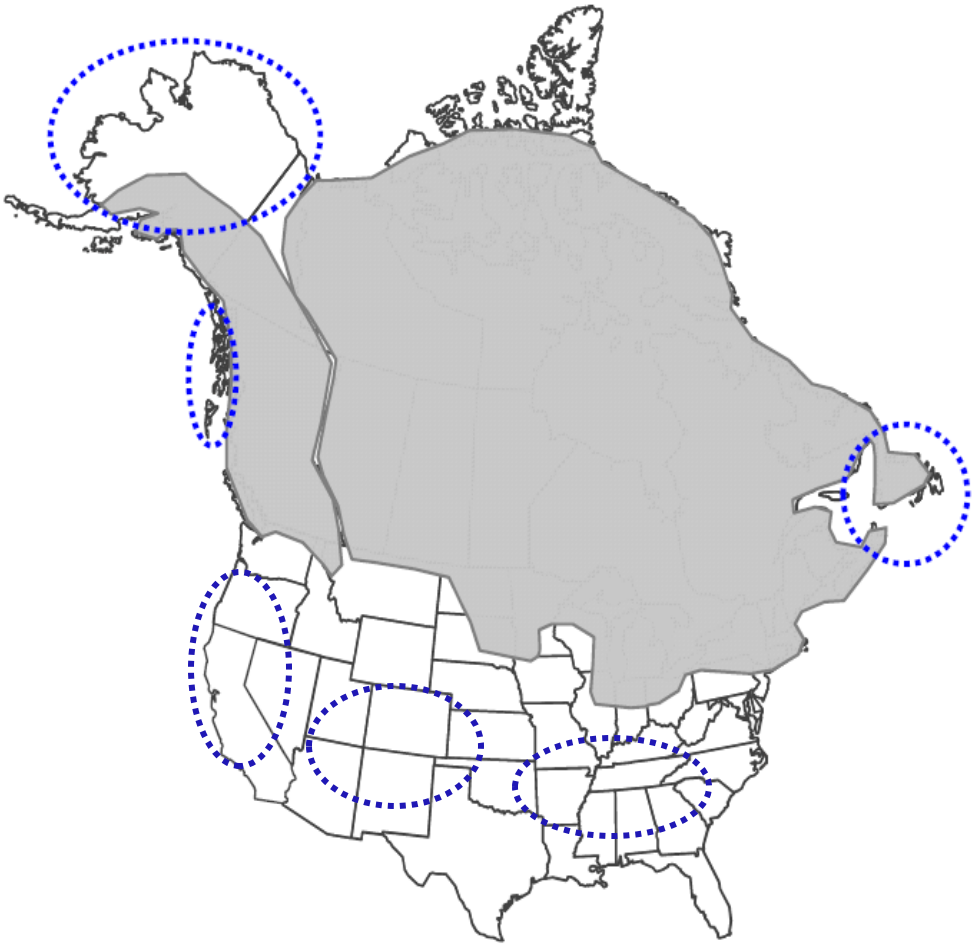


Figure 1.1. Ice coverage during the last glacial maximum (in grey) with the Laurentide (right) and Cordilleran (left) ice sheets covering most of Canada. Unglaciated areas of the United States south of the ice sheets acted as refugia along with the Atlantic Coast, Beringia in Alaska, and Haida Gwaii (dashed blue lines from right to left). Figure modified from Pielou (1991).

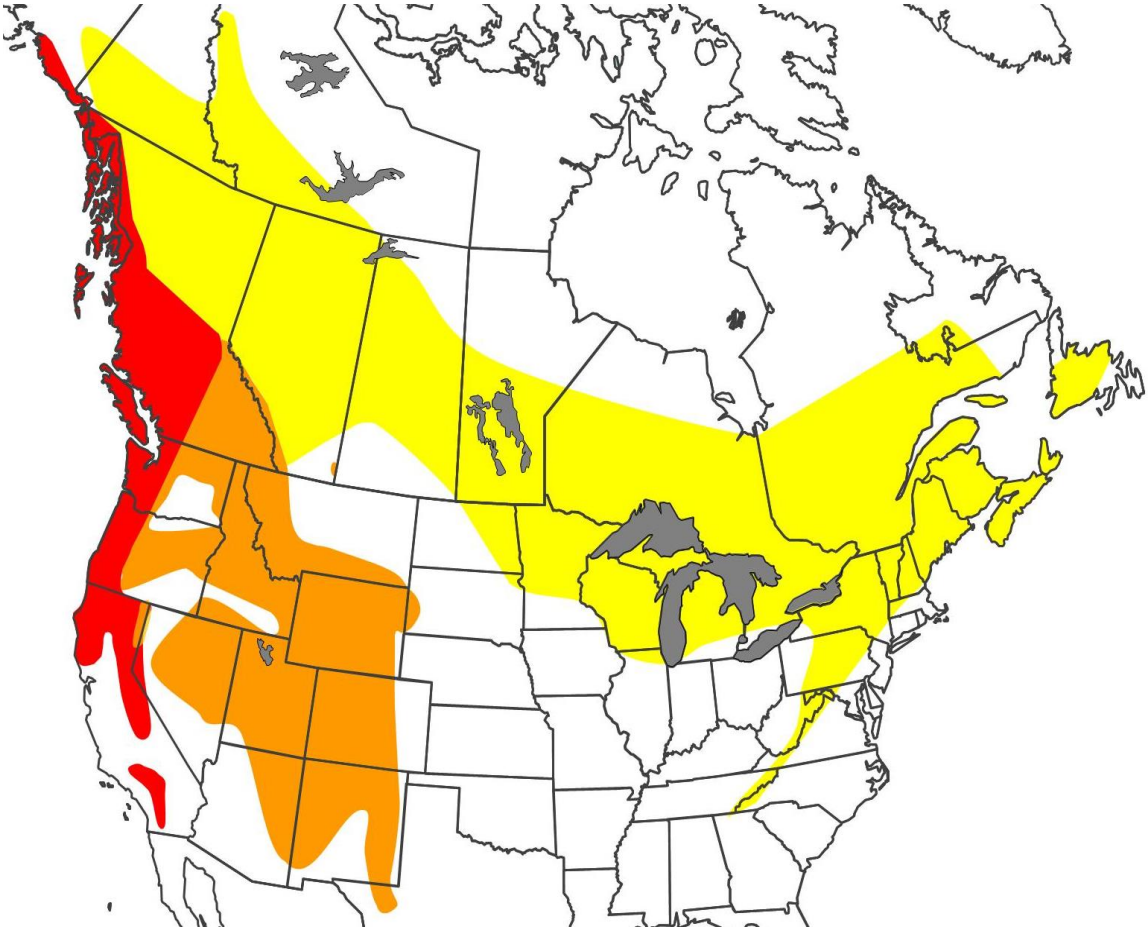


Figure 1.2. Breeding distribution of the three sapsucker species in the *varius* complex: *S. ruber* (red), *S. varius* (yellow) and *S. nuchalis* (orange).

a)  b)  c) 

d)

	S. ruber	*S. nuchalis*	*S. varius*
Red plumage	head, neck and chest	nape, crown, and throat	crown and throat
Body size	largest	intermediate	smallest
Sexually dichromatic	no	intermediate	yes
Migration	most populations resident, some migratory	mix of resident and migratory	highly migratory, no resident populations

Figure 1.3. Examples of *S. ruber* (a), *S. nuchalis* (b), and *S. varius* (c) and the major morphological differences among the three sapsucker species (d).

CHAPTER TWO

Phylogenomic analysis reveals signals of adaptive evolution of three recently evolved sister species of woodpeckers in the genus *Sphyrapicus*

A.I.J. Curtis[1] and T.M. Burg[1]

[1]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, AB, T1K 3M4

2.1 Introduction

Recent and rapid species radiations have resulted in diverse morphological, genotypic, and behavioural characteristics between newly diverged species and populations in short periods of time. Birds, in particular, have experienced rapid radiation events throughout their evolutionary history (Jarvis *et al.* 2014; Zhang *et al.* 2014). While divergence for many species dates back to the Pliocene (Avise & Walker 1998; Avise *et al.* 1998), some species origins can be dated to the Pleistocene, culminating in the last glacial maximum (LGM) when ice sheets were at their fullest extent and species were isolated in refugia (Weir & Schluter 2004). Similarly, the divergence between populations of many species (Beatty & Provan 2010; Emerson *et al.* 2010; Lait & Burg 2013) corresponds to the LGM. Genetic drift and founder effects occurred as a result of changing suitable habitat and altered the frequencies of neutral alleles. Drift combined with prolonged isolation of populations can eventually lead to speciation.

Adaptive divergence through selection for morphological, behavioural, and genetic characteristics between populations can also result in rapid radiation of species. Populations diverge due to selection for different traits (Bradic *et al.* 2013; Brown *et al.* 2010; Hohenlohe *et al.* 2010), or behavioural characteristics (Moura *et al.* 2014). Even in the presence of ongoing gene flow when individuals with specific characters have higher fitness in one environment than another, natural selection can result in speciation.

Recent radiations result in very closely related species, sometimes with unclear taxonomic distinctions. These species can have very different morphological or behavioural characteristics and still be closely related genetically (Shaffer & Thomson 2007). Contact between these recently diverged species, without the benefit of

reproductive isolation or continued selective pressure, can lead to mixing of alleles producing one panmictic population and possible lineage fusion or despeciation (Garrick *et al.* 2014). With the more recent climate changes and the subsequent reduction of suitable habitat for many species, the areas of contact between hybridizing species risk becoming greater and possibly leading to fusion. Studying the demographic history can aide us in determining how differentiated species are in terms of maintaining genetic variation with continued or increased gene flow between the species.

Delimiting species in recent radiations and determining their phylogenetic histories has become easier in recent years, especially for non-model organisms, due to advances in sequencing technology such as genotyping by sequencing (GBS) – a form of restriction-site associated DNA (RAD) sequencing. The accumulation of vast amounts of markers in the form of single nucleotide polymorphisms (SNPs) throughout the genome allows for identification of both neutral and non-neutral markers, higher possibility of uncovering rare alleles in the genome, and for the incorporation of varying rates of evolution for independent loci (Edwards *et al.* 2005; Shaffer & Thomson 2007). Studies of recent species' radiations have used genomic data from next-generation sequencing (NGS) techniques to elucidate fine-scale phylogenetic resolution typical of recent radiation such as in *Wyeomyia smithii* – the pitcher plant mosquito (Emerson *et al.* 2010) and *Cyanistes teneriffae* – the blue tits in the Canary Islands and Africa (Stervander *et al.* 2015). McCormack *et al.* (2012) also used NGS to study four groups of closely related bird species such as the king rail (*Rallus elegans*) and clapper rail (*R. longirostris*), the juncos of North and Middle America (*Junco hyemalis*, *J. phaeonotus* and *J. vulcani*), the Jamaican streamertail hummingbirds (*Trochilus polytmus* and *T. scitulus*), and the

western US white-crowned sparrows (*Zonotrichia leucophrys*). NGS has even been able to resolve the phylogenetic relationships of the closely related and recently radiated Lake Victoria cichlid species using RAD-sequencing (Wagner *et al.* 2013). Increased resolution from SNP data has also identified the taxonomic uncertainty in the redpoll finch complex (*Acanthis flammea*, *A. hornemanni* and *A. cabaret*) bringing into question the current species classification of the three species (Mason & Taylor 2015).

This study explores one recently diverged group of species, the *S. varius* superspecies complex. It is composed of three species of sapsuckers in the genus *Sphyrapicus* – *S. ruber* (red-breasted), *S. nuchalis* (red-naped), and *S. varius* (yellow-bellied). The fourth member of the *Sphyrapicus* genus (*S. thyroideus*) has previously been determined to be sister to the *S. varius* complex (Cicero & Johnson 1995; Johnson & Zink 1983) and is not included in this study. The three species in the *S. varius* complex are very closely related yet have diverse morphology. *S. ruber* is distributed from the west of the Rocky Mountains to the Pacific Coast through British Columbia (BC), Washington (WA), Oregon (OR), and California (CA), while *S. nuchalis*' range is along the interior of the Rocky Mountains from Alberta (AB) southward to New Mexico (NM) (Figure 1.2). *S. varius*, in contrast, has a much larger range from northern BC to Newfoundland, staying mainly in Canada in the west and through the prairies, but dipping down into the United States as far south as North Carolina in the east (Figure 1.2). There are three main contact zones between the species where hybridization does occur. *S. ruber* and *S. varius* have a contact zone in northern BC, *S. varius* and *S. nuchalis* hybridize in central and southern AB, while *S. ruber* and *S. nuchalis* have an extensive contact zone all along their range overlap from BC to CA.

Previous studies on the sapsucker superspecies complex have found low levels of genetic variation among the sapsuckers with one study finding that mitochondrial DNA (mtDNA) divergence between the three species was low in comparison to other groups of congeneric species (Cicero & Johnson 1995). Very low levels of genetic variation have been found between *S. ruber* and *S. varius* using both allozymes and mtDNA cytochrome *b* sequences (Cicero & Johnson 1995; Johnson & Zink 1983). A study of the *S. ruber* and *S. varius* hybrid zone in northern BC by Seneviratne *et al*. (2012) was able to differentiate between *S. ruber* and *S. varius* using the mtDNA cytochrome oxidase I (COI) region and an intron in the CHD1Z gene (a Z-linked gene), but did not include *S. nuchalis* in the study. None of the methods used in previous studies have been able to determine if these three recently radiated species have diverged into three distinct, monophyletic lineages.

Three possible hypotheses could explain this pattern. One scenario assumes that the three species evolved in isolation during the Pleistocene and following the LGM, *S. ruber* and *S. nuchalis* came into secondary contact. During secondary contact, *S. nuchalis* underwent mitochondrial capture of the *S. ruber* mitochondrial genome, resulting in some *S. nuchalis* with *S. ruber* mtDNA and mixing of nuclear DNA between the two species. A second scenario for the low genetic divergence between *S. ruber* and *S. nuchalis* involves incomplete lineage sorting (ILS). The two species may have separated so recently that they have not had sufficient time for complete lineage sorting and reciprocal monophyly to occur. The third scenario is that there were only two main lineages: *S. nuchalis* and *S. varius*, no *S. ruber*. *S. nuchalis* would have been comprised of two distinct lineages which expanded into different ecological habitats. Selection on the *S. nuchalis* birds in the wetter coastal niche resulted in the adaptive speciation of *S. ruber* and *S. nuchalis* in the

dry interior. This was reinforced genetically through assortative mating between *S. ruber* and *S. nuchalis*, as seen in the *S. ruber* and *S. nuchalis* contact zone of northern CA and southern OR (Johnson & Johnson 1985). Divergence between these two species in the third scenario is through adaptation to different habitats instead of the neutral genetic drift involved in the first two scenarios.

In this study, I use a combination of genomic data, coalescent modelling, and species distribution modelling to evaluate recent radiation in the sapsucker superspecies complex and determine which of the above hypotheses best describes their evolution. The purpose of this study is to: (1) determine if greater resolution can be found between *S. ruber* and *S. nuchalis* using SNP data obtained from GBS through genomic analyses and (2) determine the demographic history of the *S. varius* superspecies complex. Specifically, I will be using GBS to determine whether the uncertainty in genetic differentiation between *S. nuchalis* and *S. ruber* using traditional markers is a result of adaptive radiation using genomic data, or incomplete lineage sorting or recent admixture using coalescent modelling and SDM.

2.2 Materials and Methods

2.2.1 Sample Collection and DNA preparation

Samples from live birds were collected under federal, state, and provincial permits using 12 m mist nets and call play back from throughout each of the species' ranges. A blood sample (<100 µL) was taken from the brachial vein of each individual and stored in ethanol. Birds were banded and released at capture sites. Museum specimens were provided in the form of toe pad and muscle tissue stored in lysis buffer. A list of museum

samples, their collection information, and accession numbers can be found in Appendix 1. A total of 79 birds (26 *S. ruber*, 34 *S. nuchalis*, and 19 *S. varius*) were used for this study.

Blood and tissue cells were lysed with an extraction buffer [0.1 M Tris (pH 8), 0.05 M EDTA, 0.5 M NaCl, and 1% SDS, 60 µg proteinase K, and 25 µg RNase] and the DNA extracted using a standard phenol-chloroform extraction procedure (Sambrook *et al.* 1989). An ethanol precipitation and two 70% ethanol washes were performed to remove any contaminants or enzyme inhibitors from aqueous DNA. DNA was resuspended in 40 µL of 1X low TE [10 mM Tris (pH 8) and 0.1 mM EDTA] and stored at -20^{o}C.

DNA was quantified and screened for quality by loading 1 µL DNA extract on a 1% agarose gel with ethidium bromide and 1 µL lambda *Hind*III ladder (Promega, Madison, WI). A digest was performed on 10% of the samples using *EcoR*I (New England Biolabs, Ipswich, MA) to test for enzyme inhibitors in the extractions. Digests were set up in 10 µL reactions using 300-500 ng of DNA and 1 µL (20 units) *EcoR*I at 37^{o}C for two hours. Samples that cut, had high quality and high molecular weight DNA were sent to Cornell University's Institute for Genomic Diversity (IGD) for genotyping by sequencing (GBS).

2.2.2 Genotyping by sequencing

A reduced representation library (RRL) containing 79 unique barcodes for each individual was created at the IGD using the restriction enzyme *Pst*I according to the procedure for genotyping by sequencing (GBS) previously described by Elshire *et al.* (2011). Due to the lack of a reference genome for *Sphyrapicus*, *de novo* sequencing was

performed. SNPs were called using the GBS UNEAK analysis pipeline version 3.0, an extension to the JAVA program TASSEL (Bradbury *et al.* 2007). Raw data were filtered for quality by removing any reads that had incorrect or missing barcodes or restriction cut sites, any uncalled nucleotides ('N') within the first 64 bases, and any reads containing multiple barcodes or cut sites. Reads were then truncated (or padded with a polyA tail) to 64 bases then collapsed into identical tags. Tags that were comprised of less than five reads were not included in the SNP calling process. The error tolerance rate was set to 0.03 for pairwise alignment identification of SNPs with a minimum minor allele frequency (MAF) of 0.05. SNPs were not filtered based on deviations from either F_{IS} or Hardy-Weinberg. Only one SNP per read was kept and I only included SNPs where sequence data were available in more than 80% of sampled birds.

As *S. ruber* and *S. nuchalis* are so closely related and resolution between these two species has been difficult to determine using traditional techniques, I also filtered these two species together without *S. varius* to identify SNPs shared only between them and determine if more data provide greater resolution. I again used the UNEAK pipeline with the same parameters as above.

2.2.3 Genomic analyses

To determine the genetic variation between the three species using SNP data, I used the program STRUCTURE v2.3.4 (Falush *et al.* 2003; Pritchard *et al.* 2000) which uses Bayesian inference to assign individuals to K clusters based on genetic data. No *a priori* assignment of individuals to species group is required for STRUCTURE analyses.

An initial burn-in of 10,000 replicates followed by 50,000 Markov Chain Monte Carlo (MCMC) iterations was run for each K = 1-6 with four iterations for each value of K. I employed the admixture model with inferred alpha and correlated allele frequencies with a set λ of one to the model. To aid in accurate inference of genetic structure in the presence of possible weak signal from recent species radiation, I used the phenotypic species identification as sampling location priors. The estimated Ln probability of data (lnPr(X | K)) value and Q values were used to determine number of clusters and the amount of admixture. Following the recommendation by Pritchard *et al.* (2000) for hierarchical analyses of clusters to test for further structure, I sequentially removed one group and ran the analysis with the remaining groups containing SNPs found in all three species. I then compared these STRUCTURE results to that of the dataset containing SNPs found only in *S. ruber* and *S. nuchalis* to determine if the increased data provide greater resolution between the two species. The *S. ruber* and *S. nuchalis* dataset was run with the same STRUCTURE parameters as listed above with the exception of running for K = 1-4. Since both the hierarchical analyses of the dataset from the three species and the STRUCTURE run of the dataset from *S. ruber* and *S. nuchalis* showed similar results (see Results section), I continued the remaining analyses using the SNPs found in all three species.

2.2.3.1 Detection of selective and neutral evolution

To test for putative loci under selection, I used the program LOSITAN (Antao *et al.* 2008) which utilizes the F_{ST} outlier detection method proposed by Beaumont and Nichols (1996). PGDSpider v2.0.8.2 was used to format the input files for the different

programs (Lischer & Excoffier 2012). An initial run of LOSITAN was performed using all loci to determine putative loci under selection. These loci identified as outliers in this initial run are then removed from the dataset in order to calculate a mean neutral F_{ST}, without noise from the putative loci under selection, as recommended for more accurate identification of outliers (Antao *et al.* 2008). I then used all loci (including those previously identified as putative loci under selection) in a second run of LOSITAN using the newly calculated mean neutral F_{ST} to identify loci under selection with 50,000 simulations of the bisection algorithm and a false discovery rate of 0.1. LOSITAN requires *a priori* assignment of individuals to populations, which I did based on phenotypic species identification.

A discriminant analysis of principal components (DAPC) (Jombart *et al.* 2010) was performed on all 79 individuals using the software package ADEGENET (Jombart 2008; Jombart & Ahmed 2011) implemented in R (R Core Team 2013). To determine if genetic structure found between the three species in the STRUCTURE analyses is the result of neutral demographic processes or selection, DAPC was run on three datasets identified by LOSITAN: neutral loci, loci under positive selection, and loci under balancing selection. DAPC is a multivariate analysis and, unlike STRUCTURE, it does not rely on underlying population genetic assumptions, such as Hardy-Weinberg equilibrium, and is not affected by linkage disequilibrium which is beneficial when using large datasets (Jombart *et al.* 2010).

For all three datasets I first ran DAPC without *a priori* assignment of individuals to groups and used the *k*-means clustering algorithm to determine the best number of supported clusters up to a maximum of ten. This algorithm aims to minimize within group

variation while maximizing variation between groups. Bayesian Information Criterion (BIC) was used to compare the different *k* solutions for selection of the number of clusters to use in DAPC. I plotted the BIC values as a function of their associated *k* (number of clusters) to identify the best supported number of clusters for each dataset. As the BIC value is similar for $k = 1$ to $k = 3$ for both the neutral and positive selection datasets, I selected $k = 3$, as that makes the most biological sense with there being three species. The datasets were run using both the *a priori* assignment of individuals to three clusters according to species identification and the three *k*-means clusters identified by ADEGENET. In each of the DAPC analyses, 30 principal components and two discriminant functions were retained.

Pairwise F_{ST} and nucleotide differences were compared using the software ARLEQUIN (Excoffier & Lischer 2010) along with the standardized pairwise F'_{ST} calculations performed using GenAlex v.6.501 (Peakall & Smouse 2006; Peakall & Smouse 2012) for both the neutral and positive selection SNP datasets. The pairwise nucleotide difference and F'_{ST}, which corrects for variation in heterozygosity (Meirmans & Hedrick 2011), were used to control for the effect of possible variation due to within species heterozygosity. Differences between positive and neutral patterns of species' differentiation were compared using a one-tailed *t*-test and χ^2 in a 2 x 2 contingency table.

