POPULATION GENETICS AND HISTORICAL INTROGRESSION OF A NORTH AMERICAN PASSERINE WITHIN THE GENUS *SIALIA*

AARON LESLIE VEALE
Bachelor of Science (Hons), Thompson Rivers University, 2021

A thesis submitted
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

BIOLOGICAL SCIENCES

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

POPULATION GENETICS AND HISTORICAL INTROGRESSION OF A NORTH AMERICAN PASSERINE WITHIN THE GENUS *SIALIA*

AARON LESLIE VEALE

Date of Defence: August 23, 2023

Dr. T. Burg	Professor	Ph.D.
Dr. M. Reudink Thompson Rivers University Thesis Co-Supervisors	Professor	Ph.D.
Dr. C. Goater Thesis Examination Committee Member	Professor	Ph.D.
Dr. A. Iwaniuk Thesis Examination Committee Member	Professor	Ph.D.
Dr. D. Yevtushenko Thesis Examination Chair	Assistant Professor	Ph.D.

ABSTRACT

This study analyzed the genetic structure of mountain bluebirds (*Sialia currucoides*) from across their range using restriction site-associated DNA sequencing to find single nucleotide polymorphisms. I found evidence of at least four distinct genetic clusters, two of which were in Alberta. I then went on to hypothesize current and historical barriers to gene flow. In the third chapter, I examined the mountain bluebird breeding populations in Alberta and Saskatchewan for signals of current and historical hybridization. Although I did not find evidence of recent hybridization, I did detect signals of ancient introgression from both bluebird congeners. I then used both nuclear and mitochondrial DNA to examine the *Sialia* polytomy containing all three species of bluebirds. The polytomy showed signs of both incomplete lineage sorting between mitochondrial genes and mitonuclear discordance. As such, it appears that hybridization and other potential factors may obscure the evolutionary history of the entire genus.

ACKNOWLEDGEMENTS

I begin by extending my thanks to both of my supervisors, Dr. Matthew Reudink and Dr. Theresa Burg, for everything they have contributed to help make this thesis a reality. Without their guidance and expertise, I would be a ship without a rudder. When I first joined Matt's lab in 2019, I had no intentions of entering research. After two years in the Burg lab, I am more excited about academia than ever. I would also like to thank them for my new personal motto, abbreviated to the PG version of JFDI.

My thanks also to my committee members, Dr. Cam Goater and Dr. Andrew Iwaniuk. Their feedback with each step of the project proved to be invaluable. Additionally, our field season would have been very different without Andrew's generous offer allowing us to use his field truck. Likewise, I am extremely grateful to Cam for including me as a TA in the field biology course. I've never had more fun teaching before, and I loved exploring the Cypress Hills with the class.

To those who provided samples for my study, I owe you my thanks as well: Dr. Sharon Birks from the Burke Museum of Natural History and Culture; Dr. Garth Spellman at the Denver Museum of Nature and Science; Andrew Johnson from the Museum of Southwestern Biology; Dr. Catherine Dale from the Newfoundland Breeding Bird Atlas at Bird Studies Canada; Dr. Russ Dawson from the University of Northern British Columbia; Sandy Proulx; Dr. Keith Hobson from Environment and Climate Change Canada; and the passionate folks at the Mountain Bluebird Trails Conservation Society.

Thank you to the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant for making my research possible. Their funding plays an integral role in a many important and interesting projects.

I would be remiss if I did not thank my past and current members of the Burg Lab for all of their invaluable help. Thank you to Dr. Dilini Abeyrama for her abundant patience in answering my endless questions. Her role as the office advice giver was unparalleled. Thank you, Danika Schramm, and Thilini Thakshila, for all your help collecting samples. We made a formidable team as we patrolled nest box route after nest box route looking for bluebirds. Thank you to Bukola Oguntuase-Osagie and Bryan Fisher for contributing ideas and new avenues of research to explore. A big thanks goes to Dr. Juan Enciso, Dr. Linda Lait, and Dr. Brendan Graham. The collective advice and knowledge you shared was beyond anything I could have imagined.

Last but never least, thank you to all my family and friends. To my mom and dad, Wendy and Greg, I appreciate every bit of support you've been able to offer me from a province away. To my sister Lindsay, thank you for being one of my biggest supporters, no matter the endeavour. To our dog Major—well, you're trying your best. Your support has meant the world to me and allowed me to go places I never expected. Thank you to my loving partner, Alyssa, and our two energetic cats, Tig and Winnie. You mean the world to me and I cherish every moment with you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

°C	degrees Celsius
~	approximately
bam	binary alignment map
bp	base pairs
bwa	Burrows-Wheeler Alignment Tools
COI	cytochrome c oxidase subunit 1
Cyt b	cytochrome b
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EABL	eastern bluebird
F_{st}	fixation index
H_e	expected heterozygosity
H_o	observed heterozygosity
ILS	incomplete lineage sorting
K	genetic cluster
kyBP	thousand years before present
LGM	last glacial maximum
McMC	Markov chain Monte Carlo
$MgCl_2$	magnesium chloride
mM	millimolar
MOBL	mountain bluebird
mtDNA	mitochondrial DNA
myBP	million years before present
n	sample size
ND2	NADH dehydrogenase 2
nuDNA	Nuclear DNA
PC	principle coordinate
PCoA	Principle coordinate analysis
PCR	polymerase chain reaction
Q	ancestry coefficient
RADseq	restriction site-associated DNA sequencing
SNAPP	SNP and AFLP Package for Phylogenetic analysis
SNP	single nucleotide polymorphism
µM	micromolar
vcf	variant call format
WEBL	western bluebird
WOTH	wood thrush

POPULATIONS

CAB	central Alberta
CBC	central British Columbia
CCA	central California
CO	Colorado
CYP	Cypress Hills
ID	Idaho
MT	Montana
NV	Nevada
PNW	Pacific Northwest
OR	Oregon
SAB	south Alberta
SBC	southern British Columbia
SEBC	southeastern British Columbia
SER	southeastern Rockies
SK	Saskatchewan
SWON	southwestern Ontario
WA	Washington
WAB	western Alberta
WY	Wyoming

Chapter 1: General Introduction

1.1 Background

1.1.1 Speciation and population genetics

The need to categorize organisms into discrete units is innately human and may pre-date modern humans and our way of thinking. How exactly we classify species has varied drastically across time and across cultures (Berlin et al., 1973), in large part due to the complexity in defining a "species". As Charles Darwin famously wrote, "…I look at the term species, as one arbitrarily given for the sake of convenience to a set of individuals closely resembling each other" (Darwin, 1859). Yet, many attempts have been made to construct a universal definition of species; for example, the biological species concept, which was first proposed by Mayr (1942), defines a species as a group of interbreeding organisms that can produce fertile offspring. Although the biological species concept is widely applicable, it fails to account for organisms that reproduce asexually, extinct taxa with no breeding information, and organisms that are plainly different both behaviourally and morphologically but can still reproduce with each other (Ehrlich, 1961). Alternatively, the morphological species concept proposed by Cronquist (1978) defines a species as constantly and clearly distinct and can be distinguished by consistently identifiable biometrics. Morphologically species concept neglects cryptic species: species that are morphologically identical but may be behaviourally or genetically distinct. As the field of genetics continues to make advancements, the

phylogenetic species concept has risen in prominence, simply stating that a species is the smallest group of organisms with shared common ancestry (Cracraft, 1983). The discourse around what defines a species may not be resolved in the near future, but each definition plays an important role in understanding what a species—and a population—is.

At its heart, population genetics studies the process of speciation that arises through genetic variation, along with the causes of variation, like natural selection, genetic drift, and gene flow. Genetic variation among and within populations can be examined by measuring the temporal and spatial changes in allele frequencies (Moore 1977). Physical barriers, such as large bodies of water (Bertrand et al., 2014), mountain ranges (Milá et al., 2009), sky islands (Dempsey et al., 2010; Purushotham & Robin, 2016), and fragmented habitat (Bergl &Vigilant, 2007) can isolate populations causing the cessation of gene flow. Even physical distance can become a barrier to gene flow. Species with low connectivity between breeding populations may become isolated if their range is large, a process known as isolation-by-distance (IBD) (Wright, 1943). Individuals at the extremes of the range may even have suitable habitat connecting them, such as the boreal forest that spans North America. However, if individuals do not move between populations, then gene flow is unlikely to ever occur (Wright, 1943).

Once populations are separated, organisms may experience differences in selective pressures or diverge through genetic drift. Physical barriers may be contemporary, acting as an obstacle for the gene flow within an extant species, but they may also be historic. For example, repeated cycles of cooling during the Quaternary period starting ~ 2.6 million years before present (Gibbard & Head, 2009) led to the aridification and compositional change in the species community of the Guineo-

Congolian rainforest in central Africa (Anhuf et al., 2006; Vanden Abeele et al., 2021). For specialized species excluded from the altered forests, fragmentation likely occurred within their populations. Evidence of the temporary cessation in gene flow can still be detected in multiple species, such as the trees *Scorodophloeus zenkeri* (Vanden Abeele et al., 2021) and *Staudtia kamerunensis* (Vanden Abeele et al., 2023), and the okapi (*Okapia johnstoni*) (Stanton et al., 2014). Africa was not the only continent affected as average temperatures decreased worldwide. For North America, the Quaternary marks the beginning of arguably the most influential historic barrier to gene flow which fragmented populations of various taxa continent-wide.

1.1.2 Pleistocene glaciation and glacial refugia

The Quaternary can be split amongst two epochs—the first being the Pleistocene and the second being the Holocene (Gibbard & Head, 2009). Characterized by its repeated cycles of cooling and warming, the Pleistocene played host to a number of events that strengthened and contributed the speciation of taxa worldwide including North American birds. The most recent glacial maximum, known as the last glacial maximum (LGM), began ~25 kyBP (Dalton et al., 2022). Multiple studies had hypothesized that the recent glacial cycles contributed heavily to the species splits seen today (e.g., Hewitt, 1996; Rand, 1948). However, Klicka and Zink (1997) asserted that many contemporary east/west species observed today were likely already in place prior to the onset of the first glaciation, such as the blue jay and Steller's jay (*Cyanocitta cristata* and *C. stelleri*, respectively).

During the LGM, three expansive icecaps dominated North America (Hughes et al., 2013; Dalton et al., 2022). In the east, the Laurentide ice sheet stretched from Nunavut to the center of the eastern United States, and west to the eastern slope of the Rockies (Dyke et al., 2002; Dyke, 2004). In the far north, the Laurentide ice sheet met with the Innuitian ice sheet which covered parts of the Queen Elizabeth Islands (Blake, 1970; Dalton et al., 2022; England et al., 2006) On the other side of the Rocky Mountains, the Cordilleran ice sheet stretched from coastal Alaska to central Washington (Blaise et al., 1990). The timing of expansion depended on location. On Vancouver Island, the Cordilleran ice sheet reached its maximum extent ~19 kyBP (Ward et al., 2003; Al-Suwaidi et al., 2006), while western Washington reached its maximum extent ~17 kyBP (Porter & Swanson, 1998). As the creeping ice and snow pushed southward, terrestrial species were left to either adapt, move, or face extirpation.

Many northern species were forced to retreat to one or more glacial refugia (Figure 1.1), which are defined as suitable, ice-free habitats on the periphery of ice sheets (coastal regions, lowland areas, and unglaciated mountain peaks) (Holderegger & Thiel-Egenter, 2009). In northwestern North America, strong molecular evidence exists for two major refugia: Beringia and the Pacific Northwest. Beringia (now parts of Alaska, the Yukon, and the Bering Sea) is supported by various species of flora, such as white spruce (*Picea glauca*) (Anderson et al., 2006), Townsend's daisy (*Townsendia hookeri*) (Thompson & Whitton, 2006), purple saxifrage (*Saxifraga oppositifolia*) (Abbott et al., 2000b; Abbott & Comes, 2004), and entireleaf mountain avens (*Dryas integrifolia*) (Tremblay & Schoen, 1999) as well as fauna like boreal chickadee (*Poecile hudsonicus*) (Lait & Burg, 2013), brown bear (*Ursus arctos*) (Leonard et al., 2000), Alaskan marmot

(*Marmota broweri*) (Steppan et al., 1999), and Canada jay (*Perisoreus canadensis*) (Dohms et al., 2017; van Els et al., 2012). The Pacific Northwest is more complex as a refugium. Brunsfeld et al. (2001) found evidence of taxa confined by the Coast/Cascade Mountains in the west and others by the Northern Rockies in the east of the Pacific Northwest. An example of the split can be seen in the two species of blue grouse, the sooty (*Drendragapus fuliginosus*) and dusky grouse (*D. obscurus*) (Barrowclough et al., 2004). Soltis et al. (1997) also found a north (ranging from British Columbia to Oregon) and south (Oregon to California) split in various plant taxa which would later become known as the Soltis line (Brunsfeld et al., 2007). The cause of the Soltis line appears unclear. The best supported hypothesis thus far attributes the Soltis line's origins to colonization from multiple refugia within the Pacific Norwest meeting as during recolonization (Shafer et al., 2010).

Refugia east of the Rockies and towards the center of the continent (Figure 1.1) may be just as complex as the Pacific Northwest, if not more so. Studies examining falsegold groundsel (*Packera pseudaurea*) (Golden & Bain, 2000) and one-sided wintergreen (*Orthilia secunda*) (Beatty & Provan, 2010) found molecular evidence supporting a putatively cryptic refugia as far north as southwestern Alberta. The exact timing of such a refugium remains unclear, but an ice-free corridor likely existed prior to and following the LGM (Bednarski & Smith, 2007; Dyke et al., 2003). Moving south of Alberta, the interior of the continent likely appeared very different from what is known today. The xeric species of the cold and hot deserts were likely confined to low elevations, with more mesic species assemblages occurring in the highlands (Graham et al., 2020; Roberts & Hamann, 2015). Suitable habitat for southern populations of boreal

tree species, such as trembling aspen (*Populus tremuloides*), lodgepole pine (*Pinus contorta*), and black spruce (*Picea mariana*) likely occurred in the Great Plains regions and the southwest tablelands of Colorado and New Mexico (Roberts & Hamann, 2015).

Deglaciation started in earnest ~19 kyBP before present with the retreat of the Laurentide ice sheet (Dyke et al., 2003) followed therein by the Cordilleran ice sheet around ~16.5 kyBP (Porter & Swanson, 1998). By ~ 10 kyBP, boreal forest had invaded the formerly glaciated prairie provinces, connecting the distal refugia of the far north with the interior of the continent (Strong & Hills, 2005). As the glaciers continued to recede, taxa within their respective refugia were able to colonize the newly available space. Highly vagile taxa, such as large mammals and volant birds, were quickly able to spread and take advantage of the ecological vacancies. Some species, such as common raven (*Corvus corvax*) (Omland et al., 2000) and mule deer (*Odocoileus hemionus*) (Latch et al., 2009) dispersed so efficiently that multiple Pleistocene lineages demonstrate admixing, ultimately obscuring the paths they took out of their respective refugia. The progression of the Holocene (starting ~11.5 kyBP) ultimately saw many populations undergo range expansions from their Pleistocene refugia.

1.2 Secondary contact and hybridization

Wherever two allopatric populations expanded to the point of reunification, secondary contact occurred. Secondary contact can influence the history and formation of a species depending on the reproductive barriers in place and their relative completeness.

The two forms of barriers include prezygotic (physical barriers, temporal isolation, behavioural isolation, gametic isolation, or physiological incompatibility) and postzygotic (hybrid sterility or inviability) mechanisms; both are taxa-dependent and vary widely (Qvarnström et al., 2016; Veen et al., 2013). Birds, for example, use mating displays and plumage that varies widely between species to select conspecific mates (Price, 2008). In theory, assortative mating continues the process of speciation until even closely related congeners were incapable of reproducing with each other (Butlin, 1987). However, this is not always the case in nature.

When reproductive barriers break down, they may lead to the formation of a hybrid zone wherever the parental species overlap. Hybridization appears to be more commonplace than first theorized, with an estimated 25% of plant and 10% of animal species readily hybridizing, and hybrid offspring occurring anywhere between 1 in 100 to 1 in 10,000 individuals (Mallet, 2005). It is likely that many organisms interbred with close sister taxa during the early stages of speciation. Modern hominins are a prime example. There is growing evidence to support hybridization within Africa between sub-Saharan modern humans and multiple unidentified extinct hominin taxa (Lachance et al., 2012; Hsieh et al., 2016; Wall et al., 2019). Upon leaving Africa, *Homo sapiens* interbred with at least two other *Homo* species (Bae et al., 2017). Non-African groups have mosaic genomes with genetic contributions from *H. neanderthalensis* and some have additional introgression from Denisovans (no official taxonomic name as of yet) (Green et al. 2010; Sankararaman et al., 2016).

Two hybrid models will be discussed herein: the first is the bounded hybrid superiority model and the second being the tension zone (Barton and Hewitt, 1985;

Moore, 1977). In the bounded hybrid superiority model, hybrid offspring have a reproductive advantage over at least one of its parental species within the area of overlap (Moore, 1977). An example of this model is seen in the glaucous-winged (*Larus glaucascens*) x western (*L. occidentalis*) gull hybrid zone in southwestern British Columbia and the Olympic Peninsula of Washington state (Good et al., 2000). Within this narrow zone, pairings involving hybrid gulls experienced higher reproductive success rates compared to both parental species (Good et al., 2000). If hybrid fitness remains high and gene flow from the parental species stops, speciation can occur in the hybrid population. The Oxford ragwort (*Senecio squalidus*), for example, is thought to have arisen from a hybrid zone on Mt. Etna, Sicily where its two parent species overlap (Abbott et al., 2000a; James & Abbott, 2005). Molecular evidence also supports hybrid origins in the Goashan pine (*Pinus densata*) (Wang & Szmidt, 1994) and the Virgin River brittlebush (*Encelia virginensis*) (Allan et al., 1997).

Barton and Hewitt (1985) hypothesized that most hybrid zones fit better with the tension zone model rather than hybrid superiority. They hypothesized most hybrid zones are clines kept in place by the random dispersal and selection against hybrids, while pure parental populations continually immigrate into the zone. A well-studied example of tension zone is the Townsend's (*Setophaga townsendii*) X hermit (*S. townsendii*) warbler zone in the Cascade Mountains, where hybrids have lower fitness compared to either parental species (Rohwer & Wood, 1998; Rohwer et al., 2001; Wang et al., 2019). The balance of tension zones can be precarious and may change over time. Ephemeral hybrid zones are geographically small and short-lived areas of overlap typically associated with cases of range expansion and contraction or shifts in habitat (Pearson & Rohwer, 2000;

Robins et al., 2014). Paradoxically, ephemeral hybrid zones may be the most common type while also being the hardest to detect. Parental species undergoing range shifts often shift the boundary of the hybrid zone as well (Wielstra, 2019). Occasionally, one parental species invades the range of the other (Slager et al., 2020). In this scenario, the hybrid zone either collapses as both parental populations may merge back into a single species or one parent is ultimately extirpated from the region (Slager et al., 2020). Such instances can be especially hard to detect as the only evidence left is stored within the survivors' genomes.

1.3 Study species

Bluebirds are medium sized, sexually dimorphic songbirds endemic to North America. They are members of the family Turdidae (thrushes and allies). The genus consists of three species, all of which are obligate secondary cavity nesters. In all three species, males have bright blue plumage ranging partially in the UV spectrum (Figure 1.2) while female colouration is muted. Western bluebirds (*Sialia mexicana*) and eastern bluebirds (*S. sialis*) also sport dark red chests and backs to different extents. Mountain bluebirds (*S. currucoides*) completely lack the red of their counterparts, donning sky-blue plumage, with a pale breast and belly. Bluebirds have a variable diet, mainly feeding on insects during the warmer months and plant matter during the winter (Johnson & Dawson, 2020).

In the breeding season, bluebirds range throughout North America, with the exception of the far northern tundra (Figure 1.3). The breeding range of western bluebirds

spans from the southern interior of British Columbia down into central Mexico (Figure 1.3; Guinan et al., 2020). Mountain bluebirds overlap with much of this range, but their breeding range includes further north and east into the interior of the continent (Figure 1.3; Johnson & Dawson, 2020). In the east of their range, mountain bluebirds come into contact with eastern bluebirds which have the largest breeding range of all three species. Eastern bluebirds breed from Saskatchewan in the west, east to the Atlantic states and provinces, and south to Florida (Figure 1.3; Gowaty & Plissner, 2020). Disjunct populations also exist in Mexico, central America, and the Bahamas (Figure 1.3; Gowaty & Plissner, 2020).

All three species have either expanded their ranges since European colonization or become more common in areas with historically low conspecific densities (Duckworth & Badyaev, 2007; Gowaty & Plissner, 2020; Johnson & Dawson, 2020). In the west, aggressive western bluebirds have managed to expand their range eastward into western Montana, coming into secondary contact with mountain bluebirds there (Duckworth & Badyaev, 2007). The more aggressive male western bluebirds push the more placid male mountain bluebirds from nesting sites (Duckworth & Badyaev, 2007; Duckworth, 2008; Duckworth & Semenov, 2017). In the prairie provinces, mountain bluebird density has increased thanks to habitat created by European colonization and nest box programs (Johnson & Dawson, 2020), where they encounter eastern bluebirds spreading west (Gowaty & Plissner, 2020). With expanded ranges and increased secondary contact come new opportunities for hybridization.

Bluebird hybridization is a known but understudied phenomenon, with all species pairs capable of producing viable, fertile offspring (although there are no records for

eastern and western hybrids in nature). Hybrids between mountain and eastern bluebirds lack formal study but have been reported in Wisconsin (Johnson & Dawson, 2020) and Manitoba (Lane, 1969). In western Montana, both F1 and late-stage western and mountain bluebird hybrids were found by Duckworth and Semenov (2017). However, all F1 hybrids were contained with a very narrow range—pointing to an ephemeral tension zone on the move (Duckworth & Semenov, 2017). Even captive hybrids between eastern and western bluebirds have occurred, although this is unlikely to happen in nature due to largely allopatric breeding ranges (Gowaty & Plissner, 2020).

1.4 Systematics in bluebirds and close relatives

Phylogeny in *Sialia* has been largely overlooked with only a few studies using all three species for genetic analyses (Klicka et al., 2005; Voelker and Klicka, 2008). Mitochondrial DNA (mtDNA) places bluebirds in a basal position in the Turdidae phylogeny relative to other thrushes (Klicka et al., 2005). The ancestor of all bluebirds emerged during the mid-Pliocene (3.3 to 3 million years before present) (Klicka et al., 2005).

Within the genus, however, taxonomic relationships remain opaque. Early bluebird phylogenies based on morphology placed the eastern and western bluebirds as sister species (e.g., Mengel, 1970). However, Klicka et al. (2005) challenged this taxonomy using NADH dehydrogenase 2 (ND2) and the cytochrome b (cyt b) genes and found that eastern and mountain bluebirds were sister taxa. Their phylogeny was challenged in 2008 by Voelker and Klicka using the same mitochondrial genes. Voelker and Klicka (2008)

found eastern and western bluebirds to be sister to each other (matching previously phylogenies based on morphology) . Ultimately, the topology was left unresolved as a polytomy (Voelker and Klicka 2008).

1.5 Molecular markers

1.5.1 Molecular methods in population genetics

While phenotypic data like behaviour or morphology have been used in the past to infer population structure and evolutionary relationships, they are not perfect proxies. Analyzing phenotypic data can be costly, both in terms of time and funding (Avise, 1994). In many organisms, especially plants (Clausen et al., 1941), phenotypic plasticity due to environmental factors may confound morphological study. Molecular analyses only need a small contribution of DNA rather than a partial or whole organism. As the field continues to advance, molecular markers continue to be invaluable tools for studying population genetics. While phenotypic data is still important, coupling the two provides a much better picture of a species' evolutionary history.