2.2.3.2 Function of genes linked to SNPs

An important consideration in the use of LOSITAN to identify SNPs putatively under selection is that it has a tendency for type I errors (Narum & Hess 2011). To

remove any uncertainty in the identification of candidate gene function, only loci identified as positive outliers with a probability of 0.995 were included in functional Gene Ontology (GO) alignments, thereby reducing the possibility of including false positives. To determine if the loci under selection mapped to locations in the genome physically linked to candidate genes, I queried the nucleotide sequences for the SNPs under positive selection and those under balancing selection against the NCBI (National Center for Biotechnology Information) nucleotide collection database using the Basic Local Alignment Search Tool (BLASTN 2.2.32+) (Zhang *et al.* 2000) on default settings. A general nucleotide query against the NCBI database (www.ncbi.nlm.nih.gov) was used and putative GO classifications were determined based on homology using the megablast option for highly similar sequences. I also determined the expected heterozygosity of each locus for each of the three species using ARLEQUIN to determine the potential contribution each gene may have had in differentiating each species and if allele frequencies were fixed for any of the loci in specific species. Identifying the function of the candidate genes physically linked to the loci under selection will aid in determining the effect of adaptation.

2.2.4 Coalescent species tree analysis

The coalescent species tree of the *S. varius* superspecies complex was examined using the Bayesian MCMC software program SNAPP v1.1.6 (Bryant *et al.* 2012), a module of BEAST v2.1.3 (Bouckaert *et al.* 2014). SNAPP was developed specifically for SNP and amplified fragment length polymorphism (AFLP) data and estimates the species trees and parameters directly from the marker data instead of through sampling all the

gene trees (Bryant *et al.* 2012). To decrease computation time, which increases with more individuals used, I sampled a subset of six individuals from each species, in accordance with previous studies which have subsampled anywhere from one to six individuals per population and still been able to resolve phylogenetic clades (Ford *et al.* 2015; Giarla & Esselstyn 2015; Mason & Taylor 2015). SNPs identified as neutral by LOSITAN were used for determining the species tree by first removing any with missing data and MAF < 0.05 in those 18 individuals using PLINK v1.07 (Purcell *et al.* 2007). The backward and forward mutation rates (u and v) were co-estimated during the MCMC runs. All other parameters were kept at their default settings with the expected number of mutations between individuals (θ) determined by a gamma rateprior, a shape parameter (α) of 11.75, scale parameter (β) of 109.73 and a species divergence rate (λ) of 0.00765. SNAPP was run for ten million generations sampling every 1000 iterations with a burn-in of 10,000.

The output of the runs was inspected using TRACER v1.6 (Rambaut & Drummond 2007) to determine if runs converged and ensure that each parameter reached an effective sample size (ESS) > 200, after removing 10% of the trees as burn-in. Resulting tree sets were visualized using DensiTree v2.1.11 (Bouckaert 2010). I also generated the maximum clade credibility tree with TREEANNOTATOR v2.1.2 (Drummond & Rambaut 2007).

2.2.5 Species distribution modelling

Genomic data were compared against global climatic changes using species distribution modelling (SDM). Species distributions for both current climate, LGM and

the last interglacial (LIG) climates were modelled. These distributions were assessed to identify areas of possible refugia for each species and potential contact zones that may have allowed for continued gene flow between species.

2.2.5.1 Occurrence data

Occurrence data were obtained from the Global Biodiversity Inventory Facility database (GBIF; http://data.gbif.org/, accessed on August 4, 2014). Records that had sightings outside the known breeding grounds for each species and that did not have geo-references were removed. The dismo package in R (Hijmans *et al.* 2013) was used to remove some of the sample selection bias that is inherent in data retrieved from databases such as GBIF. Sample selection bias is most often geared toward easily accessible sites such as campgrounds, national parks and areas near roadways or waterways while harder to reach areas may not have comparable numbers of samples or be sampled at all. This introduces erroneous assumptions into presence-only data modelling software about abundance and presence of data in both hard and easy to access locations.

Another bias that can be introduced using global databases is through target museum sampling of known locations. One important assumption of presence-only modelling is that sampling is either random or uniform throughout the landscape, which is not the case for most GBIF data as discussed above (Dudík *et al.* 2005; Phillips *et al.* 2009; Yackulic *et al.* 2012). I used dismo to create a subsample of the occurrence points obtained from the filtered GBIF datasets to remove some of the bias created in using this data.

2.2.5.2 Environmental data

Current bioclimatic layers were downloaded from the WorldClim dataset v1.4 (Hijmans *et al.* 2005) and clipped using the SDMtoolbox (Brown 2014) to cover only North America in ArcMap v10.3 (ESRI: Redlands, CA). The PMIP2 (Paleoclimate Modelling Intercomparison Project Phase II) database (Braconnot *et al.* 2007) available from the WorldClim website (http://www.worldclim.org) provided the LGM Model for Interdisciplinary Research on Climate (MIROC) bioclimatic layers at 2.5 minute resolution. Bioclimatic data from the LIG were obtained at 30 arc-seconds resolution from WorldClim (Otto-Bliesner *et al.* 2006).

Correlated environmental variables were examined using ENMTools (Glor & Warren 2011; Nakazato *et al.* 2010; Warren *et al.* 2008, 2010; Warren & Seifert 2011) at a correlation threshold of 0.9. Those correlated variables that lent higher support to the model or had higher biological relevance to the species were kept and the others removed from the model. A total of eight bioclimatic variables containing data on annual mean temperature (Bio1), isothermality (Bio3), temperature seasonality (Bio4), max temperature of the warmest month (Bio5), annual precipitation (Bio12), precipitation seasonality (Bio15), precipitation of the driest quarter (Bio17), and precipitation of the warmest quarter (Bio18) were retained for running the model.

2.2.5.3 Modelling distribution

Seventy-five percent of the occurrence points were used for training and the remaining 25% for testing the model of species distribution using the maximum entropy

algorithm in MAXENT v3.3.3 (Phillips *et al.* 2006). The MAXENT model discriminates between the species presence points and a set of background points to determine habitat suitability. A total of 10,000 background points were used for comparison with occurrence points in the MAXENT models to determine climate suitability. The replicate run type was set to 25 cross validations, with a maximum number of 500 iterations and a convergence threshold of 0.00001. Only hinge features were selected to run the model and the regularization multiplier was set to 1. The area under the receiver operating characteristic curve (AUC) and the omission curve were used to determine overall performance of the model.

2.3 Results

2.3.1 Genomic analyses

GBS returned 261,915,468 raw reads. After removing those with errors in the barcode or cut site, 164,450,420 reads were retained as good reads and were then collapsed into 31,422,255 tags with an average of 5.23x coverage. These tags identified 7,657 SNPs in all three species of sapsuckers. Using these SNPs, the highest estimated Ln prob of data from STRUCTURE was found for K = 3, which was corroborated by the Q values (Figure 2.1a). The STRUCTURE results show both the *S. nuchalis* and *S. varius* individuals having isolated clusters with little admixture from the other species. The *S. ruber* cluster, however, has between 40-60% *S. nuchalis* ancestry. Re-running STRUCTURE in a hierarchical analysis on *S. ruber* and *S. nuchalis* individuals using the SNPs found in all three species identified K = 2 clusters that correspond to the phenotypic

identification of the species groups, with one *S. ruber* individual from central BC showing high *S. nuchalis* ancestry (Figure 2.1b).

A second dataset was created using the 8,555 SNPs found to be shared in *S. ruber* and *S. nuchalis* (filtered using only *S. ruber* and *S. nuchalis* individuals). STRUCTURE identified K=2 clusters as the highest probability with the two clusters being very similar to those identified when doing the hierarchical analysis (Fig 2.1). The same *S. ruber* individual with high *S. nuchalis* ancestry in the hierarchical analysis also showed high *S. nuchalis* ancestry in the *S. ruber* and *S. nuchalis* SNP dataset. This shows that while increasing genomic data does increase resolution between closely related individuals, as seen in these datasets that were both able to resolve the close relationship between these two species, there is an apparent threshold of resolution that can be reached for a dataset and any additional data becomes redundant.

2.3.1.1 Detection of selective and neutral evolution

Following the second LOSITAN run using the re-calculated neutral mean F_{ST} of 0.135 and all 7,657 SNPs identified as variable among all three species of sapsuckers, I identified 6,171 neutral loci. Of the 1,486 remaining loci identified as being outliers, 291 loci were determined to be under balancing selection and 1,195 under positive selection (Figure 2.2).

DAPC analysis of the neutral SNPs clearly discriminated between all three species in both the *a priori* and the *k*-means identified cluster datasets, with the 30 retained principal components explaining 50% of the variation (Figure 2.3). A *S. nuchalis*

individual from central AB clustered with the *S. varius* group in the *k* identified clusters, but with the *S. nuchalis* group using *a priori* cluster assignments. One *S. ruber* individual from central BC grouped with *S. nuchalis* sapsuckers in both the *a priori* and *k* clustering methods (Figure 2.3). This is the same *S. ruber* individual with the high *S. nuchalis* ancestry in STRUCTURE (Figure 2.1). The 30 retained principal components in the DAPC analysis on the loci under balancing selection explained 63% of the variation and showed no differentiation between the three species. The *k* value with the lowest BIC was for one cluster and the dataset with *a priori* assignment also grouped individuals into only one cluster, as is expected for conserved loci (Figure 2.4). The 30 retained principal components for the positive selection dataset explained 70% of the variation and had similar clustering to the neutral SNP dataset except the *S. nuchalis* CAB individual that clustered with *S. varius* in the neutral dataset clustered with the other *S. nuchalis* individuals in the loci under positive selection (Figure 2.5).

The pairwise F_{ST} values clearly differentiated the three species, with all three species showing significant genetic differentiation ($P < 0.05$) using both the neutral loci and the loci under positive selection datasets (Table 2.1). The standardized F'_{ST} measure showed very similar values to F_{ST} for differentiating the three species with both datasets (Table 2.2). In both the F_{ST} and F'_{ST} calculations, pairwise F_{ST} (or F'_{ST}) were higher when using the loci under positive selection. This difference between the neutral loci and the loci under positive selection F_{ST} patterns of species' differentiation was significant (*t*-test = 3.03, $P < 0.05$, d.f. = 2).

The comparison of nucleotide differences between species for both neutral and outlier loci datasets showed differentiation between the three species using both the

neutral loci and loci under positive selection with a similar pattern to the pairwise F_{ST} and F'_{ST} calculations. The loci under positive selection have more nucleotide differences at the sequence level between the species than the neutral, with the difference in the neutral and positive loci patterns of genetic differentiation being significant ($\chi^2 = 194.72$, $P < 0.0001$, d.f. = 2). *S. varius* have the greatest within species variation for both the neutral loci and loci under positive selection while *S. nuchalis* have higher variation in the neutral SNPs than those under positive selection and *S. ruber* have the least within-species variation for both datasets (Figure 2.6). The comparison between *S. ruber* and *S. nuchalis* showed lower pairwise nucleotide differences for loci under positive selection than neutral loci while the *S. varius* are distinct from the other two species for both datasets, but slightly more so from *S. nuchalis* (Figure 2.6). Between species comparisons for the neutral loci all have similar pairwise nucleotide distances while the comparisons for loci under positive selection are larger and have greater variation between species pairs, with the *S. varius/S. nuchalis* comparison showing the greatest number of nucleotide differences.

2.3.1.2 Function of genes linked to SNPs

Of the 1,195 loci under positive selection, LOSITAN identified 602 SNPs with the strongest support for positive selection (with a probability of 0.995 or higher). Twenty eight of the loci under positive selection aligned to NCBI sequences and 24 of those were located in or near genes with known function. These 24 loci are linked to candidate genes with multiple diverse putative functions such as cell signaling, structure, metabolism, immunity, tissue repair and blood clotting, cell regulation and development, cellular

transport, cranial skeletal development, and sound perception (Table 2.3). The remaining four loci aligned to locations in the genome that currently have no known function. There were three loci found to be fixed in *S. varius* and variable in the other two species, four fixed in *S. ruber*, and none fixed in *S. nuchalis* but variable in *S. varius* and *S. ruber*. There were also three loci found to be variable only in *S. varius* while both *S. ruber* and *S. nuchalis* were fixed for the same allele. The locus linked to the gene involved in sound perception was one of the three found to be fixed in *S. varius*, but variable in the other two. At this locus, *S. ruber* had a higher proportion of homozygotes favouring the allele that was fixed in *S. varius* while *S. nuchalis* had a higher proportion of homozygotes for the other allele; both *S. ruber* and *S. nuchalis* had similar numbers of heterozygotes at this locus. The other two loci that were fixed in *S. varius* but variable in *S. ruber* and *S. nuchalis* had similar genotype frequencies in both species with the greater proportion of individuals being homozygous for the allele not fixed in *S. varius*. For the four loci variable in *S. nuchalis* and *S. varius* but fixed in *S. ruber*, only one *S. nuchalis* individual (different for each locus) was variable for each locus. All other *S. nuchalis* individuals were homozygous for the same allele fixed in *S. ruber*.

Another 155 loci were identified with the strongest support (0.995 probability) for balancing selection using LOSITAN. BLAST aligned 14 of these loci to homologous NCBI sequences with functions mainly involved in cell signaling and transmembrane transport (n = 11), but also immunity, metabolism, and DNA translation (Table 2.4). One other locus under balancing selection aligned to an NCBI sequence with no known function.

2.3.2 Coalescent species tree analysis

A total of 8,990 trees were used in the SNAPP analysis to determine the species tree topology of the *S. varius* superspecies complex (Figure 2.7). The tree had a posterior probability of 69.5% and matches the mtDNA sapsucker superspecies complex phylogeny (Weir & Schluter 2004). The probability of the species tree having a ((*S. ruber*, *S. varius*), *S. nuchalis*) or ((*S. nuchalis*, *S. varius*), *S. ruber*) topology was 18.5% and 12% respectively. The maximum clade credibility is shown with the posterior probabilities of the nodes listed (Figure 2.7a). The analysis of the log file using Tracer indicated convergence of the species trees with all parameter ESS $\gg$ 200. The species tree cloudogram illustrates the wide confidence intervals in the coalescence times between the species trees and supports the hypothesis of ILS or gene flow being contributing factors to the low genetic differentiation using traditional markers (Figure 2.7b). The range of topologies with the variation in coalescence times is most likely due to the rapid, recent radiation of the three species and their continued gene flow.

2.3.3 Species distribution modelling

A total of 480 *S. varius*, 274 *S. nuchalis*, and 142 *S. ruber* occurrence points were used after correcting for sample bias with dismo. All three models performed above random (AUC = 0.5) with AUC scores of 0.90, 0.95, and 0.97 for *S. varius*, *S. nuchalis*, and *S. ruber* respectively (Appendix 2). These AUC values suggest that the models were able to accurately distinguish between occurrences and random background. The plots of the omission rates also show that the models were able to differentiate between suitable

and unsuitable habitat with the average omission rate closely resembling the models' predicted omission rates (Appendix 2). The bioclimatic variables with the greatest contributions to the species distribution models were Bio3 (isothermality – based on the mean diurnal range and the annual temperature range) for *S. nuchalis*, Bio4 (temperature seasonality – or monthly mean temperature) for *S. ruber*, and Bio5 (max temperature of the warmest month) for *S. varius*.

The projected current distribution for each species closely matched the known species' distributions (Figures 2.8a, 2.9a, 2.10a). The LGM projected distributions in general had a reduced range and a southward trend. *S. ruber* were in refugia in the south-western United States and along the coast (Figure 2.10b), *S. varius* in the south-eastern United States (Figure 2.8b), and *S. nuchalis* in the southern-most part of their range – slightly overlapping the *S. ruber* refugium in the south-western states (Figure 2.9b). A possible minimal contact zone between *S. varius* and *S. nuchalis* may have existed in New Mexico and Texas while *S. ruber* and *S. nuchalis* may also have had an extensive contact zone along the Sierra Nevada mountain range.

The MAXENT model for the LIG shows possible suitable habitat for *S. varius* sapsuckers was similar to their current distribution, but shifted northward (Figure 2.8c) which is similar to the LIG distribution of boreal forests (Otto-Bliesner *et al.* 2006). The *S. nuchalis* sapsucker LIG distribution was restricted to a small western strip of its current range in both Canada and the United States, but also included possible habitat in northern Mexico as well – unlike current suitable habitat (Figure 2.9c). Predicted habitat suitability for *S. ruber* during the LIG was very similar to current suitability and was restricted to the Pacific Coasts of Canada and the United States (Figure 2.10c). There may have been one

contact zone between all three species during the LIG in BC and another large contact zone between *S. ruber* and *S. nuchalis* along the western extent of the *S. ruber* distribution, through WA, OR and CA; similar to the contact zone in their current distribution but with a more extensive overlap.

2.4 Discussion

2.4.1 Increased genomic resolution with GBS

Previous studies have found very low genetic differentiation between *S. nuchalis* and *S. ruber* using traditional markers and have been unable to identify reciprocally monophyletic lineages (Cicero & Johnson 1995; Johnson & Zink 1983). This study, using thousands of markers from throughout the genome, has been able to distinguish among the three species using both neutral loci and loci under positive selection. Both the STRUCTURE and the DAPC plots show distinct clusters for each of the three species (Figures 2.1, 2.3, and 2.5). Similarly, the pairwise F_{ST} and F'_{ST} values showed significant genetic differentiation between all three species for both loci under positive selection and neutral loci datasets (Tables 2.1 and 2.2). The loci under balancing selection that would be more likely to be conserved, in contrast, showed no differentiation among the three species (Figure 2.4). Interestingly, the addition of the approximately 900 SNPs found solely in *S. ruber* and *S. nuchalis* did not increase the resolution between these two species any more than the hierarchical structure analysis using the SNPs found in all three of the species (Figure 2.1b and c). This is in contrast to the RAD-sequencing study on the Lake Victoria cichlid radiation where Wagner *et al.* (2013) found that increasing the

number of SNPs used subsequently increased the resolution between species. The difference in that study was that the increased number of SNPs were due to decreased filtering stringency allowing for loci with missing data for more individuals. Our study maintained the same threshold of missing data and the increase in SNPs came from decreasing the number of individuals used when filtering (only the two species instead of all three). This illustrates that there is a threshold to the amount of information attainable using NGS techniques and that understanding the study species and questions being addressed is essential to assist in bioinformatics pathways. Simply increasing the number of SNPs may not provide any further information on the study system, whereas increasing the number of rare alleles might.

2.4.2 Positive and neutral genomic patterns of differentiation

Both the neutral SNPs and the loci under positive selection support isolation of all three species. STRUCTURE differentiated all three species using SNPs shared among them (Figure 2.1). The DAPC analyses were able to genetically differentiate all three species using both the neutral loci and loci under positive selection datasets with only two individuals clustering with a species different from their phenotype (Figures 2.3 and 2.5). Those two individuals were from contact zones shared by the species they clustered with, where hybridization could explain their genetic similarity to the sympatric species. Similarly, both datasets were able to genetically differentiate the three species using F_{ST}, F'_{ST} (Tables 2.1 and 2.2), and pairwise nucleotide differences (Figure 2.6).

The genomic data from the two datasets (neutral loci and loci under positive selection) support the hypothesis that the species have evolved while in isolation due to neutral genetic drift and adaptive speciation. Other species that have evolved due to adaptation and selection from local environments also showed similar greater genetic differentiation in the loci under positive selection than the neutral loci (Chavez-Galarza *et al.* 2013; Hamlin & Arnold 2015; Moura *et al.* 2014). While the DAPC plot using the loci under positive selection was similar to the DAPC plot of the neutral loci, the pairwise F_{ST}, F'_{ST}, and nucleotide diversity showed distinct differences between the neutral loci and loci under positive selection datasets (Figure 2.6, Tables 2.1 and 2.2). This discrepancy may be due to the methodology of the DAPC analysis focusing on maximizing between-group variability, which would allow for better visualization and differentiation of clusters (Jombart *et al.* 2010). The genetic differentiation visible in the neutral loci using pairwise comparisons indicates either neutral genetic drift during isolation, possible linkage of the neutral loci to the loci under selection resulting in divergence hitchhiking, or some combination of both of these scenarios (Cruickshank & Hahn 2014). However, the stronger differentiation between species using the loci under positive selection compared to the neutral loci in the pairwise F_{ST} and F'_{ST} calculations indicates that not all the genetic variation is a result of neutral processes, but that selection played a role. Adaptation of sapsuckers to different habitats during both the LIG and the LGM followed by their subsequent adaptation to current habitats upon expansion out of glacial refugia may have resulted in selection for candidate genes that were linked to neutral loci which would then show similar patterns of divergence. While the pressure of adaptive speciation could have affected the genomic structure of the neutral loci, it is likely that the species naturally diverged through neutral drift and adaptive selection

while separated during the Pleistocene. Benham and Witt (2016) did not find evidence of adaptation along a geographic gradient, but instead found that geographic isolation of populations in different environments resulted in both neutral and adaptive divergence between hummingbird populations in the Andes. The similar genomic patterns of the two datasets and the strong signal of divergence seen in some of the loci under positive selection support the hypothesis that sapsuckers have evolved both through neutral genetic drift and selection on individuals due to differing ecological pressures from local environments while in isolation.

The measures of pairwise nucleotide diversity provide further evidence for evolution via drift and selection. Since F_{ST} is a relative measure of divergence, variation in within-species heterozygosity can bias the between species pairwise F_{ST} (Charlesworth 1998; Cruickshank & Hahn 2014). Measures of absolute diversity, such as nucleotide diversity (Nei & Li 1979) and F'_{ST} are not affected by within-species variation. The strong pattern of positive selection visible between *S. varius* and the other two species may be a result of strong selection to low interbreeding, especially between the phenotypically similar *S. varius* and *S. nuchalis*. The reduced interspecific nucleotide differences between *S. ruber* and *S. nuchalis* at the loci under positive selection compared to the neutral loci may be an indicator of recurrent selection throughout a history of continued gene flow between the two species (Cruickshank & Hahn 2014) and supports the close relatedness of these two species. Linked selection and recurrent selection along with continued gene flow and adaptation explain the rapid species radiation seen in the SNAPP coalescent modelling where loci resistant to introgression maintain their high variation between species resulting in their early coalescence times.

2.4.3 Functional annotation of SNPs under selection

Using traditional sequencing data, screening the genome for loci linked to genes involved in local adaptation is difficult in non-model organisms, however, GBS data were able to identify outlier loci that were potentially linked to candidate genes under positive selection for each sapsucker species. A few of these genes may have had a role in adaptive speciation among the three incipient species due to pre-zygotic barriers in the form of reproductive isolation, such as the locus linked to a gene involved in sound perception. Although no study has yet been conducted on vocalization between the three species of sapsuckers, an analysis of drumming patterns in *S. varius* concluded that sapsucker acoustic drumming displays are performed on substrates that provide the loudest and longest sound possible to provide the greatest range (Eberhardt 1997). Drumming has a similar function to passerine song in woodpeckers - a territorial acoustic display for deterring conspecific competitors and advertisement for potential mates during breeding season – and could therefore be involved in reproductive isolation between the species (Wilkins & Ritchison 1999). Interestingly, the locus linked to the candidate gene involved in sound perception was one of the three loci variable in *S. ruber* and *S. nuchalis* and fixed in *S. varius*, with *S. ruber* having a higher proportion of individuals with *S. varius* genotype than *S. nuchalis* (Table 2.3). This could be explained by the distribution of *S. nuchalis* being central to, and overlapping with, the other two species, requiring *S. nuchalis* to evolve to differentiate conspecific sounds from interspecific (Figures 2.8, 2.9, and 2.10).