1.5.2 Mitochondrial genes

Mitochondria play a pivotal role in cellular respiration and are crucial for the maintenance of living systems. Most eukaryotic cells contain a plethora of mitochondria which are independently capable of fusing, movement and divisions (Bereiter-Hann & Voth, 1994). MtDNA has served as a facet in genetic studies for a number of reasons.

Unlike nuclear markers, mtDNA is haploid and uniparentally inherited from the maternal side in most birds. MtDNA can be ideal for phylogenetic and biodiversity studies as it has a much faster mutation rate (considered to be an order of magnitude higher) (Brown et al., 1979; Ballard and Whitlock, 2004). Mitochondrial genomes are also smaller than nuclear genomes, being about 1/10,000 the size of the shortest animal nuclear genome (Ballard & Whitlock, 2004). The shorter length makes mtDNA a cost-effective option for sequencing and primer development. In fact, the development of universal primers allowed for the amplification of mitochondrial genes across a range of species (Kocher et al., 1989; Sorenson et al., 1998).

1.5.3 Nuclear markers—single nucleotide polymorphisms

As advantageous as mtDNA is, Voelker and Klicka (2008) suggested that the bluebird polytomy may be best resolved using nuclear markers rather than mtDNA. Nuclear DNA (nuDNA) is biparentally inherited, meaning each diploid organism inherits one copy of each marker from each paent, and undergoes recombination. While recombination and biparental inheritance make tracing lineages more difficult than with mtDNA, the larger effective population size offers counterbalance to the limitations of mtDNA (Freeland et al., 2012).

Single-nucleotide polymorphisms (SNPs) are mutations located at single base pair positions found throughout the genome (Jehan & Lakhanpaul, 2006). SNPs have become widely adopted in many population and speciation studies due to their low mutation rates which in turn reduce the chances of homoplasy (Brito & Edwards, 2009; Brumfield et al.,

2003). They can be categorized into either transversions (C/G, A/T, C/A, and T/G) or transitions (C/T and G/A). SNPs can be tetrallelic or triallelic, but most are biallelic (Hayward et al., 2012).

1.6 Thesis goals

Current literature on the genetics of the mountain bluebird is not well described. The purpose of my research is to expand what is known regarding the genetics of these widespread songbirds and where they fit within the genus *Sialia*. As such, my four primary research goals were to: 1) establish if mountain bluebirds show discrete population structure and identify potential candidate genes, 2) use any structuring to inform which glacial refugia mountain bluebirds used during the LGM, 3) detect signals ongoing hybridization in bluebirds from the western prairie provinces, and 4) resolve the polytomy among extant *Silia* species. To accomplish my objectives, I analyzed DNA from blood, tissue, and feathers from wild caught bluebirds and existing museum samples.

1.6.1 Population structure and glacial refugia

To determine the presence or absence of discrete breeding groups within the mountain bluebird range, I used next-generation sequencing to find unlinked SNPs. Each SNP corresponds to a locus; thus, covering multiple SNPs acts as a proxy for markers from across the nuclear genome. I then used these markers to analyze gene flow between

populations and assess potential physical barriers to it. The relationship between populations also provided potential insights into which glacial refugia bluebirds used during the LGM and where they colonized first upon their expansion. Finally, I also used an annotated thrush genome to identify potential candidate genes that may correspond to local adaptations among breeding populations.

Much is known regarding the breeding habits of mountain bluebirds (see Haecker, 1948; Herlugson, 1981; Johnson & Dawson, 2020; McArthur et al., 2017) but their genetic structure remains nebulous. As per my knowing, this is the first time a population level study has been performed in the species. Given the large breeding range of mountain bluebirds (from Alaska to New Mexico), I predicted that discrete breeding groups will be detected. I also predicted that signals of significant IBD will be detected as birds from the extremes of the range are unlikely to ever undergo gene flow with each other.

1.6.2 Detecting both ongoing and ancient hybridization

Based on previous examples of known hybridization, I analyzed mountain bluebirds from the western prairie provinces along with western and eastern bluebird samples in order to detect and model any intermediate genotypes. I initially focused on contemporary hybridization and later expanded my research to include signals of ancient introgression as well (which had not been studied prior). If western and eastern bluebirds are expanding into the western prairie provinces, they are likely doing so in initially small numbers. As Hubbs' (1955) principle stipulates, hybridization rates should increase when

the numerically inferior western and eastern bluebird breed with the more common mountain bluebird. Therefore, I predicted that signals of hybridization would be detectable in areas where at least two bluebird species had been known to overlap.

1.6.3 Phylogenetics of bluebirds

Understanding the polytomy within the *Sialia* phylogeny may provide important insights into how the three species rapidly evolved. The bluebird phylogeny has been addressed using mitochondrial genes as markers, yet the contrasting results from the few studies available have created an unresolved polytomy (Klicka et al., 2005; Voelker and Klicka, 2008). Voelker and Klicka (2008) suggested the use of nuDNA may resolve the before mentioned polytomy. As such, I created a phylogeny using only nuclear markers and then compared this to three gene trees created using mtDNA. I predict that the two western bluebird species (mountain and western) will be more closely related based off their significant range overlap.

1.7 Thesis Organization

My thesis consists of four chapters. Chapter 1 provides an general overview for the biological concepts relevant to this study (e.g., speciation, glacial refugia, hybridization) and describes the molecular methods used to examine these subjects. Chapter 2 analyzes population structure from across the mountain bluebird breeding range, relates it to possible glacial refugia, and identifies possible candidate genes that

separate the groups. Chapter 3 looks for evidence of ongoing hybridization within bluebirds from the western prairies, as well as evidence for ancient hybridization, and attempts to resolve the bluebird polytomy using two different types of molecular marker. Chapter 4 then summarizes the results of chapters 2 and 3 and discusses what the study of mountain bluebirds tells us about the species history pre and post LGM, how hybridization alters genomes, and what is causing the bluebird polytomy.

1.8 Literature Cited

Abbott, R. J., & Comes, H. P. (2004). Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (purple saxifrage). New Phytologist, 161(1), 211-224.

Abbott, R. J., James, J. K., Irwin, J. A., & Comes, H. P. (2000a). Hybrid origin of the Oxford ragwort, *Senecio squalidus* L. Watsonia, 23(1), 123-138.

Abbott, R. J., Smith, L. C., Milne, R. I., Crawford, R. M., Wolff, K., & Balfour, J. (2000b). Molecular analysis of plant migration and refugia in the Arctic. Science, 289(5483), 1343-1346.

Allan, G. J., Clark, C., & Rieseberg, L. H. (1997). Distribution of parental DNA markers in *Encelia virginensis* (Asteraceae: Heliantheae), a diploid species of putative hybrid origin. Plant systematics and evolution, 205, 205-221.

Al-Suwaidi, M., Ward, B.C., Wilson, M.C., Hebda, R.J., Nagorsen, D.W., Marshall, D., Ghaleb, B., Wigen, R.J. & Enkin, R.J. (2006). Late Wisconsinan Port Eliza Cave deposits and their implications for human coastal migration, Vancouver Island, Canada. Geoarchaeology: An International Journal, 21(4), 307-332.

Anderson, L. L., Hu, F. S., Nelson, D. M., Petit, R. J., & Paige, K. N. (2006). Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proceedings of the National Academy of Sciences, 103(33), 12447-12450.

Anhuf, D., Ledru, M.P., Behling, H., Da Cruz Jr, F.W., Cordeiro, R.C., Van der Hammen, T., Karmann, I., Marengo, J.A., De Oliveira, P.E., Pessenda, L. & Siffedine, A. (2006). Paleo-environmental change in Amazonian and African rainforest during the LGM. Palaeogeography, Palaeoclimatology, Palaeoecology, 239(3-4), 510-527.

Avise, J. C. (1994). Molecular markers, natural history and evolution. Chapman and Hall, New York.

Bae, C. J., Douka, K., & Petraglia, M. D. (2017). On the origin of modern humans: Asian perspectives. Science, 358(6368), eaai9067.

Ballard, J., & Whitlock, M. (2004) The incomplete natural history of mitochondria. Molecular Ecology., 13 (4): 729-744. 10.1046/j.1365-294X.2003.02063.x.

Barrowclough, G. F., Groth, J. G., Mertz, L. A., & Gutiérrez, R. J. (2004). Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). Molecular Ecology, 13(7), 1911-1922.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. Annual review of Ecology and Systematics, 16(1), 113-148.

Beatty, G. E., & Provan, J. I. M. (2010). Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. Molecular Ecology, 19(22), 5009-5021.

Bednarski, J. M., & Smith, I. R. (2007). Laurentide and montane glaciation along the Rocky Mountain Foothills of northeastern British Columbia. Canadian Journal of Earth Sciences, 44(4), 445-457.

Bereiter-Hann, J., & Voth, M. (1994) Dynamics of mitochondria inliving cells: shape changes, dislocations, fusion, and fission of mitochondria. Microscopy Research Techniques, 27, 198 –219.

Bergl, R. A., & Vigilant, L. (2007). Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (*Gorilla gorilla diehli*). Molecular Ecology, 16(3), 501-516.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. American Anthropologist, 75(1), 214-242.

Bertrand, J.A.M., Bourgeois, Y.X.C., Delahaie, B., Duval, T., García-Jiménez, R., Cornuault, J., Heeb, P., Milá, B., Pujol, B. & Thébaud, C. (2014). Extremely reduced dispersal and gene flow in an island bird. Heredity, 112(2), 190-196.

BirdLife International and Handbook of the Birds of the World. (2021). Bird species distribution maps of the world. Version 2021.1. Available at http://datazone.birdlife.org/species/requestdis.

Blaise, B., Clague, J. J., & Mathewes, R. W. (1990). Time of maximum Late Wisconsin glaciation, west coast of Canada. Quaternary Research, 34(3), 282-295.

Blake Jr, W. (1970). Studies of glacial history in Arctic Canada. I. Pumice, radiocarbon dates, and differential postglacial uplift in the eastern Queen Elizabeth Islands. Canadian Journal of Earth Sciences, 7(2), 634-664.

Briner, J.P., Kaufman, D.S., Werner, A., Caffee, M., Levy, L., Manley, W.F., Kaplan, M.R. & Finkel, R.C., (2002). Glacier readvance during the late glacial (Younger Dryas) in the Ahklun Mountains, southwestern Alaska. Geology, 30(8), 679-682.

Brito, P. H., & Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica, 135, 439-455.

Brown, W., M., George, M., & Wilson, A., C., (1979). Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America. 76 (4): 1967-1971. 10.1073/pnas.76.4.1967.

Brumfield, R. T., Beerli, P., Nickerson, D. A., & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology & Evolution, 18(5), 249-256.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). Comparative phylogeography of northwestern North America: a synthesis. Special Publication-British Ecological Society, 14, 319-340.

Butlin, R. (1987). Speciation by reinforcement. Trends in Ecology & Evolution, 2(1), 8-13.

Clausen, J., Keck, D. D., & Hiesey, W. M. (1941). Regional differentiation in plant species. The American Naturalist, 75(758), 231–250. doi:10.1086/280955

Cracraft, J. (1983). Species concepts and speciation analysis. Current Ornithology, 1, 159-187.

Cronquist, A. (1978). Once again, what is a species? Biosystematics in agriculture. In Beltsville Symposia in Agricultural Research (Vol. 2, pp. 3-20).

Dalton, A. S., Stokes, C. R., & Batchelor, C. L. (2022). Evolution of the Laurentide and Innuitian ice sheets prior to the Last Glacial Maximum (115 ka to 25 ka). Earth-Science Reviews, 224, 103875.

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

Dempsey, Z. W., Burg, T. M., & Goater, C. P. (2019). Found, forgotten, and found again: systematics and distribution of Cooper's Rocky Mountain snail (*Oreohelix cooperi*) on a sky island in the Canadian Prairies. Canadian Journal of Zoology, 97(9), 833-840.

Duckworth, R. A., & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. Proceedings of the National Academy of Sciences, 104(38), 15017-15022.

Duckworth, R. A., & Semenov, G. A. (2017). Hybridization associated with cycles of ecological succession in a passerine bird. The American Naturalist, 190(4), E94-E105.

Dohms, K. M., Graham, B. A., & Burg, T. M. (2017). Multilocus genetic analyses and spatial modeling reveal complex population structure and history in a widespread

resident North American passerine (*Perisoreus canadensis*). Ecology and Evolution, 7(23), 9869-9889.

Dyke, A. S. (2004). An outline of North American deglaciation with emphasis on central and northern Canada. Developments in Quaternary Sciences, 2, 373-424.

Dyke, A. S., Andrews, J. T., Clark, P. U., England, J. H., Miller, G. H., Shaw, J., & Veillette, J. J. (2002). The Laurentide and Innuitian ice sheets during the last glacial maximum. Quaternary Science Reviews, 21(1-3), 9-31.

Dyke, A. S., & Evans, D. J. (2003). Ice-marginal terrestrial landsystems: northern Laurentide and Innuitian ice sheet margins. Glacial Landsystems, 24, 143-165.

Ehrlich, P. R. (1961). Has the biological species concept outlived its usefulness? Systematic Zoology, 10(4), 167-176.

England, J., Atkinson, N., Bednarski, J., Dyke, A. S., Hodgson, D. A., & Cofaigh, C. Ó. (2006). The Innuitian Ice Sheet: configuration, dynamics and chronology. Quaternary Science Reviews, 25(7-8), 689-703.

Freeland, J. R., Kirk, H., & Peterson, S.D. (2012). Molecular Ecology (Second Edition). Chichester, West Sussex, UK: Wiley-Blackwell.

Gibbard, P. L., & Head, M. J. (2009). IUGS ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. Quaternaire. Revue de l'Association française pour l'étude du Quaternaire, 20(4), 411-412.

Golden, J. L., & Bain, J. F. (2000). Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packera* (Asteraceae) species in southwestern Alberta. Evolution, 54(5), 1566-1579.

Good, T. P., Ellis, J. C., Annett, C. A., & Pierotti, R. (2000). Bounded hybrid superiority in an avian hybrid zone: effects of mate, diet, and habitat choice. Evolution, 54(5), 1774-1783.

Gowaty, P. A. & Plissner, J. H. (2020). Eastern bluebird (*Sialia sialis*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

Graham, M. R., Santibáñez-López, C. E., Derkarabetian, S., & Hendrixson, B. E. (2020). Pleistocene persistence and expansion in tarantulas on the Colorado Plateau and the effects of missing data on phylogeographical inferences from RADseq. Molecular Ecology, 29(19), 3684-3701.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y. & Hansen, N.F. (2010). A draft sequence of the Neandertal genome. science, 328(5979), 710-722.

Guinan, J. A., P. A. Gowaty, & E. K. Eltzroth (2020). Western bluebird (*Sialia mexicana*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.wesblu.01

Haecker, F. W. (1948). A nesting study of the Mountain Bluebird in Wyoming. Condor 50:216–219.

Hayward, A., Morgan, J. D., & Edwards, D. (2012). Reviews; SNP discovery and applications in *Brassica napus*. Journal of Plant Biotechnology, 39(1), 49.

Herlugson, C. J. (1981). Nest site selection in Mountain Bluebirds. Condor, 83:252-255.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. Biological journal of the Linnean Society, 58(3), 247-276.

Holderegger, R., & Thiel-Egenter, C. (2009). A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. Journal of Biogeography, 36(3), 476-480.

Hsieh, P., Woerner, A. E., Wall, J. D., Lachance, J., Tishkoff, S. A., Gutenkunst, R. N., & Hammer, M. F. (2016). Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. Genome Research, 26(3), 291-300.

Hughes, P. D., Gibbard, P. L., & Ehlers, J. (2013). Timing of glaciation during the last glacial cycle: evaluating the concept of a global 'Last Glacial Maximum'(LGM). Earth-Science Reviews, 125, 171-198.

James, J. K., & Abbott, R. J. (2005). Recent, allopatric, homoploid hybrid speciation: the origin of *Senecio squalidus* (Asteraceae) in the British Isles from a hybrid zone on Mount Etna, Sicily. Evolution, 59(12), 2533-2547.

Jehan, T., & Lakhanpaul, S. (2006) Single nucleotide polymorphism (SNP)–methods and applications in plant genetics: A review. Indian Journal of Biotechnology. 4, 435-459.

Johnson, L. S. & Dawson, R. D. (2020). Mountain bluebird (*Sialia currucoides*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

Klicka, J., Voelker, G., & Spellman, G. M. (2005). A molecular phylogenetic analysis of the "true thrushes" (Aves: Turdinae). Molecular Phylogenetics and Evolution, 34(3), 486-500.

Klicka, J., & Zink, R. M. (1997). The importance of recent ice ages in speciation: a failed paradigm. Science, 277(5332), 1666-1669.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences, 86(16), 6196-6200.

Lachance, J., Vernot, B., Elbers, C.C., Ferwerda, B., Froment, A., Bodo, J.M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R. & Zhang, K. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell, 150(3), 457-469.

Lait, L. A., & Burg, T. M. (2013). When east meets west: population structure of a high-latitude resident species, the boreal chickadee (*Poecile hudsonicus*). Heredity, 111(4), 321-329.

Lane, J. (1969). Hybridism in the eastern and mountain bluebirds. Blue Jay, 27(1), 18-21.

Latch, E. K., Heffelfinger, J. R., Fike, J. A., & Rhodes Jr, O. E. (2009). Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic glacial refugia and postglacial recolonization. Molecular Ecology, 18(8), 1730-1745.

Leonard, J.A., Wayne, R.K., & Cooper, A. (2000). Population genetics of ice age brown bears. Proceedings of the National Academy of Sciences of the United States of America, 97, 1651–1654.

Mallet, J. (2005). Hybridization as an invasion of the genome. Trends in Ecology & Evolution, 20(5), 229-237.

Mayr, E. (1942). Birds collected during the Whitney South Sea Expedition. 48, Notes on the Polynesian species of *Aplonis*. The American Museum of Natural History; no. 1166.

McArthur, S. L., McKellar, A. E., Flood, N. J., & Reudink, M. W. (2017). Local weather and regional climate influence breeding dynamics of mountain bluebirds (*Sialia currucoides*) and tree swallows (*Tachycineta bicolor*): a 35-year study. Canadian Journal of Zoology, 95(4), 271-277.

Mengel, R. M. (1970). The North American Central Plains as an isolating agent in bird speciation. Pleistocene and recent environments of the central Great Plains, 279-340.

Milá, B., Wayne, R. K., Fitze, P., & Smith, T. B. (2009). Divergence with gene flow and fine-scale phylogeographical structure in the wedge-billed woodcreeper, *Glyphorynchus spirurus*, a Neotropical rainforest bird. Molecular Ecology, 18(14), 2979-2995.

Milot, E., Gibbs, H. L., & Hobson, K. A. (2000). Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Molecular Ecology, 9(6), 667-681.

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52(3), 263-277.

Omland, K. E., Tarr, C. L., Boarman, W. I., Marzluff, J. M., & Fleischer, R. C. (2000). Cryptic genetic variation and paraphyly in ravens. Proceedings of the Royal Society of London. Series B: Biological Sciences, 267(1461), 2475-2482.

Pearson, S. F., & Rohwer, S. (2000). Asymmetries in male aggression across an avian hybrid zone. Behavioral Ecology, 11(1), 93-101.

Porter, S. C., & Swanson, T. W. (1998). Radiocarbon age constraints on rates of advance and retreat of the Puget lobe of the Cordilleran ice sheet during the last glaciation. Quaternary Research, 50(3), 205-213.

Price, T. (2008). Speciation in birds. Roberts and Company.

Purushotham, C. B., & Robin, V. V. (2016). Sky island bird populations isolated by ancient genetic barriers are characterized by different song traits than those isolated by recent deforestation. Ecology and Evolution, 6(20), 7334-7343.

Qvarnström, A., Ålund, M., McFarlane, S. E., & Sirkiä, P. M. (2016). Climate adaptation and speciation: particular focus on reproductive barriers in *Ficedula* flycatchers. Evolutionary Applications, 9(1), 119-134.

Rand, A. L. (1948). Glaciation, an isolating factor in speciation. Evolution, 314-321.

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. Proceedings of the Royal Society B: Biological Sciences, 282(1804), 20142903.

Robins, J. H., Tintinger, V., Aplin, K. P., Hingston, M., Matisoo-Smith, E., Penny, D., & Lavery, S. D. (2014). Phylogenetic species identification in *Rattus* highlights rapid

radiation and morphological similarity of New Guinean species. PloS One, 9(5), e98002.

Rohwer, S., Bermingham, E., & Wood, C. (2001). Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. Evolution, 55(2), 405-422.

Rohwer, S., & Wood, C. (1998). Three hybrid zones between hermit and Townsend's warblers in Washington and Oregon. The Auk, 115(2), 284-310.

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Current Biology, 26(9), 1241-1247.

Slager, D. L., Epperly, K. L., Ha, R. R., Rohwer, S., Wood, C., Van Hemert, C., & Klicka, J. (2020). Cryptic and extensive hybridization between ancient lineages of American crows. Molecular Ecology, 29(5), 956-969.

Shafer, A. B., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology, 19(21), 4589-4621.

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution, 206, 353-373.

Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., & Mindell, D. P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution, 12(2), 105-114.

Stanton, D.W., Hart, J., Galbusera, P., Helsen, P., Shephard, J., Kümpel, N.F., Wang, J., Ewen, J.G. & Bruford, M.W. (2014). Distinct and diverse: range-wide phylogeography reveals ancient lineages and high genetic variation in the endangered okapi (*Okapia johnstoni*). PLoS One, 9(7), p.e101081.

Steppan, S. J., Akhverdyan, M. R., Lyapunova, E. A., Fraser, D. G., Vorontsov, N. N., Hoffmann, R. S., & Braun, M. J. (1999). Molecular phylogeny of the marmots (Rodentia: Sciuridae): tests of evolutionary and biogeographic hypotheses. Systematic Biology, 48(4), 715-734.

Strong, W.L. & Hills, L.V. (2005), Late-glacial and Holocene palaeovegetation zonal reconstruction for central and north-central North America. Journal of Biogeography, 32: 1043-1062. https://doi.org/10.1111/j.1365-2699.2004.01223.x

Thompson, S. L., & Whitton, J. (2006). Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (*Townsendia hookeri*). Molecular Ecology, 15(11), 3389-3400.

Tremblay, N. O., & Schoen, D. J. (1999). Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. Molecular Ecology, 8(7), 1187-1198.

van Els, P., Cicero, C., & Klicka, J. (2012). High latitudes and high genetic diversity: phylogeography of a widespread boreal bird, the gray jay (*Perisoreus canadensis*). Molecular Phylogenetics and Evolution, 63(2), 456-465.

Vanden Abeele, S., Janssens, S. B., Piñeiro, R., & Hardy, O. J. (2021). Evidence of past forest fragmentation in the Congo Basin from the phylogeography of a shade-tolerant tree with limited seed dispersal: *Scorodophloeus zenkeri* (Fabaceae, Detarioideae). BMC Ecology and Evolution, 21(1), 1-12.

Vanden Abeele, S., Matvijev, K., Hardy, O.J., Assumani, D.M., Angoboy Ilondea, B., Beeckman, H., Bouka, G.U., Boupoya, C.A., Deklerck, V., Flot, J.F. and Gillet, J.F., 2023. Genetic breaks caused by ancient forest fragmentation: phylogeography of *Staudtia kamerunensis* (Myristicaceae) reveals distinct clusters in the Congo Basin. Tree Genetics & Genomes, 19(3), p.27.

Veen, T., Faulks, J., Tyler, F., Lloyd, J., & Tregenza, T. (2013). Diverse reproductive barriers in hybridising crickets suggests extensive variation in the evolution and maintenance of isolation. Evolutionary Ecology, 27, 993-1015.

Voelker, G., & Klicka, J. (2008). Systematics of *Zoothera* thrushes, and a synthesis of true thrush molecular systematic relationships. Molecular Phylogenetics and Evolution, 49(1), 377-381.

Wall, J.D., Ratan, A., Stawiski, E., Kim, H.L., Kim, C., Gupta, R., Suryamohan, K., Gusareva, E.S., Purbojati, R.W., Bhangale, T. & Stepanov, V. (2019). Identification of African-specific admixture between modern and archaic humans. The American Journal of Human Genetics, 105(6), 1254-1261.

Ward, B. C., Wilson, M. C., Nagorsen, D. W., Nelson, D. E., Driver, J. C., & Wigen, R. J. (2003). Port Eliza cave: North American West Coast interstadial environment and implications for human migrations. Quaternary Science Reviews, 22(14), 1383-1388.

Wang, S., Rohwer, S., Delmore, K., & Irwin, D. E. (2019). Cross-decades stability of an avian hybrid zone. Journal of Evolutionary Biology, 32(11), 1242-1251.

Wang, X. R., & Szmidt, A. E. (1994). Hybridization and chloroplast DNA variation in a Pinus species complex from Asia. Evolution, 48(4), 1020-1031.

Wielstra, B. (2019). Historical hybrid zone movement: more pervasive than appreciated. Journal of Biogeography, 46(7), 1300-1305.

Wilson, J.S. & Pitts, J.P. (2010) Phylogeographic analysis of the nocturnal velvet ant genus *Dilophotopsis* (Hymenoptera:Mutillidae) provides insights into diversification in the Nearctic deserts. Biological Journal of the Linnean Society, 101, 360–375.

Wright, S. (1943). Isolation-by-distance. Genetics, 28(2), 114.

Figure 1.1 A map of the glaciated North America ~25,000 years before present at the start of LGM (shp files from Dalton et al., 2022). Major refugia are labeled and circled to outline approximate boundaries. Ice free corridor in southwestern Alberta marked with blue hash lines based on Shafer et al. (2010). Contemporary geographic features that may have been refugia for mesic species are labeled and shaded in purple.

Figure 1.2 Male breeding plumage of each bluebird (a) mountain (b) eastern (c) western. Western bluebird photo obtained from wikicommons.

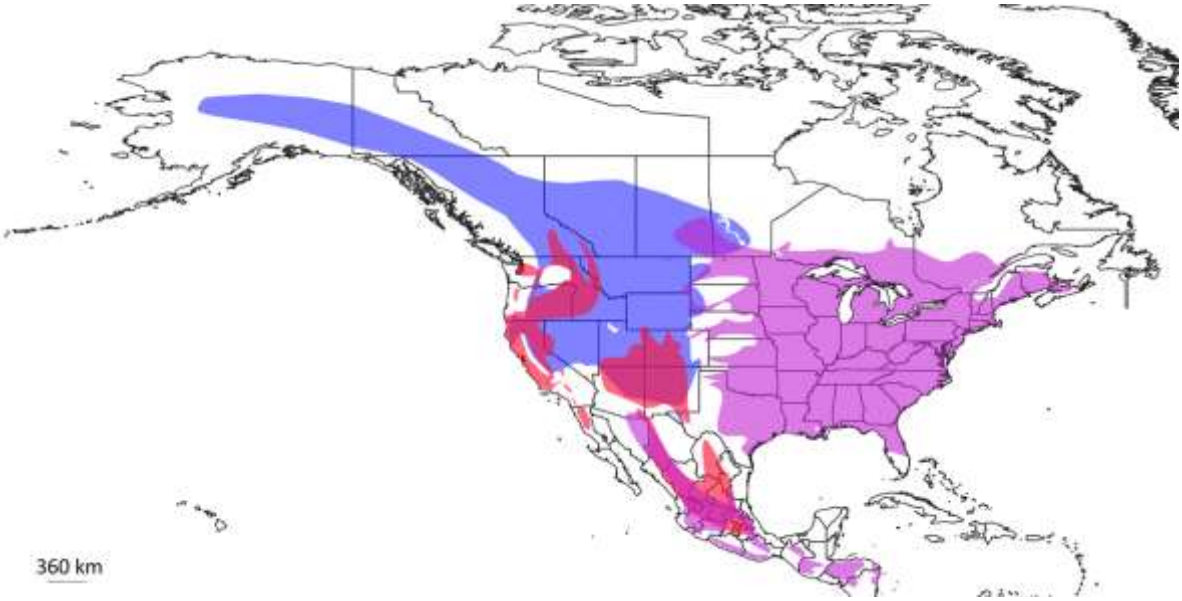

Figure 1.3 Breeding ranges of *S. mexicana* (red), *S. currucoides* (blue) and *S. sialia* (purple). Based on shp files provided by BirdLife International and Handbook of the Birds of the World (2021).

Chapter 2: Neutral markers reveal subtle population structure within a widespread songbird

Abstract

Understanding how both contemporary and historical physical barriers influence gene flow is key to understanding how resilient a species is to future disruptions. Mountain bluebirds are sexually dimorphic songbirds endemic to western North America. Within their large breeding range, mountain bluebirds experience a wide variety of environmental conditions and elevations that may disrupt gene flow and create population structure. Using single nucleotide polymorphisms obtained through restriction site-associated DNA sequencing, I determined the presence of at least four distinct breeding groups. Based on this structure, I then determined which factors may be responsible for the reduced connectivity and the overall history of each population going back to the last glacial maximum. Finally, I identified five candidate genes under balancing selection and three under diversifying selection. I provided valuable insights into the connectivity and genetic variation present in mountain bluebirds across most of their range.

2.1 Introduction

Population genetics techniques are an important tool for conservation as it allows us to distinguish breeding populations and determine their level of connectivity. To ascertain breeding population delineations, we can examine genetic variation both between and within populations by measuring the temporal and spatial changes in allele frequency (Moore, 1977). For many species, the most influential factor influencing changes in allele frequency is the dispersal and connectivity of populations (Slatkin, 1985). A positive relationship between geographic distance and genetic distance is often seen in nature (Vekemans & Hardy, 2004). Populations separated by long distances and low connectivity, for instance, may find it difficult or impossible to exchange alleles, a process known as isolation-by-distance (Wright, 1943). Consequently, individuals that are closer to each other should also be the most genetically related.

Isolation-by-distance is not the only physical barrier to gene flow, however. Reduced gene flow in larger and more vagile species can occur when their range is divided by geographic features such as mountain ranges, fragmented habitat, or water bodies (see Vignieri, 2005; Sánchez-Montes et al., 2018). Once separated, stochastic events may eliminate alleles in some isolated populations but not others, leading to genetic drift. Selective pressures experienced by different populations may also vary between barriers, leading to local adaptations (e.g., Savolainen et al., 2007; Bradshaw & Holzapfel, 2001). Geographic features are also typically longer lived, adding the potential of not only contemporary effects on gene flow, but historic as well. One of the most

influential barriers to gene flow in many contemporary species was the last glacial maximum (LGM).

Starting ~ 2.58 million years before present (myBP), the world shifted towards a period of global cooling and glacial expansion known as the Quaternary (Gibbard et al., 2009). The Quaternary can be split between two epochs; the Pleistocene (starting ~ 2.58 myBP) and the Holocene (starting ~11.7 thousand years before present (kyBP) (Gibbard et al., 2009). For the duration of the Pleistocene, the planet experienced repeated cycles of glacial expansion (known as glacial maxima) and interglacial periods. By ~25 kyBP, the earth was locked in the last glacial maximum (LGM) with North America being dominated by three expansive ice sheets (Dalton et al., 2023). The largest of the three—the Laurentide icesheet—stretched from the Atlantic Coast of Canada, down into the Great Lakes, and the west to the Rockies (Dyke et al., 2002; Dyke, 2004; Dalton et al., 2023). In the North, the Laurentide icesheet coalesced with the Innuitian icesheet which covered the Queen Elizabeth Islands in the far north of Nunavut and the Northwest Territories (Dalton et al., 2022). West of the Rockies, the Cordilleran ice sheet ran the coast of Alaska, down the Alexander Archipelago, partially covering Haida Gwaii and Vancouver Island, and stretched into central Washington (Blaise et al., 1990). Terrestrial species in the north faced a binary choice; either move to more suitable habitat or face annihilation.

As the climate continued to cool, many species were forced to retreat into small pockets of suitable habitat known as refugia (Haffer, 1969), where individual population became isolated, undergoing independent evolution. In central Alaska, strong molecular evidence from multiple boreal species supports a refugia known as Beringia (Abbott et

al., 2000; Abbott & Comes, 2004; Anderson et al., 2006; van Els et al., 2012; Lait & Burg, 2013; Leonard et al., 2000; Steppan et al., 1999; Thompson & Whitton, 2006; Tremblay & Schoen, 1999). Further south, the Pacific Northwest shows evidence of multiple refugia (Barrowclough et al., 2004; Brunsfeld et al., 2007; Shafer et al., 2010; Soltis et al., 1997). Taxa in the Pacific Northwest typically split along a north/south axis (Soltis et al., 1997; Brunsfeld et al., 2007) or an east/west split along the Rockies and Coast/Cascade Mountains (e.g., Barrowclough et al., 2004). Although glaciated during the LGM, ice free corridors east of the Rockies in Alberta likely existed (Bednarski & Smith, 2007; Dalton et al., 2023; Dyke et al., 2003; Shafer et al., 2010). Molecular studies suggest that the area may have acted as a cryptic refugia for some species in the far southwest of Alberta (Beatty & Provan, 2010; Golden & Bain, 2000). In the southwestern interior of the continent, the xeric conditions of today were likely confined to low elevations, allowing the spread of mesic species at higher elevations (Graham et al., 2020; Roberts & Hamann, 2015). Boreal tree species, such as trembling aspen (*Populus tremuloides*), lodgepole pine (*Pinus contorta*), white spruce (*Picea glauca*), and black spruce (*Picea mariana*) likely inhabited what is now the Great Plains and the southwest tablelands of Colorado and New Mexico (Roberts & Hamann, 2015).

Around 19 kyBP, the eastern ice sheets began to retreat (Dyke et al., 2003). The Cordilleran ice sheet in the west followed soon after ~16.5 kyBP (Porter & Swanson, 1998). Newly opened habitat was quickly seized upon by taxa expanding from their former refugia. Some taxa were able to colonize so rapidly that secondary contact occurred and admixing of multiple Pleistocene lineages is evident despite their separation during the LGM (Latch et al., 2009; Omland et al., 2000). While well-studied in some

species (e.g., black-capped chickadees (*Poecile atricapillus*) (Hindley et al., 2018), others have received little attention on the matter. The current study used contemporary population genetics techniques to elucidate possible glacial refugia for one such songbird found throughout western North America today.

The mountain bluebird (*Sialia currucoides*) is partially migratory songbird within the family Turdidae. Mature males are clad in vibrant azure blue plumage on their head and back, sky blue plumage on their chests and upperparts, and white underparts (Figure 2.1b). Females have more variation, being gray on the back, with blue tinges to both their wing and tail feathers, and either gray or rufous chests (Figure 2.1c). Mountain bluebirds are secondary cavity nesters, preferring dry open woodland, often at higher elevations, with short grasses or other forms of low-lying vegetation (Johnson & Dawson, 2020). As such, breeding birds have a high affinity for ecoregions such as aspen parkland, scrubland, and coniferous forest opened by intense wildfire activity (Johnson & Dawson, 2020). Mountain bluebirds can be adaptable in their habitat requirements; provided a cavity is available, they will even nest in badland and tundra environments within cliff faces and rock formations (Johnson & Dawson, 2020). With the widespread adoption and maintenance of nest boxes, mountain bluebirds have also expanded into grassland and savannah-like habitats (Johnson & Dawson, 2020). Prior to colonization by Europeans, mountain bluebirds were believed to be rare in the prairies (Johnson & Dawson, 2020). Their current breeding range now occupies a large majority of western North America—as far north as Alaska, east to central Manitoba, and south to New Mexico (Johnson & Dawson, 2020). Mountain bluebirds have relatively high fidelity to their breeding sites, with most adults returning consistently to the same site year after year to breed

(Herlugson, 1981; RDD et al., unpublished, found in Johnson & Dawson, 2020). Natal philopatry, however, is rare. In Indian Head, Saskatchewan, out of over 1,000 nestlings banded during the 1960s and 70s, Scott (1987) found that < 1% returned as breeding adults. RDD et al. (unpublished, found in Johnson & Dawson, 2020) monitored the dispersal of nestlings from a site in central British Columbia. Here, males were found to have a mean dispersal distance of 2.4 km ± 1.9 SD while females travelled 3.0 km ± 2.2 SD on average.

The primary aim of this study was to determine the extent of population structure within mountain bluebird breeding populations. Based on isolation-by-distance, species are predicted to show positive correlations between geographic distance and genetic distance. Mountain bluebirds have an expansive range (Figure 2.1a) and most adults return to the same breeding site with relatively little movement between breeding seasons (Herlugon, 1981; Dawson et al., unpublished, found in Johnson & Dawson, 2020). While young birds rarely return to their natal territory, they typically disperse within a short distance (≤ 3 km) from it (Dawson et al., unpublished, found in Johnson & Dawson, 2020; Scott, 1987). Thus, I predicted that population structure due to isolation-by-distance. The second aim was to determine if contemporary and past geographic barriers create discontinuity in gene flow. Past barriers to gene flow, such as the glaciation of North America and resulting fragmentation of suitable habitat, should create discrete breeding groups. Other potential barriers continue to persist today (e.g., the Rocky Mountains, discontinuous habitat) and should have the same effect of disrupting the movement of individuals. I predict that these barriers will also impact gene flow between breeding groups. In population genetics, neutral markers are typically used as they

theoretically do not respond to selective pressures and are therefore good proxies for the demographic and the evolutionary history of a species (Luikart et al., 2003). However, evidence suggests that non-neutral markers responding to adaption may provide important insights into processes like local adaptation (Kirk & Freeland, 2011). My last goal, if population structure was present, was to identify some of the potential candidate genes that may contribute to separation. To my knowledge, is the first study of its kind to analyze the genetic population structure of mountain bluebirds.

2.2 Methods

2.2.1 Sample collection and DNA extractions

Fieldwork occurred from mid-May through late June of 2022 to coincide with the mountain bluebird breeding season (Johnson & Dawson, 2020). Birds were located at nest boxes and the location of sampled birds was noted using a GPS. To sample the birds, I used song playback to entice individuals out their next boxes while mist nests were set on either side. Once caught, each individual was visually assessed for injury or signs of acute stress. In cases of stress or injury, I pulled a rectrix and the individual was promptly released. For non-stressed individuals, I collected biometrics (e.g., body mass, tarsus length, bill length, wing chord, etc.) and a ~ 50 µl blood sample via brachial venipuncture. I also collected back up samples using filter paper. Both samples were then transferred to 99% ethanol for storage at ambient temperature, then transferred to -20ºC upon return to the lab. In total 82 mountain bluebirds were sampled from four locations in Alberta and one from Saskatchewan. An additional 35 samples were acquired from nest

boxes around southern Alberta provided by the Mountain Bluebird Trail Society. Dr. Ken Otter and Dr. Matthew Reudink provided 10 samples a piece from two sites in British Columbia. Finally, muscle tissue samples from outside of Canada were obtained from the Denver Museum of Nature and Science and the University of Washington Burke Museum adding 15 and 96, respectively.

Of the 248 samples acquired, 157 were selected based on quality and geographic representation for restriction enzyme-associated sequencing. In total, the 157 samples represented 14 geographic sites from across western North America (Figure 2.1a). DNA was then extracted from the samples using a modified salting-out DNA extraction protocol (Miller et al., 1988).

2.2.2 RADseq

A form of ddRADseq (double digest restriction enzyme-associated DNA sequencing) was used and modified to include a third enzyme (Peterson et al., 2012). The extracted DNA samples were digested using the restriction enzymes *MspI* (4 base-pair cutter), *NsiI* (6 base-pair cutter) and *PstI* (6-base pair cutter) following Abed et al. (2019). Library preparation occurred at Université Laval and samples were then sequenced on an Illumina NovaSeq 6000 with paired end reads at Génome Québec.

2.2.3 Data processing and filtering

Following sequencing, the raw reads were analyzed with Fastqc/0.11.5, generating quality reports for each forward and reverse read to ensure successful

sequencing and spot any major errors (Andrews, 2011). Raw reads obtained from Génome Québec were then demultiplexed with Sabre/1.00, assigning reads to individuals. Finally, Cutadapt (Martin, 2011) was used to remove adapter and barcode sequences.

Reads were then run through Stacks/2.3e, a program designed for restriction-enzyme based data, to genotype and call variant sites for each individual (Catchen et al., 2013). An annotated reference genome for the Swainson's thrush (*Catharus ustulatus*) (Termignoni-Garcia et al., 2022) was available at chromosome level of assembly, so the reference-based pipeline was used in Stacks. The reference genome was first indexed using samtools/0.1.2 (Danecek et al., 2021). Next, alignment to the reference genome of forward and reverse reads was achieved using the Burrows-Wheeler Alignment Tool (bwa) while also marking low quality alignments as secondary (Li & Durbin, 2009). Samtools/0.1.2 was used again to sort and index the resulting binary alignment map (bam) files produced through alignment. Variants were then called using default parameters in the Stacks/2.3e ref_map pipeline to sort and align the files.

A variant call format (vcf) file was subsequently produced using the populations function within the reference-based pipeline in Stacks/2.3e. VCFtools/0.1.16 (Danecek et al., 2011) was then used to filter the vcf files for both SNPs and individuals. SNPs with more than 10% missing data were removed, following the subsequent removal of individuals with more than 30% missing data prior to downstream analyses. The filtered dataset had 53,501 linked SNPs and 80 individuals.

The populations function in Stacks was then used a second time with additional parameters, namely the *write-single-snp* command and a minimum minor allele

frequency (min-maf) of 0.05. The *write-single-snp* command randomly selects one when multiple SNPs are present within the same locus of the same individual, while the min-maf is applied to the metapopulation and selects sites that meet or exceed the minimum minor allele threshold. Selecting a single SNP per locus reduces redundancy within the data and accounts for potential linkage disequilibrium. Employing the min-maf parameter eliminates any one-off alleles that may arise due to sequencing error. Following the second round of filtering, 19,259 unlinked SNPs were left for the 80 individuals. As neutral markers are hypothesized to be equally affected by demography and the evolutionary history of a species (Luikart et al., 2003), an additional data set was produced using loci that were not under either directional or balancing selection. To find putatively neutral loci, Bayescan v2.1 (Foll & Gaggiotti, 2008) was used to verify the 19,259 SNPs and determined 6,998 loci to be neutral based off a log q-value. The other 12,258 markers were under balancing selection and three were under diversifying selection.

2.2.4 Population Analyses with neutral markers

Pairwise F_{ST} and the corresponding P-values were calculated in Arlequin v3.5.2.2 (Excoffier & Lisher, 2010). Arlequin was also used to calculate mean observed heterozygosity and expected heterozygosity for each population. A Benjamini-Hochberg correction was applied to the P-values to account for multiple comparisons (15 iterations) (Benjamini & Hochberg, 1995).

Next, I constructed an ancestry matrix using STRUCTURE 2.3.4 (Pritchard, 2000). I used a burnin value of 100,000 and a Markov-chain Monte Carlo (MCMC) value

of 300,000. I ran STRUCTURE for K=1-6, 10 iterations each. The subsequent results were collected and ran with STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt, 2012). Finally, the plots were visualized using a Clumpak server (Kopelman, 2015).

Lastly, a Principal Coordinate Analyses (PCoA) was performed to visualize genetic distances and potential population structure within the dataset. A genetic distance matrix was created using the pairwise genetic distances between each individual in Adegenet package in R-Studio/2021.09.0 (Jombart, 2011). The excel add-in, GenAlEx v6.5, was then used to visualize and create the PCoA based on genetic distance matrix calculated in Adegenet (Peakall & Smouse 2006, 2012).

2.2.5 Isolation-by-distance with neutral markers

To test for isolation-by-distance, a Mantel test for matrix correspondence was performed with 99 permutations using GenAlEx v6.5, ascertaining if there was a significant correlation ($P<0.05$) between geographic distance and genetic distance. Prior to the Mantel test, genetic distances had been calculated for the PCoA and were used again to test for isolation-by-distance. The pairwise geographic distance matrix between all 80 individuals was generated in GenAlEx v6.5 using sample location coordinates.

2.2.6 Phylogeography with neutral markers

To establish the phylogenetic relationships amongst the metapopulations, the SNAPP add-on (Bryant et al., 2012) was used to create a species tree in BEAST v2.4.0 (Bouckaert et al., 2014). Species trees are created in SNAPP using a coalescent model for

unlinked biallelic SNPs, a process which is very computationally demanding. A similar approach as Younger et al. (2016) was used to decrease the computational demand by randomly selecting two individuals per population. Mutation rates were calculated as part of the MCMC using the default parameters. The MCMCs were run for 5 million iterations and convergence was checked in Tracer v1.6 (Drummond & Rambaut, 2007). As outlined in the Tracer's user manual, an Effective Sample Size (ESS) >200 is an acceptable value. An ESS value of 369 was achieved. Five thousand and one maximum likelihood species trees were generated using TreeAnnotator v2.7.3 (Drummond & Rambaut, 2007) with a 10% burn-in using and median tree heights. Finally, Figtree v.1.4.4 (Rambaut, 2018) was used to visualize the resulting species tree.

2.2.7 Identifying causes of balancing selection

The largest subset of markers (12,296) was determined to be under balancing selection. To check for population structure in just the balancing markers, a PCoA was generated using GenAlEx. A multi-loci redundancy analysis (RDA) was performed using the R package vegan/2.6.4 (Oksanen et al., 2022). Three variables were used to model the impact of environment on the dataset. The first was mean temperature during the hottest quarter of the year as this is when birds are on the breeding ground. The second was mean precipitation levels during the warmest quarter of the year, when increased precipitation has been correlated with increased nest failure (McArthur et al., 2017). Both were obtained using the R package bioclim/0.3.0 (Booth et al., 2014). The final variable, elevation, was obtained using google earth coordinates from each sampling location based on lat and long. An RDA for each variable was generated, followed by varying

combinations of two environmental variables, and finally all three. Since RDAs do not allow for missing data, missing genotypes were filled in using the most common genotype noted (Forester et al., 2018). Each RDA was then plotted using ggplot2 (Wickham, 2016). Each RDA triplot used symmetrical scaling to scale the SNP and individual scores by the square root of the eigenvalues for easier visualization. Candidate genes for local adaptation were selected based on ordination space, having a standard deviation of 3.25 (p=0.001).

2.2.8 Candidate genes under diversifying selection

While separating the neutral markers from those under selection, Bayescan determined three SNPs to be under diversifying selection (counted as being in the top 99 percentile). The position of these SNPs was obtained through Rstudio/2023.04.21. The markers were then selected from the annotated Swainson's thrush reference genome to search for genes located in the corresponding areas.

2.3 Results

2.3.1 Population analyses with neutral markers

Pairwise F_{ST} (for neutral loci only) ranged from 0.013 to 0.079 (Table 2.1). All pairwise comparisons showed significance relative to each other. Following the Benjamini-Hochberg correction to reduce the rate of false positives, all comparisons remained significant. When mean observed and expected heterozygosity were calculated, western Alberta had the lowest at 0.136 and 0.162 while the Cypress Hills had the highest

at 0.255 and 0.355 respectively. Overall, the mean observed and expected heterozygosities were similar when compared within each population, with the notable exception of the Cypress Hills which differed by 0.1 (Table 2.2).