Evidence of assortative mating in the northern CA and southern OR contact zone between *S. ruber* and *S. nuchalis* (Johnson & Johnson 1985; Trombino 1998) and the

ability of *S. ruber* and *S. nuchalis* to discriminate between interspecific and conspecific drums in sympatry, though not in allopatry (Trombino 1998), could be indicative of sexual selection due to sound perception. Other woodpeckers species have shown similar patterns of increased mate choice discrimination when found in sympatry with the competing congener though they show no difference in reactions to interspecific drums compared to conspecific drums in areas of allopatry (Stark 2002; Stark *et al.* 1998), while *Drosophila subquinaria* have increased discrimination for conspecific mates in populations sympatric with their congener, *D. recens*, than in allopatric populations (Rundle & Dyer 2015). This increased preference for conspecific mate choice in sympatric populations indicates that positive selection is occurring in areas of contact between the two species to reduce interbreeding with the congener. A study on the *S. ruber* and *S. varius* hybrid zone, in contrast, found no difference in the reactions of either species to the other's mating calls (Seneviratne *et al.* 2012). Another study found no evidence of discrimination between interspecific and conspecific drums in allopatric *S. varius* and *S. nuchalis* populations – although they did not study drumming in sympatric populations where greater discrimination between conspecific and interspecific drumming patterns had previously been found (Stark 2002; Stark *et al.* 1998). This may be that *S. varius* diverged earlier and did not have much contact with the other two species except in small areas of habitat overlap (Figures 2.8, 2.9, and 2.10). In contrast, the continued areas of sympatry of *S. nuchalis* and *S. ruber* may have resulted in selective pressure to reduce interspecific mating through sensory perception of each species to interspecific sounds and resulting in selection on that locus.

The two other loci found to be fixed only in *S. varius* aligned to a region of the genome possibly involved in cranial skeletal construction and heart muscle development known as the DiGeorge syndrome critical region (NCBI accession number XM_009910311). Mutations in the DiGeorge critical region result in facial anomalies and heart defects in humans with at least one gene transcript being strongly expressed in human adult skeletal and muscle heart tissue (Gong *et al.* 1997; Robin & Shprintzen 2005). As this region is involved in cranial skeletal construction, it is possible that selection on this trait has aided the rapid radiation of these species. Selection on the loci involved in cranial structure may have played a role in the nesting and foraging strategies of the three species. While sapsuckers nest in trees that have been softened either due to decay, preferred by *S. varius* and *S. nuchalis* (Dobkin *et al.* 1995; Losin *et al.* 2006; Savignac & Machtans 2006), or death, preferred by *S. ruber* (Joy 2000), the hardness of those trees varies for each species. *S. varius* prefer nesting in mixed wood forests, specifically trembling aspen or maple tree stands (Savignac & Machtans 2006; Tozer *et al.* 2011) and forage on living aspen, maple, and birch trees (Eberhardt 2000; Savignac & Machtans 2006). *S. ruber* instead nest in old growth conifer forests in dead and decaying Douglas-fir, pine, and hemlock trees (Backhouse 2005; Joy 2000), while they still forage at live deciduous and conifer trees (Backhouse 2005). The nesting and foraging patterns of *S. nuchalis* are a mix of the other two species. While they nest in aspen stands like *S. varius* (Losin *et al.* 2006; Smith 1982), they have also been found in old growth conifer forests like *S. ruber*, and nest in larch and birch trees (McClelland & McClelland 2000; Tozer *et al.* 2011). The aspen trees preferred by *S. varius* are hardwood while the coniferous trees preferred by *S. ruber* and some populations of *S. nuchalis* are softwood trees. The woodpecker family has developed specialized cranial structure through natural

selection in order to cope with the impact to their brain case during drilling and drumming (Bock 1999). Adaptation to nesting in the harder wooded aspen and maple stands may have played a role in the selection for a single allele in *S. varius* at this locus. Conversely, the *S. nuchalis* and *S. ruber* have similar habitats, so selection has resulted in a higher proportion of homozygotes for the other allele at this locus that codes for cranial development. Similar influences of nesting and foraging strategies on genomic population structure were found in killer whales, corresponding to different behavioural characteristics (Moura *et al.* 2014). Populations that inhabit close geographic locations but different ecotypes are more differentiated than populations that are further apart geographically but share an ecotype (Moura *et al.* 2014). Similarly, the choice of trees for nest building and sap-well drilling may have aided the differentiation of sapsuckers through adaptation and selection on candidate genes possibly associated with skull structure development.

The other 25 loci under positive selection and linked to candidate genes had diverse functions (Table 2.3). Most were involved in cell cycle processes and regulation, membrane transport, development, and signaling (Table 2.3). While bird species historically have undergone rapid radiation, the genomic conservation of function, synteny, and sequence is generally much higher in birds than in mammals (Zhang *et al.* 2014). A recent genomic study covering multiple avian orders determined that the evolution of certain Gene Ontology (GO) categories, such as spinal cord development, microtubule-based processes, cell cycle processes, and transmembrane processes, were faster in birds (specifically Neoaves) than in mammals while those genes involved in brain function and neural function evolve faster in mammals (Zhang *et al.* 2014). This

could be due to the reduced genome sizes of birds and the adaptations necessary for flight. These GO categories were some of the ones that the BLAST search aligned the 28 loci to in this study, supporting the identification of these loci as outliers and the possibility of rapid evolution of these regions of the genome. The GO classifications of the loci under positive selection further support the hypothesis that these three species were also under natural selective pressures and have undergone adaptive evolution along with neutral evolution.

2.4.4 Coalescent modelling supports rapid radiation

The species tree with the highest support from the coalescent model had the same topology as the mtDNA COI region from Weir and Schluter (2004), with *S. ruber* and *S. nuchalis* as sister taxa to *S. varius*. The overall pattern supports the previous findings of the close relationship between *S. ruber* and *S. nuchalis* using allozyme and mitochondrial markers (Cicero & Johnson 1995; Johnson & Zink 1983), however, SNP data have identified distinct genetic clusters for each species. Irrespective of the cytonuclear discrepancy, the GBS and mtDNA COI both show rapid radiation. Similar results have been found in cichlid fish inhabiting extreme soda lakes in Africa (Ford *et al.* 2015), which has resulted in rapid adaptive radiation with continued hybridization and gene flow, where the species trees support the recent branching of closely related species previously un-identifiable by traditional markers. In this study, the coalescence times for many of the trees varied (Figure 2.7) and the identification of two alternative tree topologies (*S. ruber*,(*S. varius, S. nuchalis*)) (12.5%), and (*S. nuchalis*,(*S. varius, S. ruber*) (18.5%) is indicative of either incomplete lineage sorting or gene flow between the

species (Bryant *et al.* 2012). These alternative topologies place a closer relationship between *S. ruber* and *S. varius* or *S. nuchalis* and *S. varius* and goes against the hypothesis that *S. ruber* evolved from the adaptation of a second *S. nuchalis* lineage after expansion. In this hypothesis we would expect to see all the *S. ruber* species trees most closely related to *S. nuchalis*, but the existence of the two alternative topologies indicate that both *S. ruber* and *S. nuchalis* most likely diverged from the same ancestral lineage and retain different ancestral alleles. The recent coalescence times seen in the tree along with the clear identification of distinct clusters by both STRUCTURE and DAPC (Figures 2.1, 2.3, 2.5, and 2.7), support the hypothesis that the species have diverged rapidly due to adaptive speciation and neutral drift in the presence of continued gene flow and admixture resulting in the low genetic differentiation between *S. ruber* and *S. nuchalis*. With adaptive speciation, we would expect to see short branch lengths indicative of recent radiation and variation in patterns of genetic differentiation between neutral loci and loci under positive selection (Brown *et al.* 2010; Ford *et al.* 2015; Hamlin & Arnold 2015). The short branch lengths of many of the *S. ruber* and *S. nuchalis* trees in the cloudogram support the recent radiation in this complex (Figure 2.7). Additionally, the pattern of increased differentiation between the species using loci under positive selection compared to neutral loci lends support to the theory of adaptive radiation among the species (Figures 2.3 and 2.5, Tables 2.1 and 2.2). Finally, hybridization between species can affect the evolution of different parts of the genome at differing rates, as MacLeod *et al.* (2015) found in the marine iguanas in the Galapagos Islands where estimates of divergence times for nuclear DNA were much more recent than that for mtDNA. In a study on boreal bird superspecies complexes, the *Sphyrapicus* complex had the most recent split (*S. ruber* and *S. nuchalis*) of all nine superspecies studied (Weir &

Schluter 2004). It is possible that dating the divergence times of sapsuckers using genomic data may show different estimates from the mtDNA as a result of their hybridization and adaptive radiation; unfortunately, as the tree did not include outgroups in this study, divergence dating was not performed.

2.4.5 SDM and potential contact zones

While the three sapsucker species have all inhabited different ecological niches throughout the LIG, LGM, and current time period, there have been instances of possible overlap between the species throughout their histories. In particular, *S. ruber* and *S. nuchalis* appear to have maintained range overlap throughout the duration of their histories (Figures 2.9, and 2.10). This may explain the close genetic relationship between the two species, in addition to their recent divergence. The two species may not have had sufficient time in isolation for complete lineage sorting and reproductive isolation to occur, resulting in closely related species with shared ancestral alleles that are able to hybridize. This hybridization between the two diverging species may have promoted adaptive speciation through the sharing of adaptively advantageous alleles between species while maintaining the areas of the genome that are resistant to introgression, possibly aiding the speciation process (Abbott *et al.* 2013). Johnson and Johnson (1985) found evidence for asymmetric mating between *S. ruber* and *S. nuchalis* in the northern CA contact zone, with *S. nuchalis* females showing higher preference for the redder *S. ruber* males, which supports the hypothesis that gene flow occurred in contact zones during the LGM and LIG.

The isolation of *S. varius* through much of the last 130,000 years is evidenced in their (mostly) allopatric distribution during both the LIG and LGM and distinct genetic clustering (Figures 2.1, 2.3, 2.5, 2.8, 2.9, and 2.10). During the LGM, *S. varius* may have had a small area of overlap with *S. nuchalis*, however, in the LIG there does not appear to be any overlap (Figures 2.8c and 2.9c). While in allopatric isolation, *S. varius* adapted to the local environment and underwent genetic drift, but not complete reproductive isolation. Following secondary contact, incomplete reproductive isolation allowed for gene flow between the species but now selection for specific traits maintained the differentiation between species. This is supported by the genomic patterns of differentiation visible in the neutral SNPs and those under positive selection described earlier.

This study supports a combination of the three hypotheses for evolution of the *S. varius* superspecies complex. The three species were historically isolated with limited contact for gene flow, as evidenced by both the species distribution modelling of suitable habitat during the LIG and the LGM and the genetic data. *S. varius* most likely evolved in isolation, with only a small area of possible overlap with *S. nuchalis* during the LGM, as it is the most genetically distinct from the other two species. *S. ruber* and *S. nuchalis* had large areas that were mostly isolated with areas of overlapping suitable habitat through much of the LIG and the LGM where gene flow between the two most recently radiated species may have occurred. The SNAPP analysis shows that all three species' lineages have undergone rapid and recent radiation and the variation in branch lengths implies ILS or admixture. The genetic data show that they are distinct species, making the hypothesis that ILS is responsible for the cytonuclear discord unlikely, as all three species have clear,

distinct clades with GBS data, and lending support to the secondary admixture hypothesis. Since ILS is unlikely due to distinct genetic clusters, the paraphyletic mtDNA is most likely a result of admixture between *S. ruber* and *S. nuchalis*. The hypothesis that the *S. ruber* clade is evolved from an isolated *S. nuchalis* lineage is supported by the low nucleotide diversity of *S. ruber* (Figure 2.6). This could be a result of a population of *S. nuchalis* that became isolated, lost allelic diversity through the founder effect, and evolved due to local selection for the warm and wet climate along the Coast. Additionally, the higher pairwise difference between *S. ruber* and *S. nuchalis* from the loci under positive selection compared to the neutral loci (Tables 2.1 and 2.2) show that adaptation aided in the evolution of the *S. ruber* clade. However, the variation in topologies from the coalescent model shows that *S. ruber* has unique alleles that are not a subset of *S. nuchalis*, but instead are more closely related to *S. varius*. This instead supports the hypothesis that admixture and hybridization account for the paraphyletic mtDNA structure of *S. nuchalis* and its close relation to *S. ruber*. All three species showed evidence of adaptive speciation as well as neutral drift with possible gene functions of selected loci playing a role in niche specialization and possible reproductive isolation. Therefore, it appears that the *S. varius* superspecies complex has evolved partially in isolation, but with continued gene flow between the species through areas of limited contact throughout their demographic history, while their rapid divergence has been maintained and driven through natural selection by their local habitats and continued hybridization.

Future studies on gene expression analyses through RNA-seq data may provide further insight into how these species have diverged due to selection, especially with

regards to plumage variation and sexual selection. Also, divergence dating of the species, not possible in the present study due to limited resources, using genomic data and outgroup species would be advantageous to clarify differences in nuclear and mitochondrial patterns of genetic differentiation. Finally, an in depth study on the hybrid zone between *S. nuchalis* and *S. ruber*, possibly using hybrid capture genomic data, would be beneficial in clarifying potential adaptive genomic regions under selection between the two species resulting in the distinct morphological variation between *S. ruber* and *S. nuchalis*. This study is the first to use genomic data to resolve the species lineages of the sapsuckers in the *S. varius* complex and identify genomic patterns of variation between the closely related and recently radiated species and lays the foundation for future genomic and gene expression studies on this complex. I determined that both genetic drift and adaptation to local environments in isolation has driven divergence and identified possible candidate genes under positive selection within the three species that may have been some of the mechanisms of divergence to drive speciation in the presence of continued gene flow in the *S. varius* superspecies complex.

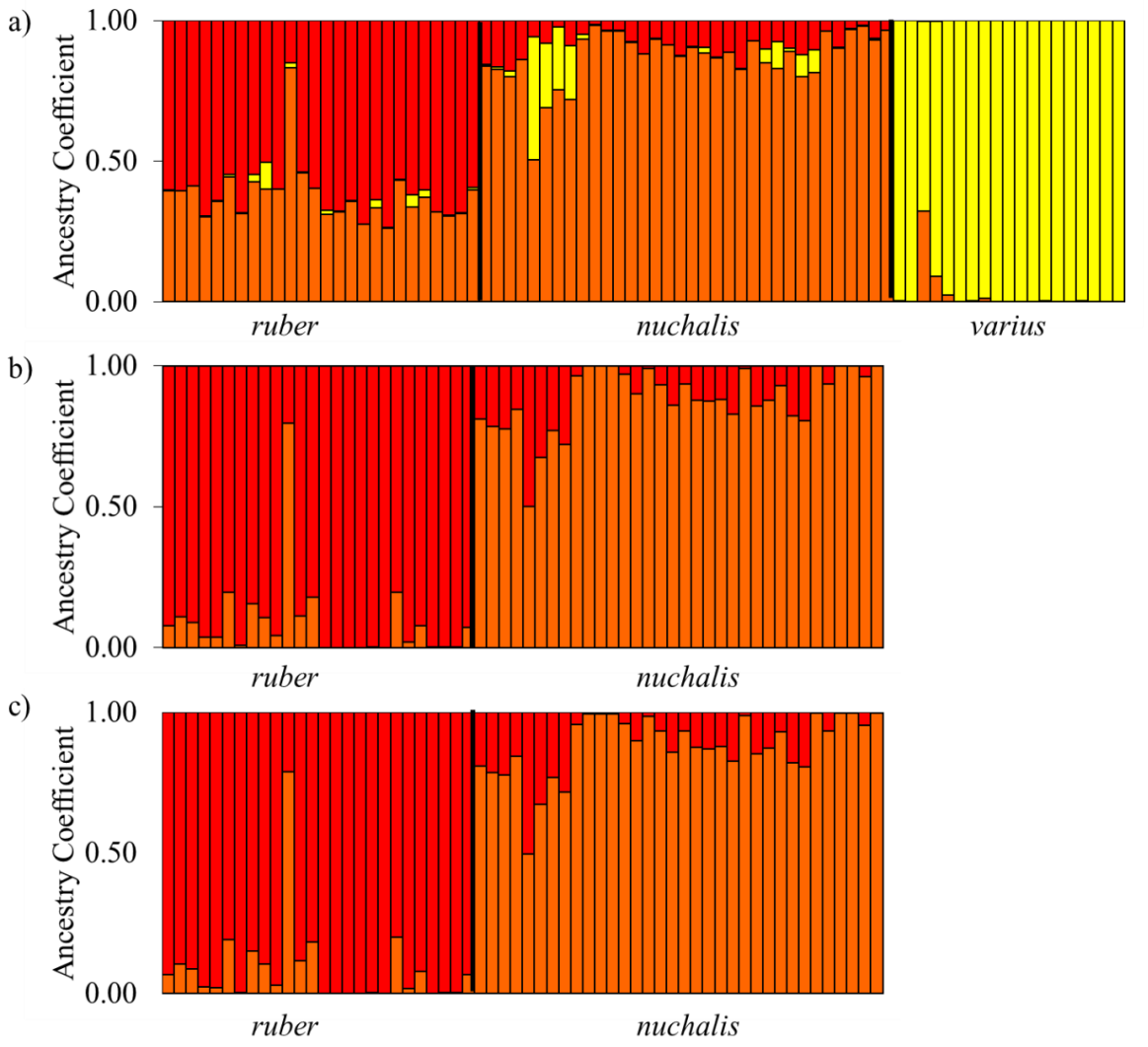


Figure 2.1. STRUCTURE plot of assignment of ancestry coefficient values for all individuals for K = 3 clusters using all 7,657 SNPs (a), and only *S. ruber* and *S. nuchalis* individuals using the 7,657 SNPs found in all three species (b), and using the 8,555 SNPs found to be shared between those two species (c).

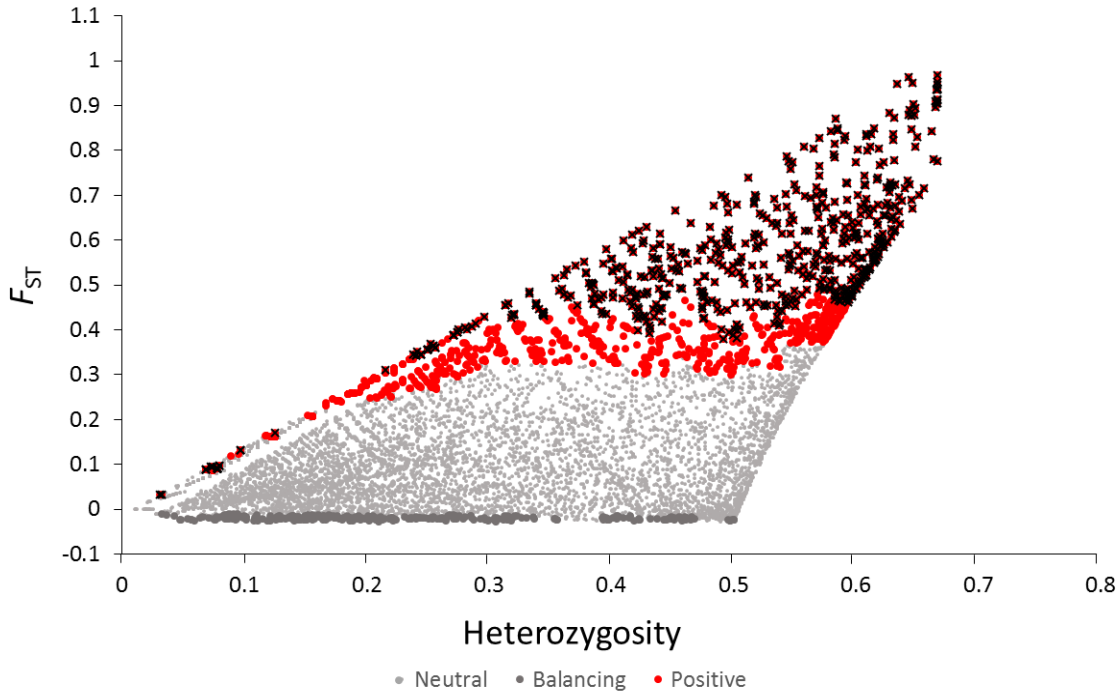


Figure 2.2 Outlier detection plot from LOSITAN based on neutral expectance of F_{ST}. Loci under positive selection are identified by red circles and those loci under balancing by grey. Possible candidate gene loci (black X) are the loci identified as loci under positive selection outliers with 99.5% probability (used in the gene ontology analyses).

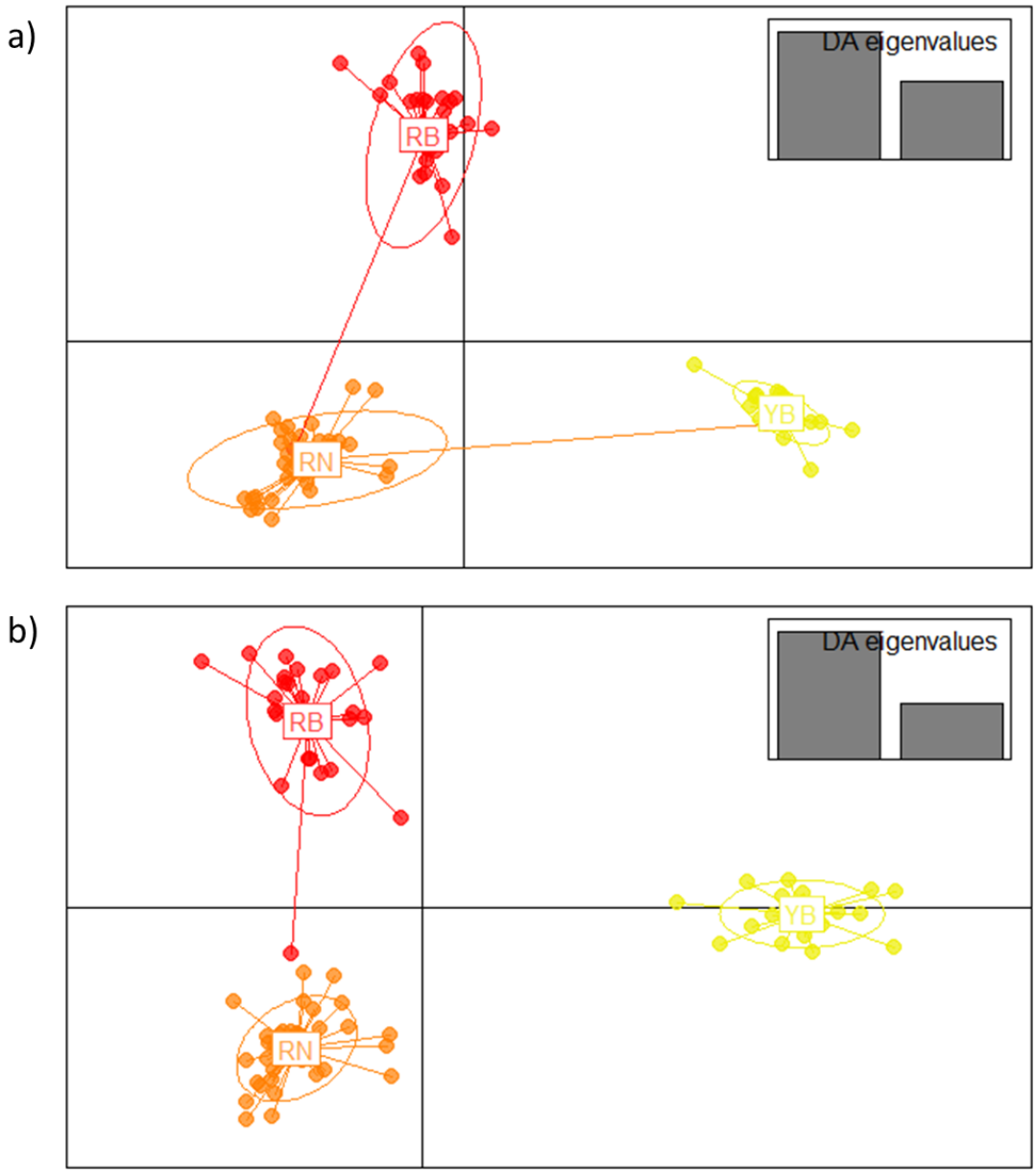


Figure 2.3 Scatter plot of DAPC of the neutral loci dataset. Clusters shown are identified using (a) the *k*-means clustering in ADEGENT and (b) clusters assigned *a priori*. Yellow circles (cluster YB) are the *S. varius* individuals according to phenotypic species identification, orange (RN) are *S. nuchalis* and red (RB) are *S. ruber*. The two discriminant eigenvalues are plotted in the bar graph inserts and indicate the amount of variance described by each. The principal components retained for each plot (30) contain 50.9% of conserved variance.