The PCoA appeared to support the results of the pairwise Fst values, clearly separating birds from western Alberta and the Cypress Hills into their own clusters (Figure 2.2a). The birds from Saskatchewan also formed a distinct cluster, although sample size was small (n=2). The individuals from the four remaining populations, southern British Columbia, the Pacific Northwest, Nevada and the southeastern Rockies, form a looser cluster; however, individuals from each population do cluster together. The first three PC accounted for similar amounts of variation with PC1 accounting for 1.92%, followed by the second and third accounting for 1.75% and 1.69% respectively.

Lastly, the ancestry matrix I generated using STRUCTURE did not show splitting at the initial population estimates (K=2, 3 ,4). The best clustering was K=6 (Figure 2.3b), as shown by the Pr(K) (Figure 2.3c). The Pacific Northwest, central California, and Nevada formed one cluster, while western Alberta, the Cypress Hills, and the southeastern Rockies each formed their own clusters and southern British Columbia and Saskatchewan showed evidence of admixture. When K=6 in STRUCTURE, southern British Columbia, western Alberta, the Cypress Hills, and southeastern Rockies each formed their own cluster while the Pacific Northwest and central California formed a single cluster. Nevada and Saskatchewan both showed evidence of admixture. Saskatchewan individuals had Q=0.5 with a fifth cluster and 0.2-0.25 with southeastern Rockies and 0.25-0.3 with western Alberta.

2.3.2 Isolation-by-distance with neutral markers

The results of a Mantel test for matrix correspondence were significant (P-value = 0.010). Therefore, geographic distance and genetic distance show a weak but positive correlation (R^2=0.0219, Figure 2.5). There were no individuals that were obvious outliers along the slope of the trendline.

2.3.3 Phylogeographic analysis with neutral markers

I visualized the phylogenetic relationships using an unrooted tree based on maximum likelihood probabilities. Early nodes are strongly supported, with posterior probability of 1 for both the first and second (Figure 2.6). Of the clades examined, western Alberta was the most distantly related, followed by the Cypress Hills. As the cladogram progresses, the posterior probabilities decrease slightly but remain high at 0.9991 reflecting the Pacific Northwest and southeastern Rockies complexes as sister to each other. The final topology yields southern British Columbia sister to the Pacific Northwest and the southeastern Rockies being sister to Nevada.

2.3.4 Population analyses with balancing markers

The PCoA generated using balancing markers failed to yield clear population structure (Figure 2.3.b). The first PC accounted for the most genetic variation between populations (1.46%), closely followed by the second and third accounting for 1.44% and 1.43% respectively.

2.3.5 Redundancy analyses using balancing markers

All RDA models using the three environmental variables failed to account for the variance within the markers under balancing selection. Seven markers associated with temperature were found to be above the strict 3.25 SD ($p \leq 0.001$) threshold. When visualized as a triplot, none of the models were able to resolve discrete structure between populations based on the environmental variables used, only the plot using all three variables seeming to provide a north/south split (Figure 2.4).

The three markers identified as being under diversifying selection and the seven balancing markers associated with temperature were manually indexed and determined using a subset of the Swainson's thrush annotated reference genome originally used for alignment (Table 2.3).

2.3.6 Candidate genes under balancing selection

Of the seven SNPs potentially responding to temperature, five were linked to known genes. Two SNPs were located on chromosome 6 at positions 21, 616, 044 and 52, 958, 759. The first was mapped within the Neurexin-3 gene and the second was within the SCUBE2 gene. The next marker was on chromosome 12 at position 13, 513, 798 and was within the SCAPER gene. Another marker was mapped to chromosome 17 at position 12, 393, 077 corresponding to the MYH7B gene. The final marker was located to the Z chromosome at position 14, 800, 422 within the MRPS30 gene.

2.3.7 *Candidate genes under diversifying selection*

The first diversifying SNP identified was located at position 39, 844, 386 on chromosome 1. When indexed, this SNP was determined to be within the NFKB inhibitor interacting Ras like 1 gene. The second SNP was located at position 468, 964 chromosome six and corresponds to the gene responsible for the isocitrate dehydrogenase (NAD(+)) 3 catalytic subunit alpha. The final marker was located at position 6, 768, 863 on the chromosome 11. This position was found to correspond to the aryl hydrocarbon receptor-like gene.

2.4 Discussion

2.4.1 Population structure and contemporary barriers to gene flow

Mountain bluebirds are widespread throughout discontinuous habitat across western North America (Johnson & Dawson, 2020). Based on their range, I predicted a positive relationship reflective of isolation-by-distance. Using a Mantel test, I found a significant but weak positive correlation between the geographic distance and genetic distance between mountain bluebird populations (P-value=0.010, R^2=0.0219). IBD may help to explain why sites with no apparent physical barriers other than distance (e.g., southern British Columbia and the Pacific Northwest) appear distinct in some analyses but not others. Southern British Columbia occasionally clustered with the Pacific Northwest in the PCoA and when I ran K=5 in STRUCTURE; when I ran K=6 and the pairwise FST, the two sites showed distinct structuring from each other. Given the

inconsistencies, southern British Columbia and the Pacific Northwest are best left as a complex at this time.

The weak positive correlation between genetic and geographic distances indicates other factors are inhibiting gene flow between populations. My second prediction stated that the Rocky Mountains and discontinuous habitat would also isolate breeding sites and contribute to population structuring. My results support these predictions. At least four distinct clusters appear supported by the results of the distance-based pairwise F_{st}, PCoA and ancestry matrices. The four consistently distinct populations were: the Pacific complex (composed of Washington, Oregon, Idaho, central California, and possibly British Columbia); western Alberta (consisting of southwestern Alberta and central Alberta); the Cypress Hills (southeastern Alberta); and the southeast Rockies (southeast Montana, Wyoming, and Colorado). Admixture between the Pacific complex and Nevada was also detected in the ancestry matrices. However, Nevada individuals were split in the PCoA between the Pacific Northwest and the southeastern Rockies without a clear barrier. Saskatchewan also formed a discrete cluster in the PCoA and ancestry matrices, but could not be included in the pairwise F_{st} due to its small sample size.

Out of four distinct clusters (the Pacific complex, western Alberta, the Cypress Hills, and the southeastern Rockies), three were east of the Rocky Mountains and one was west. The Pacific complex showed very little admixture with the eastern breeding groups (although there appeared to be slight gene flow with western Alberta at some point based on the PCoA and ancestry matrices). The Rocky Mountains restrict gene flow in other songbirds such as yellow warblers (*Setophaga petechia*) (Milot et al., 2000) and mountain chickadees (*Poecile gambeli*) (Spellman et al., 2007), as well as birds

considered to be highly mobile like red-tailed hawks (*Buteo jamaicensis*) (Hull et al., 2008). Although evidence supports mountain bluebirds breeding at higher elevations (~4,270 m) in the Rockies (Haecker, 1948), these are likely rare instances and not sustained breeding groups capable of connecting populations.

Interestingly, western Alberta and the Cypress Hills also appeared to be separate breeding clusters despite their relatively proximity to each other (< 300 km). The Cypress Hills are an elevated (~1250 m) region surrounded by semi-arid grasslands on all sides (Sauchyn, 1990). The aptly named sky islands are excellent habitat for breeding bluebirds with relic populations of lodgepole pine (*Pinus contorta*), trembling aspen (*Populus tremuloides*), and white spruce (*Picea glauca*) providing ample nesting sites (Sauchyn, 1990). It is likely that the xeric grassland surrounding the hills act as a strong barrier between the western Albertan birds and those in the Cypress Hills. Breeding attempts in the interim habitat are unlikely to be able to support large enough populations to bridge the gap in gene flow.

2.4.2 Glacial refugia, expansion and recolonization post-LGM

Like many species, mountain bluebirds were likely displaced from their northern range by the spread of the ice sheets during the LGM. The Pacific Northwest, Nevada, and the southeastern Rockies are likely the oldest populations of mountain bluebirds among those I sampled. Given their affinity for nesting in stands of trembling aspen (*Populus tremuloides*) (Johnson & Dawson, 2020), it seems logical that refugia suited for aspen may have doubled as refugia for mountain bluebirds. During the LGM, trembling

aspen had three main populations: one in the coastal cascades; a second spanning the east slope of the cascades, through to the Sierra Nevadas and the Northern Rocky Mountains; and a third along the eastern U.S. Rocky Mountains (Montana, Wyoming, and Colorado) (Bagley et al., 2020). The second and third populations of trembling aspen fit well with the structuring seen in mountain bluebirds and would have allowed for the connectivity in gene flow observed over long distances. The Sierra Nevada, eastern Cascades, and Rockies also provide elevation as preferred by mountain bluebirds.

The intermediate Nevada populations may also make more sense given other molecular studies. Nevada showed admixture with the Pacific complex in both ancestry matrixes and appeared partially cluster with both the Pacific and southeastern groups. The phylogeographic analysis, however, placed Nevada as the sister group to the southeastern Rockies. Mitochondrial evidence in grasshopper mice (*Onychomys spp.*) found populations in the Great Basin experienced an east/west split: western individuals were genetically similar to those in the Columbia Basin while eastern individuals within the Colorado Plateau grouped with the Wyoming Plateau and the Great Plains (Riddle & Honeycutt, 1990). Nevada may therefore represent the contact point where the Pacific complex comes into contact with birds from the southeastern Rockies.

The next oldest sites are likely western Alberta and the Cypress Hills breeding populations. During the LGM, Alberta was covered by the Laurentide icesheet, as was the rest of Canada and the northern United States. However, pollen profiles show that as early as 14 kyBP, the Cypress Hills and portions of southern Alberta were possibly dominated by boreal forest with either shrubland or a mixture of *Populus/Salix/Poaceae* parkland along the southern edge (Strong & Hills, 2005). Although mountain bluebirds

have a diverse habitat range today, they are completely absent in coniferous forests with dense canopy cover (Johnson & Dawson, 2020). Mountain bluebirds have a natural preference for aspen parkland, savannah-like habitat, and stands of forests that have been opened through wildfire regimes (Johnson & Dawson, 2020). Both western Alberta and the Cypress Hills were likely among some of the first locations in the north to be repopulated by expanding bluebird populations. Less than 2,000 years later, (12 kyBP) the area around the Cypress Hills became dominated by less suitable xeric grasslands, ultimately separating the sky island from the rest of the cordilleran forest in the west (Strong & Hill, 2005).

The Saskatchewan breeding birds may be the most nebulous of all. Due to their small sample size after filtering, only two of six individuals remained making most statistical analyses impossible due to the small sample size. In the PCoA, the two samples grouped closely together with each other but were distanced from the next closest location (the Cypress Hills in southeast Alberta). When K=5, Saskatchewan showed contributions from western Alberta, the southeastern Rockies and southern British Columbia. While genetic contributions from southern British Columbia seem unlikely due to distance, gene flow from western Alberta and the southeast Rockies are more plausible. Both birds sampled were from eastern Saskatchewan, which approaches the edge of the mountain bluebird's range. Mountain bluebirds were likely rare in parts of the province dominated by prairie prior to European colonization (Johnson & Dawson, 2020). As an edge population, it is possible the mountain bluebirds in Saskatchewan have seen multiple extinction and recolonization events as metapopulations blink in and out of existence. Scott and Lane (1977) recorded the movement of a female mountain bluebird

and possibly her mate 210 km from Saskatchewan into Manitoba (210 km from the edge of the range) during the breeding season after experiencing storm-related nest failure. Migrants from western Alberta and southeastern Rockies are therefore not impossible.

2.4.3 Genes of interest under balancing and directional selection

I detected around 12,200 markers under balancing selection—a signal associated with high allelic diversity (Fijarczyk & Babik, 2015). High allelic diversity may be maintained by factors including habitat heterogeneity leading to local adaptation, overdominance or heterozygote advantage, and frequency dependent selection (Fijarczyk & Babik, 2015). Studies on avian candidate genes have typically been focused on either migration (e.g., Chakarov et al., 2013; Peterson et al., 2013; Ramos et al., 2017) or plumage morphology (e.g, Chakarov et al., 2013; Walsh et al., 2012). Of the candidate genes I identified based on the associated SNPs (Table 2.3), only a few have received attention in bird-based study systems (primarily in domestic chickens (*Gallus gallus* var. *domesticus*)). The Neurexin-3 gene, for example, was associated with survivorship following outbreaks of the HPAI H7N2 virus in commercial laying chickens (Drobik-Czwarno et al., 2018). Despite being associated with habitat heterogeneity in mountain bluebirds, it is possible that diversity in this gene is maintained through overdominance or heterozygote advantage as part of an immune response, although as to how remains unclear. The MYH7B gene is involved in the development of embryonic chicken hearts, where is involved in the development of the myocardium (Warkman et al., 2012). This gene is conserved throughout vertebrate embryological development and is expressed at similar stages between organisms (Warkman et al., 2012). The MRPS30 gene has also been studied in chickens. A ubiquitous housekeeping gene, the highest activity of

MRPS30 occurs in the heart and muscles and codes for ribosomal subunit proteins (Davies et al., 2012). The expression levels of MRPS30 also remain relatively stable in birds undergoing heat stress (Cedraz de Oliveira et al., 2017; Gromboni et al., 2020). With mountain bluebirds experiencing a large range of temperatures during the breeding season, those in warmer environments are likely at greater risk of heat stress than individuals who are not. Maintaining the products of genes responsible for protein synthesis would be critical at high temperatures, thus variation within the overall population may be maintained by differences in breeding site temperature.

The three candidates under diversifying selection also yielded little study in avian model systems. The NFKB inhibitor interacting Ras like 1 gene is an orthologous gene not yet understood within humans (Murphy et al., 2006). It belongs to a family of immune system genes that are functionally inactive through inhibition until the phosphorylation of the inhibitor gene and subsequent cleavage (Yamamoto and Gaynor, 2004). Previous work in the red-headed bunting (*Emberiza bruniceps*) has suggested that NFKB genes play an important role in defense against the various pathogens encountered and transmitted during migration (Tiwari et al., 2023).

NAD(+) also has an ortholog in humans. NAD(+) encodes the alpha subunit protein of an isozyme of NAD(+)-dependent isocitrate dehydrogenase (Murphy et al 2006), though again its role in birds remains unknown. The last gene, aryl hydrocarbon-like receptor gene has been responsible for gene regulation of metabolic enzymes, immune response, stem cell maintenance, and cellular differentiation (Gutiérrez-Vázquez & Quintana, 2018).

2.5 Conclusions

Overall, I have provided the first analyses into the population structure of mountain bluebirds using RAD-seq techniques. By using a subset of neutral markers across the entire genome, I found evidence of population structure, with at least four groups experiencing different levels of gene flow. Isolation-by-distance appears to play a weak but significant role limiting gene flow between distant populations. Despite mountain bluebirds' tolerance for nesting at high elevations (Haecker, 1948), I also found evidence that the northern Rocky Mountains influence gene flow and connectivity between populations. However, the Rocky Mountains may break down as a barrier in more temperate climates such as the American southwest. Gene flow connectivity between mountain bluebirds on either side of the southern Rockies via Nevada suggests that valleys and other corridors exist between breeding groups. Likewise, unsuitable habitat also inhibits gene flow, as seen by the divisions between western Alberta and the Cypress Hills. During the LGM, mountain bluebirds would have been concentrated in areas of suitable habitat within the Pacific Northwest and Interior Southwest where open aspen parkland and subalpine forests were predominant. Areas farther north, such as Beringia and the putative southwestern Albertan refugia likely lacked suitable habitat for mountain bluebirds at the time. To get a better understanding of population structure within mountain bluebirds and their genetics, samples from the extreme north and extreme south, as well as additional samples from peripheral populations in Saskatchewan, are likely required as well as the addition of whole-genome sequencing techniques. Lastly, I identified five putative candidate genes under balancing selection and three under diversifying selection. In terms of balancing selection causes, I found

SNPs corresponding to immune response rather than those linked directly to environmental variables. Overall, my study has provided valuable insights into the connectivity and genetic variation present in mountain bluebirds across most of their range. This information may prove invaluable as populations of this dynamic bird continue to shift in the face of climate change.

2.6 Literature Cited

Abbott, R. J., & Comes, H. P. (2004). Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (purple saxifrage). New Phytologist, 161(1), 211-224.

Abbott, R. J., Smith, L. C., Milne, R. I., Crawford, R. M., Wolff, K., & Balfour, J. (2000). Molecular analysis of plant migration and refugia in the Arctic. Science, 289(5483), 1343-1346.

Abed A, Légaré G, Pomerleau S, St-Cyr J, Boyle B, Belzile FJ. 2019 Genotyping-by-Sequencing on the Ion Torrent Platform in Barley. Methods Molecular Biology. 1900:233-252.

Anderson, L. L., Hu, F. S., Nelson, D. M., Petit, R. J., & Paige, K. N. (2006). Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. Proceedings of the National Academy of Sciences, 103(33), 12447-12450.

Barrowclough, G. F., Groth, J. G., Mertz, L. A., & Gutiérrez, R. J. (2004). Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). Molecular Ecology, 13(7), 1911-1922.

Beatty, G. E., & Provan, J. I. M. (2010). Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. Molecular Ecology, 19(22), 5009-5021.

Bednarski, J. M., & Smith, I. R. (2007). Laurentide and montane glaciation along the Rocky Mountain Foothills of northeastern British Columbia. Canadian Journal of Earth Sciences, 44(4), 445-457.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statistical Society B, 57, 289– 300.

Blaise, B., Clague, J. J., & Mathewes, R. W. (1990). Time of maximum Late Wisconsin glaciation, west coast of Canada. Quaternary Research, 34(3), 282-295.

Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. Diversity and Distributions, 20(1), 1-9.

Bradshaw, W. E., & Holzapfel, C. M. (2001). Genetic shift in photoperiodic response correlated with global warming. Proceedings of the National Academy of Sciences, 98(25), 14509-14511.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). Comparative phylogeography of northwestern North America: a synthesis. Special Publication-British Ecological Society, 14, 319-340.

Cedraz de Oliveira, H., Pinto Garcia, A. A., Gonzaga Gromboni, J. G., Vasconcelos Farias Filho, R., Souza do Nascimento, C., & Arias Wenceslau, A. (2017). Influence of heat stress, sex and genetic groups on reference genes stability in muscle tissue of chicken. PloS One, 12(5), e0176402.

Chakarov, N., Jonker, R. M., Boerner, M., Hoffman, J. I., & Krüger, O. (2013). Variation at phenological candidate genes correlates with timing of dispersal and plumage morph in a sedentary bird of prey. Molecular Ecology, 22(21), 5430-5440.

Dalton, A. S., Stokes, C. R., & Batchelor, C. L. (2022). Evolution of the Laurentide and Innuitian ice sheets prior to the Last Glacial Maximum (115 ka to 25 ka). Earth-Science Reviews, 224, 103875.

Davies, S. M., Sanchez, M. I. L., Narsai, R., Shearwood, A. M. J., Razif, M. F., Small, I. D., Whelan, J., Rackham, O. & Filipovska, A. (2012). MRPS27 is a pentatricopeptide repeat domain protein required for the translation of mitochondrially encoded proteins. FEBS Letters, 586(20), 3555-3561.

Drobik-Czwarno, W., Wolc, A., Fulton, J. E., Arango, J., Jankowski, T., O'Sullivan, N. P., & Dekkers, J. C. M. (2018). Identifying the genetic basis for resistance to avian influenza in commercial egg layer chickens. Animal, 12(7), 1363-1371.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dyke, A. S. (2004). An outline of North American deglaciation with emphasis on central and northern Canada. Developments in Quaternary Sciences, 2, 373-424.

Dyke, A. S., Andrews, J. T., Clark, P. U., England, J. H., Miller, G. H., Shaw, J., & Veillette, J. J. (2002). The Laurentide and Innuitian ice sheets during the last glacial maximum. Quaternary Science Reviews, 21(1-3), 9-31.

Earl, Dent A. and vonHoldt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2), 359-361.

Fijarczyk, A., & Babik, W. (2015). Detecting balancing selection in genomes: limits and prospects. Molecular Ecology, 24(14), 3529-3545.

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. Molecular Ecology, 27(9), 2215–2233.

Gibbard, P. L., & Head, M. J. (2009). IUGS ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. Quaternaire. Revue de l'Association française pour l'étude du Quaternaire, 20(4), 411-412.

Golden, J. L., & Bain, J. F. (2000). Phylogeographic patterns and high levels of chloroplast DNA diversity in four *Packera* (Asteraceae) species in southwestern Alberta. Evolution, 54(5), 1566-1579.

Graham, M. R., Santibáñez-López, C. E., Derkarabetian, S., & Hendrixson, B. E. (2020). Pleistocene persistence and expansion in tarantulas on the Colorado Plateau and the effects of missing data on phylogeographical inferences from RADseq. Molecular Ecology, 29(19), 3684-3701.

Gromboni, J. G. G., Cedraz de Oliveira, H., Botelho Diniz Marques, D., Amândio Pinto Garcia Junior, A., Vasconcelos Farias Filho, R., Fernando Gromboni, C., Machado Souza, T, & Arias Wenceslau, A. (2020). Influence of heat stress on reference genes stability in heart and liver of two chicken genotypes. Plos One, 15(2), e0228314.

Gutiérrez-Vázquez, C. and Quintana F.J. (2018). Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. Immunity. 48 (1): 19–33.

Haffer, J. (1969). Speciation in Amazonian forest birds. Science, 165(3889), 131–137.

Haecker, F. W. (1948). A nesting study of the Mountain Bluebird in Wyoming. Condor 50:216–219.

Hindley, J., Graham, B. A., & Burg, T. M. (2018). Pleistocene glacial cycles and physical barriers influence phylogeographic structure in Black-capped Chickadees (*Poecile atricapillus*), a widespread North American passerine. Canadian Journal of Zoology, 96(12), 1366-1377.

Herlugson, C. J. (1981). Nest site selection in Mountain Bluebirds. Condor, 83:252-255.

Hull, J. M., Ernest, H. B., Harley, J. A., Fish, A. M., & Hull, A. C. (2009). Differential migration between discrete populations of juvenile Red-tailed Hawks (*Buteo jamaicensis*). The Auk, 126(2), 389-396.

Johnson, L. S. and Dawson, R. D. (2020). Mountain Bluebird (*Sialia currucoides*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

Kirk, H., & Freeland, J. R. (2011). Applications and implications of neutral versus non-neutral markers in molecular ecology. International Journal of Molecular Sciences, 12(6), 3966-3988.

Klicka, J., & Zink, R. M. (1997). The importance of recent ice ages in speciation: a failed paradigm. Science, 277(5332), 1666-1669.

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

Lait, L. A., & Burg, T. M. (2013). When east meets west: population structure of a high-latitude resident species, the boreal chickadee (*Poecile hudsonicus*). Heredity, 111(4), 321-329.

Latch, E. K., Heffelfinger, J. R., Fike, J. A., & Rhodes Jr, O. E. (2009). Species-wide phylogeography of North American mule deer (*Odocoileus hemionus*): cryptic

glacial refugia and postglacial recolonization. Molecular Ecology, 18(8), 1730-1745.

Leonard, J.A., Wayne, R.K., & Cooper, A. (2000). Population genetics of ice age brown bears. Proceedings of the National Academy of Sciences of the United States of America, 97, 1651–1654.

Lovette, I. J. (2005). Glacial cycles and the tempo of avian speciation. Trends in Ecology & Evolution, 20(2), 57-59.

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity. Molecular Ecology, 19(15), 3038-3051.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics, 4(12), 981–994.

McArthur, S. L., McKellar, A. E., Flood, N. J., & Reudink, M. W. (2017). Local weather and regional climate influence breeding dynamics of mountain bluebirds (*Sialia currucoides*) and tree swallows (*Tachycineta bicolor*): a 35-year study. Canadian Journal of Zoology, 95(4), 271-277.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, 16(3), 1215.

Moore, W. S. (1977). An evaluation of narrow hybrid zones in vertebrates. The Quarterly Review of Biology, 52(3), 263-277.

Murphy M, Brown G, Wallin C, et al. (2006) Gene Help: Integrated Access to Genes of Genomes in the Reference Sequence Collection. [Updated 2022 Nov 4]. In: Gene Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK3841/

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J. (2022). _vegan: Community Ecology Package_. R package version 2.6-4, <https://CRAN.R-project.org/package=vegan>.

Omland, K. E., Tarr, C. L., Boarman, W. I., Marzluff, J. M., & Fleischer, R. C. (2000). Cryptic genetic variation and paraphyly in ravens. Proceedings of the Royal Society of London. Series B: Biological Sciences, 267(1461), 2475-2482.

Peterson, M. P., Abolins-Abols, M., Atwell, J. W., Rice, R. J., Milá, B., & Ketterson, E. D. (2013). Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus *Junco*. F1000Research, 2.115-131

Porter, S. C., & Swanson, T. W. (1998). Radiocarbon age constraints on rates of advance and retreat of the Puget lobe of the Cordilleran ice sheet during the last glaciation. Quaternary Research, 50(3), 205-213.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Rambaut, A. (2018) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Riddle, B. R., & Honeycutt, R. L. (1990). Historical biogeography in North American arid regions: an approach using mitochondrial-DNA phylogeny in grasshopper mice (genus *Onychomys*). Evolution, 44(1), 1-15.

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. Proceedings of the Royal Society B: Biological Sciences, 282(1804), 20142903.

Sánchez-Montes, G., Wang, J., Arino, A. H., & Martínez-Solano, Í. (2018). Mountains as barriers to gene flow in amphibians: Quantifying the differential effect of a major mountain ridge on the genetic structure of four sympatric species with different life history traits. Journal of Biogeography, 45(2), 318-331.

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. Annual Reviews of Ecology, Evolution, and Systematics, 38, 595-619.

Schneider, R. R. (2013). Alberta's natural subregions under a changing climate: past, present, and future.

Scott, L., & Lane, J. (1974). Mountain Bluebird travels 130 miles to renest. Blue Jay 32, 44-45.

Shafer, A. B., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology, 19(21), 4589-4621.

Slatkin, M. 1985 Rare alleles as indicators of gene flow. Evolution 39 (1): 53-65.

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution, 206, 353-373.

Spellman, G. M., Riddle, B., & Klicka, J. (2007). Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to Quaternary climate change. Molecular Ecology, 16(5), 1055-1068.

Steppan, S. J., Akhverdyan, M. R., Lyapunova, E. A., Fraser, D. G., Vorontsov, N. N., Hoffmann, R. S., & Braun, M. J. (1999). Molecular phylogeny of the marmots (Rodentia: Sciuridae): tests of evolutionary and biogeographic hypotheses. Systematic Biology, 48(4), 715-734.

Stiles, D. (2014). Long distance Mountain Bluebird recoveries. Willet, 27, 2.

Strong, W.L. and Hills, L.V. (2005), Late-glacial and Holocene palaeovegetation zonal reconstruction for central and north-central North America. Journal of Biogeography, 32, 1043-1062.

Termignoni-Garcia, F., Kirchman, J. J., Clark, J., & Edwards, S. V. (2022). Comparative population genomics of cryptic speciation and adaptive divergence in Bicknell's and gray-cheeked thrushes (Aves: *Catharus bicknelli* and *Catharus minimus*). Genome Biology and Evolution, 14(1), evab255.

Thompson, S. L., & Whitton, J. (2006). Patterns of recurrent evolution and geographic parthenogenesis within apomictic polyploid Easter daises (*Townsendia hookeri*). Molecular Ecology, 15(11), 3389-3400.

Tiwari, J., Sur, S., Naseem, A., Rani, S., & Malik, S. (2023). Photoperiodic modulation of melatonin receptor and immune genes in migratory redheaded bunting. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 279, 111381.

Tremblay, N. O., & Schoen, D. J. (1999). Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. Molecular Ecology, 8(7), 1187-1198.

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. Molecular Ecology, 13(4), 921-935.

van Els, P., Cicero, C., & Klicka, J. (2012). High latitudes and high genetic diversity: phylogeography of a widespread boreal bird, the gray jay (*Perisoreus canadensis*). Molecular Phylogenetics and Evolution, 63(2), 456-465.

Vignieri, Sacha N. 2005Streams over mountains: influence of riparian connectivity on gene flow in the Pacific jumping mouse (*Zapus trinotatus*). Molecular Ecology 14 (7): 1925-1937.

Walsh, N., Dale, J., McGraw, K. J., Pointer, M. A., & Mundy, N. I. (2012). Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. Proceedings of the Royal Society B: Biological Sciences, 279(1726), 58-66.

Warkman, A. S., Whitman, S. A., Miller, M. K., Garriock, R. J., Schwach, C. M., Gregorio, C. C., & Krieg, P. A. (2012). Developmental expression and cardiac transcriptional regulation of Myh7b, a third myosin heavy chain in the vertebrate heart. Cytoskeleton, 69(5), 324-335.

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

Wright, S. 1943 Isolation-by-distance. Genetics 28 (2): 114.

Yamamoto, Y., & Gaynor, R. B. (2004). IκB kinases: key regulators of the NF-κB pathway. Trends in biochemical sciences, 29(2), 72-79.

Younger, J. L., Clucas, G. V., Kao, D., Rogers, A. D., Gharbi, K., Hart, T., & Miller, K. J. (2017). The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. Molecular Ecology, 26(15), 3883-3897.

Table 2.1 Pairwise F_{ST} visualized as a heatmap between all study populations with sample size ≥5. All values retained significant following Benjamini-Hochberg corrections for multiple comparisons (Benjamini & Hochberg, 1995). P-values <0.05 *, <0.001**, <0.0001***.

	SBC	PNW	NV	WAB	CYP	SER
SBC		***	***	***	**	***
PNW	0.186		***	***	**	***
NV	0.038	0.023		***	**	***
WAB	0.037	0.013	0.048		***	***
CYP	0.062	0.029	0.061	0.079		***
SER	0.029	0.015	0.03	0.026	0.045	

Table 2.2 Mean expected and observed heterozygosity across all study populations with sample size ≥5.

Population	**H_o**	**H_e**
SBC	0.21117	0.24945
PNW	0.19605	0.22592
NV	0.24929	0.29183
WAB	0.13617	0.16237
CYP	0.25495	0.35480
SER	0.19571	0.22510

Table 2.3 Three markers and their corresponding candidate genes noted as being under diversifying selection based on global Fst scores found in Bayescan.

Chromosome	SNP position	Fst	Candidate	Selection	Function in prior bird studies
1	39,844,386	0.44744	NFKB inhibitor interacting Ras like 1	Diversifying	Correlated to immune response during migration
6	468,964	0.29073	isocitrate dehydrogenase (NAD(+)) 3 catalytic subunit alpha	Diversifying	Unknown
11	6,768,863	0.29037	aryl hydrocarbon receptor-like	Diversifying	Unknown
6	21,616,044	0.030335	Neurexin-3	Balancing	Associated with immune response against avian flu
6	52,958,759	0.030329	SCUBE2	Balancing	Unknown
12	13,513,798	0.030325	SCAPER	Balancing	Unknown
17	12,393,077	0.030321	MYH7B	Balancing	Myocardial formation in embryos
Z	14,800,422	0.030329	MRPS30	Balancing	Ribosomal housekeeping gene

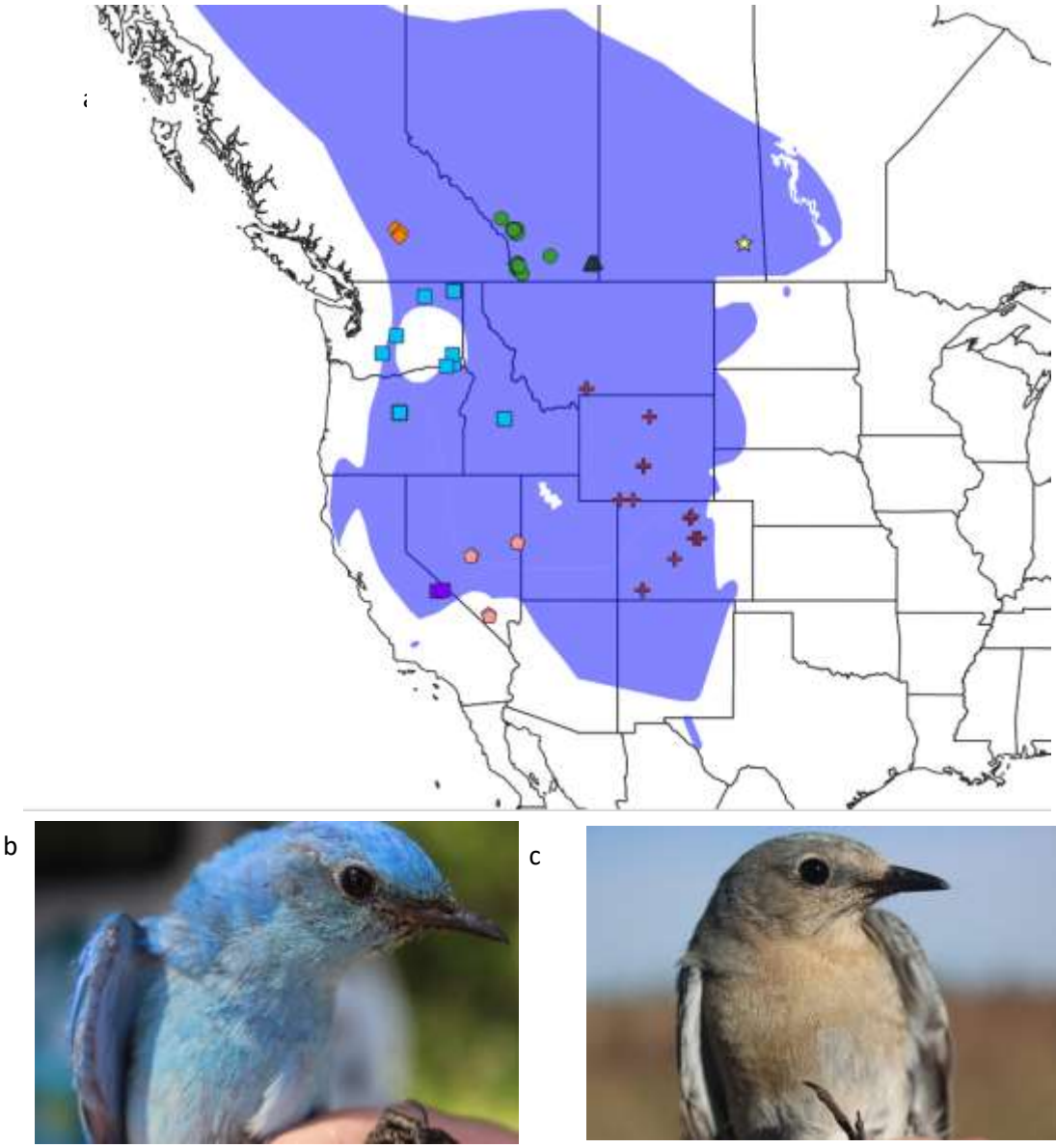

Figure 2.1 a) Map of North America overlayed with the breeding range of the mountain bluebird. Sample sites are those retained after filtering: SBC (orange diamonds), PNW (blue squares), CCA (purple squares), NV (peach pentagons), WAB (green circles), CYP (midnight blue triangles), SER (maroon crosses), and SK (yellow stars). **b)** Male bluebird sampled in the CYP. **c)** Female bluebird sampled in the CYP.

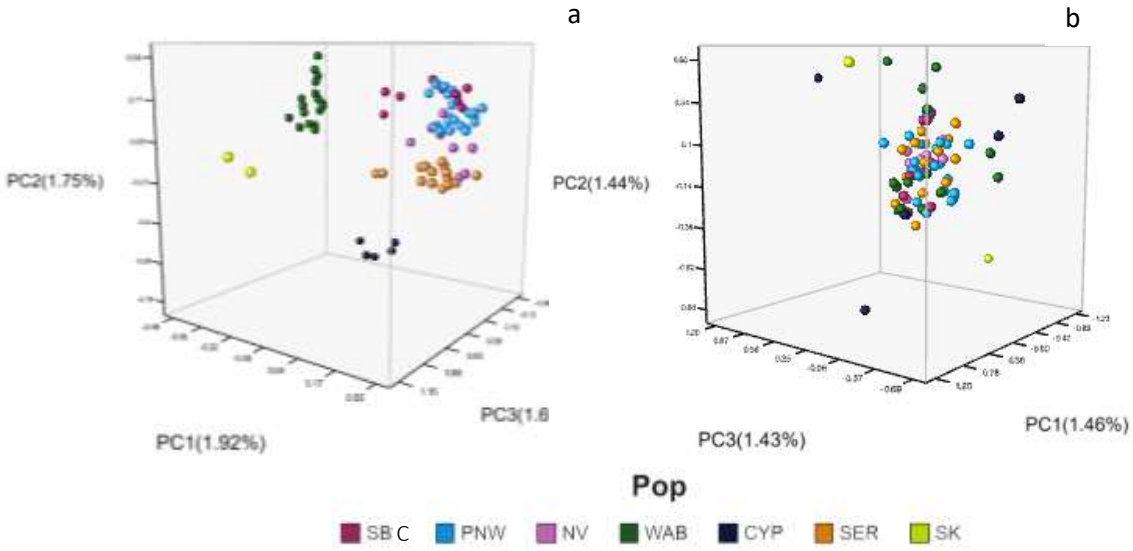

Figure 2.2 PCoA showing first three PCs accounting for the most variation a) Using neutral markers (n=6,998) from sites remaining after filtering. b) Using balancing markers (n=12,296) from sites remaining after filtering.

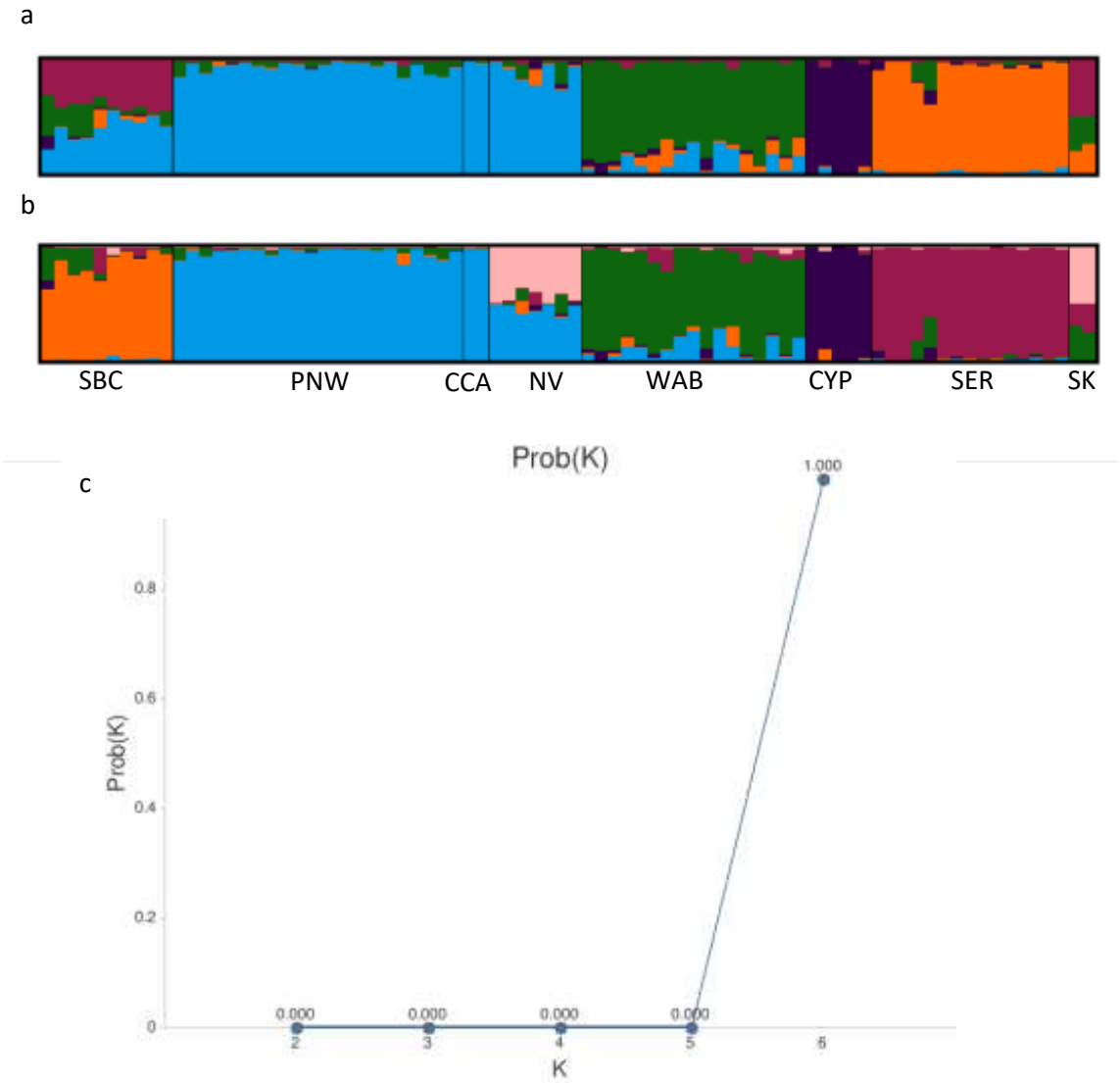

Figure 2.3 Ancestry matrices based on 80 individuals from 8 populations using 6,998 SNPs. a) STRUCTURE Plot (K=5) b) STRUCTURE plot (K=6). c) Best K as indicated by Pr(K).

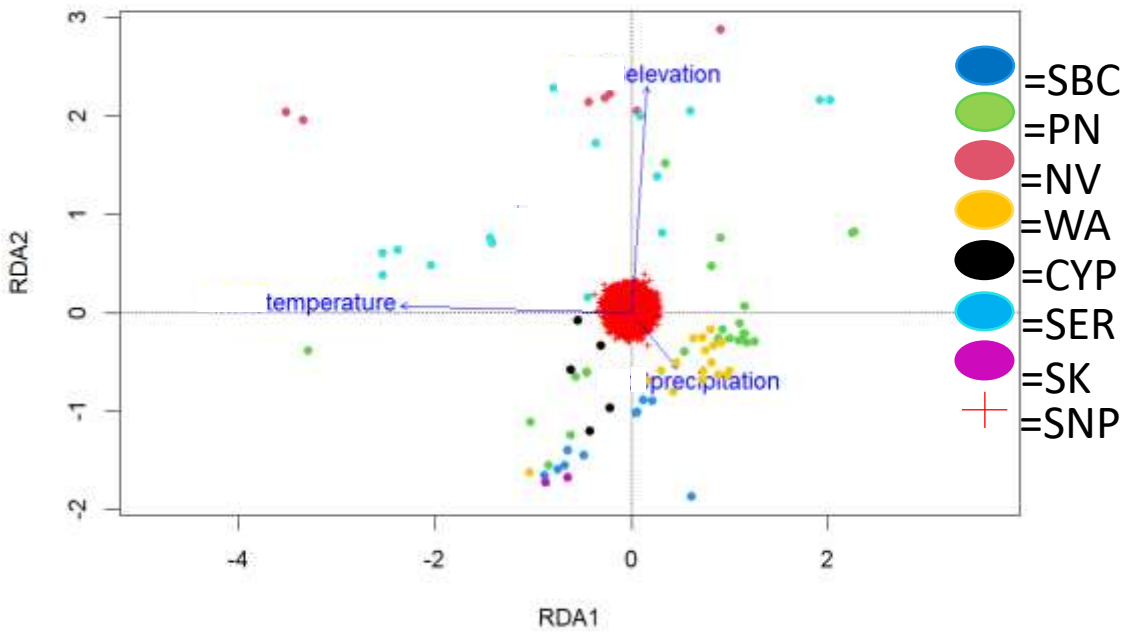

Figure 2.4 Triplot using all three environmental variables. Although the association between SNPs and the variables were not significant, using three in combination divided the northern populations from southern.

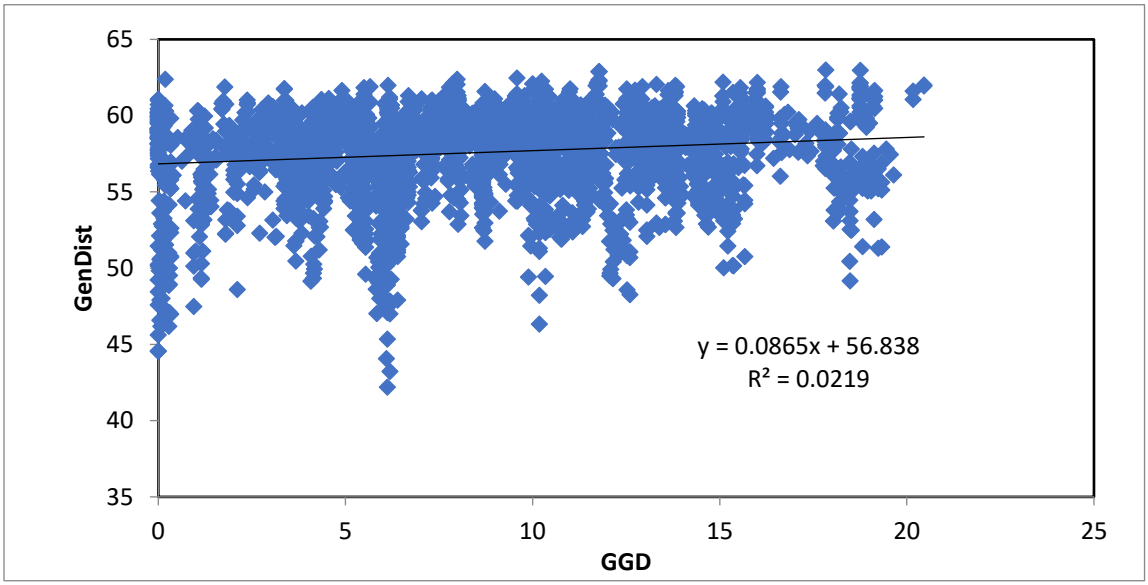

Figure 2.5 Mantel test for genetic distance as a response variable to geographic distance using neutral marker dataset.