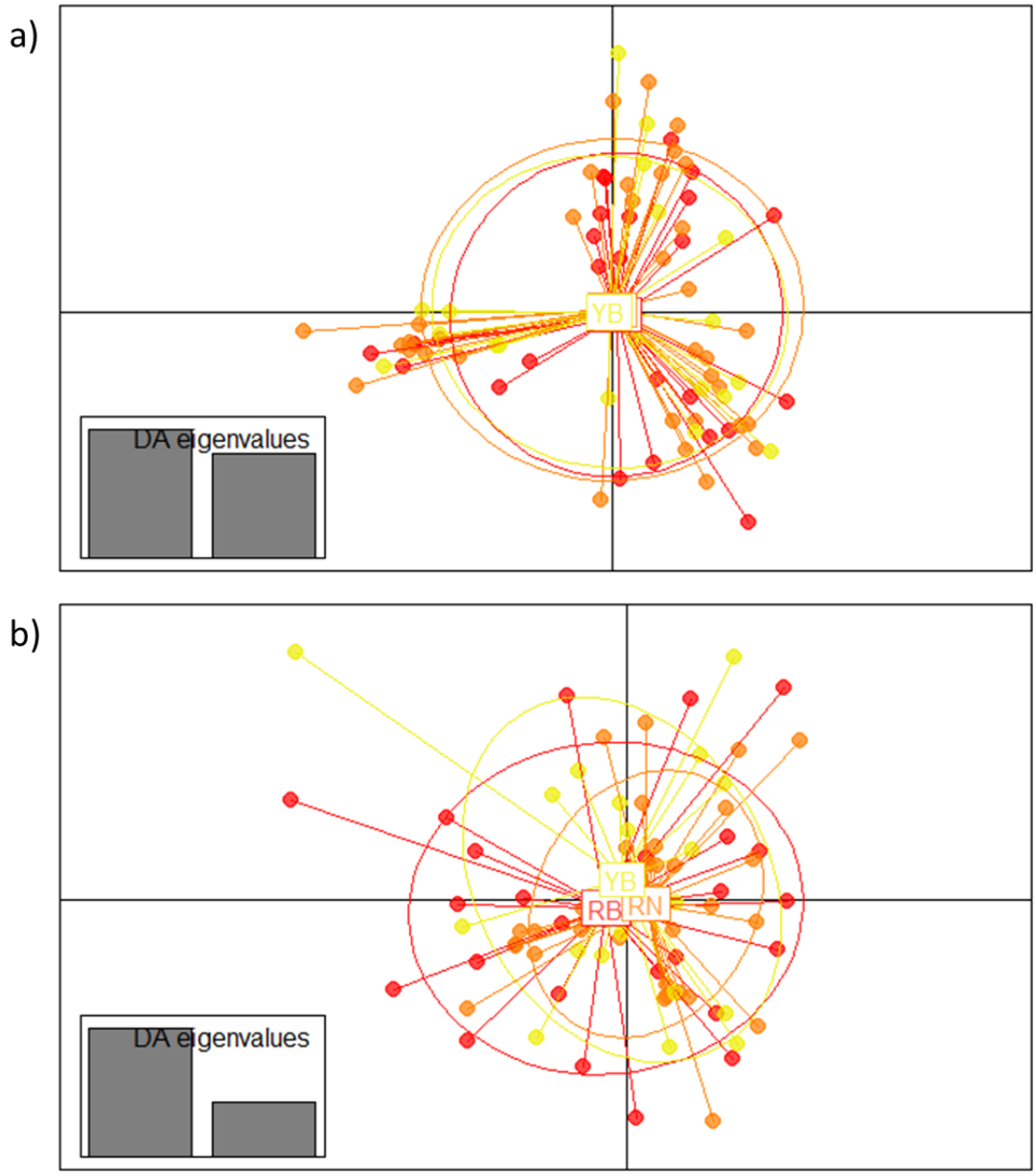


Figure 2.4 DAPC of the loci under balancing selection. Clusters were identified using (a) the *k*-means assignment in ADEGENT and (b) *a priori* assignment according to phenotype. Colours and grouping of individuals correspond to those in Fig. 2.3. The proportion of variance each of the two discriminant eigenvalues explains is plotted in the bar graph inserts. The principal components retained for each plot (30) explain 64.4% of conserved variance.

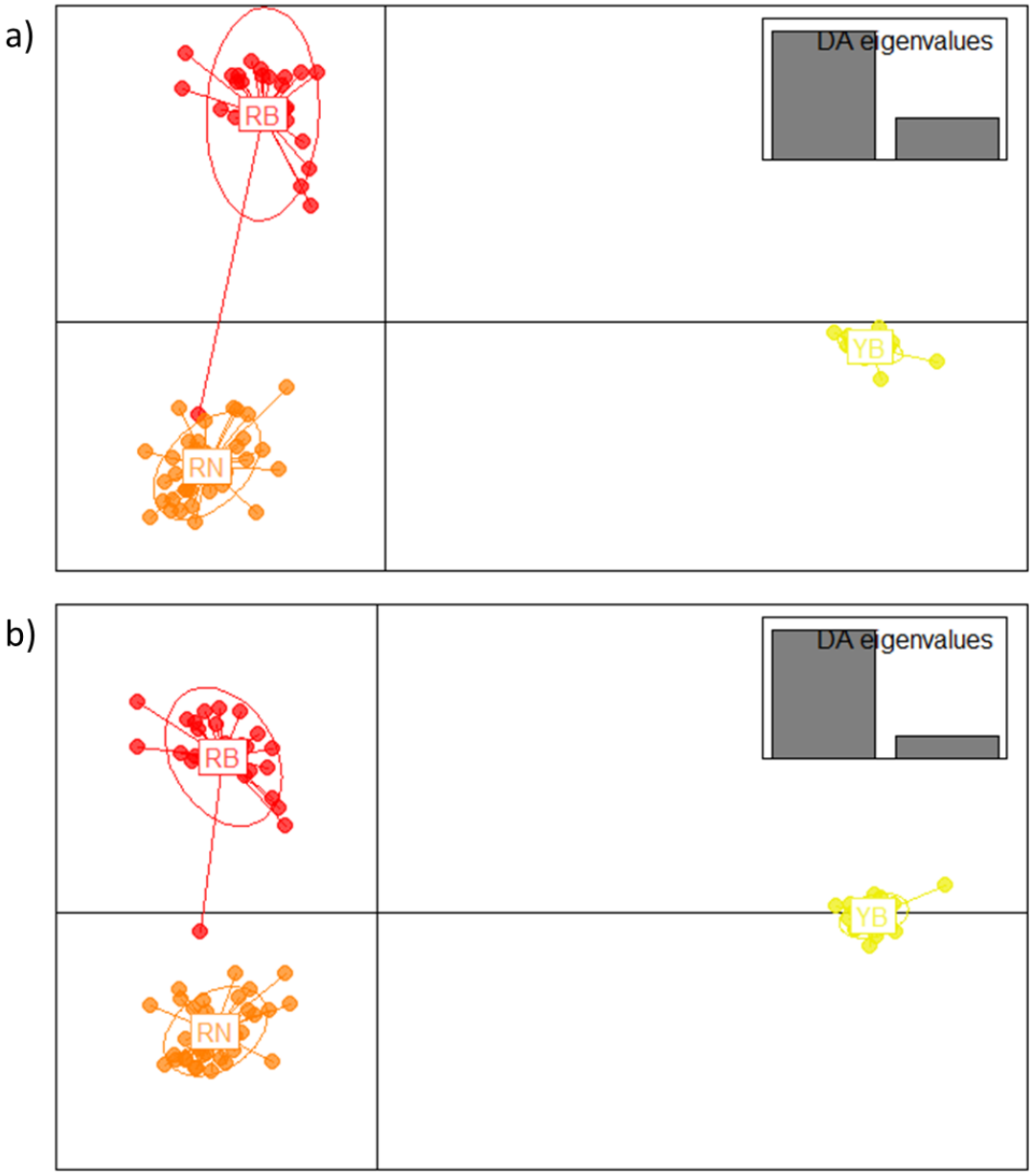


Figure 2.5 Discriminant analysis of the loci under positive selection. Clusters were identified using (a) the *k*-means *a posteriori* assignment in ADEGENT and (b) *a priori* assignment according to phenotype. Colours and grouping of individuals correspond to those in Fig. 2.3. Bar graph inserts depict the proportion of variance each of the two discriminant eigenvalues explains. The principal components retained for each plot (30) contain 73.7% of conserved variance.

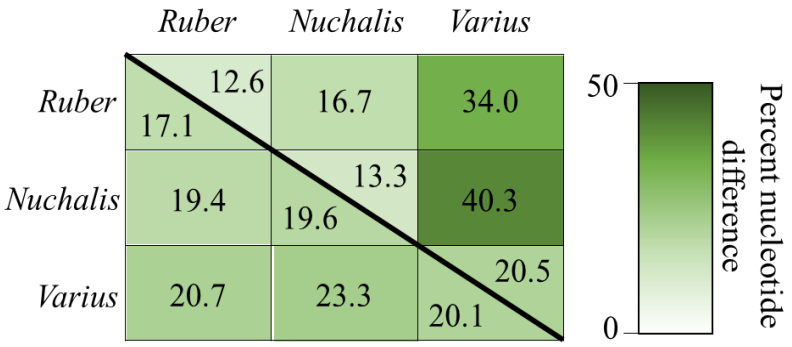


Figure 2.6 Pairwise comparisons of nucleotide differences for loci under positive selection (above diagonal) and neutral loci (below diagonal). Within species differences are along the diagonal. Numbers are recorded as the percent of the sequence that is different between the two species (or within the species).

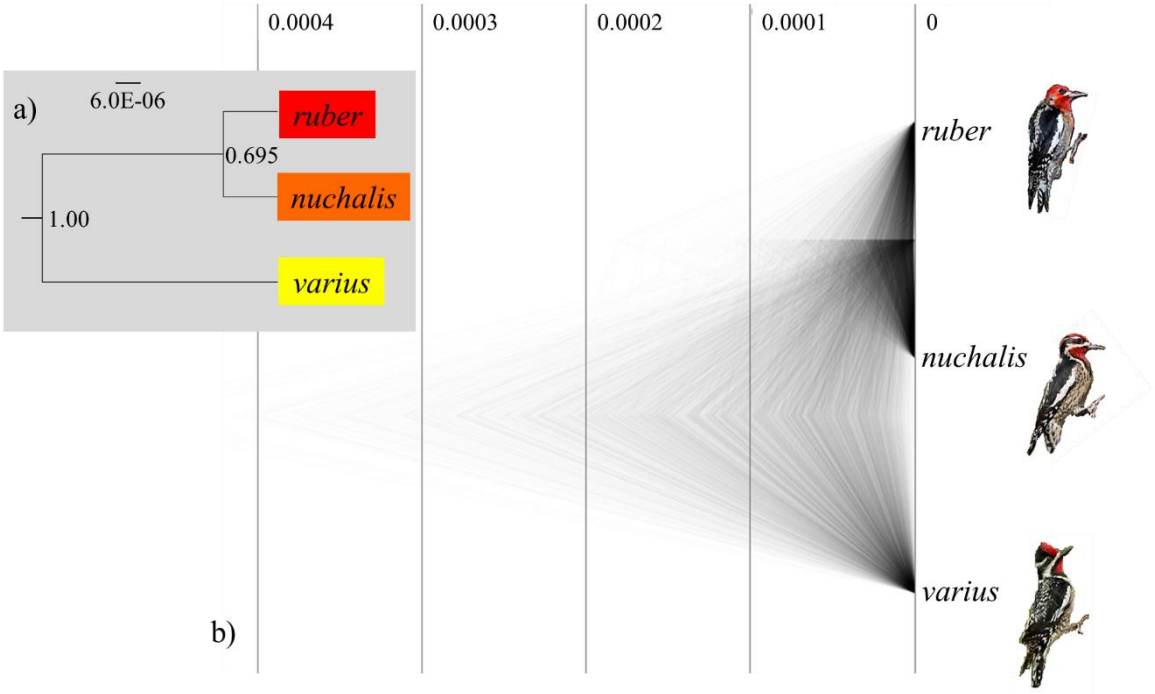


Figure 2.7 SNAPP coalescent species tree analysis of 8,990 trees. The maximum clade credibility tree with the posterior probability of each node identified is the grey insert (a) and the species tree cloudogram is shown (b).

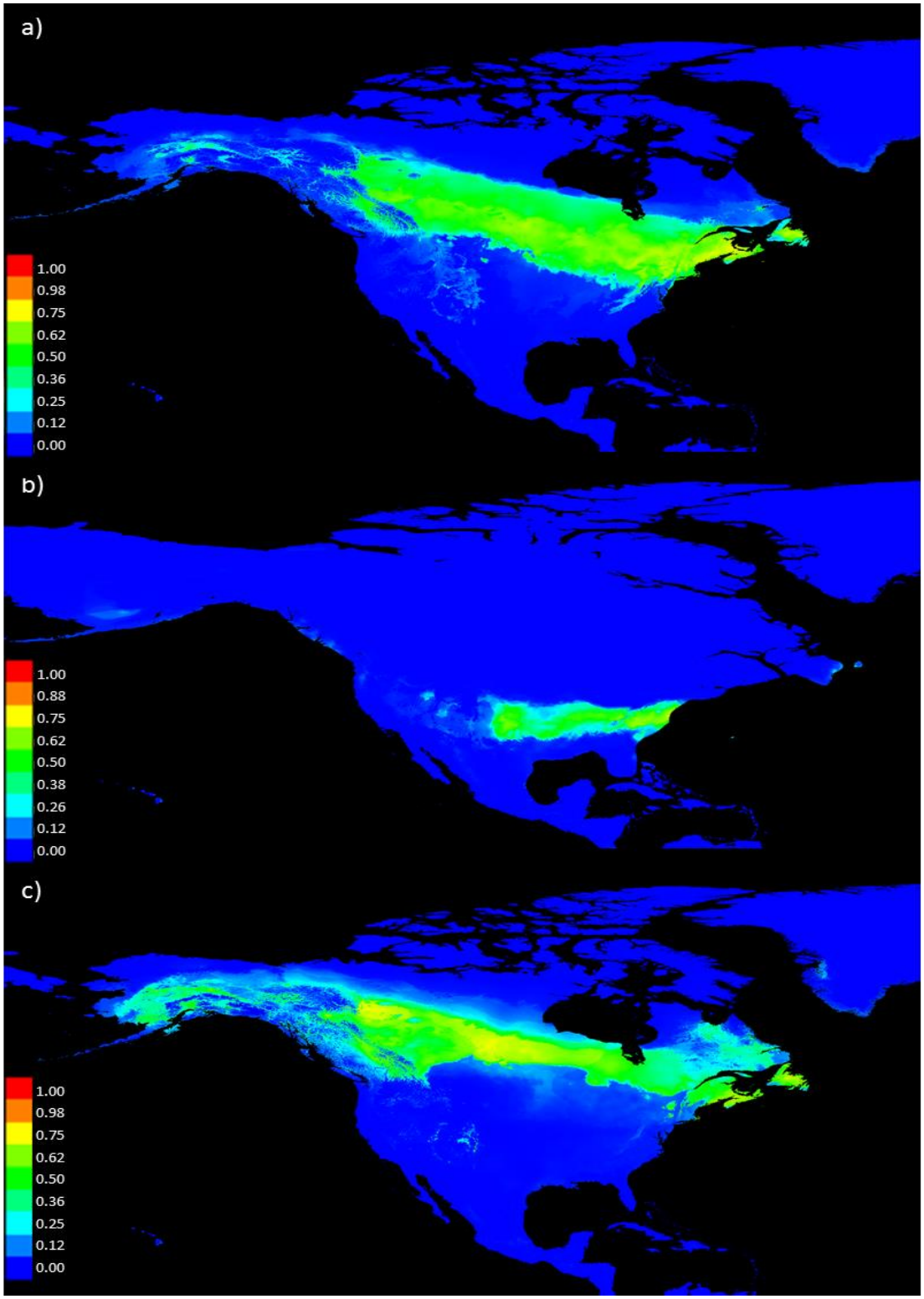


Figure 2.8 MAXENT species distribution modelling of suitable habitat for *S. varius*. The top map is the current species distribution (a), (b) is the prediction of suitable habitat during the LGM, and (c) is prediction during the LIG. Areas with least probability of suitable habitat are shaded dark blue.

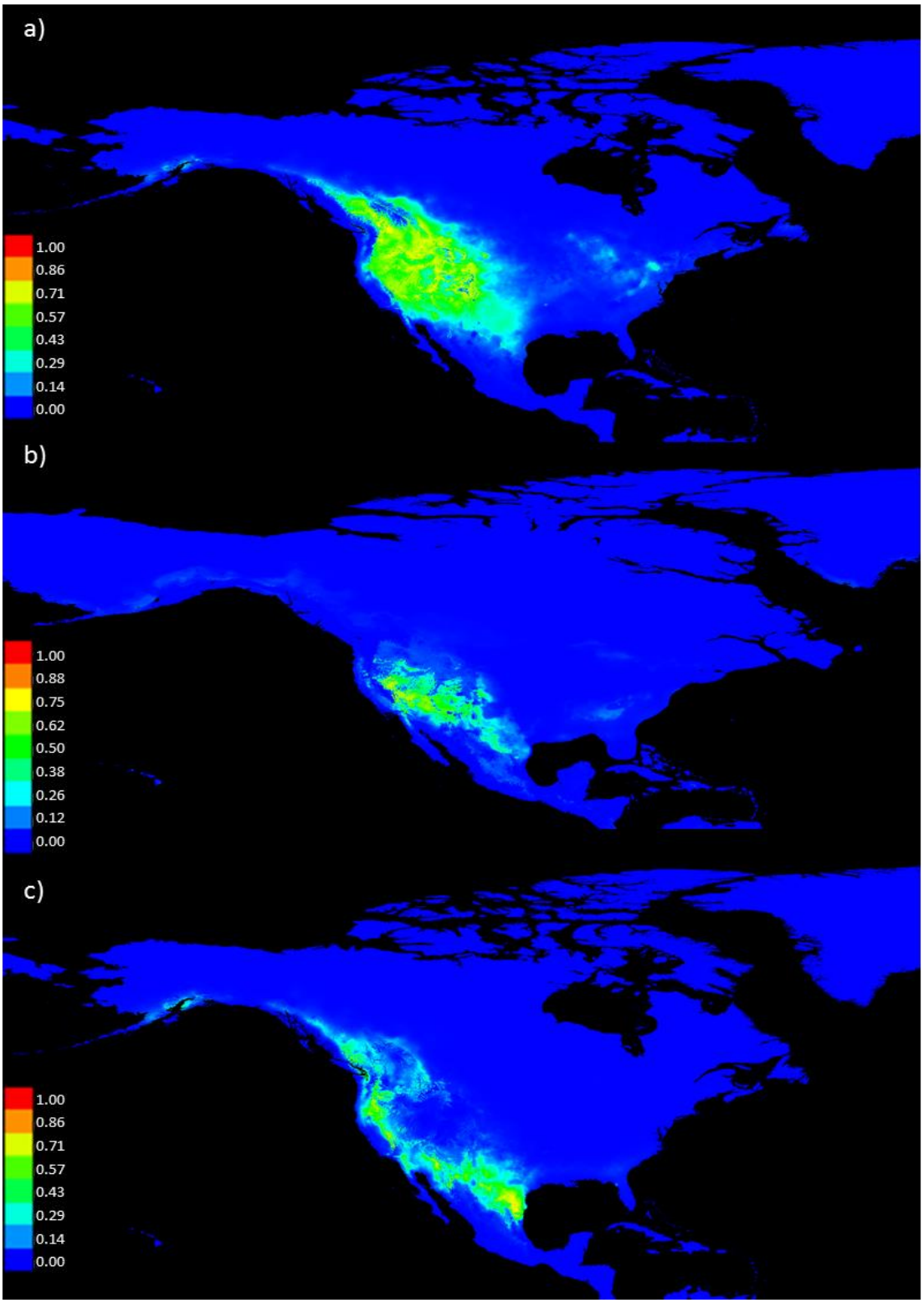


Figure 2.9 MAXENT species distribution modelling of suitable habitat for *S. nuchalis*. The top map is the current species distribution (a), (b) is the prediction of suitable habitat during the LGM, and (c) is prediction during the LIG. Areas with least probability of suitable habitat are shaded dark blue.