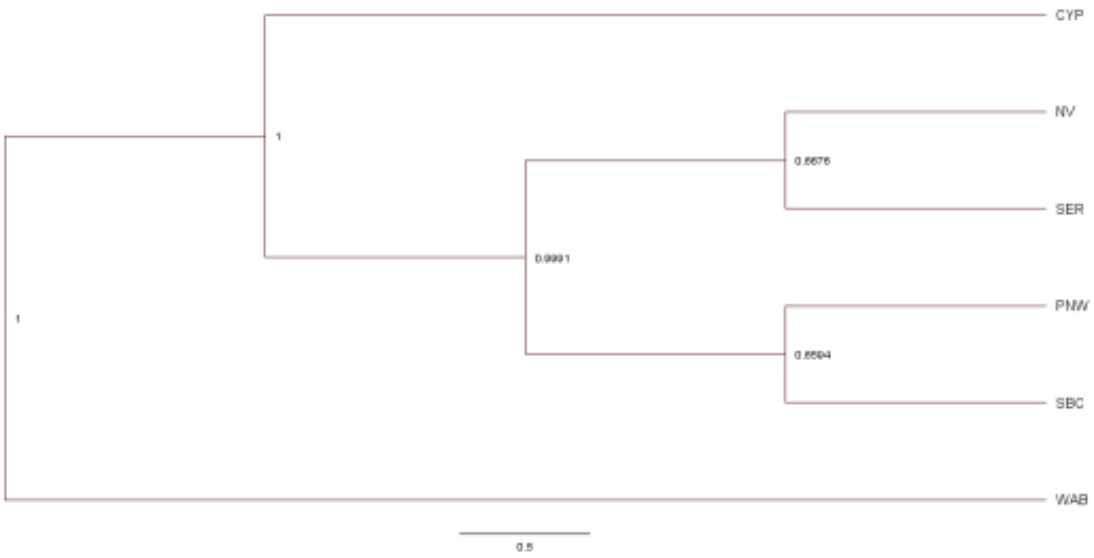

Figure 2.6 Maximum likelihood tree using median height from 4500 retained unrooted trees. Topology based on two random with bootstrap values generated for each node.

Chapter 3: RADseq and mitochondrial sequencing reveal clear mitonuclear discordance and signals of historic introgression within the genus *Sialia*

Abstract

Speciation is often discussed as a discrete process that results from the cessation of gene flow between lineages. However, growing evidence suggests that many species experience hybridization with congeners at some point in their evolutionary history. One such lineage is the North American bluebirds (genus *Sialia*). While consisting of only three species, the exact phylogenetic relationships within this small group remain unclear. In this study, I found signatures of historical hybridization amongst three lineages of bluebirds while failing to find contemporary hybrids. I then attempted to address the unresolved polytomy within the well-known North American genus using nucleic markers obtained through restriction enzyme-associated DNA sequencing. Although I could not determine the exact phylogeny using both mitochondrial and nuclear markers, I did find substantial evidence of mitonuclear discordance between the two sets of markers. I also found signs of incomplete lineage sorting within the mitochondrial genes themselves. Overall, bluebirds appear to show rapid speciation with continued gene flow through occasional hybridization.

3.1 Introduction

A species' ability to adapt and survive within their environment is reliant on genetic variation. Hybridization introduces novel genes or alleles of existing genes between populations and species, which may be beneficial under selective pressure (Mallet, 2007; Abbott et al., 2013). Hybridization is relatively common, with an estimated 25% of plant and 10% of animal species readily hybridizing, and hybrid offspring accounting for between 1 in 100 to 1 in 10,000 individuals (Mallet, 2005). Contemporary hybrid zones provide an incomplete picture for an organism's genetic history, as the effects of introgression on parental species can vary with time and space (Abbott et al., 2013). Increasing evidence of geographically small and short-lived hybrid zones, known as ephemeral hybrid zones, supports a changing dynamic across many hybrid zones over time (Hewitt, 1988; Arntzen & Wallis, 1991; Krosby & Rohwer, 2010; Stacy et al., 2016). Ephemeral hybrid zones have been linked to cases of range expansion or shifts in habitat (Pearson & Rohwer, 2000; Robins et al., 2014) as interspecies pairings are thought to increase in frequency when one species is rarer and therefore numerically disadvantaged (e.g., may not have an established population or is at the edge of its range). A numerical imbalance leading to increased hybridization is known as Hubbs' principle or the desperation hypothesis (Hubbs, 1955).

Ephemeral hybrid zones may be the most common type of hybrid zone, yet also the most difficult to detect. Unlike other hybrid zones maintained by a balance between dispersal and a selection against hybrids, ephemeral hybrid zones shift as ranges expand

and contract until a new equilibrium is reached and hybridization events diminish. Some species show a complex history of hybridization. For example, multiple studies of modern humans (*Homo sapiens*) have found evidence supporting hybridization within Africa between sub-Saharan Africans and multiple unknown extinct hominin taxa (Hammer et al., 2011; Lachance et al., 2012; Hsieh et al., 2016; Wall et al., 2019). Modern humans interbred with at least two other *Homo* species upon expanding out of Africa (Bae et al., 2017) as well. Some groups of modern *Homo sapiens* have mosaic genomes with genetic contributions from Neanderthal and Denisovan (Green et al., 2010; Sankararaman et al., 2016). Given the overlap and mobility of many closely related species, those that underwent similar expansions in their ranges likely also have complex genetic histories.

Because hybridization can affect both a population's nuclear DNA (nDNA) and mitochondrial DNA (mtDNA), discordance can appear between the two. Unlike nDNA, mtDNA is maternally inherited and does not undergo recombination in most animal species (Birky, 2001). The maternal mtDNA can be rapidly introgressed into species' genomes through hybridization (Chan & Levin, 2005; Krosby & Rohwer, 2010; Toews et al., 2011) and may persist in the gene pool long after the initial hybridization event. Therefore, one commonly proposed explanation for mitonuclear discordance is historical introgression.

Bluebirds (*Sialia spp.)* are charismatic songbirds within the family Turdidae (thrushes and allies). All bluebirds are sexually dimorphic, with males exhibiting bright UV-blue plumage while females are typically duller. Western and eastern bluebirds (*S. mexicana* and *S. sialis*) also have red on their chests and backs to varying extents, while

mountain bluebirds (*F*) lack the red of its counterparts, instead exhibiting sky-blue plumage with a paler breast and belly. During the breeding season, bluebirds can be found throughout much of North America, with the exception of the far north in the center and eastern parts of the continent (Figure 3.1). The western most species, the western bluebird, inhabits western North America, stretching from southern interior British Columbia down into central Mexico (Guinan et al., 2020). Western bluebirds exist in sympatry with the mountain bluebird throughout most of its range; however, the latter's range extends further north and east into the continent, and it typically breeds at higher elevations (Johnson & Dawson, 2020). Breeding mountain bluebirds are rare as far north as Alaska and the Yukon, while their range extends east into southwest Manitoba, where their range overlaps with the eastern bluebirds. Eastern bluebirds have largest breeding range of all three species, stretching from Manitoba in the west, east to the Atlantic coast, and south to the Florida, with disjunct populations in Mexico, central America, and the Bahamas (Gowaty & Plissner, 2020).

The ranges of all three bluebird species have expanded over the last century. During the past 30 years, western bluebirds have expanded their range eastward into western Montana, where they overlap with mountain bluebirds (Duckworth & Badyaev, 2007). As western bluebirds continue to expand their range, they supplant the mountain bluebirds, forcing them to move to higher latitudes or nest sites outside of the expansion (Duckworth & Badyaev, 2007; Duckworth, 2008; Duckworth & Semenov, 2017). For their part, mountain bluebirds have spread east into the prairie provinces following European colonization and the implementation of nest box programs (Johnson & Dawson, 2020). Simultaneously, eastern bluebirds are spreading west, particularly in

Manitoba and Saskatchewan (Gowaty & Plissner, 2020). All bluebird species pairs can produce viable, fertile hybrid offspring. In western Montana, Duckworth and Semenov (2017) found demonstrated hybridization was more common than once thought in bluebirds. Over 15 years, the found multiple instances of western bluebirds interbreeding with mountain bluebirds. Interestingly, Duckworth and Semenov (2017) noted the presence of F_1 hybrids overlapped with the front of the western bluebird range expansion—evidence of a shifting ephemeral hybrid zone. Examples of mountain and eastern bluebird hybridization are not as well documented and typically occur when both are out of or at the edge of their respective ranges. Hybrids have been reported in Wisconsin, (Johnson & Dawson, 2020) and Manitoba (Lane, 1969). Lastly, hybridization between eastern and western bluebirds appears unlikely given they share very limited overlap in their breeding ranges (Figure 3.1). Gowaty and Plissner (2020) noted that while there are no records in the wild, hybrids have occurred in captivity.

Despite encompassing only three species, the phylogenetics of North American bluebirds remains opaque. Prior to genetic analyses, phylogenetic trees based on morphology placed the eastern and western bluebirds as sister species to each other and mountain bluebirds in a basal position (e.g., Mengel, 1970). The morphological phylogeny was challenged in 2005 by Klicka et al. using concatenated mitochondrial DNA (COI and ND2 genes) which placed mountain bluebirds as sister to eastern bluebirds. However, this was again challenged in 2008 when a broader analysis including more members of the family Turdidae was conducted by Voelker and Klicka. Once again, the phylogeny resembled the original with western and eastern bluebirds as sister

taxa, but the authors concluded it was best to leave the taxonomic relationships within *Sialia* as a polytomy until nDNA could be included.

This study seeks to answer three main questions. First, is there ongoing hybridization among bluebird species within western prairie provinces, specifically Alberta and Saskatchewan. Although mountain bluebirds are by far the most common breeder throughout much of the south and central part of the provinces, both western and eastern bluebirds appear to be increasing in frequency (albeit in low numbers). According to Hubbs' principle, the numerical imbalance should increase the chances of hybridization. Second, is there evidence of past introgression in mountain bluebirds? As mountain bluebirds currently hybridize with modern congeners, past hybridization events are not only possible but likely. Finally, can the bluebird polytomy be resolved using both mtDNA and nDNA from restriction enzyme-associated DNA sequencing (RADseq) data (an approach not yet used in bluebirds). Voelker and Klicka (2008) stated that they believed the inclusion of nDNA was critical to solving unresolved relationships. I predict my phylogenies will support a novel topology with the two western most species of bluebird sharing a more recent common ancestor.

3.2 Methods

3.2.1 Sample collection and DNA extraction

Sample collection took place starting mid-May through late June in 2022 where multiple bluebird species had been previously observed (eBird, 2021). Nest box routes

were used to target high density nesting areas and active nest locations of sampled birds were recorded using a GPS. Song playback was used to entice individuals out their next boxes while two mist nests were placed to capture adult birds. Sampled individuals were quickly examined for acute stress and physical injury. If determined to be stressed or injured, handling time was minimized and a rectrix was pulled in place of blood sampling and the individual was promptly released. Remaining healthy individuals were measured for biometrics (e.g., body mass, tarsus length, bill length, wing chord, etc.) and then a small (~ 50 µl) blood sample was extracted via the brachial vein on the wing. Filter paper and cotton were applied to stop the bleeding and the filter paper provided a secondary sample. Primary and back-up samples were placed in sample tubes containing 99% ethanol and stored ambient temperature for transport back to the lab. They were then transferred to -20ºC. In addition, tissue samples from multiple museums and past studies were used to supplement collection. Overall, 189 bluebird samples were used based on location, potential DNA yield, and species representation (consisting of 157 mountain, 17 western, and 15 eastern). DNA was then extracted from the samples using a modified salting-out DNA extraction protocol (Miller et al., 1988) and DNA concentrations was measured using a NanoDrop® ND-1000 UV-Vis spectrophotometer.

3.2.2 RADseq

To prepare the samples, a modified ddRADseq (double digest restriction enzyme-associated DNA sequencing) protocol (changed to include a third enzyme) was used (Peterson et al., 2012). Extracted DNA samples were digested with three restriction enzymes *MspI* (4 base-pair cutter), *NsiI* (6 base-pair cutter) and *Pst I* (6-base pair cutter)

and a library constructed following Abed et al.(2019). Samples were then sent to Genome Quebec for sequencing on an Illumina NovaSeq 6000 with paired end reads at Génome Québec.

3.2.3 Data processing

After, raw reads were analyzed with Fastqc/0.11.5 (Andrews, 2011), and quality reports for each forward and reverse read were generated to note sequence quality. Génome Québec then sent raw sample reads which were demultiplexed using Sabre/1.00, assigning reads to individuals. The program Cutadapt (Martin, 2011) was then used to trim remaining adapter and barcode sequences.

Sequencing reads were analyzed using Stacks/2.3e, which is designated for restriction-enzyme based data, to genotype and call variant sites for each individual sample (Catchen et al., 2013). Using an annotated Swainson's thrush (*Catharus ustulatus*) genome (GenBank reference: GCF_009819885.2), the sequences were aligned using the reference-based pipeline in Stacks (Termignoni-Garcia et al., 2022). The Swainson's thrush genome was used as they are also members of the family Turdidae, and had chromosome level assembly. Indexing of the reference genome was done using Samtools/0.1.2 (Danecek et al., 2021). Alignment of forward and reverse reads was done using the Burrows-Wheeler Alignment Tool (bwa) (Li & Durbin, 2009). Samtools/0.1.2 was used a second time to sort and index the binary alignment map (bam) files created from the alignment. Sequence variants from mountain bluebirds in Alberta, British

Columbia, and Saskatchewan were then called using default parameters in the Stacks/2.3e ref_map pipeline to sort and align the files.

3.2.4 Data filtering – western and mountain bluebirds

A variant call format (vcf) file was produced for western and mountain bluebirds using the populations function within the reference-based pipeline in Stacks/2.3e. Initial filtering selected loci that were present in >2 populations (-p 2) and occurred in >50% of the individuals across all populations (-R 0.5). To account for one-off SNPs introduced through sequencing error, a minimum allele frequency of 5% was used (-min-maf 0.05). As many downstream analyses software assume unlinked SNPs, the --write-single-snp command was used to select a random SNP per locus. A total of 36,977 SNPs remained following initial filtering.

To further eliminate missing data, the vcf file was filtered again using VCFtools/0.1.16 (Danecek et al., 2011) for both SNPs and individuals. SNPs with more than 30% missing data were removed, followed by the subsequent removal of individuals with more than 35% missing data prior. The refiltered dataset contained 6,160 unlinked SNPs and 57 individuals (7 western bluebirds and 50 mountain bluebirds).

3.2.5 Data filtering – eastern and mountain bluebirds

A second variant call format (vcf) file was subsequently produced using the populations function within the reference-based pipeline in Stacks/2.3e using eastern

bluebirds. The same parameters were used for initial filtering (-p 2, -R 0.5, -min-maf 0.05, --write-single-snp). A total of 42,950 SNPs remained following initial filtering.

Again, the vcf file was refiltered to reduce missing data using VCFtools/0.1.16 (Danecek et al., 2011) for both SNPs and individuals. SNPs with more than 30% missing data were removed, followed by the subsequent removal of individuals with more than 35% missing data prior. The refiltered dataset contained 7,477 unlinked SNPs and 56 individuals (6 eastern bluebirds and 50 mountain bluebirds).

3.2.6 Early-stage hybrid identification

For both datasets, an ancestry matrix was constructed using STRUCTURE 2.3.4 (Pritchard, 2000). The program used a burnin value of 100,000 and a Markov-chain Monte Carlo (MCMC) value of 300,000. STRUCTURE was run for K = 2 - 4 with 10 runs each. The subsequent results ran with STRUCTURE HARVESTER v.0.6.94 (Earl & vonHoldt, 2012) to assign ancestry proportions to each individual. Finally, the plots were visualized using a Clumpak server (Kopelman, 2015).

Principal Coordinate Analyses (PCoA) were also performed to visualize genetic distances for both datasets. Genetic distance matrices were created using the pairwise Euclidean genetic distances to calculate dissimilarities between each individual in Adegenet package in R-Studio/2021.09.0 (Jombart, 2011). GenAlEx v/6.5 was used to visualize and create the PCoA based on genetic distance matrix (Peakall & Smouse 2006).

3.2.7 Data filtering – bluebird with outgroup

To search for signs of ancient introgression and create a phylogeny using nDNA, a subset of samples with roughly equal sample sizes for each species was used. The subset data set included western (n=6), eastern (n=6), and mountain bluebirds (n=8), as well as wood thrush (*H. mustelina*) (n=8) as an outgroup. Sequence variants shared between all bluebirds and wood thrush were then called using default parameters in the Stacks/2.3e ref_map pipeline to sort and align the files. Populations was then used to filter the samples and create a vcf file. Only loci present in >80% of the individuals across all four species (-R 0.8) were included. Following initial filtering, 271,621 linked SNP remained. Vcftools was then used to filter SNPs with more than 10% missing data, followed by the subsequent removal of individuals with more than 10% missing data prior (however this did not exclude any of the original 28 individuals). The data set contained 151,717 linked SNPs and 28 individuals.

3.2.8 Ancient introgression using D-statistics and f_4-ratio

To test for signs of possible historical introgression using the filtered data set of 151,717 linked SNPs, Dsuite v0.5 r44 (Malinsky et al., 2021) was used to calculate f_4 – ratio and Patterson's D statistics (Patterson et al., 2018). Patterson's D (also known as the ABBA BABA statistic), treats biallelic SNPs from three ingroup taxa and one outgroup as either the derived allele ('B') or the ancestral ('A'). The number of SNPs in the 'ABBA' configuration roughly equal those in the 'BABA', with a significant deviation from zero evidence of gene flow (Patterson et al., 2018). The first analysis assumed

western and mountain bluebirds were sister species to each other and attempted to detect possible gene flow between mountain and eastern bluebirds. The second analysis used the tree generated from the multi-species coalescent model in SNAPP after it was converted into a NEWICK format using Figtree (Rambaut, 2018).

3.2.9 mtDNA sequencing

To supplement the available bluebird mitochondrial data available for public download from Genbank, extracted DNA was amplified with universal cytochrome b (Cytb) (n=11) barcoding primers L15213 (5' – GMCGAGGAHTCTACTAYGGCTC – 3') and H15914 (5' – GGTTGTTCTACTGGTTGGCT – 3') (Sorenson et al., 1999). A 704 bp region of the Cytb locus was amplified in 25 µl reactions with GoTaq flexi buffer, 2.5 mM $MgCl_2$, 0.1 mM dNTP, 0.5 µM of each primer, 0.1 units of GoTaq flexi DNA polymerase. An Eppendorf master cycle was used under the following conditions: 1 cycle of 94°C for 2 minutes, 50°C for 45 seconds, and 72°C for 1 minute; 37 cycles of 94°C for 30 seconds, 50°C for 45 seconds, and 72°C for 1 minute; with a final cycle at 72°C for 5 minutes. Some extracted samples were also amplified for the NADH dehydrogenase subunit 2 (ND2) loci (n=3) using universal ND2 barcoding primers ND2 L5215 (5' –TATCGGGCCCATACCCCGAATAT – 3') (Sorenson et al., 1999) and H5783 (5' – CCTARGTGRGAGATRGAKGAG – 3') (Hackett, 1996). A 568 bp region of the ND2 locus was amplified using the same PCR mix (substituting the appropriate primers) and the same PCR conditions, with the exception of a 55°C annealing temperature. Amplified DNA was visualized on 0.8% agarose gel stained with ethidium bromide and sent to McGill University for sequencing.

MEGA 7.0 (Kumar et al., 2016) was used to check, align, and trim the sequences. The amplified Cytb and ND2 sequences were supplemented with mountain, (Klicka et al., 2005), western (Klicka et al., 2005; Miller et al., 2007; Olsson et al., 2013), and eastern bluebird sequences (Barker, 2004; Klicka et al., 2005; Sangster et al., 2010; Winker & Pruett, 2006; Zuccon & Ericson, 2010). Three genera of basal Turdidae were used to root the trees, including the most basal member, the Grandala (*Grandala coelicolor*) (Price et al., 2014), two extant species of *Stizorhina* (Sangster et al., 2010) and three solitaires (*Myadestes spp*) (Barber & Klicka, 2010; Klicka et al., 2005; Lovette & Bermingham, 2000; Miller et al., 2007; Olsson et al., 2013; Sly et al., 2011; Voelker & Spellman, 2004; Zuccon & Ericson, 2010).

Additionally, Cytochrome c oxidase 1 (COI) gene sequences for all three bluebird species were available through Genbank and were used to create a third gene tree with the Townsend's solitaire (*M. townsendii*) as an outgroup (all samples from Kerr et al., 2007; Tavares & Baker, 2008). Based on available literature, the COI gene has not been used in a phylogenetic study to date which includes all three bluebird species. All samples were aligned and trimmed in MEGA 7.0 (Kumar et al., 2016) producing 583 bp sequences for analyses.

3.2.10 mtDNA phylogenetic analyses

Alignments were used to create Bayesian phylogenetic trees using BEAST2 v2.7.4 (Bouckaert et al., 2014). The Grandala, considered to be the most basal member of Turdidae, was used to root both the Cytb and ND2 gene trees. Control files were

generated in BEAUti v2.7.4 using the bModel test v 1.3.3 extension to select the site model. The clock model was set to default (strict clock), along with the priors (Yule model with a uniform birth rate), and the MCMC updates (totaling 10^6 iterations, saving every 10^3 steps). TreeAnotator v2.7.4 was used to select a maximum clade credibility tree with the default settings (10% burn in, 0.0 posterior probability limit, with common ancestor heights). The consensus tree with posterior probabilities was then visualized using Figtree v1.4.4 (Rambaut, 2018). Densitree v2.7.4 was also used to map the consensus trees against all possible trees that remained following burn in.

3.2.11 nDNA phylogenetic filtering and analyses

To establish the phylogenetic relationships between bluebird species using nDNA, the vcf file containing all three bluebird species and wood thrush was converted to NEXUS file using the python script vcf2phylip v2.0 (Ortiz, 2019). The SNAPP add-on (Bryant et al., 2012) was used to create a species tree in BEAST v2.4.0 (Bouckaert et al., 2014). Species trees are created in SNAPP using a coalescent model for unlinked biallelic SNPs, a process which is very computationally demanding. A similar approach as Younger et al. (2016) was used to decrease the computational demand by randomly selecting a subset of individuals per population (n=4). Mutation rates were set to the default parameters (lambda, u, and v = 1/x), and were calculated as the MCMC ran. The MCMCs were run for 1 million iterations and convergence was checked in Tracer v1.6 (Drummond & Rambaut, 2007). As outlined in the Tracer's user manual, an Effective Sample Size (ESS) >200 is an acceptable value and an ESS value of 560.7 was obtained. Five thousand and one maximum likelihood species trees were generated using

TreeAnnotator v2.7.3 (Drummond & Rambaut, 2007) with a 10% burn-in using median tree heights. Figtree v.1.4.4 (Rambaut, 2018) was used to visualize the resulting species tree.

For confirmation, a rooted maximum likelihood tree with accompanying bootstrap values was also generated using the SVDquartets function in Paup* v4.0a (Swofford, 2002). A multispecies coalescence model was used to evaluate 10^6 random quartets while performing 100 bootstrap replications. A PCoA was also generated using the same individuals and the 151,717 associated SNPs to provide a better idea of the accuracy.

3.3 Results

3.3.1 Early-stage hybrid identification

The PCoA examining the recent ancestry of mountain bluebirds from western Canada with eastern bluebirds (Figure 3.2a) did not detect any recent hybrids. Each species formed a distinct cluster. The first PC (17.9%) accounted for the most genetic variation between the two species, followed by the second PC accounted for 1.80%. The absence of mountain and eastern bluebird hybrids was also confirmed through the ancestry matrix in STRUCTURE (Figure 3.2b). When K=2, no individual exhibited any signs of an admixed genotype. The same remained true at K=3 and 4, with over-splitting occurring in mountain bluebirds.

The PCoA examining recent ancestry of western Canadian mountain and western bluebirds (Figure 3.2c) did not detect any recent hybrids. The first PC (9.14%) accounted

for the most genetic variation between the two species, followed by the second PC accounting for 1.97%. The ancestry matrix in STRUCTURE (Figure 3.2d) supported PCoAs findings. When K=2, no individuals exhibited any signs of an admixed genotype. Only K=4 showed admixture within one western bluebird, but this is likely due to a higher proportion of missing data within that individual (percent missing was ~30% compared to <10% for the rest).