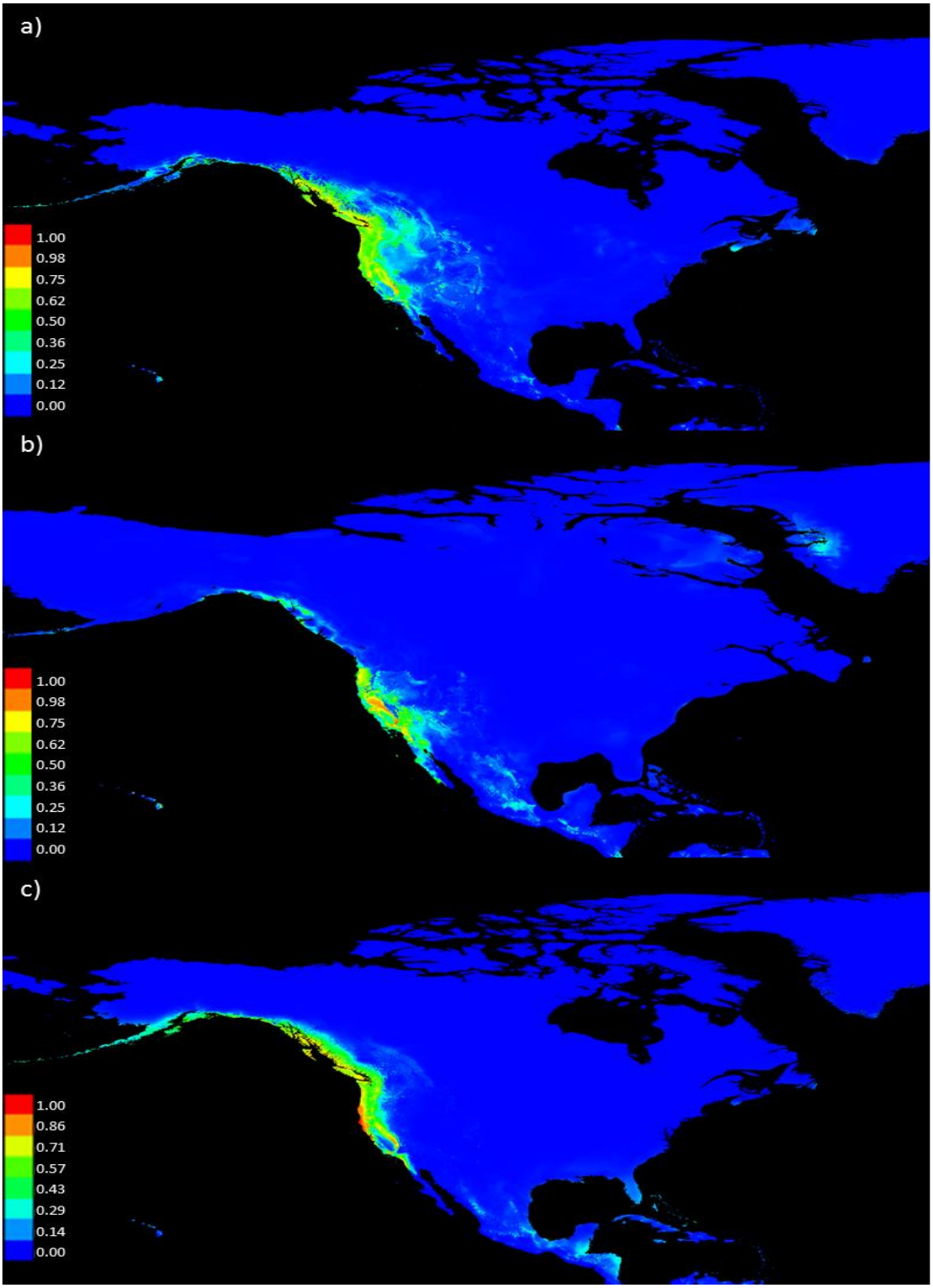


Figure 2.10 MAXENT species distribution modelling of suitable habitat for *S. ruber*. The top map is the current species distribution (a), (b) is the predicted suitable habitat during the LGM, and (c) is suitable habitat during the LIG. Areas with higher suitability, and therefore greater probability of occurrences, are in red with least probability being dark blue.

Table 2.1 Pairwise F_{ST} comparisons between species for loci under positive selection (above diagonal) and neutral loci (below diagonal). All comparisons are significant ($P < 0.05$).

	Ruber	*Nuchalis*	*Varius*
Ruber	-	0.22	0.52
Nuchalis	0.05	-	0.60
Varius	0.10	0.15	-

Table 2.2 Pairwise F'_{ST} comparisons between species for loci under positive selection (above diagonal) and neutral loci (below diagonal). All comparisons are significant ($P < 0.001$).

	Ruber	*Nuchalis*	*Varius*
Ruber	-	0.18	0.51
Nuchalis	0.08	-	0.63
Varius	0.08	0.14	-

Table 2.3. Putative function of the genes linked to the 28 SNPs under positive selection that aligned to the nucleotide database using BLASTN. The NCBI accession number for each gene and expected heterozygosity of each locus for each species is also provided.

SNP ID	Gene Product	Putative Function	NCBI Accession Number	H_e		
				S. ruber	*S. nuchalis*	*S. varius*
TP7548	Transforming growth factor beta regulator 4 (TBRG4)	Signaling	XM_009907786.1	0.43	0.25	0.24
TP12075	Uncharacterized LOC104551532	Unknown	XM_010202937.1	0.00	0.03	0.50
TP18389	Coagulation factor III (thromboplastin, tissue factor) (F3)	Repair	XM_002191596.3	0.05	0.17	0.40
TP28963	Doublecortin domain containing 2B (DCDC2B)	Signaling, Microtubule based	XM_009908515.1	0.00	0.00	0.21
TP33355	Alpha-kinase 2 (ALPK2)	Regulation	XM_009907218.1	0.00	0.03	0.26
TP33439	Nitric oxide synthase 2, inducible (NOS2)	Development/Immunity	XM_009903179.1	0.32	0.20	0.35
TP55351	WW domain binding protein 2 (WBP2)	Unknown	XM_009897937.1	0.19	0.10	0.46
TP57671	Genome assembly scaffold152	Unknown	LK064797.1	0.09	0.03	0.46

SNP ID	Gene Product	Putative Function	NCBI Accession Number	H_e		
				S. ruber	*S. nuchalis*	*S. varius*
TP68708	Plectin	Structure	XM_009897126.1	0.15	0.09	0.49
TP75293	Four and a half LIM domains 3 (FHL3)	Regulation	XM_009908388.1	0.44	0.31	0.13
TP76274	Insulin receptor (INSR)	Metabolism	XM_010017251.1	0.06	0.16	0.11
TP81961	Transcriptional adaptor 1 (TADA1)	DNA Transcription	XM_005492271.1	0.13	0.24	0.24
TP88558	Transmembrane O-methyltransferase-like (LOC103041557)	Sensory perception of sound	XM_007247683.1	0.43	0.43	0.00
TP92798	DiGeorge syndrome critical region gene 14	Development	XM_009910311.1	0.36	0.40	0.00
TP94325	MET proto-oncogene, receptor tyrosine kinase (MET), transcript variant X2	Development/Repair	XM_010413520.1	0.10	0.10	0.50
TP100248	Sorting nexin 33 (SNX33)	Signaling/Structure/ Regulation	XM_009956141.1	0.51	0.33	0.07
TP101898	Nucleoporin 160kDNA	Transport	XM_009900007.1	0.20	0.06	0.23

SNP ID	Gene Product	Putative Function	NCBI Accession Number	H_e		
				S. ruber	*S. nuchalis*	*S. varius*
TP110775	DiGeorge syndrome critical region gene 14	Development	XM_009910311.1	0.34	0.38	0.00
TP111918	Multiple EGF-like-domains 6	Regulation	XM_009911312.1	0.00	0.04	0.51
TP124241	Haplotype Ppa4 FIB7 beta fibrinogen gene	Repair	KF803992.0	0.29	0.09	0.33
TP122576	Phosphofurin acidic cluster sorting protein 2	Transport/Regulation	XM_009898259.1	0.00	0.00	0.21
TP137897	Carnitine O-octanoyltransferase	Metabolism	XR_751756.1	0.00	0.03	0.51
TP138244	Eukaryotic translation elongation factor 2	Protein synthesis	EU738634.1	0.10	0.07	0.46
TP148915	Cyclin E2 (CCNE2)	Regulation	XM_009902935.1	0.00	0.00	0.51
TP158906	BAC chrom z	Unknown	AC202801.3	0.51	0.19	0.13
TP168398	Unknown gene	Unknown	LK064835.1	0.28	0.19	0.23
TP169658	Lactacte dehydrogenase (LDH) gene, exon2, and partial cds	Metabolism	KJ455271.1	0.12	0.26	0.39
TP171589	Leucine zipper, putative tumor suppressor 2	Regulation	XM_009908198.1	0.29	0.16	0.20

Table 2.4. Putative function of the genes linked to the 15 SNPs under balancing selection that aligned to the nucleotide database using BLASTN. The NCBI accession number for each gene and expected heterozygosity of each locus for each species is also provided.

SNP ID	Gene Product	Putative Function	NCBI Accession Number	H_e		
				S. ruber	*S. nuchalis*	*S. varius*
TP5434	Voucher NHMD 130713 lactate dehydrogenase (LDH) gene	Metabolism	KJ455268.1	0.31	0.24	0.15
TP7045	Spectrin beta chain, non-erythrocytic 1-like	Regulation	XM_009900006.1	0.31	0.14	0.00
TP10887	Spectrin, beta, non-erythrocytic 5	Signaling	XM_014780398.1	0.13	0.15	0.17
TP54535	Leucine rich repeat containing 3 (LRRC3)	Immunity	XM_009899477.1	0.30	0.28	0.34
TP76935	Ankyrin repeat domain 40 (ANKRD40)	Signaling	XM_009911525.1	0.08	0.12	0.12
TP82443	Long-chain fatty acid transport protein 1-like	Transport	XM_008935455.1	0.22	0.13	0.19
TP108189	Eukaryotic translation initiation factor 2D (EIF2D)	DNA translation	XM_009897160.1	0.11	0.15	0.16

SNP ID	Gene Product	Putative Function	NCBI Accession Number	H		
				S. ruber	*S. nuchalis*	*S. varius*
TP117988	Transmembrane protein 38B	Transport and Signaling	LK065078.1	0.21	0.15	0.16
TP118139	Inhibin beta C chain-like	Signaling	XM_010134295.1	0.34	0.29	0.30
TP125093	Tripartite motif containing 35 (TRIM35)	Unknown	XM_009895465.1	0.29	0.33	0.31
TP140180	Regulatory solute carrier protein, family 1, member 1 (RSC1A1)	Signaling	XM_009911343.1	0.10	0.03	0.12
TP145656	Transmembrane protein 201 (TMEM201)	Transmembrane transport and Signaling	XM_011585276.1	0.32	0.32	0.31
TP157149	Patched domain containing 4 (PTCHD4)	Transmembrane transport and Signaling	XM_008947999.1	0.27	0.26	0.30
TP160622	NIMA-related kinase 9 (NEK9)	Signaling	XM_014941928.1	0.13	0.09	0.07
TP173078	WD repeat-containing protein 24-like	Signaling	XM_009897095.1	0.22	0.17	0.23

CHAPTER THREE

Genomic structure of the central Alberta hybrid zone between *Sphyrapicus varius* and *S. nuchalis*

A.I.J. Curtis[1] and T.M. Burg[1]

[1]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, AB, T1K 3M4

3.1 Introduction

The evolution of a species often culminates in reproductive isolation through behavioural, ecological, or morphological barriers such as varied displays for mate attraction, filling different niches, and pre- or postzygotic apomorphies (Gompert & Buerkle 2009; Shaffer & Thomson 2007). These characteristics aid in the identification and delimitation of species. When reproductive isolation is incomplete, gene flow in the form of introgressive hybridization between species can occur at contact zones. Introgression has the potential for adaptive speciation through the introduction of new, potentially beneficial alleles that are selected for (Barton 2001; Hamilton *et al.* 2013a, b; Staubach *et al.* 2012). Not all of the genome introgresses at the same rate across species due to differing recombination rates, local selective pressures, and varying success of hybrids and potential backcrosses (Funk & Omland 2003). The study of hybrid zones allows evolutionary biologists to identify the effect of gene flow and incomplete reproductive isolation on speciation.

Reproductive isolation is not limited to barriers found solely within parental individuals. Reduced fitness of hybrid individuals can result in hybrid breakdown where selection is against hybrid individuals and the hybrid zone is maintained by continual immigration of parental species into the zone and interspecific mating, creating a tension hybrid zone (Barton & Hewitt 1985; Freeland *et al.* 2011; Seneviratne *et al.* 2016; Seneviratne *et al.* 2012). Alternatively, fitness of hybrids can be high in the hybrid zone due to the location of these zones in areas of an environmental gradient between parental niches, where hybrid superiority results in greater fitness of the hybrid phenotype over either parental (Barton 2001; Freeland *et al.* 2011; Hamilton *et al.* 2013b; Milne & Abbott

2008; Moore 1977). Therefore, the amount of introgression seen in a hybrid population is dependent on the fitness of hybrids and their habitat. Studying hybrid fitness can aid in the identification of potential barriers to reproduction between the species, which also helps identify divergent traits in parental forms and possible forces driving speciation.

The *Sphyrapicus varius* superspecies complex is composed of three species of sapsuckers with distinct geographic distributions that overlap at the edges of their ranges where hybrid zones are formed (Figure 3.1 insert). Yellow-bellied sapsuckers (*S. varius*) have the widest distribution across northern North America from Alaska to Newfoundland and have been found to be genetically distinct from the other two species (Cicero & Johnson 1995; Johnson & Zink 1983). *Sphyrapicus ruber* (red-breasted sapsucker) is distributed along the Pacific Coast of North America from British Columbia (BC) to California (CA) and is the most phenotypically distinct. *Sphyrapicus nuchalis* (red-naped sapsucker) is most closely related to *S. ruber* genetically, but is phenotypically similar to *S. varius* (Cicero & Johnson 1995; Johnson & Zink 1983; Weir & Schluter 2004). It is distributed along the interior of the Rocky Mountains from central Alberta (CAB) down to New Mexico. These species, though distinct, are recently diverged and closely related. Recent genomic data have found that the species evolved while in isolation through both neutral genetic processes and positive selection (Chapter 2), however, they have not evolved to the point of reproductive isolation. Instead, *S. varius* and *S. ruber* hybridize in the contact zone in northern BC (Seneviratne *et al.* 2016; Seneviratne *et al.* 2012); *S. ruber* and *S. nuchalis* hybridize along their extensive contact zone from BC to CA (Johnson & Johnson 1985); and *S. nuchalis* and *S. varius* hybrids

can be found in central and southwestern AB (Figure 3.1). These contact zones provide instances of possible gene flow and introgression among the three species.

Previous work on sapsucker hybrid zones has found that *S. ruber* and *S. varius* form a tension hybrid zone in BC maintained by selection, with no apparent preference for mate discrimination between the two species within the hybrid zone (Seneviratne *et al.* 2012). In contrast, the contact zone between *S. ruber* and *S. nuchalis* shows signs of asymmetric mating with the higher proportion of hybrid crosses occurring between *S. ruber* males and *S. nuchalis* females, most likely a result of female preference for males with more red colouring (Johnson & Johnson 1985; Trombino 1998). Additionally, the hybrid zone varies with changes in climate and is predicted to result in decreased hybridization between the two species based on the reduced suitable habitat predicted for the species using future climate projections (Billerman *et al.* 2016). The hybrid zone between *S. nuchalis* and *S. varius*, in contrast, has not been studied genetically, but has been studied and found to be dynamic (Hudon 2001).

This study uses genomic data, specifically genotyping by sequencing (GBS) a form of restriction-site associated DNA sequencing (RADseq), to study the genomic structure of hybrids and "pure" individuals within the central Alberta (CAB) hybrid zone between *S. varius* and *S. nuchalis* (Figure 3.1). To date, little work has been done on this hybrid zone to determine how it is maintained and the effect of interspecific introgression within the contact zone on the genomic structure of the parent populations. I use GBS to obtain genomic resolution of both hybrid and parental individuals within the hybrid zone and compare them to individuals from allopatric populations to determine if the zone is composed of mainly early generation hybrids or advanced generation hybrids and

backcrosses. This will aid in determining the amount of introgression occurring within the hybrid zone and the possible introgression of interspecific alleles in populations outside the hybrid zone. I hypothesize that, due to *S. nuchalis* having more contact with both *S. varius* and *S. ruber* throughout its history (Chapter 2), it will have evolved to be more selective in conspecific mate choice than *S. varius* which will result in a higher proportion of *S. varius* individuals mating with hybrids. Therefore, I predict that hybrid individuals will have a higher proportion of *S. varius* ancestry than *S. nuchalis*.

3.2 Materials and Methods

3.2.1 Sample collection and sequencing

A mix of museum and lab collected samples were used in this study. Museum specimens were provided in the form of toe pad and muscle tissue stored in lysis buffer. A list of museum samples, their collection information, and accession numbers can be found in Appendix 3. Samples from live birds were collected using 12 m mist nets and call play back from throughout each of the species' ranges. A blood sample (<100 μL) was taken from the brachial vein of each individual and stored in ethanol. Birds were banded and released at capture sites. DNA was extracted using a standard phenol-chloroform extraction procedure and 63 samples (34 *S. nuchalis*, 19 *S. varius*, and 10 hybrids) were sent to Cornell University's Institute for Genomic Diversity (IGD) for genotyping by sequencing (GBS), following the procedure in Elshire *et al.* (2011) with the restriction enzyme *Pst*I. An additional 84 samples (39 *S. nuchalis*, 25 *S. varius*, and 20 hybrids) were sent to the Genomic Sequencing and Analysis Facility (GSAF) at the

University of Texas for double digest RADseq using the restriction enzyme pair *Nla*III and *MluC*I following the procedure in Peterson *et al.* (2012).

3.2.2 SNP calling

Due to the lack of a reference genome available for *Sphyrapicus*, *de novo* SNP calling was performed. For the Cornell dataset, I used the GBS UNEAK analysis pipeline version 3.0, which is an extension of the JAVA program TASSEL (Bradbury *et al.* 2007), to filter reads and call SNPs. Quality filtering removed any reads with incorrect, missing, or multiple restriction cut sites or barcodes. Reads were then truncated to 64 bases and aligned into identical sequence tags. A threshold of a minimum of five reads per tag was set for inclusion in the SNP calling process with the error tolerance rate set to 0.03 and a minimum minor allele frequency (MAF) of 0.05 for pairwise alignment identification of SNPs. A maximum of one SNP per read was kept to aid in minimizing linkage disequilibrium in downstream analyses.

The UTexas data were filtered, demultiplexed, and cleaned using the STACKS v1.09 process_radtags pipeline (Catchen *et al.* 2013; Catchen *et al.* 2011). The denovo_map pipeline in STACKS was then used to identify SNPs *de novo* with both the number of reads required to create a stack and the number of mismatches allowed between loci set to four. UTexas SNPs were then filtered for a MAF of 0.05 (as per the Cornell SNPs) using PLINK v1.07 (Purcell *et al.* 2007). Only one SNP per read was kept for downstream genomic analyses.

3.2.3 Genomic analysis of hybrid zone individuals

To determine the proportion of parent species ancestry assigned to individuals in the hybrid zone, I used the program ADMIXTURE v1.2.3 (Alexander *et al.* 2009). ADMIXTURE is similar to the program STRUCTURE used in Chapter 2 and also assigns individuals to clusters based on their genotype. Unlike STRUCTURE, however, ADMIXTURE does not use a Markov Chain Monte Carlo (MCMC) method but instead it uses the maximum likelihood of the model, leading to faster performance of ADMIXTURE than STRUCTURE with large SNP datasets. The proportion of ancestry from either of the parent species of both hybrids and "pure" individuals in the hybrid zone will show the reproductive success of hybrid individuals within the zone. If hybrid individuals are backcrossing with individuals of the parent species more frequently than with other hybrids, then the proportion of ancestry within the hybrid zone will reflect a skewed ancestry to either parent. If the hybrid zone is maintained by parent species migrating into the hybrid zone and mating with members of the other species to only produce F1 hybrids (no advanced generation hybrids), then the ancestry coefficients for hybrids will be 50% ancestry from both species.

ADMIXTURE was run on the two different datasets (Cornell and UTexas) separately as they both contain different loci and different individuals. The results were then combined after the analyses to view the whole study group together. Each ADMIXTURE analysis was run for populations K = 1-4, using a quasi-Newton algorithm for accelerated convergence (Zhou *et al.* 2011) and a 5-fold cross-validation. The default block relaxation algorithm was used to perform point estimation with the termination criterion set for the analyses to stop running when the change in the log-likelihood of

point estimations between iterations increased by <0.0001. The number of clusters that best fit the data was determined by the K value with the lowest cross-validation error and individual *Q* estimates were used to make bar graphs of individual ancestry. No *a priori* assignments of individuals to groups was performed in the ADMIXTURE analyses.

3.2.4 Genomic structure of hybrid zone

To determine the proportion of F1 and advanced generation hybrids within the CAB hybrid zone, I compared the hybrid index score to the interspecific heterozygosity score for each individual sampled within the CAB contact zone using the program INTROGRESS v1.22 (Gompert & Buerkle 2009, 2010). INTROGRESS is a software package implemented in R (R Core Team 2013) that uses a maximum-likelihood approach to determine the hybrid index of each individual. The hybrid index estimates the proportion of alleles in admixed individuals that are inherited from parental populations by comparing the allele frequencies of admixed individuals to the allele frequencies of *a priori* identified parental populations (Buerkle 2005). The interspecific heterozygosity is the calculated mean number of loci in an individual that have one allele from each parent species (Gompert & Buerkle 2010). A pure parental individual would have a hybrid index of either zero (*S. nuchalis*) or one (*S. varius*) and interspecific heterozygosity of close to zero. Hybrid individuals should have higher interspecific heterozygosity (more loci with alleles from both species) and a range of HINDEX scores, with F1 hybrids having a HINDEX = 0.5 and interspecific heterozygosity = 1.

I graphed the interspecific heterozygosity against the hybrid index for each individual to distinguish pure individuals from F1 hybrids or advanced generation backcrosses. *S. varius* and *S. nuchalis* individuals from allopatric populations with high assignment of ancestry to their respective species in ADMIXTURE were used as *a priori* parental populations in the INTROGRESS analysis. All individuals located within the CAB hybrid zone (n = 60), regardless of their phenotypic identification as pure or hybrid individuals, were used in the analysis to determine the genetic composition of individuals within the hybrid zone.

3.3 Results

3.3.1 Genomic analysis

After filtering the Cornell dataset, there were 11,311 SNPs shared between *S. varius* and *S. nuchalis* while the UTexas dataset contained 1,638 SNPs. ADMIXTURE identified K = 2 for having the lowest cross-validation error, and therefore the optimal number of clusters for both the UTexas and the Cornell datasets.

The ADMIXTURE plot differentiated the individuals outside of the hybrid zones into two distinct clusters – one for *S. varius* and one for *S. nuchalis* with minimal admixture within these two groups (Figure 3.2). *S. nuchalis* individuals (identified by phenotype) within the CAB hybrid zone are more admixed than the allopatric *S. nuchalis* outside the zone. The *S. nuchalis* individuals within the hybrid zone are also more admixed than the *S. varius* individuals within the zone. *S. varius* within the CAB hybrid

zone show little or no admixture, with the exception of two individuals that have 48% and 32% *S. nuchalis* ancestry (Figure 3.2).

The hybrid individuals within the contact zone have a mix of assignment to both parent species' ancestry. The amount of admixture to the different species changes throughout the hybrid zone, with those individuals located further north having a higher proportion of *S. varius* ancestry and those to the south having a greater proportion of *S. nuchalis* ancestry (Figures 3.1 and 3.2). In general, the proportion of ancestry to either species is indicative of advanced generation hybrids.

3.3.2 Genomic structure of hybrids

The individuals within the hybrid zone exhibited low overall interspecific heterozygosity (0.03-0.31; Figure 3.3). None of the individuals phenotypically identified as *S. varius* (*n* = 12) or *S. nuchalis* (*n* = 22) had interspecific heterozygosity of zero, indicating that these individuals have some of their genes sharing an allele from the other species. Similarly, the hybrid index identified no individuals as pure *S. varius* (HINDEX = 1.0) or *S. nuchalis* (HINDEX = 0.0) (Figure 3.3). In general, however, the non-hybrid individuals had HINDEX scores more associated with their respective species. The allelic ancestry of the hybrids within the contact zone are not skewed towards one species or the other, but instead range across the hybrid index between the two species, similar to the results for hybrids found in ADMIXTURE. There are no F1 generation hybrids (HINDEX = 0.5, interspecific heterozygosity = 1.0) or even any early

generation hybrids. Instead the low interspecific heterozygosity shows that the individuals within the hybrid zone are advanced generation hybrids.