3.3.2 Measures of ancient introgression

Dsuite was able to recover signs of possible ancient introgression within the bluebird genus. When using the Klicka et al. (2005) phylogeny with ((eastern, mountain) western) Patterson's D statistic differed significantly from zero (D=0.1740, $p=2.3x10^{-16}$) (Table 3.1). The f_4-ratio suggested ~19% of the biallelic SNPs measured did not fit the expected pattern of inheritance (AA(BB)), but (AB(BA)).

When the phylogeny was changed to match the novel configuration recovered by PAUP*, with ((western, mountain) eastern), Patterson's D remained significant (D=0.0448, $p=2.385x10^{-3}$). Again, the f_4-ratio suggested possible admixture, however, it was much lower; ~2% of the biallelic SNPs measured did not fit the expected pattern of inheritance.

3.3.3 mtDNA phylogenetic analyses

In total, 10,001 trees for each gene were generated in BEAST2, from which 9001 were retained for further analyses. For each tree, the maximum clade credibility tree is shown with posterior probabilities in the nodes (Figures 3.3-3.5). The log files for Cytb, ND2, COI in Tracer revealed convergence of the species tree with ESS>200 (ESS = 2641, 2448, and 902, respectively). While the ND2 tree supports the same findings as Klicka et al. (2005) with eastern bluebirds sister species to mountain, the Cytb tree does not. Instead, the Cytb gene tree concurs with the COI gene tree, which together places eastern bluebirds as sister species to western, and supports the later findings of Voelker and Klicka (2008).

3.3.4 nDNA phylogenetic analyses

As with each mitochondrial gene, a total of 1001 trees were generated for the SNAPP analysis, with 901 being retained following a 10% burn in. The tree, containing the three bluebirds and the wood thrush as an outgroup, retained strong posterior probabilities of 1.0 throughout (Figure 3.6a). Again, an ESS >200 was achieved. The species arrangement was consistent with the previous findings of Klicka et al. (2005), where eastern bluebirds are sister species to mountain bluebirds using the concatenated ND2 and Cytb genes.

However, the phylogeny generated using the SVD quartets method indicates a different topology. Using wood thrush as an outgroup to root the tree recovers a novel topology of western bluebirds as the sister group to mountain bluebirds (Figure 3.6c). This topology is also supported by the results of the PCoA produced by the genetic

distance matrix of all four species (Figure 3.6b). The first PC accounts for 53.31%, clearly separating out the members of the genus *Sialia* from wood thrush. The second PC only accounts for 4.78% of the difference, and it clusters western with mountain bluebirds and separates them from eastern bluebirds.

3.4 Discussion

Based on the dataset there is no evidence to support contemporary hybridization within mountain bluebirds with its congeners in the western prairie provinces. This includes both F_1 and advanced stage hybrids. However, there were clear signs of ancient introgression present using Patterson's D-statistic and the f_4-ratio. Ancient introgression may also explain the mitonuclear discordance occurring between the phylogenies when nDNA is compared to mtDNA (although alternate causes cannot be ruled out as of yet, like mitochondrial genes being under selection).

3.4.1 Contemporary hybridization in Alberta

Contrary to my predictions, there does not appear to be evidence of contemporary hybridization in the Albertan population of mountain bluebirds. Although Hubbs' principle (1955) predicts that hybridization should increase in frequency when species are numerically imbalanced and has been observed in bluebirds previously (Duckworth and Semenov, 2017), there are likely multiple reasons for the lack of early hybrids. During their study in neighboring Montana, Duckworth and Semenov (2017) were able to monitor the same nesting sites for 15 years. During this time, they were able to collect

samples from around 1,297 pairs (2,594 individuals) of nesting bluebirds. Still, they detected only 15 early-stage hybrids despite having a much larger sample size and found that heterospecific pairings accounted for only 0.23% of the pairings overall. This suggests that while hybridization in bluebirds does occur occasionally and is increased in localized instances (i.e. western Montana), it is not as sporadic as in other bird species, such as mountain chickadees (*Poecile gambeli*) and black-capped chickadees (*P. atricapulus*) (Grava et al., 2012; Grabenstein et al., 2022) or Cabot's and elegant terns (*Thalasseus acuflavida* and *T. elegans*, respectively) (Velarde & Rojo, 2012; Dufour et al., 2017). Similarly, records of hybridization between mountain and eastern bluebirds are rare, suggesting infrequent hybridization.

The first ever breeding record of both eastern and western bluebirds in Alberta are also more recent compared to other locations within their ranges. The Royal Alberta Museum places the first record of eastern bluebirds breeding in 1977, while the first western bluebird breeding record was even later in 1984. In contrast, the breeding history of western bluebirds in Montana are much older. Western bluebirds were once more common throughout the Pacific Northwest and into western Montana but were extirpated in the 19th and 20th centuries when timber harvesting reduced available nest sites (Guinan et al., 2020). Where the birds in Montana appear to be reclaiming a mix of old and new territory, the Alberta birds likely consist of wayward vagrants. In fact, western bluebirds have been known to form loose migratory flocks with mountain bluebirds in places like Washington (Herlugson, 1975). It is possible that, rather than being an extension of the expansion seen in Montana, vagrant birds arrive with flocks alongside the more common mountain bluebird and either form conspecific breeding pairs or move on to an area with

higher conspecific density without breeding. The same is likely true for eastern bluebirds within Alberta and Saskatchewan. Records of eastern and mountain bluebird hybrids from Manitoba (e.g., Lane, 1969) tend to occur where both species is at their range limits, but still have higher densities than what is observed in Alberta. Overall, likely owing to both very low population densities and small relative sample sizes, it appears that hybridization is either incredibly rare and thus challenging to detect, or not occurring in the first place.

3.4.2 Signals of ancient introgression

Both coalescence models created using PAUP* and SNAPP resolved two different phylogenies despite using the same dataset. Wide-spread introgression (more than 10 individuals per generation) violates the assumptions of most coalescence models and can lead to the altering of tree topologies (Zhang et al., 2011). If sustained introgression occurs between non-sister taxa, (i.e., between western and mountain bluebirds based on the 2005 phylogeny), tree topologies may become distorted (Mallet et al., 2016). This phenomenon has been observed in many rapidly speciated taxa, including African mosquitos (Fontaine et al., 2015), *Heliconius* butterflies (Heliconius Genome Consortium 2012), African cichlids (Cui et al., 2013), and Galapagos finches (Lamichhaney et al., 2015). Mallet and colleagues (2016) note that examples of widespread hybridization between non-sister taxa often follow rapid species radiation within the group. Although not as speciose as the previously listed taxa, bluebirds would fit this description as well. Klicka et al. (2005) placed their split from the proto-*Sialia* ancestor to the mid-Pliocene, around 3 million years before present. With hybrids still

occurring today, past bouts of hybridization events in ephemeral hybrid zones appear likely.

Evidence of ancient hybridization events also appear supported by the results of Patterson's D and the f_4-ratio statistics (Table 3.1). When mountain bluebirds are sister taxa to eastern bluebirds, Dsuite detected more bi-allelic SNPs in the ABBA configuration than expected between western and mountain bluebirds despite the two presumably being non-sister taxa ($D=1.741 \times 10^{-1}$, P-value$=2.3 \times 10^{-16}$). The *f4*-ratio inferred ~19% of the genome sampled could have resulted from an admixture event. When the phylogeny was changed to match the results of SVD quartets, Dsuite also detected a significant ABBA pattern being present between eastern and mountain bluebirds ($D= 4.480 \times 10^{-2}$, P-value = 0.0023) albeit with a greatly reduced *f4*-ratio inferring ~2.3% arising from an admixture event. The difference in *f4*-ratios makes sense given the current and likely past range overlap. Western and mountain bluebirds are the two western-most species and have ranges overlapping in much of the same territory. The two bluebirds are often only separated by elevation differences which is a dynamic observed in other songbirds (e.g., mountain and black-capped chickadees (Hill & Lein, 1988)). In contrast, eastern and mountain bluebirds have been in contact for a much shorter amount of time, leading to fewer hybridization opportunities. In general, a higher number of hybridization events corresponds with greater opportunity for more of the genome to be introgressed.

Of course, there are alternative possibilities to introgression to explain the overlap of shared genome. Western and mountain bluebirds may share a more recent common ancestor and therefore are sister taxa to each other. Looking at Figure 3.2a compared to

3.2c, the PC accounting for the most genetic distance is lower between western and mountain bluebirds than it is for eastern and the mountain bluebirds (9.17% versus 17.90%). This remains true even in the presence of an outgroup and with much tighter filtering used to produce the PCoA (Figure 3.6c). Western and mountain bluebirds appear to share more of their nDNA which supports the findings of SVD quartets phylogeny (Figure 3.7). Another alternative is that bluebirds speciated rapidly, as suggested by Klicka et al. (2005) leading to incomplete lineage sorting as different mitochondrial lines persist. Evidence for a rapid speciation can be seen in Figure 3.6a, where the branch lengths of each bluebird species is much shorter than that of the outgroup (wood thrush). Ultimately, this inconsistency likely means the current polytomy cannot be readily resolved based on the findings of this study alone.

3.4.3 Possible causes of mitonuclear discordance within the bluebird genus

The species trees with the highest support for each mitochondrial gene and the coalescence models for nDNA reveal an interesting contradiction amongst the bluebird topologies. Only the ND2 gene and the SNAPP coalescence tree (Figure 3.6a) support the same topology as Klicka et al. (2005) where eastern and mountain bluebirds are sister to each other. Both the Cytb and COI gene trees (Figures 3.4 and 3.5, respectively) support the later phylogeny of Voelker and Klicka (2008), with eastern and western bluebirds as sister taxa. Voelker and Klicka (2008) attributed the differences between *Sialia* spp. to the inclusion of multiple samples from *Myadestes* spp. in their analyses. To account for the inclusion of an outgroup altering the lengths of the branches and changing the topology, multiple outgroups were used in the current study. The final multiple species

coalescence tree created with SVD quartets in PAUP*, produced a completely new topology, with mountain and western bluebird as sister taxa (albeit weaker bootstrap support when rooted with an outgroup).

Disagreement between nDNA and mtDNA based phylogenies is not uncommon and can be found in many species (DeSalle & Giddings, 1986; Patton & Smith, 1994; Avise, 2000; Linnen & Farrell, 2007; Gompert et al., 2008; Obertegger et al., 2018; Myoshu, 2023). Possible explanations for the different topology of each tree include: 1) the different mutation rates of nuDNA versus mtDNA, 2) biparental (nuDNA) versus uniparental (mtDNA) inheritance, 3) coalescence time since effective population size is larger in nuDNA than in mtDNA, 4) creating a gene tree versus a species tree, and 5) hybridization and introgression through continued gene flow following speciation. Uniparentally inherited DNA (i.e., mtDNA and chloroplast DNA) can admix at a faster rate than nuclear DNA when gene flow between hybridizing populations is separated by prezygotic barriers (Chan & Levin, 2005). Chan and Levin (2005) found 42 studies where the apparent introgression of mtDNA or chloroplast DNA was more extensive than the introgression of nDNA. Duckworth and Semenov (2017) noted that of the hybrid pairs in Montana, all consisted of a male western and a female mountain bluebird. If past ephemeral hybrid zones existed between the species pair with the same hybridization dynamic, it is unlikely western bluebird mtDNA would have the chance to rapidly admix into the mountain bluebird population. This may partially explain why a phylogeny where western and mountain bluebird are sister species is not produced with mitochondrial markers.

Aside from hybridization, mitonuclear discordance can also be indicative of incomplete lineage sorting (ILS). ILS occurs when copies of ancestral genes do not coalesce to a common ancestral state prior to a speciation event. Although mtDNA may be viewed as a neutral marker, mitochondrial genes are still subject to the same microevolutionary forces as nDNA, albeit without recombination. Selection in mtDNA has been observed in a number of organisms, including *Drosphila* (Nigro, 1994), copepods (Schizas et al., 2001), and mice (Takeda et al., 2000). In members of the genus *Parus*, Zink (2005) found that the ND2 gene fit a mildly deleterious allele model, with silent point mutations persisting longer than those that altered the amino acid composition, which were quickly selected against. As a single molecule, strong direct selection for an allele of a gene indirectly selects for all of the mitochondrial genes within the plasmid (Ballard & Whitlock, 2004). Additionally, parts of the nuclear genome may also select mitochondrial genes. In six mammal species, mtDNA-encoded residues proximal to nDNA-encoded residues within the Cytochrome c oxidase (complex IV) were found to evolve more rapidly than more distal residues, indicating they were under indirect positive selection (Schmidt et al., 2001). Overall, trying to infer phylogenetic histories from single genes may risk heavy biases as beneficial alleles persist past speciation events (Ballard & Whitlock, 2004).

3.5 Conclusions

Overall, there does not appear to be ongoing hybridization within the western prairie provinces' bluebird populations as I predicted and as suggested by Hubbs' principle. I did detect signals of ancient introgression, though, using both Patterson's D-

statistic and the f_4-ratio. The statistical tests show a significant departure in biallelic SNPs from the expected phylogeny, indicating continued gene flow between mountain bluebirds and its congeners after speciation. This may further be supported by the clear mitonuclear discordance within the genus *Sialia* when using mitochondrial genes versus nuclear loci. Unfortunately, I was unable to resolve the polytomy within the bluebird clade using two different types of genetic markers (suggested by Voelker and Klicka (2008)). I was also unable to determine the exact cause of the ILS, but my results, along with those of past bluebird studies, support a rapid species radiation combined with subsequent historical hybridization events previously unknown within *Sialia*. The presence of both historical and contemporary hybridization in bluebirds further supports the idea that speciation is not always a discrete process; speciation is often punctuated by bouts of gene flow between congeners.

3.6 Literature Cited

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R. and Butlin, R.K. (2013). Hybridization and speciation. Journal of Evolutionary Biology, 26(2), 229-246.

Abed A, Légaré G, Pomerleau S, St-Cyr J, Boyle B, Belzile FJ. 2019 Genotyping-by-Sequencing on the Ion Torrent platform in barley. Methods Molecular Biology. 1900:233-252. doi: 10.1007/978-1-4939-8944-7_15. PMID: 30460569.

Andrews, S. (2011). FastQC—a quality control tool for high throughput sequence data [Online]. Babraham Bioinformatics.

Arntzen, J. W., & Wallis, G. P. (1991). Restricted gene flow in a moving hybrid zone of the newts *Triturus cristatus* and *T. marmoratus* in western France. Evolution, 45(4), 805-826.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. Molecular Ecology, 13(4), 729-744.

Barber, B. R., & Klicka, J. (2010). Two pulses of diversification across the Isthmus of Tehuantepec in a montane Mexican bird fauna. Proceedings of the Royal Society B: Biological Sciences, 277(1694), 2675-2681.

Barker, F. K. (2004). Monophyly and relationships of wrens (Aves: Troglodytidae): a congruence analysis of heterogeneous mitochondrial and nuclear DNA sequence data. Molecular Phylogenetics and Evolution, 31(2), 486-504.

Bae, C. J., Douka, K., & Petraglia, M. D. (2017). On the origin of modern humans: Asian perspectives. Science, 358(6368), eaai9067.

Birky Jr, C. W. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annual Review of Genetics, 35(1), 125-148.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M.A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Computational Biology, 10(4), e1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing

gene trees in a full coalescent analysis. Molecular biology and evolution, 29(8), 1917-1932.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124-3140.

Chan, K. M., & Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution, 59(4), 720-729.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. Bioinformatics, 30(23), 3317-3324.

Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. Evolution, 67(8), 2166-2179.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M. & Li, H. (2021). Twelve years of SAMtools and BCFtools. Gigascience, 10(2), p.giab008.

DeSalle, R., & Giddings, L. V. (1986). Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian Drosophila. Proceedings of the National Academy of Sciences, 83(18), 6902–6906.

Duckworth, R. A., & Semenov, G. A. (2017). Hybridization associated with cycles of ecological succession in a passerine bird. The American Naturalist, 190(4), E94-E105.

eBird. 2021. eBird: An online database of bird distribution and abundance [web application]. eBird, Cornell Lab of Ornithology, Ithaca, New York.

Fontaine, M. C., Pease, J. B., Steele, A., Waterhouse, R. M., Neafsey, D. E., Sharakhov, I. V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., & Besansky, N. J. (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science, 347(6217), 1258524.

Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in Lycaeides. Molecular Ecology, 17(24), 5231-5244.

Gowaty, P. A. & Plissner, J. H. (2020). Eastern bluebird (*Sialia sialis*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.easblu.01

Grabenstein, K. C., Otter, K. A., Burg, T. M., & Taylor, S. A. (2022). Sympatry leads to reduced body condition in chickadees that occasionally hybridize. Ecology and Evolution, 12(4), e8756.

Grava, A., Grava, T., Didier, R., Lait, L.A., Dosso, J., Koran, E., Burg, T.M., & Otter, K.A. (2012). Interspecific dominance relationships and hybridization between black-capped and mountain chickadees. Behavioral Ecology, 23(3), 566-572.

Guinan, J. A., P. A. Gowaty, and E. K. Eltzroth (2020). Western bluebird (*Sialia mexicana*), version 1.0. In Birds of the World (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.wesblu.01

Hackett, S. J. (1996). Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). Molecular Phylogenetics and Evolution, 5(2), 368-382.

Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94-98.

Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. Trends in ecology & evolution, 3(7), 158-167.

Hsieh, P., Woerner, A. E., Wall, J. D., Lachance, J., Tishkoff, S. A., Gutenkunst, R. N., & Hammer, M. F. (2016). Model-based analyses of whole-genome data reveal a complex evolutionary history involving archaic introgression in Central African Pygmies. Genome Research, 26(3), 291-300.

Hubbs, C. L. (1955). Hybridization between fish species in nature. Systematic Zoology, 4(1), 1-20.

Johnson, L. S. and R. D. Dawson (2020). Mountain bluebird (*Sialia currucoides*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

Kerr, K. C., Stoeckle, M. Y., Dove, C. J., Weigt, L. A., Francis, C. M., & Hebert, P. D. (2007). Comprehensive DNA barcode coverage of North American birds. Molecular Ecology Notes, 7(4), 535-543.

Klicka, J., Voelker, G., & Spellman, G. M. (2005). A molecular phylogenetic analysis of the "true thrushes" (Aves: Turdinae). Molecular Phylogenetics and Evolution, 34(3), 486-500.

Krosby, M., & Rohwer, S. (2010). Ongoing movement of the hermit warbler X Townsend's warbler hybrid zone. PloS One, 5(11), e14164.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution, 33(7), 1870-1874.

Lachance, J., Vernot, B., Elbers, C. C., Ferwerda, B., Froment, A., Bodo, J. M., Lema, G., Fu, W., Nyambo, T.B., Rebbeck, T.R., & Tishkoff, S. A. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell, 150(3), 457-469.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.J., Wang, C., Zamani, N., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature, 518(7539), 371-375.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754-1760.

Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. Evolution, 61(6), 1417-1438.

Lovette, I. J., & Bermingham, E. (2000). C-mos variation in songbirds: molecular evolution, phylogenetic implications, and comparisons with mitochondrial differentiation. Molecular Biology and Evolution, 17(10), 1569-1577.

Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-fast D-statistics and related admixture evidence from VCF files. Molecular Ecology Resources, 21(2), 584-595.

Mallet, J. (2005). Hybridization as an invasion of the genome. Trends in Ecology & Evolution 20(5), 229-237.

Mallet, J. (2007). Hybrid speciation. Nature, 446(7133), 279-283.

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? BioEssays, 38(2), 140-149.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal, 17(1), 10-12.

Mengel, R. M. (1970). The North American Central Plains as an isolating agent in bird speciation. Pleistocene and recent environments of the central Great Plains, 279-340.

Miller, M. J., Bermingham, E., & Ricklefs, R. E. (2007). Historical biogeography of the new world solitaires (*Myadestes spp*). The Auk, 124(3), 868-885.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, 16(3), 1215.

Myoshu, H. (2023). A boundary between occurrences of mitochondrial cytochrome b haplotypes as well as mito-nuclear discordance in the house mouse *Mus musculus* from the northern area of Tochigi Prefecture, Japan. Mammalian Biology, 103(2), 175-186.

Toews, D. P., Brelsford, A., & Irwin, D. E. (2011). Hybridization between Townsend's *Dendroica townsendi* and black-throated green warblers *D. virens* in an avian suture zone. Journal of Avian Biology, 42(5), 434-446.

Nigro, L. (1994). Nuclear background affects frequency dynamics of mitochondrial DNA variants in *Drosophila simulans*. Heredity, 72(6), 582-586.

Obertegger, U., Cieplinski, A., Fontaneto, D., & Papakostas, S. (2018). Mitonuclear discordance as a confounding factor in the DNA taxonomy of monogonont rotifers. Zoologica Scripta, 47(1), 122-132.

Olsson, U., & Alström, P. (2013). Molecular evidence suggests that the enigmatic Sulawesi endemic Geomalia heinrichi belongs in the genus Zoothera (Turdidae, Aves). Chinese Birds, 4, 155-160.

Ortiz EM. (2019) vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Zenodo.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D., (2012). Ancient admixture in human history. Genetics, 192(3).1065-1093.

Patton, J. L., & Smith, M. F. (1994). Paraphyly, polyphyly, and the nature of species boundaries in pocket gophers (genus *Thomomys*). Systematic Biology, 43(1), 11-26.

Peakall, R. O. D., & Smouse, P. E. (2006). GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6(1), 288-295.

Pearson, S. F., & Rohwer, S. (2000). Asymmetries in male aggression across an avian hybrid zone. Behavioral Ecology, 11(1), 93-101.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One, 7(5), e37135.

Purcell, K. L., Verner, J., & Oring, L. W. (1997). A Comparison of the breeding ecology of birds nesting in boxes and tree cavities. The Auk, 114(4), 646–656. doi:10.2307/4089284

Rambaut, A. (2018) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/

Robbins, T. R., Walker, L. E., Gorospe, K. D., Karl, S. A., Schrey, A. W., McCoy, E. D., & Mushinsky, H. R. (2014). Rise and fall of a hybrid zone: implications for the roles of aggression, mate choice, and secondary succession. Journal of Heredity, 105, 226-236.

Sangster, G., Alström, P., Forsmark, E., & Olsson, U. (2010). Multi-locus phylogenetic analysis of Old World chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level (Aves: Muscicapidae). Molecular Phylogenetics and Evolution, 57(1), 380-392.

Schizas, N. V., Chandler, G. T., Coull, B. C., Klosterhaus, S. L., & Quattro, J. M. (2001). Differential survival of three mitochondrial lineages of a marine benthic copepod exposed to a pesticide mixture. Environmental Science & Technology, 35(3), 535-538.

Schmidt, T. R., Wu, W., Goodman, M., & Grossman, L. I. (2001). Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. Molecular Biology and Evolution, 18(4), 563-569.

Sly, N. D., Townsend, A. K., Rimmer, C. C., Townsend, J. M., Latta, S. C., & Lovette, I. J. (2011). Ancient islands and modern invasions: disparate phylogeographic histories among Hispaniola's endemic birds. Molecular Ecology, 20(23), 5012-5024.

Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., & Mindell, D. P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution, 12(2), 105-114.

Stacy, E. A., Johansen, J. B., Sakishima, T., & Price, D. K. (2016). Genetic analysis of an ephemeral intraspecific hybrid zone in the hypervariable tree, *Metrosideros polymorpha*, on Hawaiin Island. Heredity, 117(3), 173-183.

Swofford, D.L. (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), 4.0 edn, Sinauer, Sunderland, MA.