In order to determine if the low interspecific heterozygosity for the hybrid individuals and the higher than expected interspecific heterozygosity for the pure individuals are a result of widespread recombination or incomplete lineage sorting of ancestral traits, I identified a subset of 206 near species-specific loci. These markers (hereafter referred to as nearly diagnostic loci) had an allele-frequency differential (δ) over 0.9 and were compared to all the markers. The species-specific markers identified three *S. varius* as pure (HINDEX = 1.0, interspecific heterozygosity = 0); two *S. nuchalis* and one hybrid individual were also identified as pure *S. nuchalis*. Although the interspecific heterozygosity of individuals varied between the loci with $\delta > 0.9$ and all the loci, the over-all patterns were similar (Figure 3.3). This indicates that widespread hybridization and not ancestral polymorphism is responsible for the higher interspecific heterozygosity seen in the pure individuals using all SNPs.

3.4 Discussion

3.4.1 Introgression or ancestral polymorphism?

The individuals in the central and southern Alberta hybrid zone have low interspecific heterozygosities (Figure 3.3). Diagnostic markers are those loci that are species specific and were identified as having an allele differential between the two species of $\delta > 0.9$, indicating that they can differentiate the two species. If the pattern in the hybrid zone is similar using nearly diagnostic loci to the pattern using all markers,

then deviation in interspecific heterozygosity from what is expected for pure individuals and hybrid individuals within the contact zone is most likely a result of historical admixture and not ILS, since diagnostic markers – by nature – should preclude ILS (Hamilton *et al.* 2013b; Stolting *et al.* 2013). This is the case in this study, where similar patterns between the data with all loci versus the data with nearly diagnostic loci indicate that the introgression seen within the hybrid zone is a result of admixture and not ancestral polymorphism. The complete differentiation between the *S. nuchalis* and *S. varius* individuals from allopatric populations using ADMIXTURE also supports this conclusion.

3.4.2 Genomic structure of hybrids

The presence of many advanced generation hybrids with low interspecific heterozygosity and a range of HINDEX scores indicates a well-established hybrid zone with many hybrid individuals backcrossing with pure parent individuals. However, there is a dearth of F1 generation individuals, which would be expected in the presence of heterospecific mate pairs.

Previous genetic studies on a hybrid zone between *S. ruber* and *S. varius* and one between *S. ruber* and *S. nuchalis* in British Columbia also found a lack of F1 generation hybrids (Seneviratne *et al.* 2016; Seneviratne *et al.* 2012). In those studies, however, genetic hybrids in those zones were a minority (Seneviratne *et al.* 2016), whereas this study identified most individuals located within the hybrid zone as hybrids. The lack of F1 and early generation hybrids in the other sapsucker hybrid zones was determined to be

a result of selection against hybrids (Seneviratne *et al.* 2016). The abundance of advanced generation hybrids in the SAB and CAB hybrid zone suggests if selection is present, it is weak. Hybrid zones made up of mostly late generation hybrids can be a result of selection on novel hybrid adaptations from an environmental gradient (Hamilton *et al.* 2013b; Milne & Abbott 2008; Pinheiro *et al.* 2016). The central and southern Alberta hybrid zones are located along an environmentally transitional habitat from the mountains to the foothills which could result in selection on intermediate phenotypes of advanced generation hybrids.

Overall, there is no particular parent species that the hybrids are uniformly skewed towards throughout the zone. The results show that neither *S. nuchalis* nor *S. varius* are more predisposed to conspecific mate choice due to historical sympatric and parapatric distribution with closely related species, as I had predicted of *S. nuchalis*. Instead, the skew towards parent ancestry varies across the zone from a north to south direction with the individuals towards the north of the zone having a higher proportion of *S. varius* ancestry and those individuals towards the south of the zone having a higher proportion of *S. nuchalis* ancestry. This could simply be a result of the prevalence of pure individuals available in each section. The range for *S. varius* connects with the zone from the north of the hybrid zone while the range of *S. nuchalis* comes up from the south end of the zone (Figure 3.1). This would then make it more likely from a mating perspective for hybrids to be mating with pure *S. nuchalis* in the south and *S. varius* in the north mainly as they are limited in their mate choice.

The central and southern Alberta hybrid zone is not predominantly maintained by selection for intermediate phenotypes across an environmental cline where hybrid

phenotypes are superior to parental within the zone. Instead, the zone follows the sapsucker hybrid zone tradition of being a tension hybrid zone. The lack of F1 hybrid individuals indicates selection against the early generation hybrids and maintenance of the zone through dispersal of parental genotypes into the zone to pair with advanced generation hybrids (Cabria *et al.* 2011; Hamilton *et al.* 2013b; Milne & Abbott 2008; Pinheiro *et al.* 2016; Quintela *et al.* 2010; Seneviratne *et al.* 2016). Parental phenotypes appear to have the advantage within the zone so that those hybrids that are phenotypically similar to the parent species, though still slightly intermediate in phenotype, are selected for over the intermediate F1 phenotype. Alternatively, the lack of F1 hybrids in this study may be a result of small sample size or that F1 hybrids are likely present in the zone, but at a low frequency and are more likely to backcross with parental individuals, producing advanced generation hybrids (Cabria *et al.* 2011; Vali *et al.* 2010).

The lack of F1 generation hybrids in addition to the clear differentiation between *S. varius* and *S. nuchalis* outside of the hybrid zone (Figure 3.2) indicate that the concern of increased hybridization and introgression and the subsequent loss of genetic differentiation and speciation due to decreased suitable habitat via global warming or other anthropogenic effects is rather minimal in these two hybridizing species. Though the species have not evolved to the point of reproductive isolation, other factors such as mate choice and possible selection against hybrids appear to play a role in maintaining the species boundaries during contact. Additionally, it is likely that the species have evolved through bouts of allopatric and sympatric distribution throughout their histories where occasions of hybridization have resulted in adaptive divergence between them (Chapter 2). The skew of individuals towards *S. varius* in the north and *S. nuchalis* in the south

may be also a result of selection by the environment on individuals most suited for that habitat (namely, individuals more phenotypically similar to the pure individuals found in that environment). Future studies involving transect sampling to further investigate the north to south trend of skewing towards the available parental phenotype and look at the frequency of F1 and early generation hybrids in the zone would be beneficial to understanding the interaction of the environment with regards to selection against F1 and early generation hybrids in the hybrid zone.

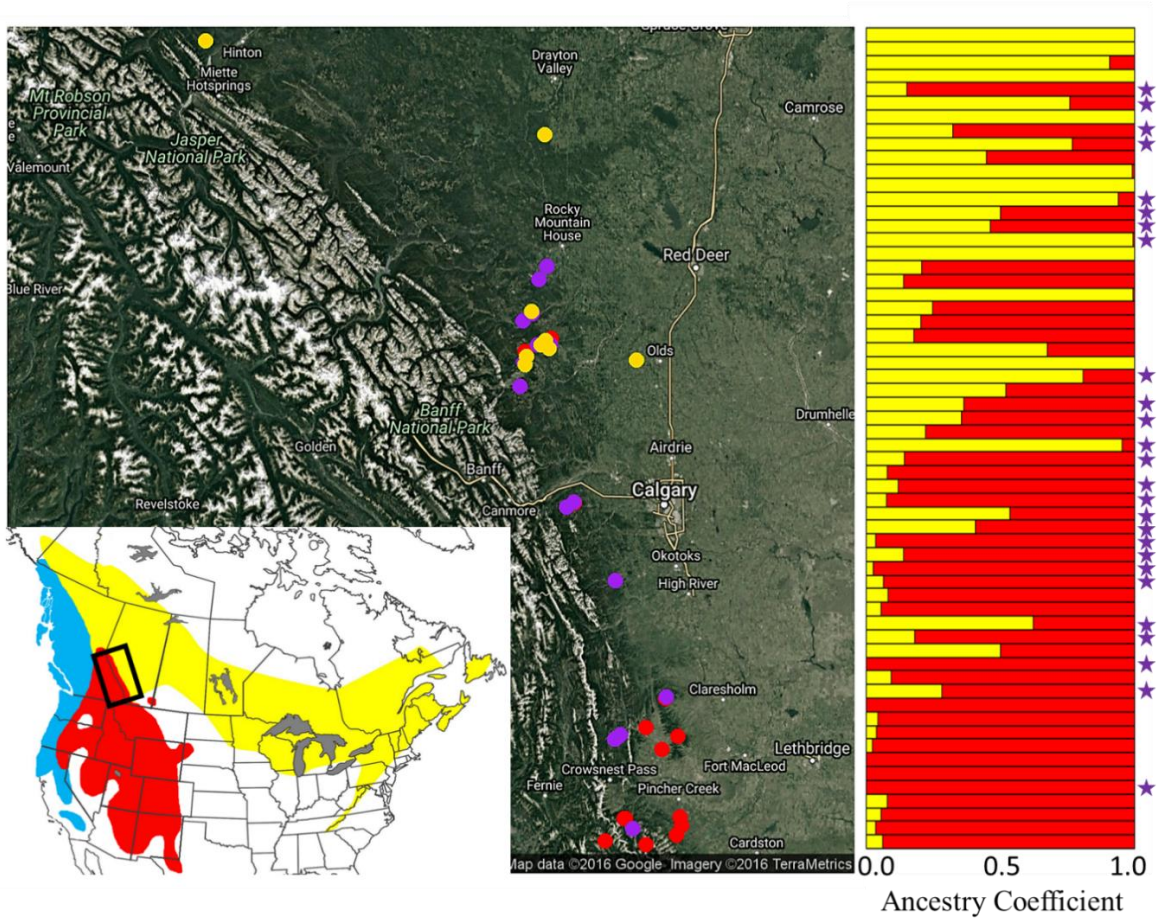


Figure 3.1. Sampling locations of individuals phenotypically identified as *S. nuchalis* (red circles), *S. varius* (yellow circles), and hybrids (purple circles) within the central and southern Alberta hybrid zone. The ADMIXTURE plot on the right indicates the proportion of the genotype that is *S. varius* ancestry (yellow) and *S. nuchalis* ancestry (red). Individuals in the plot are arranged by ascending sampling latitude, with the most southern individual at the bottom of the plot and the most northern at the top. Individuals marked with a purple star correspond to those hybrid individuals in purple circles on the map. The area sampled for the hybrid zone is indicated by the black rectangle on the map insert showing the ranges of the three closely related *Sphyrapicus* species.

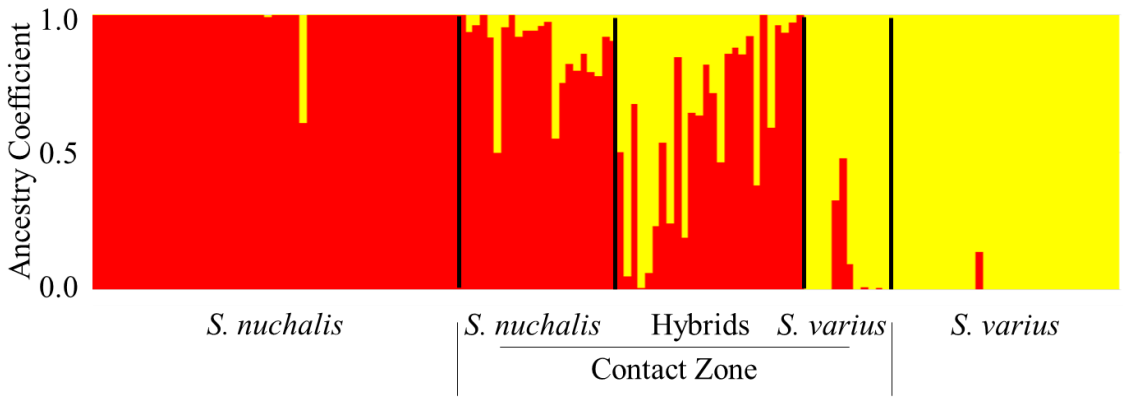


Figure 3.2. ADMIXTURE plot of the Q values for each individual sampled. The red indicates *S. nuchalis* ancestry and the yellow is *S. varius* ancestry. The individuals within the contact zone are divided based on the phenotype identity of the bird at capture.

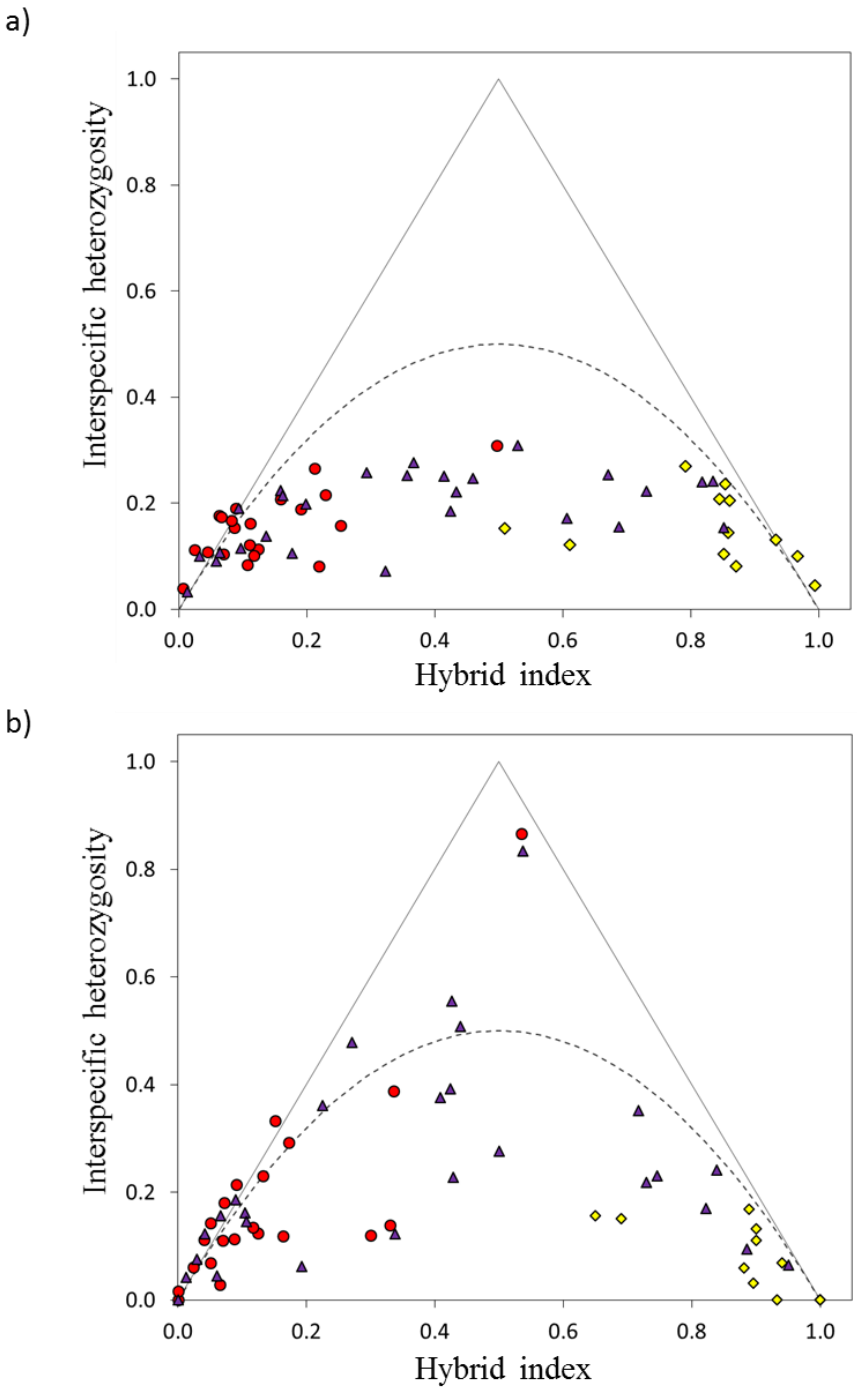


Figure 3.3. Interspecific heterozygosity over hybrid index (HINDEX) scores of each individual in the hybrid zone using (a) all loci and (b) nearly diagnostic loci ($\delta > 0.9$). Red circles are *S. nuchalis*, yellow diamonds *S. varius*, and purple triangles are hybrid individuals. A HINDEX score of 0.0 indicates pure *S. nuchalis* genetic ancestry and a score of 1.0 is pure *S. varius*. Individuals located on the solid line are back-crossed individuals. The dotted line indicates the expected distribution for individuals if mating was random throughout the zone.

CHAPTER FOUR

General Discussion

A.I.J. Curtis[1] and T.M. Burg[1]

[1]Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, AB, T1K 3M4

4.1 Conclusions

GBS provided clear resolution of the three sapsucker species in the *Sphyrapicus* superspecies complex using both neutral and non-neutral markers. It also showed greater capacity to distinguish genetically between the two closest related sapsucker species (*S. ruber* and *S. nuchalis*) over the previous traditional methods. Finally, GBS was able to determine that the hybrid zone between *S. varius* and *S. nuchalis* was a tension zone (with advanced generation hybrids) maintained by immigration of parental phenotypes into the zone.

4.2 Phylogenomics of the sapsucker superspecies complex

Sapsuckers have a long history of parapatric and sympatric distribution where continued contact between pairs of species has resulted in continued admixture. During the LGM, *S. ruber* had a possible contact zone with *S. nuchalis* along the Sierra Nevada mountain range. Prior to that, during the LIG, the suitable habitat for both species was similar to their current extensive contact zone. Prolonged contact is one possible explanation for why the *S. ruber* and *S. nuchalis* genomes are so similar and difficult to differentiate using traditional methods (Cicero & Johnson 1995; Johnson & Zink 1983). Similar results can be found in the cichlid fish of Lake Victoria and the lizards of the Teiidae family in North and South America where recent and rapid radiation has resulted in the inability of traditional methods to resolve the phylogenetic relationship of these species. As found in this study, NGS was able to provide greater resolution of monophyletic lineages in both studies (Tucker *et al.* 2016; Wagner *et al.* 2013).

The complete distinction of the three sapsucker species into separate clusters using both neutral and non-neutral markers indicates that ILS is not a likely cause of the close genetic relationship among the three species. Incomplete lineage sorting would show greater assignment of ancestry to interspecific groups of the sapsucker individuals. The coalescent modelling tree in Chapter 2 shows limited admixture or ILS in some of the genes studied; however, the individuals in this study showed high ancestry in STRUCTURE to the cluster associated with their phenotypic species identification, and the DAPC clustered most individuals according to the phenotypic identification as well. Additionally, the hybrid data support the hypothesis that the introgression and close relationship between *S. varius* and *S. nuchalis* was due to admixture and not ancestral polymorphism. Similar results were found in a study on hybridization in greater and lesser scaup where the ducks showed evidence of hybridization and introgression in sympatric populations, but the advanced generation hybrids were the result of gene flow and not ILS due to the high probability assignment of allopatric populations to their respective species (Lavretsky *et al.* 2016).

The parapatric and sympatric distributions of the species throughout their history would most likely have prolonged or prevented the speciation process if not for adaptation playing a role in their divergence. Areas of contact between diverging species can result in the sharing of alleles between the two species so long as reproductive isolation has not occurred. These areas of potential hybridization and introgression slow the accumulation of segregating alleles and traits that lead to speciation. However, selection for specific traits or against hybrids can still result in speciation in the presence of introgression, as seen in spruce trees, marine iguanas, and this study on sapsuckers (De

La Torre *et al.* 2014; MacLeod *et al.* 2015). The fairly recent radiation of the sapsucker species complex, especially that of the two closest species (*S. nuchalis* and *S. ruber*) in comparison to other boreal superspecies complexes (Weir & Schluter 2004), supports the conclusions that the three species have diverged through adaptive evolution. Leading edge populations moving into new territory due to the recession and advance of the ice sheets during the Pleistocene would have adapted to their new habitats. Genes that are under positive selection can potentially diverge more quickly than those that diverge through neutral processes, resulting in other parts of the genome diverging through hitchhiking (Brown *et al.* 2010; Ford *et al.* 2015; MacLeod *et al.* 2015).

The pressures of selection reduced the occurrence of hybridization and introgression between the species, possibly due to behavioural selection and/or sexual selection. Hybrid individuals in areas of contact are still produced, but the proportion of hybrids to pure parent individuals in the zone would be skewed towards pure individuals as interspecific mating would be less common due to sexual selection or phenotypic selection. My data suggest that hybrid individuals with intermediate characteristics were selected against while those individuals that did survive backcrossed to parental individuals due to the increased likelihood of finding a pure individual within the contact zone. This would result in the advanced generation hybrid zone we see in the central Alberta hybrid zone between *S. nuchalis* and *S. varius* in this study and between other closely related species that exhibit incomplete reproductive isolation (Hamilton *et al.* 2013b; Stolting *et al.* 2013).

4.3 Genomic structure of the hybrid zone

The ancestry of admixed individuals within the hybrid zone was a mixture across the hybrid index. Contrary to my prediction, there was no skew towards either *S. nuchalis* or *S. varius* ancestry, which would indicate a reduced stringency for intraspecific mate choice, with hybrid backcrosses to both species (Cahill *et al.* 2015; Hamilton *et al.* 2013b; Vali *et al.* 2010). Instead, the skew towards backcrosses with one parent species changed based on geographic location. Hybrids found more north in the contact zone showed a skew towards *S. varius* ancestry while those found in the south showed a skew towards *S. nuchalis*. This could be due to the increased probability of finding *S. varius* in the north and *S. nuchalis* in the south where their distributions naturally meet with the zone and they would be more likely to be found. Therefore, my prediction that the preference of females for redder *S. nuchalis* males resulting in a skew towards *S. nuchalis* ancestry was incorrect. It appears that sapsuckers in the central Alberta contact zone are more opportunistic than meticulous in mate choice, where they will mate with the more abundant species present – or the species which they are most likely to encounter. This has been found for other species too, where the predominant species is more likely to be chosen as a mate, whether it be hetero- or conspecific (Cabria *et al.* 2011; Lavretsky *et al.* 2016; Vali *et al.* 2010).

4.4 Consequences of gene flow

The sapsucker species have genetic support for their taxonomic identification however, they are still recently radiated and hybridize between species pairs (Weir &

Schluter 2004). They have evolved in the presence of continued gene flow resulting in adaptive divergence interspersed with periods of allopatric speciation. Outside of sympatric populations, there is little introgression between the three species. This has similarly been found in European mink (Cabria *et al.* 2011) and red wolves (Bohling *et al.* 2016) where previous hypotheses had predicted that continued hybridization between the rarer species with the more abundant species would result in extirpation of the less common species. Instead, both studies identified a dearth of hybrids within the hybrid zone, indicating that natural isolating mechanisms were working at maintaining species boundaries (Bohling *et al.* 2016; Cabria *et al.* 2011). The maintenance of "pure" allopatric populations along with the lack of early generation hybrids indicates that gene flow between the species is low. This leads to the conclusion that hybridization in sapsuckers is not a large concern. The proportion of hybrids and the resulting introgression of interspecific genomic data into allopatric populations is minimal and of low concern with regards to conservation management in the face of reduced habitat due to anthropogenic effects.

4.5 Future Work

This study used GBS to successfully differentiate the three species of sapsuckers. However, I was unable to accurately determine divergence dates of the three species due to the lack of an outgroup. Future studies could use *S. thyroides*, which is sister to the other three sapsuckers studied here, to estimate the divergence times of the three species. Other studies using genomic data to identify loci under positive selection and neutral loci have found that estimates of divergence dates are more recent using loci under positive

selection than mtDNA divergence dates (MacLeod *et al.* 2015). Presently, the divergence times between sapsuckers estimated using mtDNA molecular clock are between approximately 800 ka to 1.5 Ma (for *S. varius* from the other two) and 250 ka to 500 ka for *S. nuchalis* and *S. ruber* (Weir & Schluter 2004). Using genomic data and identifying loci under positive selection from neutral loci may provide different information on the divergence estimates of these three species due to their adaptive radiation and continued hybridization; however, without the *S. thyroides* outgroup I was unable to do this in the current study.

While the present study was able to genetically differentiate the *S. ruber* individuals from the *S. nuchalis*, the historical relationship between the two species is still unclear, especially with regards to the large plumage variation between these species in spite of the recent divergence. Further studies on the contact zone between these two species and the method of mate selection – if it is still asymmetric and assortative as determined previously (Johnson & Johnson 1985) would be beneficial in understanding what role plumage variation plays in sapsucker speciation. The selection by females for males with more red may have played a role in driving adaptation through sexual selection (Johnson & Johnson 1985). Identifying plumage genes and whether they are under selection and how they vary among sapsuckers would aide in determining the role of sexual selection on these three species. Additionally, understanding the amount of hybridization and introgression currently occurring between these two species would assist in determining the effect that hybridization due to decreased suitable habitat from anthropogenic causes has on these two species.

Finally, additional data on the central Alberta hybrid zone between *S. nuchalis* and *S. varius* would enable us to better understand the dynamics within the contact zone. The lack of F1 hybrids in the present study is irregular in comparison to previous studies on other sapsucker hybrid zones (Seneviratne *et al.* 2016; Seneviratne *et al.* 2012) and possibly a result of restricted sample selection due to the necessity to isolate large amounts of high quality DNA for NGS. Another study involving a transect across the hybrid zone with more individuals may clarify the pattern of skewed hybrid index towards *S. nuchalis* or *S. varius* being contingent upon the location of the hybrid individuals closer to either parent species. Additionally, the present study had limited *S. varius* individuals outside of the contact zone due to the inability to extract good quality DNA, which may have limited the resolution of hybrid structure within the zone. Increased understanding into how and why the species hybridize may also help further our understanding of how mates are chosen and the possible effect of increased introgression between the three species and if the species can retain their separation in the face of continued or increased gene flow.

REFERENCES

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, Butlin RK, Dieckmann U, Eroukhmanoff F, Grill A, Cahan SH, Hermansen JS, Hewitt G, Hudson AG, Jiggins C, Jones J (2013) Hybridization and speciation. *Journal of Evolutionary Biology* **26**, 229-246.

Aitken KEH, Martin K (2007) The importance of excavators in hole-nesting communities: availability and use of natural tree holes in old mixed forests of western Canada. *Journal of Ornithology* **148**, S425-S434.

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655-1664.

Amadon D (1966) The superspecies concept. *Systematic Biology* **15**, 245-249.

Angeloni F, Wagemaker CAM, Jetten MSM, den Camp H, Janssen-Megens EM, Francoijs KJ, Stunnenberg HG, Ouborg NJ (2011) De novo transcriptome characterization and development of genomic tools for *Scabiosa columbaria L.* using next-generation sequencing techniques. *Molecular Ecology Resources* **11**, 662-674.

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

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics* **9**, 323.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography - The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proceedings of the Royal Society B-Biological Sciences* **265**, 457-463.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society B-Biological Sciences* **265**, 1707-1712.

Backhouse F (2005) *Woodpeckers of North America* Firefly Book Ltd., Richmond Hill, ON.

Bacon CD, McKenna MJ, Simmons MP, Wagner WL (2012) Evaluating multiple criteria for species delimitation: an empirical example using Hawaiian palms (Arecaceae: *Pritchardia*). *BMC Evolutionary Biology* **12**, 17.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3**, 7.

Barton NH (2001) The role of hybridization in evolution. *Molecular Ecology* **10**, 551-568.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113-148.

Beatty GE, Provan J (2010) Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. *Molecular Ecology* **19**, 5009-5021.

Beatty GE, Provan J (2011) Phylogeographic analysis of North American populations of the parasitic herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late glacial refugia. *Journal of Biogeography* **38**, 1585-1599.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**, 1619-1626.

Benham PM, Witt CC (2016) The dual role of Andean topography in primary divergence: functional and neutral variation among populations of the hummingbird, *Metallura tyrianthina*. *BMC Evolutionary Biology* **16**.

Billerman SM, Murphy MA, Carling MD (2016) Changing climate mediates sapsucker (Aves: *Sphyrapicus*) hybrid zone movement. *Ecology and Evolution* **Early View**.

Bock WJ (1999) Functional and evolutionary morphology of woodpeckers. *Ostrich* **70**, 23-31.

Bohling JH, Dellinger J, McVey JM, Cobb DT, Moorman CE, Waits LP (2016) Describing a developing hybrid zone between red wolves and coyotes in eastern North Carolina, USA. *Evolutionary Applications* **9**, 791-804.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372-1373.

Braconnot P, Otto-Bliesner B, Harrison S, Joussaume S, Peterchmitt JY, Abe-Ouchi A, Crucifix M, Driesschaert E, Fichefet T, Hewitt CD, Kageyama M, Kitoh A, Laine A, Loutre MF, Marti O, Merkel U, Ramstein G, Valdes P, Weber SL, Yu Y, Zhao Y (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last

Glacial Maximum - Part 1: experiments and large-scale features. *Climate of the Past* **3**, 261-277.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Bradic M, Teotonio H, Borowsky RL (2013) The population genomics of repeated evolution in the blind cavefish *Astyanax mexicanus*. *Molecular Biology and Evolution* **30**, 2383-2400.

Brown JL (2014) SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* **5**, 694-700.

Brown JL, Maan ME, Cummings ME, Summers K (2010) Evidence for selection on coloration in a Panamanian poison frog: a coalescent-based approach. *Journal of Biogeography* **37**, 891-901.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**, 1917-1932.

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes* **5**, 684-687.

Cabria MT, Michaux JR, Gomez-Moliner BJ, Skumatov D, Maran T, Fournier P, de Luzuriaga JL, Zardoya R (2011) Bayesian analysis of hybridization and introgression between the endangered European mink (*Mustela lutreola*) and the polecat (*Mustela putorius*). *Molecular Ecology* **20**, 1176-1190.

Cahill JA, Stirling I, Kistler L, Salamzade R, Ersmark E, Fulton TL, Stiller M, Green RE, Shapiro B (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology* **24**, 1205-1217.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and genotyping goci de novo from short-read sequences. *G3-Genes Genomes Genetics* **1**, 171-182.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution* **15**, 538-543.

Chavez-Galarza J, Henriques D, Johnston JS, Azevedo JC, Patton JC, Munoz I, De La Rua P, Pinto MA (2013) Signatures of selection in the Iberian honey bee (*Apis*

mellifera iberiensis) revealed by a genome scan analysis of single nucleotide polymorphisms. *Molecular Ecology* **22**, 5890-5907.

Cicero C, Johnson NK (1995) Speciation in sapsuckers (*Sphyrapicus*): 3. Mitochondrial-DNA sequence divergence at the cytochrome-b locus. *Auk* **112**, 547-563.

Clark PU, Alley RB, Pollard D (1999) Climatology - Northern hemisphere ice-sheet influences on global climate change. *Science* **286**, 1104-1111.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The Last Glacial Maximum. *Science* **325**, 710-714.

Clark PU, Mix AC (2002) Ice sheets and sea level of the Last Glacial Maximum. *Quaternary Science Reviews* **21**, 1-7.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**, 3133-3157.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

De La Torre AR, Roberts DR, Aitken SN (2014) Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular Ecology* **23**, 2046-2059.

DeSalle R, Rosenfeld JA (2013) *Phylogenomics: a primer* Garland Science, Taylor & Francis Group, LLC, New York, NY.

Dobkin DS, Rich AC, Pretare JA, Pyle WH (1995) Nest-site relationships among cavity-nesting birds of riparian and snowpocket aspen woodlands in the northwestern Great Basin. *Condor* **97**, 694-707.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**, 214.

Dudík M, Phillips SJ, Schapire RE (2005) Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems*, 323-330.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ (2002) The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews* **21**, 9-31.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene History of the Laurentide Ice Sheet. *Géographie physique et Quaternaire* **41**, 237.

Eberhardt LS (1997) A test of an environmental advertisement hypothesis for the function of drumming in yellow-bellied sapsuckers. *Condor* **99**, 798-803.

Eberhardt LS (2000) Use and selection of sap trees by yellow-bellied sapsuckers. *Auk* **117**, 41-51.

Edwards SV, Jennings WB, Shedlock AM (2005) Phylogenetics of modern birds in the era of genomics. *Proceedings of the Royal Society B-Biological Sciences* **272**, 979-992.

Eisen JA, Fraser CM (2003) Phylogenomics: Intersection of evolution and genomics. *Science* **300**, 1706-1707.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* **6**, 10.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16196-16200.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Ford AGP, Dasmahapatra KK, Rueber L, Gharbi K, Cezard T, Day JJ (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology* **24**, 3421-3440.

Freeland JR, Kirk H, Petersen S (2011) *Molecular Ecology*, 2nd edn. John Wiley & Sons, Ltd., West Sussex, UK.

Fulton RJ, Prest VK (1987) Introduction: The Laurentide ice sheet and its significance. *Géographie physique et Quaternaire* **41**, 181-186.

Funk DJ, Omland KE (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.

Garrick RC, Benavides E, Russello MA, Hyseni C, Edwards DL, Gibbs JP, Tapia W, Ciofi C, Caccone A (2014) Lineage fusion in Galapagos giant tortoises. *Molecular Ecology* **23**, 5276-5290.

Giarla TC, Esselstyn JA (2015) The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology* **64**, 727-740.

Glor RE, Warren D (2011) Testing ecological explanations for biogeographic boundaries. *Evolution* **65**, 673-683.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology* **18**, 1207-1224.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources* **10**, 378-384.

Gong WL, Emanuel BS, Galili N, Kim DH, Roe B, Driscoll DA, Budarf ML (1997) Structural and mutational analysis of a conserved gene (DGSI) from the minimal DiGeorge syndrome critical region. *Human Molecular Genetics* **6**, 267-276.

Gugger PF, Sugita S (2010) Glacial populations and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews* **29**, 2052-2070.

Hamilton JA, Lexer C, Aitken SN (2013a) Differential introgression reveals candidate genes for selection across a spruce (*Picea sitchensis* x *P. glauca*) hybrid zone. *New Phytologist* **197**, 927-938.

Hamilton JA, Lexer C, Aitken SN (2013b) Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology* **22**, 827-841.

Hamlin JAP, Arnold ML (2015) Neutral and selective processes drive population differentiation for *Iris hexagona*. *Journal of Heredity* **106**, 628-636.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 183-195.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Hijmans RJ, Phillips SJ, Leathwick J, Elith J (2013) dismo: Species distribution modeling. In: *R package version 0.9-3*.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel ddaptation in threespine stickleback using sequenced RAD tags. *Plos Genetics* **6**, 23.

Hudon J (2001) Effect of human disturbance on genetic diversity: case study using sapsuckers in Kananaskis Country. *Provincial Museum of Alberta*.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320-1331.

Johnson NK, Johnson CB (1985) Speciation in sapsuckers (*Sphyrapicus*). 2. Sympatry, hybridization, and mate preference in *Sphyrapicus ruber daggetti* and *Sphyrapicus nuchalis*. *Auk* **102**, 1-15.

Johnson NK, Zink RM (1983) Speciation in sapsuckers (*Sphyrapicus*) .1. Genetic differentiation. *Auk* **100**, 871-884.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 15.

Joy JB (2000) Characteristics of nest cavities and nest trees of the Red-breasted Sapsucker in coastal montane forests. *Journal of Field Ornithology* **71**, 525-530.

Kalcounis MC, Brigham RM (1998) Secondary use of aspen cavities by tree-roosting big brown bats. *Journal of Wildlife Management* **62**, 603-611.

Lait LA, Burg TM (2013) When east meets west: population structure of a high-latitude resident species, the boreal chickadee (*Poecile hudsonicus*). *Heredity* **111**, 321-329.

Lavretsky P, Peters JL, Winker K, Bahn V, Kulikova I, Zhuravlev YN, Wilson RE, Barger C, Gurney K, McCracken KG (2016) Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. *Molecular Ecology* **25**, 661-674.

Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298-299.

Losin N, Floyd CH, Schweitzer TE, Keller SJ (2006) Relationship between aspen heartwood rot and the location of cavity excavation by a primary cavity-nester, the Red-naped Sapsucker. *Condor* **108**, 706-710.

MacLeod A, Rodriguez A, Vences M, Orozco-terWengel P, Garcia C, Trillmich F, Gentile G, Caccone A, Quezada G, Steinfartz S (2015) Hybridization masks speciation in the evolutionary history of the Galapagos marine iguana. *Proceedings of the Royal Society B-Biological Sciences* **282**.

Margold M, Jansson KN, Kleman J, Stroeven AP, Clague JJ (2013) Retreat pattern of the Cordilleran ice sheet in central British Columbia at the end of the last glaciation reconstructed from glacial meltwater landforms. *Boreas* **42**, 830-847.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology* **24**, 3009-3025.

Mayr E (1963) *Animal species and evolution* The Belknap Press, Cambridge, MA.

McClelland BR, McClelland PT (2000) Red-naped sapsucker nest trees in northern rocky mountain old-growth forest. *Wilson Bulletin* **112**, 44-50.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution* **66**, 526-538.

McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT (2012) Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution* **62**, 397-406.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources* **11**, 5-18.

Metzker ML (2010) Applications of next-generation sequencing: Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**, 550-553.

Milne RI, Abbott RJ (2008) Reproductive isolation among two interfertile Rhododendron species: low frequency of post-F-1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology* **17**, 1108-1121.

Moore WS (1977) An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52**, 263-277.

Moura AE, Kenny JG, Chaudhuri R, Hughes MA, A JW, Reisinger RR, de Bruyn PJ, Dahlheim ME, Hall N, Hoelzel AR (2014) Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Molecular Ecology* **23**, 5179-5192.

Nakazato T, Warren DL, Moyle LC (2010) Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* **97**, 680-693.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11 Suppl 1**, 184-194.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5269-5273.

Otto-Bliesner BL, Marsha SJ, Overpeck JT, Miller GH, Hu AX, Mem CLIP (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751-1753.

Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW (2010) Conservation genetics in transition to conservation genomics. *Trends in Genetics* **26**, 177-187.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* **7**, e37135.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S (2009) Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* **19**, 181-197.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* University of Chicago Press, Chicago, IL.

Pinheiro F, Gouveia T, Cozzolino S, Cafasso D, Cardoso-Gustavson P, Suzuki RM, Palma-Silva C (2016) Strong but permeable barriers to gene exchange between sister species of *Epidendrum*. *American Journal of Botany* **103**, 1472-1482.

Pritchard JK, Stephens M, Peter D (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Purcell S *PLINK v1.07.* Available from http://pngu.mgh.harvard.edu/purcell/plink/

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.

Puritz JB, Addison JA, Toonen RJ (2012) Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. *Plos One* **7**, 10.

Quintela M, Thulin CG, Hoglund J (2010) Detecting hybridization between willow grouse (*Lagopus lagopus*) and rock ptarmigan (*L. muta*) in central Sweden through Bayesian admixture analyses and mtDNA screening. *Conservation Genetics* **11**, 557-569.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A, Drummond AJ (2007) *Tracer v1.4.* Available from http://beast.bio.ed.ac.uk/Tracer

Rissler LJ (1995) The influence of yellow-bellied sapsuckers on local insect community structure. *Wilson Bulletin* **107**, 746-752.

Rissler LJ, Apodaca JJ (2007) Adding more ecology into species delimitation: Ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology* **56**, 924-942.

Robin NH, Shprintzen RJ (2005) Defining the clinical spectrum of deletion 22q11.2. *The Journal of Pediatrics* **147**, 90-96.

Rundle HD, Dyer KA (2015) Reproductive character displacement of female mate preferences for male cuticular hydrocarbons in *Drosophila subquinaria*. *Evolution* **69**, 2625-2637.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning* Cold Spring Harbor Laboratory Press New York.

Savignac C, Machtans CS (2006) Habitat requirements of the yellow-bellied sapsucker, *Sphyrapicus varius*, in boreal mixedwood forests of northwestern Canada. *Canadian Journal of Zoology* **84**, 1230-1239.

Seneviratne SS, Davidson P, Martin K, Irwin DE (2016) Low levels of hybridization across two contact zones among three species of woodpeckers (*Sphyrapicus* sapsuckers). *Journal of Avian Biology* **47**.

Seneviratne SS, Toews DPL, Brelsford A, Irwin DE (2012) Concordance of genetic and phenotypic characters across a sapsucker hybrid zone. *Journal of Avian Biology* **43**, 119-130.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology* **56**, 896-906.

Short LL, Morony JJ (1970) A second hybrid Williamsons X red-naped sapsucker and an evolutionary history of sapsuckers. *Condor* **72**, 310.

Smith KG (1982) On habitat selection of Williamson's and "red-naped" yellow-bellied sapsuckers. *Southwestern Naturalist* **27**, 464-466.

Stark RD (2002) *An analysis of eastern Nearctic woodpecker drums*, PhD Thesis, Ohio State University.

Stark RD, Dodenhoff DJ, Johnson EV (1998) A quantitative analysis of woodpecker drumming. *Condor* **100**, 350-356.

Staubach F, Lorenc A, Messer PW, Tang K, Petrov DA, Tautz D (2012) Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *Plos Genetics* **8**, 13.

Stervander M, Carlos Illera J, Kvist L, Barbosa P, Keehnen NP, Pruisscher P, Bensch S, Hansson B (2015) Disentangling the complex evolutionary history of the western Palearctic blue tits (*Cyanistes spp*.) - phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Molecular Ecology* **24**, 2477-2494.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16**, 608-613.

Stolting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C (2013) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**, 842-855.

Tozer DC, Nol E, Burke DM (2011) Quality of mature aspen and maple forests for breeding yellow-bellied sapsuckers (*Sphyrapicus varius*). *Canadian Journal of Zoology* **89**, 148-160.

Trombino CL (1998) *Species interactions in the hybrid zone between red-breasted (Sphyrapicus ruber) and red-naped (Sphyrapicus nuchalis) sapsuckers: Fitness consequences, reproductive character displacement, and nest site selection*, PhD Thesis, Northern Illinois University.

Tucker DB, Colli GR, Giugliano LG, Hedges SB, Hendry CR, Lemmon EM, Lemmon AR, Sites JW, Jr., Pyron RA (2016) Methodological congruence in phylogenomic

analyses with morphological support for teiid lizards (Sauria: Teiidae). *Molecular Phylogenetics and Evolution* **103**, 75-84.

Vali U, Dombrovski V, Treinys R, Bergmanis U, Daroczi SJ, Dravecky M, Ivanovski V, Lontkowski J, Maciorowski G, Meyburg BU, Mizera T, Zeitz R, Ellegren H (2010) Widespread hybridization between the greater spotted eagle *Aquila clanga* and the lesser spotted eagle *Aquila pomarina* (Aves: Accipitriformes) in Europe. *Biological Journal of the Linnean Society* **100**, 725-736.

Wagner CE, Keller I, Wittwer S, Selz OM, Mwaiko S, Greuter L, Sivasundar A, Seehausen O (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.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**, 2868-2883.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**, 607-611.

Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21**, 335-342.

Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society B-Biological Sciences* **271**, 1881-1887.

Wilkins HD, Ritchison G (1999) Drumming and tapping by Red-bellied Woodpeckers: Description and possible causation. *Journal of Field Ornithology* **70**, 578-586.

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Campbell Grant EH, Veran S, O'Hara RB (2012) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution* **4**, 236-243.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-1320.

Zhang J, Chiodini R, Badr A, Zhang GF (2011) The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics* **38**, 95-109.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* **7**, 203-214.

Zhou H, Alexander D, Lange K (2011) A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and Computing* **21**, 261-273.

Appendix 1. Sample location and museum accession numbers of museum specimens used in chapter 2. Each sample was given a Project ID determined by their phenotypic identification. The latitude (N) and longitude (W) where available are provided in decimal format.

Project ID	Location	Accession Number	Museum	Decimal Latitude	Decimal Longitude
rbsaWA013	Carnation, King Co. WA	UWBM# 80431 AMN 04	Burke Museum, BG TB		
rbsaWA015	Bellevue, King Co. WA	UWBM# 85279 JMSE 004	Burke Museum, BG TB		
rbsaWA017	Acme, Whatcom Co. WA	UWBM# 88794 MLD 094	Burke Museum, BG TB		
rbsaWA013	Carnation, King Co. WA	UWBM# 80431 AMN 04	Burke Museum, BG TB		
rbsaWA014	Seattle, King Co. WA	UWBM# 81654 GHL 005	Burke Museum, BG TB		
rbsaWA015	Bellevue, King Co. WA	UWBM# 85279 JMSE 004	Burke Museum, BG TB		
rbsaWA017	Acme, Whatcom Co. WA	UWBM# 88794 MLD 094	Burke Museum, BG TB		
rnsaCAB001	James River, bridge on 584, AB	Z94.13.9	Royal Alberta Museum	51.9	-115

rnsaCAB002	James River, Wilson James Campsite, AB	Z94.13.10	Royal Alberta Museum	51.833	-115.22
rnsaCAB003	James River, Wilson James Campsite, AB	Z94.13.11	Royal Alberta Museum	51.833	-115.22
rnsaCAB004	9 miles west of Bearberry, Improvement District 10, AB	Z95.11.14	Royal Alberta Museum	51.867	-115.08
rnsaCO005	Park Co. Colorado	UWBM# 56359 GAV 858	Burke Museum, BG TB	39.4402	-105.77
rnsaCO006	Gunnison, Gunnison Co. Colorado	UWBM# 56363 GAV 862	Burke Museum, BG TB		
rnsaNM003	San Marcial, NM	cat #MSB Birds 28952, NK: 170195	Museum of Southwest Biology	33.6339	-107.01
rnsaNM008	Black Range, NM	cat #MSB Birds 29241, NK: 170616	Museum of Southwest Biology	32.9438	-107.7
rnsaNM012	Guadalupe Mountains, Dark Canyon, NM	cat #MSB Birds 29305, NK: 35770	Museum of Southwest Biology	32.1309	-104.71
rnsaSAB016	Porcupine Hills	Z02.14.9	Royal AB museum	50.0172	-114.05
rnsaWA001	Easton, Kittitas Co. WA	UWBM# 49988 CSW 3959	Burke Museum, BG TB		

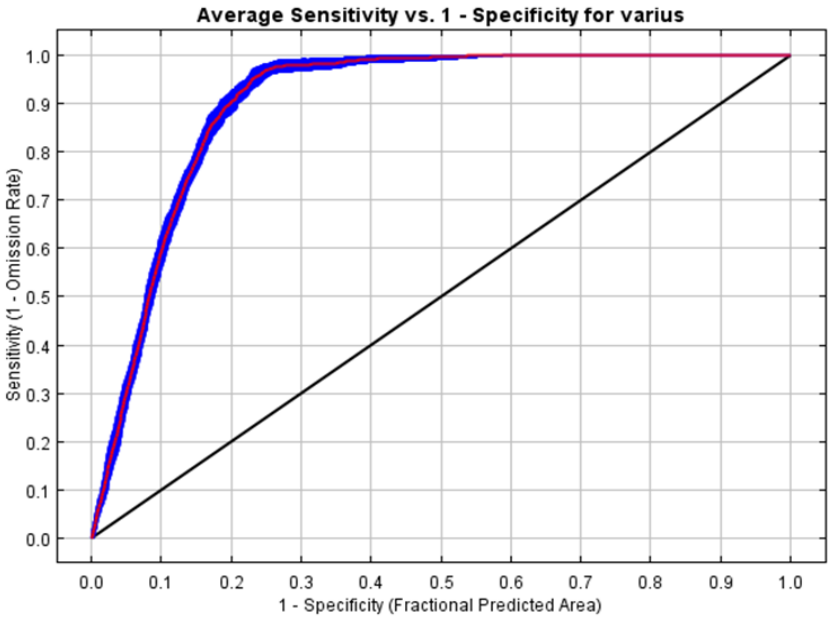
rnsaWA003	Kittitas Co. WA	UWBM# 49001 CSW 4984	Burke Museum, BG TB		
rnsaWA004	Kittitas Co. WA	UWBM# 49002 CSW 4985	Burke Museum, BG TB		
rnsaWA006	Kirkland, King Co. WA	UWBM# 85284 SEZ 045	Burke Museum, BG TB		
rnsaxybsaCAB001	James River area	Z09.8.1	Royal Alberta Museum	51.8685	-115.04
rnsaxybsaCAB002	James River area	Z09.8.2	Royal Alberta Museum	51.6554	-115.26
rnsaxybsaCAB005	James River area	Z09.8.5	Royal Alberta Museum	52.0256	-115.16
rnsaxybsaCAB009	James River area	Z10.3.4	Royal Alberta Museum	51.8678	-115.01
rnsaxybsaCAB011	James River area	Z10.3.6	Royal Alberta Museum	51.8685	-115.02
rnsaxybsaCAB014	Clearwater River area	Z10.3.9	Royal Alberta Museum	51.989	-115.23
rnsaxybsaCAB018	James River area	Z10.3.14	Royal Alberta Museum	51.8681	-115.11

rnsaxybsaCAB023	1.5 miles south Strachan, Improvement District 10, AB	Z95.9.1	Royal Alberta Museum	52.2	-115.1
rnsaxybsaCAB024	Cow Lake area, Improvement District 10, AB	Z95.9.7	Royal Alberta Museum	52.267	-115.03
rnsaxybsaCAB030	James River area, west of Sundre	Z11.5.10	Royal Alberta Museum	51.7774	-115.23
ybsaCAB013	James River area	Z10.3.13	Royal Alberta Museum	51.8059	-115.2
ybsaCAB021	Brule area	Z11.4.30	Royal Alberta Museum	53.4005	-117.87
ybsaCAB022	James River area, west of Sundre	Z11.5.2	Royal Alberta Museum	51.89	-115.05
ybsaCAB025	Clearwater River area	Z11.5.9	Royal Alberta Museum	52.039	-115.16
ybsaCAB027	James River area, west of Sundre	Z11.5.14	Royal Alberta Museum	51.849	-115.02
ybsaCAB028	4 miles West of Alder Flats, AB	Z02.15.7	Royal Alberta Museum	52.93	-115.05
ybsaNC003	Dillingham, Buncombe Co. NC	UWBM# 86868 RBB 538	Burke Museum, BG TB		

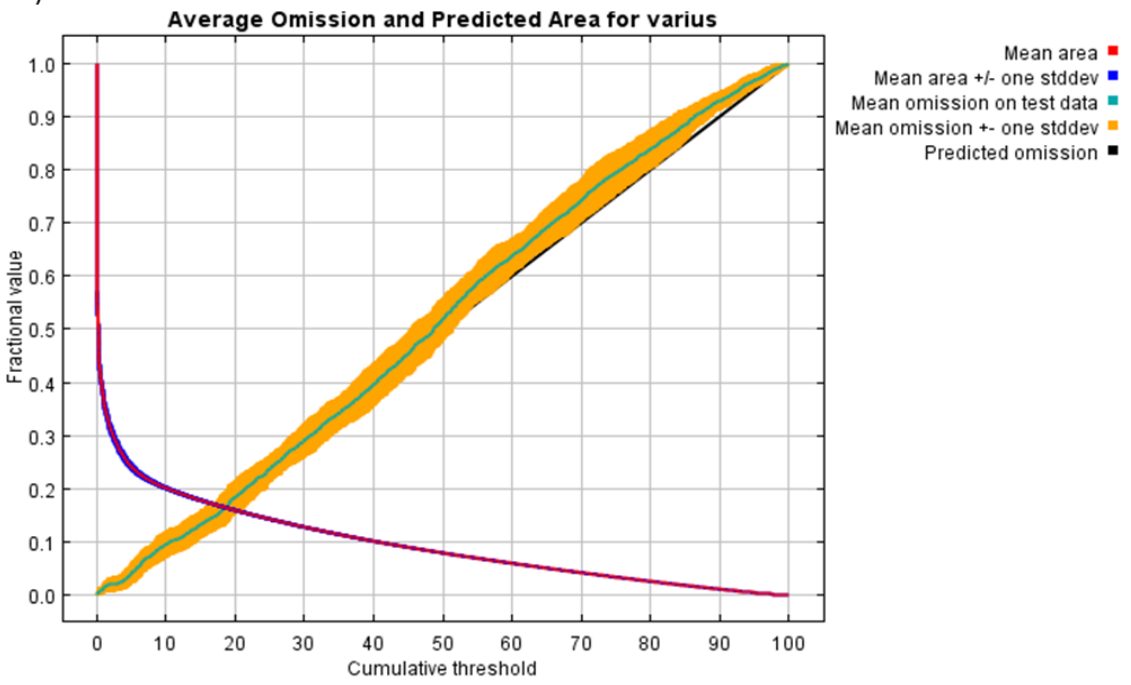
ybsaNC004	Murchison, Buncombe Co. NC	UWBM# 86869 RBB 539	Burke Museum, BG TB		
ybsaSK010	Rouleau, SK	accn20030 exn	Royal Saskatchewan Museum	50.183	-104.933
ybsaSK011	Regina, SK	accn20040 exn 001	Royal Saskatchewan Museum	50.450	-104.617

Appendix 2. Figures illustrating the fit of the species distribution models from Chapter 2.

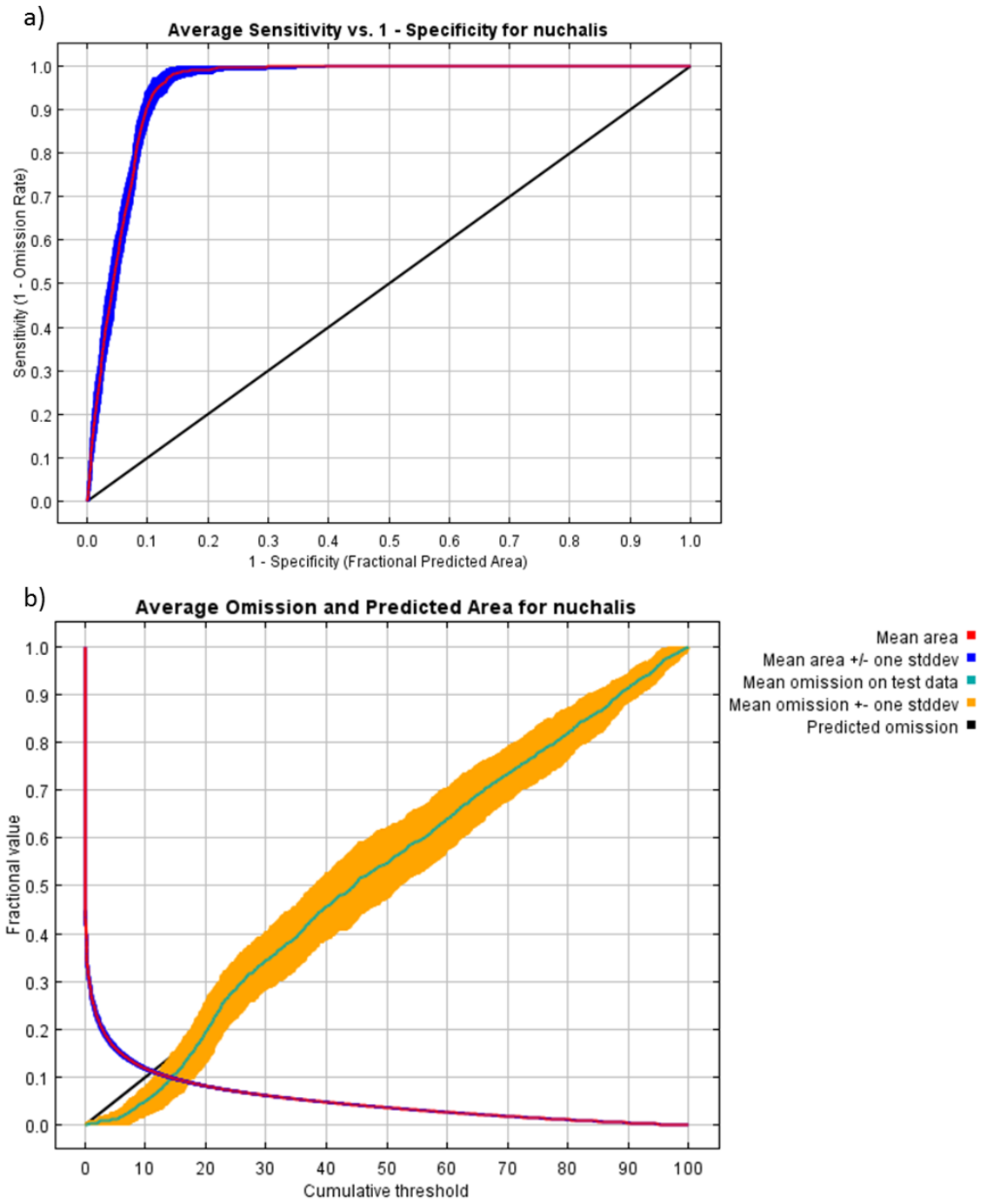
a)



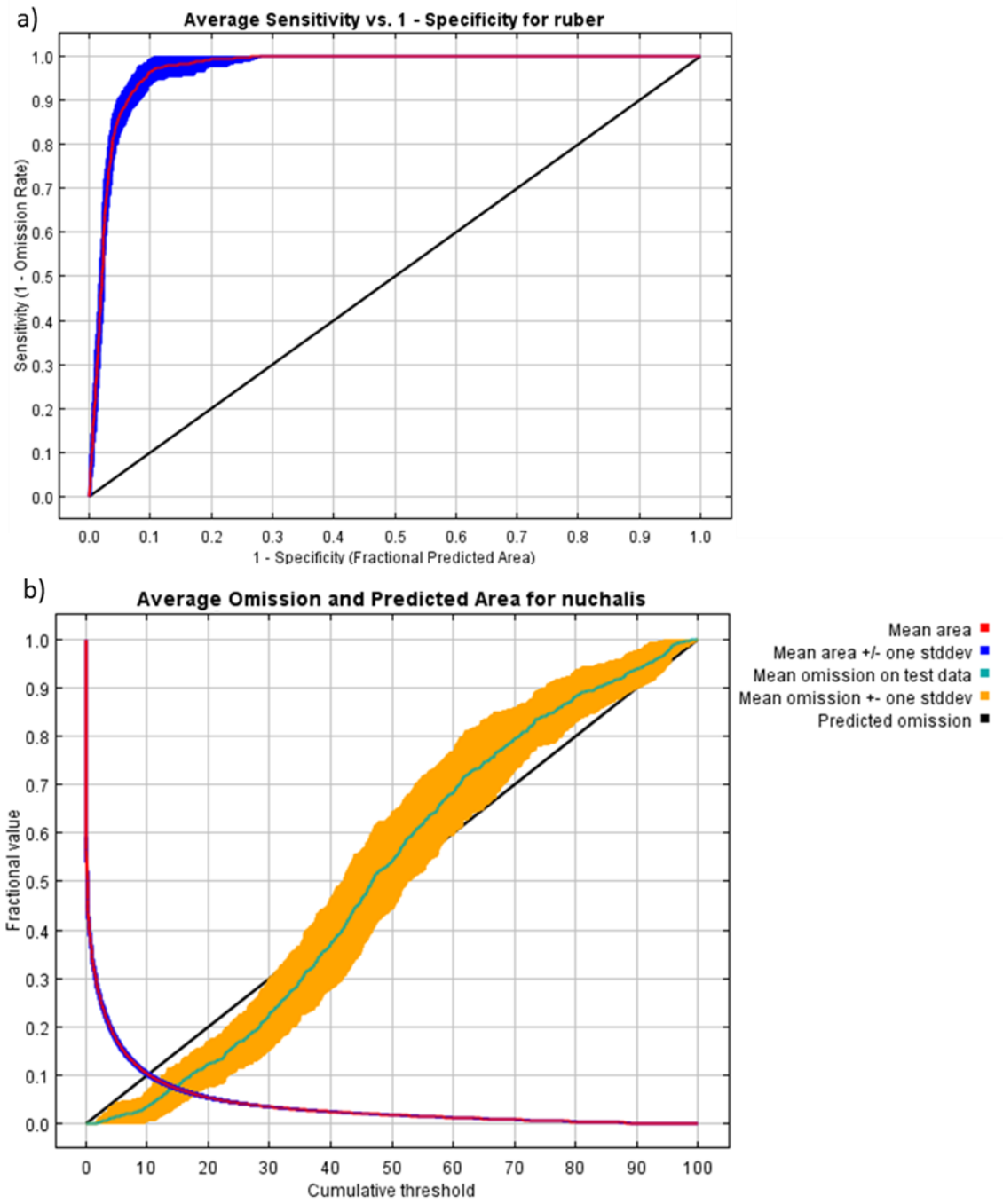
b)



Appendix 2.1 Area Under the Curve (AUC) and omission rate figures for analysis of the MAXENT model prediction for *S. varius*. The AUC plot (a) shows the mean AUC (0.90) in red with ± one standard deviation (blue line). The black line indicates a random model prediction. The omission rates are shown in (b) with the test data omission rate (turquoise line), the mean ± one standard deviation of the test data (yellow line), and the predicted omission rate (black line).



Appendix 2.2 Area Under the Curve (AUC) and omission rate figures for analysis of the MAXENT model prediction for *S. nuchalis*. The AUC plot (a) shows the mean AUC (0.95) in red with ± one standard deviation (blue line). The black line indicates a random model prediction. The omission rates are shown in (b) with the test data omission rate (turquoise line), the mean ± one standard deviation of the test data (yellow line), and the predicted omission rate (black line).



Appendix 2.3 Area Under the Curve (AUC) and omission rate figures for analysis of the MAXENT model prediction for *S. ruber*. The AUC plot (a) shows the mean AUC (0.97) in red with ± one standard deviation (blue line). The black line indicates a random model prediction. The omission rates are shown in (b) with the test data omission rate (turquoise line), the mean ± one standard deviation of the test data (yellow line), and the predicted omission rate (black line).

Appendix 3. Sample location and museum accession numbers of museum specimens used in the central Alberta hybrid zone study between *S. varius* and *S. nuchalis*. Each sample was given a Project ID determined by their phenotypic identification. The latitude (N) and longitude (W) where available are provided in decimal format. Individuals within the hybrid zone are marked (*) while all others are individuals from allopatric populations.

	Project ID	Location	Band Number	Decimal Latitude	Decimal Longitude	Museum
*	rnsaCAB001	James River, bridge on 584, AB	Z94.13.9	51.9000	-115.0000	Royal Alberta Museum
*	rnsaCAB002	James River, Wilson James Campsite, AB	Z94.13.10	51.8330	-115.2170	Royal Alberta Museum
*	rnsaCAB003	James River, Wilson James Campsite, AB	Z94.13.11	51.8330	-115.2170	Royal Alberta Museum
*	rnsaCAB003	James River, Wilson James Campsite, AB	Z94.13.11	51.8330	-115.2170	Royal Alberta Museum
*	rnsaCAB004	9 miles west of Bearberry, Improvement District 10, AB	Z95.11.14	51.8670	-115.0830	Royal Alberta Museum
*	rnsaCAB004	9 miles west of Bearberry, Improvement District 10, AB	Z95.11.14	51.8670	-115.0830	Royal Alberta Museum
*	rnsaCAB005	James River, Improvement District 10, AB	Z95.12.4	51.7670	-115.2170	Royal Alberta Museum
	rnsaCO004	Gunnison, Gunnison Co. Colorado	UWBM# 53394 GAV 257	38.8412	-106.4783	Royal Alberta Museum

	rnsaCO005	Park Co. Colorado	UWBM# 56359 GAV 858	39.4402	-105.7698	Royal Alberta Museum
	rnsaCO006	Gunnison, Gunnison Co. Colorado	UWBM# 56363 GAV 862			Royal Alberta Museum
	rnsaNM003	San Marcial, NM	cat #MSB Birds 28952, NK: 170195	33.6339	-107.0084	Royal Alberta Museum
	rnsaNM008	Black Range, NM	cat #MSB Birds 29241, NK: 170616	32.9438	-107.7040	Royal Alberta Museum
	rnsaNM017	Guadalupe Mountains, Dark Canyon, NM	cat #MSB Birds 29309, NK: 35775	32.1116	-104.7387	Royal Alberta Museum
*	rnsaSAB016	Porcupine Hills, AB	Z02.14.9	50.0172	-114.0522	Royal Alberta Museum
*	rnsaSAB017	Beaver Creek, Porcupine Hills, AB	Z96.18.9	49.8167	-113.9500	Royal Alberta Museum
	rnsaWA001	Easton, Kittitas Co. WA	UWBM# 49988 CSW 3959			Royal Alberta Museum
	rnsaWA003	Kittitas Co. WA	UWBM# 49001 CSW 4984			Royal Alberta Museum
	rnsaWA004	Kittitas Co. WA	UWBM# 49002 CSW 4985			Royal Alberta Museum

	rnsaWA004	Kittitas Co. WA	UWBM# 49002 CSW 4985			Royal Alberta Museum
	rnsaWA005	Cle Elum, Kittitas Co. WA	UWBM# 63675 MAM 18	47.1333	-120.9000	Royal Alberta Museum
	rnsaWA006	Kirkland, King Co. WA	UWBM# 85284 SEZ 045			Royal Alberta Museum
*	rnsaxybsaCAB001	James River area, AB	Z09.8.1	51.8685	-115.0437	Royal Alberta Museum
*	rnsaxybsaCAB002	James River area, AB	Z09.8.2	51.6554	-115.2607	Royal Alberta Museum
*	rnsaxybsaCAB005	James River area, AB	Z09.8.5	52.0256	-115.1573	Royal Alberta Museum
*	rnsaxybsaCAB009	James River area, AB	Z10.3.4	51.8678	-115.0068	Royal Alberta Museum
*	rnsaxybsaCAB011	James River area, AB	Z10.3.6	51.8685	-115.0156	Royal Alberta Museum
*	rnsaxybsaCAB014	Clearwater River area, AB	Z10.3.9	51.9890	-115.2330	Royal Alberta Museum
*	rnsaxybsaCAB018	James River area, AB	Z10.3.14	51.8681	-115.1149	Royal Alberta Museum

*	rnsaxybsaCAB023	1.5 miles south Strachan, Improvement District 10, AB	Z95.9.1	52.2000	-115.1000	Royal Alberta Museum
*	rnsaxybsaCAB024	Cow Lake area, Improvement District 10, AB	Z95.9.7	52.2670	-115.0330	Royal Alberta Museum
*	rnsaxybsaCAB025	James River, Improvement District 10, AB	Z95.12.7	51.7670	-115.2170	Royal Alberta Museum
*	rnsaxybsaCAB026	James River, Improvement District 10, AB	Z95.12.6	51.7670	-115.2170	Royal Alberta Museum
*	rnsaxybsaCAB030	James River area, west of Sundre, AB	Z11.5.10	51.7774	-115.2274	Royal Alberta Museum
*	rnsaxybsaCAB033	Vicary Creek Area, AB	Z12.2.7	49.8215	-114.4246	Burke Museum, BG TB
*	rnsaxybsaCAB034	Vicary Creek Area, AB	Z12.2.20	49.7995	-114.4678	Burke Museum, BG TB
*	rnsaxybsaCAB035	Porcupine Hills, AB	Z12.2.27	50.0243	-114.0447	Royal Saskatchewan Museum
*	rnsaxybsaCAB036	Porcupine Hills, AB	Z12.2.30	50.0235	-114.0457	Royal Saskatchewan Museum
*	rnsaxybsaCAB037	Kananaskis Country, AB	Z12.3.1	51.0265	-114.8734	Royal Saskatchewan Museum

*	rnsaxybsaCAB038	Kananaskis Country, AB	Z12.3.2	51.0526	-114.8117	Royal Saskatchewan Museum
*	rnsaxybsaCAB039	Kananaskis Country, AB	Z12.3.9	50.6437	-114.4661	Royal Saskatchewan Museum
*	ybsaCAB013	James River area, AB	Z10.3.13	51.8059	-115.2022	Royal Saskatchewan Museum
*	ybsaCAB021	Brule area, AB	Z11.4.30	53.4005	-117.8685	Royal Saskatchewan Museum
*	ybsaCAB022	James River area, west of Sundre, AB	Z11.5.2	51.8900	-115.0475	Royal Saskatchewan Museum
*	ybsaCAB025	Clearwater River area, AB	Z11.5.9	52.0390	-115.1604	Burke Museum, BG TB
*	ybsaCAB027	James River area, west of Sundre, AB	Z11.5.14	51.8490	-115.0182	Burke Museum, BG TB
	ybsaNC003	Dillingham, Buncombe Co. NC	UWBM# 86868 RBB 538			Burke Museum, BG TB
	ybsaNC004	Murchison, Buncombe Co. NC	UWBM# 86869 RBB 539			Burke Museum, BG TB
	ybsaSK007	Grand Coulee, SK	accn16687 exn 001	50.4333	-104.8167	Burke Museum, BG TB

ybsaSK008	Ramada Inn, Regina, SK	accn16698 exn 002	50.4500	-104.6167	Museum of Southwest Biology
ybsaSK009	Regina, SK	accn17618 exn 002	50.4500	-104.6167	Museum of Southwest Biology
ybsaSK010	Rouleau, SK	accn20030 exn	50.1833	-104.9333	Museum of Southwest Biology
ybsaSK011	Regina, SK	accn20040 exn 001	50.4500	-104.6167	Burke Museum, BG TB
ybsaSK012	RSM, Regina, SK	accn18288 exn 001	50.4500	-104.6167	Burke Museum, BG TB
ybsaSK013	RSM, Regina, SK	accn16729 exn 001	50.4500	-104.6167	Burke Museum, BG TB
ybsaSK014	Regina, SK	accn17134 exn 002	50.4500	-104.6167	Burke Museum, BG TB