Takeda, K., Takahashi, S., Onishi, A., Hanada, H., & Imai, H. (2000). Replicative advantage and tissue-specific segregation of RR mitochondrial DNA between C57BL/6 and RR heteroplasmic mice. Genetics, 155, 777– 783.

Tavares, E. S., & Baker, A. J. (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evolutionary Biology, 8(1), 1-14.

Termignoni-Garcia, F., Kirchman, J. J., Clark, J., & Edwards, S. V. (2022). Comparative population genomics of cryptic speciation and adaptive divergence in Bicknell's and gray-cheeked thrushes (Aves: *Catharus bicknelli* and *Catharus minimus*). Genome Biology and Evolution, 14(1), evab255.

Velarde, E., & Rojo, P. (2012). Presumed hybrid elegant x Cabot's terns *Thalasseus elegans* X *T. acuflavida* in Isla Rasa, Gulf of California, Mexico. Marine Ornithology, 40, 25-29.

Voelker, G., & Klicka, J. (2008). Systematics of *Zoothera* thrushes, and a synthesis of true thrush molecular systematic relationships. Molecular Phylogenetics and Evolution, 49(1), 377-381.

Voelker, G., & Spellman, G. M. (2004). Nuclear and mitochondrial DNA evidence of polyphyly in the avian superfamily Muscicapoidea. Molecular Phylogenetics and Evolution, 30(2), 386-394.

Wall, J. D., Ratan, A., Stawiski, E., Kim, H. L., Kim, C., Gupta, R., Suryamohan, K., Gusareva, E.S., Purbojati, R.W., Bhangale, T., & Peterson, A. S. (2019). Identification of African-specific admixture between modern and archaic humans. The American Journal of Human Genetics, 105(6), 1254-1261.

Winker, K., & Pruett, C. L. (2006). Seasonal migration, speciation, and morphological convergence in the genus *Catharus* (Turdidae). The Auk, 123(4), 1052-1068.

Zhang, C., Zhang, D. X., Zhu, T., & Yang, Z. (2011). Evaluation of a Bayesian coalescent method of species delimitation. Systematic Biology, 60(6), 747-761.

Zink, R. M. (2005). Natural selection on mitochondrial DNA in *Parus* and its relevance for phylogeographic studies. Proceedings of the Royal Society B: Biological Sciences, 272(1558), 71–78.

Zuccon, D., & Ericson, P. G. (2010). The *Monticola* rock-thrushes: phylogeny and biogeography revisited. Molecular Phylogenetics and Evolution, 55(3), 901-910.

Table 3.1 Summary table of D-statistics and f_4-ratio using two different phylogenies. Individuals and SNPs were filtered for missing data >10% (n=28, 151,717 SNPs).

((P1,P2)P3)	((EABL,MOBL)WEBL)	((WEBL,MOBL)EABL)
D-statistic	0.174074	0.0447788
Z-score	9.04054	3.04097
p-value	2.3e-16	0.0023582
f_4-ratio	0.193713	0.0238484
BBAA	517.685	672.822
ABBA	672.822	517.685
BABA	473.31	473.31

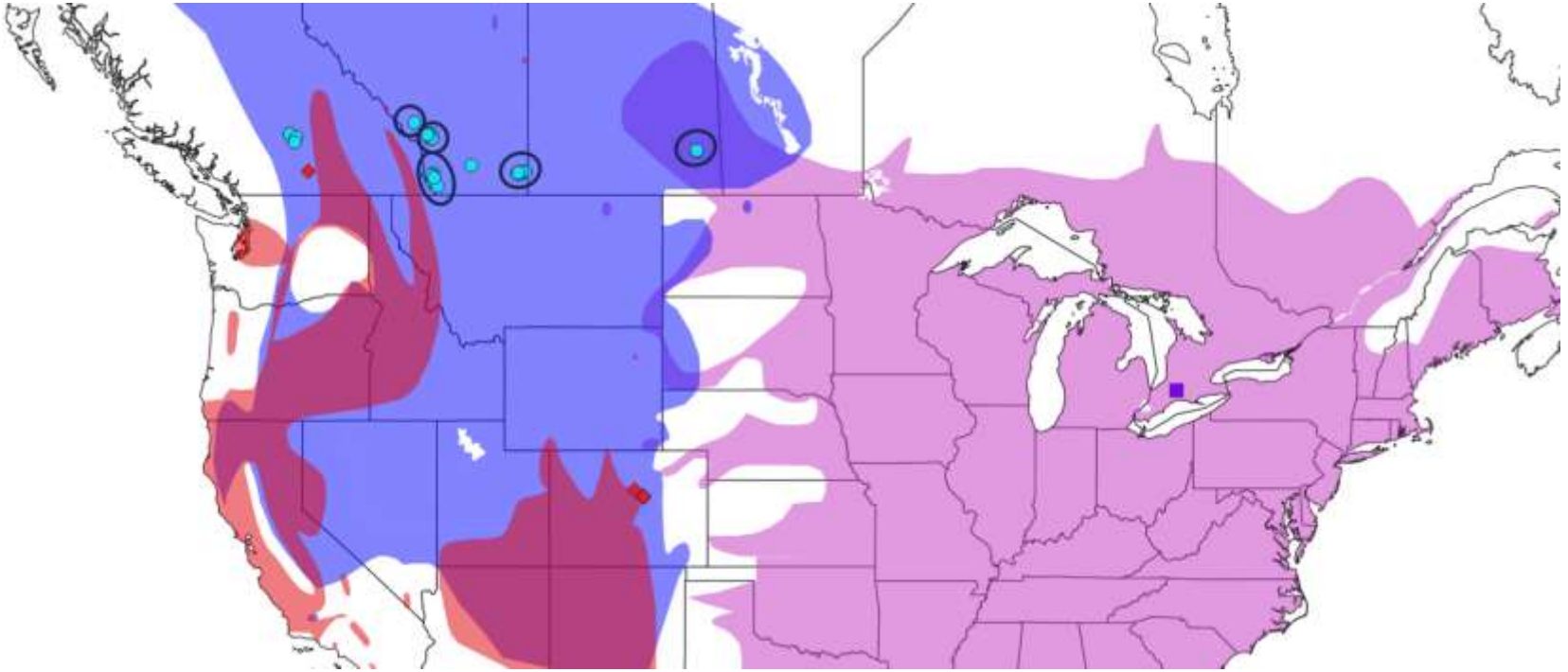

Figure 3.1 Breeding ranges for all three species of bluebird: western (red), mountain (blue), and eastern (purple). Samples used in recent hybridization analyses for each species: western (red diamonds), mountain (blue circles), and eastern (purple square). Field sites are represented by black circles.

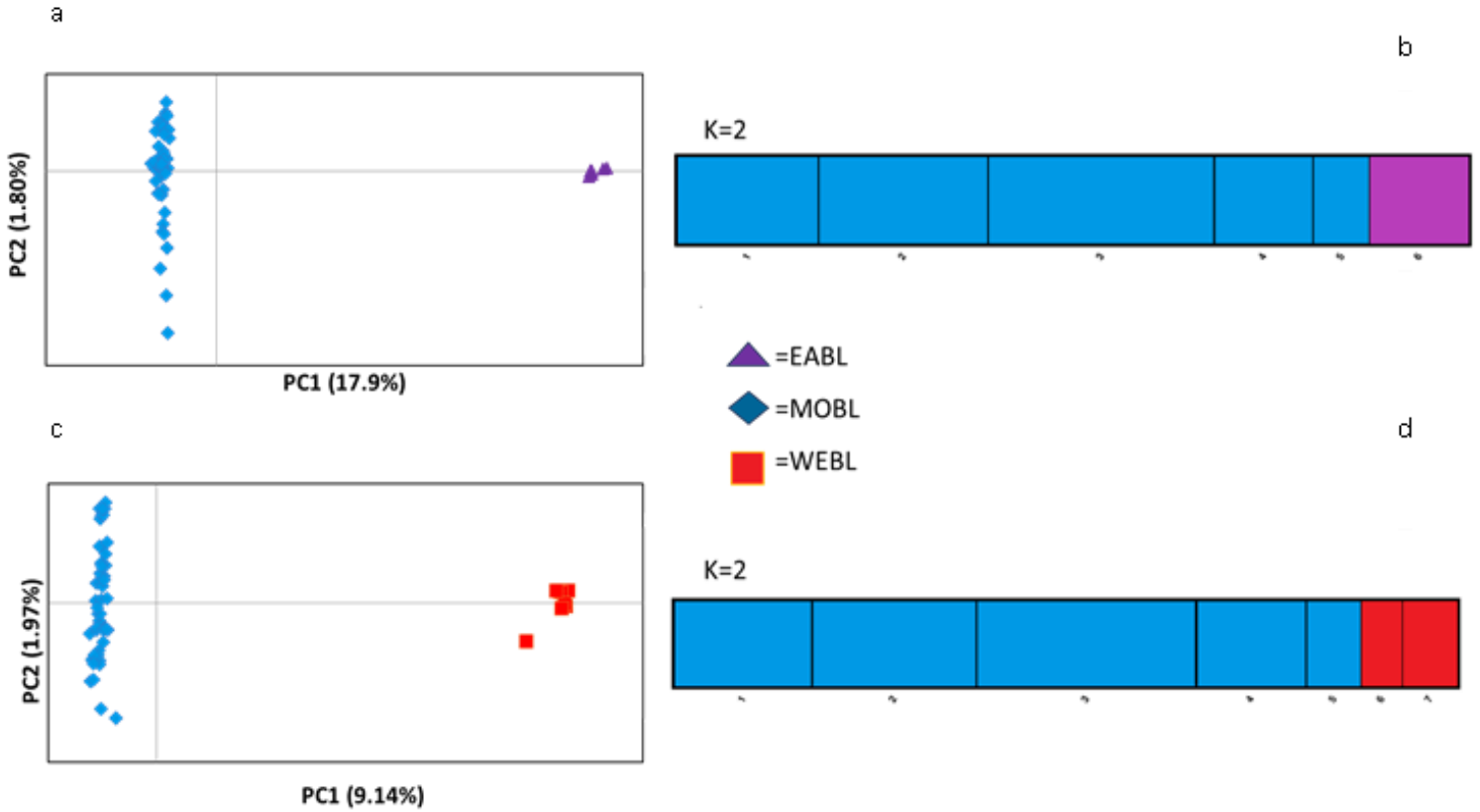

Figure 3.2 **a)** PCoA of genetic distance between mountain bluebirds from Alberta, British Columbia, and Saskatchewan and eastern bluebirds from southwestern Ontario. Individuals were filtered for >35% missing data (n=56) and unlinked SNPs were filtered for >30% data (7,477). **b)** STRUCTURE Ancestry matrix using the same individuals. **c)** PCoA of genetic distance between mountain bluebirds from Alberta, British Columbia, and Saskatchewan and western bluebirds from southeastern British Columbia and Colorado. Individuals were filtered for >35% missing data (n=57) and unlinked SNPs were filtered for >30% data (6,160). **d)** Accompanying STRUCTURE Ancestry matrix using the same individuals.

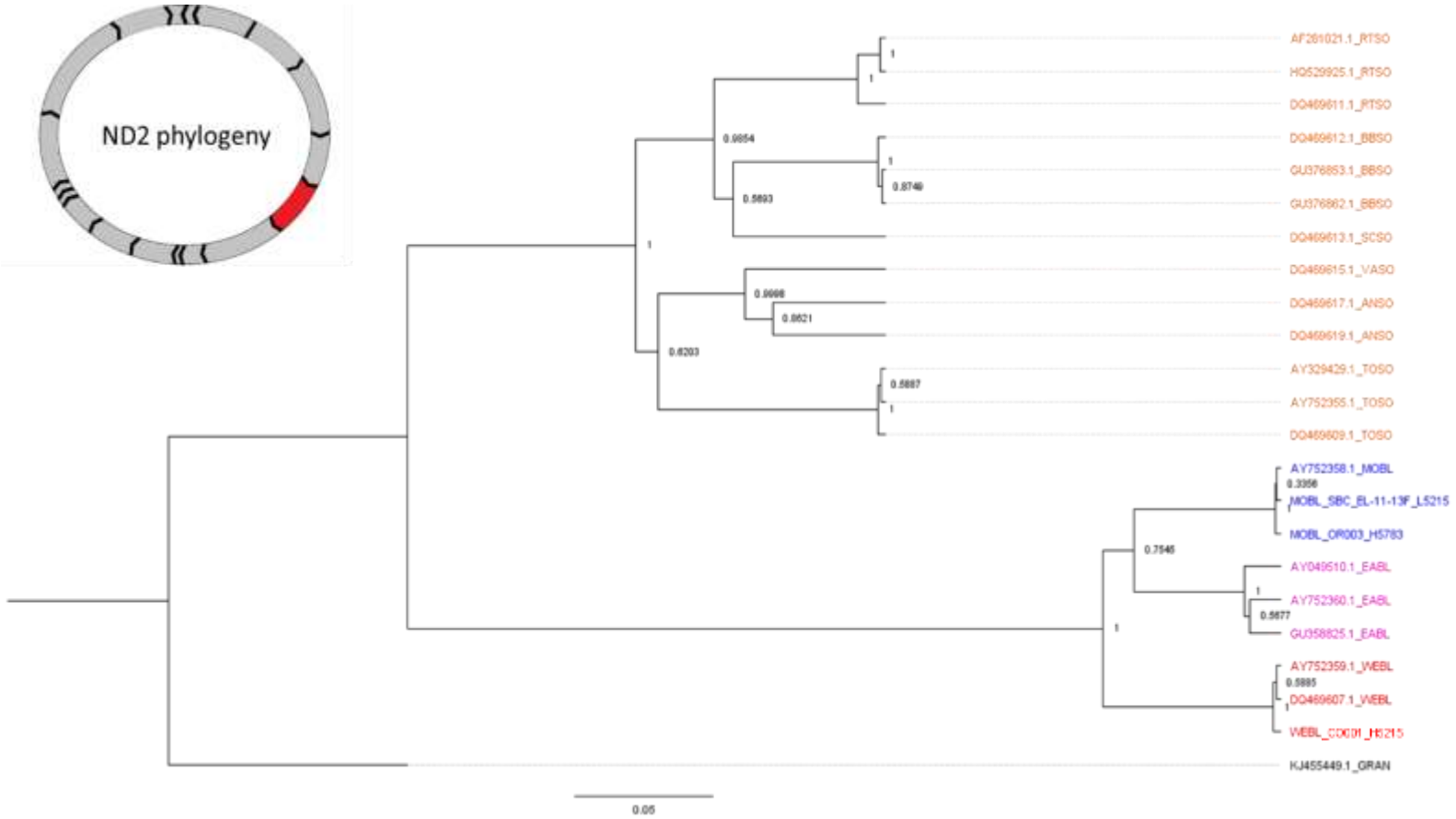

Figure 3.3 Rooted maximum likelihood tree using common ancestor heights for the ND2 gene in mountain (MOBL, blue), western (WEBL, red), and eastern (EABL, purple) bluebirds. The grandala (Gran), Townsend' s solitaire (TOSO), varied solitaire (VASO), slate-coloured solitaire (SCSO), brown-backed solitaire (BBSO), and rufous-throated solitaire (RTSO) used for outgroup. Sequences taken from GenBank start with the accession number, followed by the species. Posterior probabilities are visible at each node.

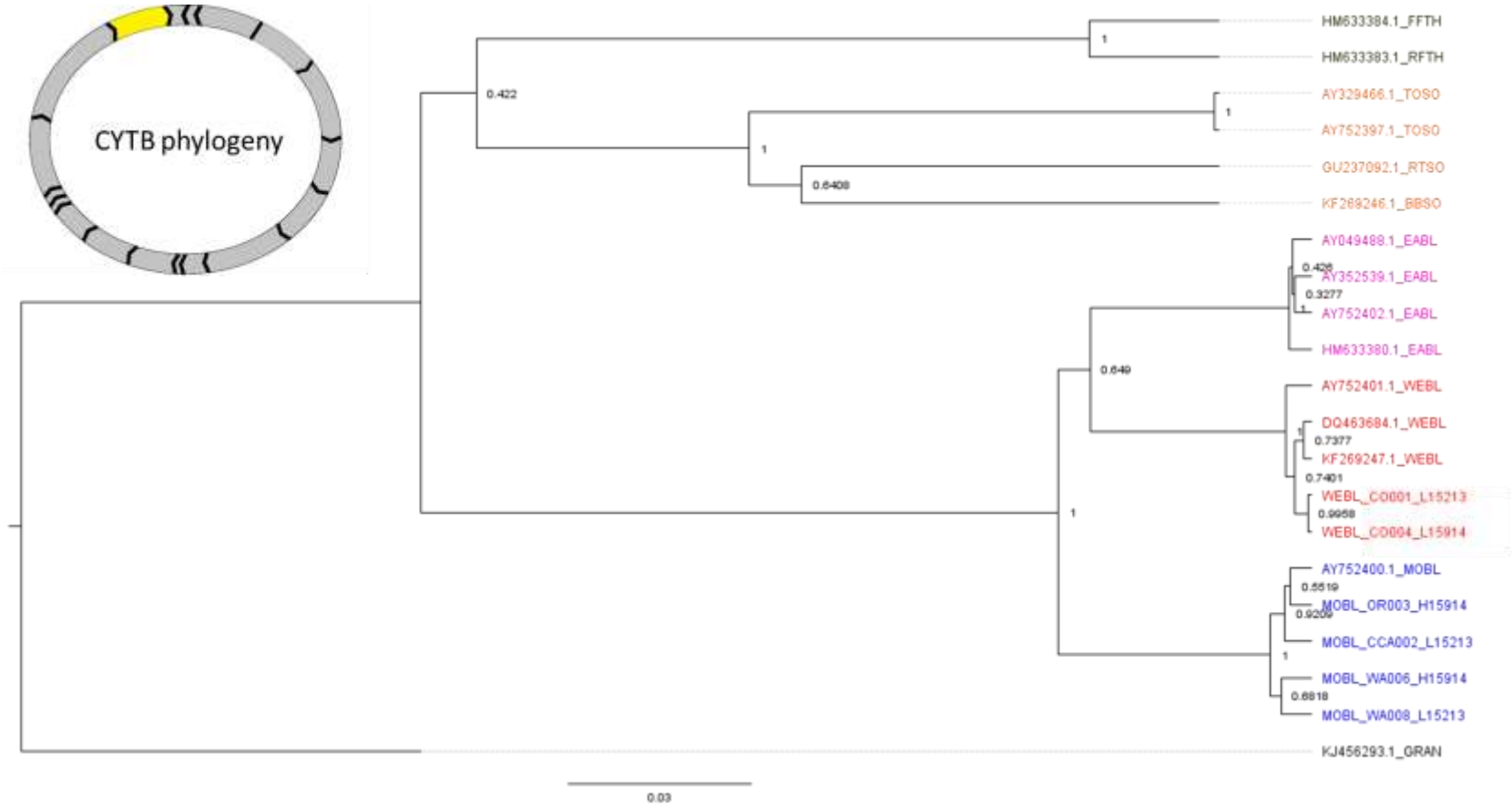

Figure 3.4 Rooted maximum likelihood tree using common ancestor heights for the Cyt b gene in mountain (MOBL, blue), western (WEBL, red), and eastern (EABL, purple) bluebirds. The grandala (Gran), Townsend's solitaire (TOSO), rufous-throated solitaire (RTSO), brown-backed solitaire (BBSO), rufous flycatcher-thrush (RFTH), and Finch's flycatcher-thrush (FFTH) used for outgroup. Sequences taken from Genbank start with the accession number, followed by the species. Posterior probabilities added at each node.

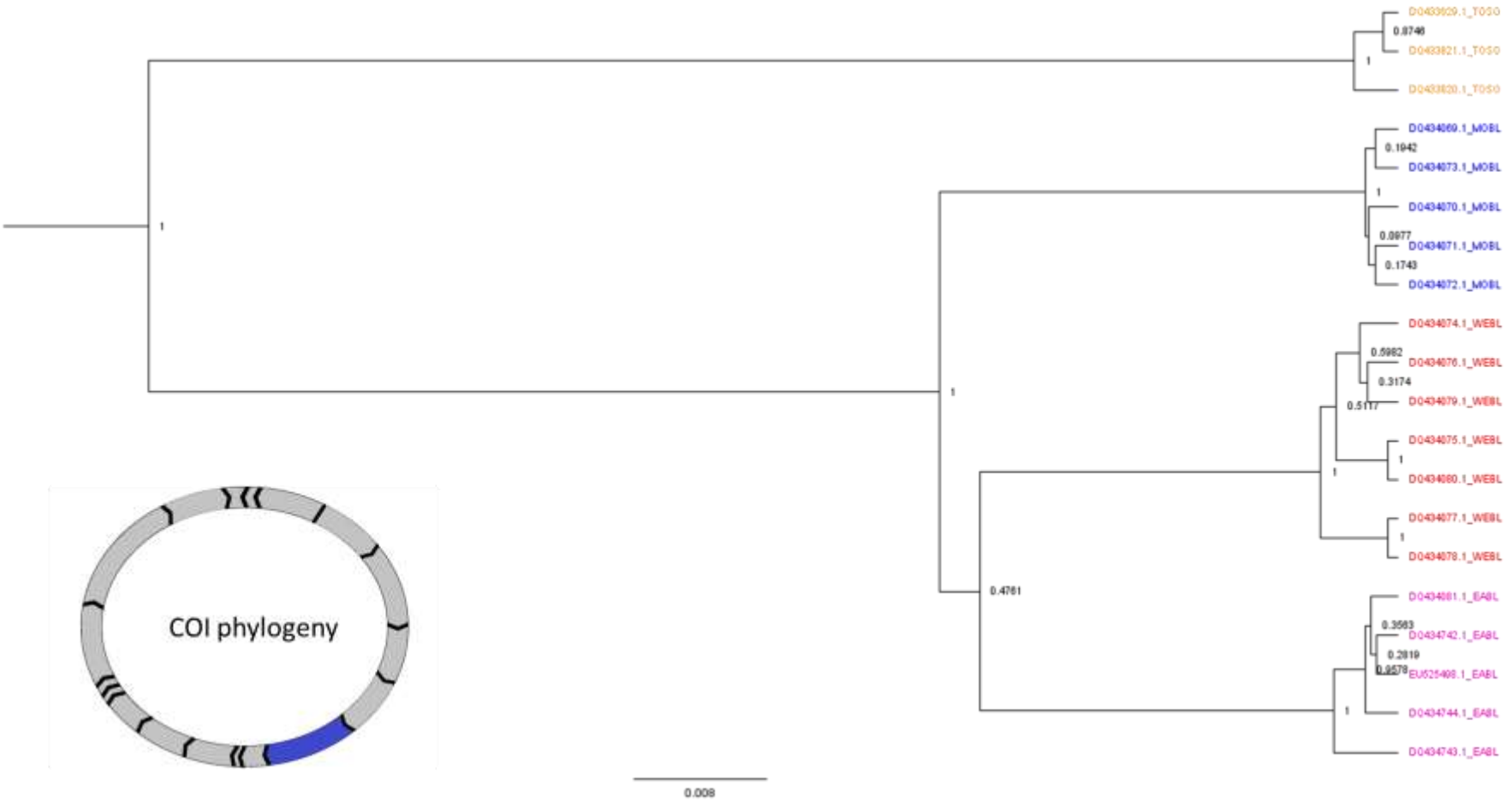

Figure 3.5 Rooted maximum likelihood tree using common ancestor heights for the COI gene in mountain (MOBL, blue), western (WEBL, red), and eastern (EABL, purple) bluebirds, with Townsend' s solitaire (TOSO) used for an outgroup. All sequences include their Genbank accession number followed by the species name. Posterior probabilities are visible at each node.

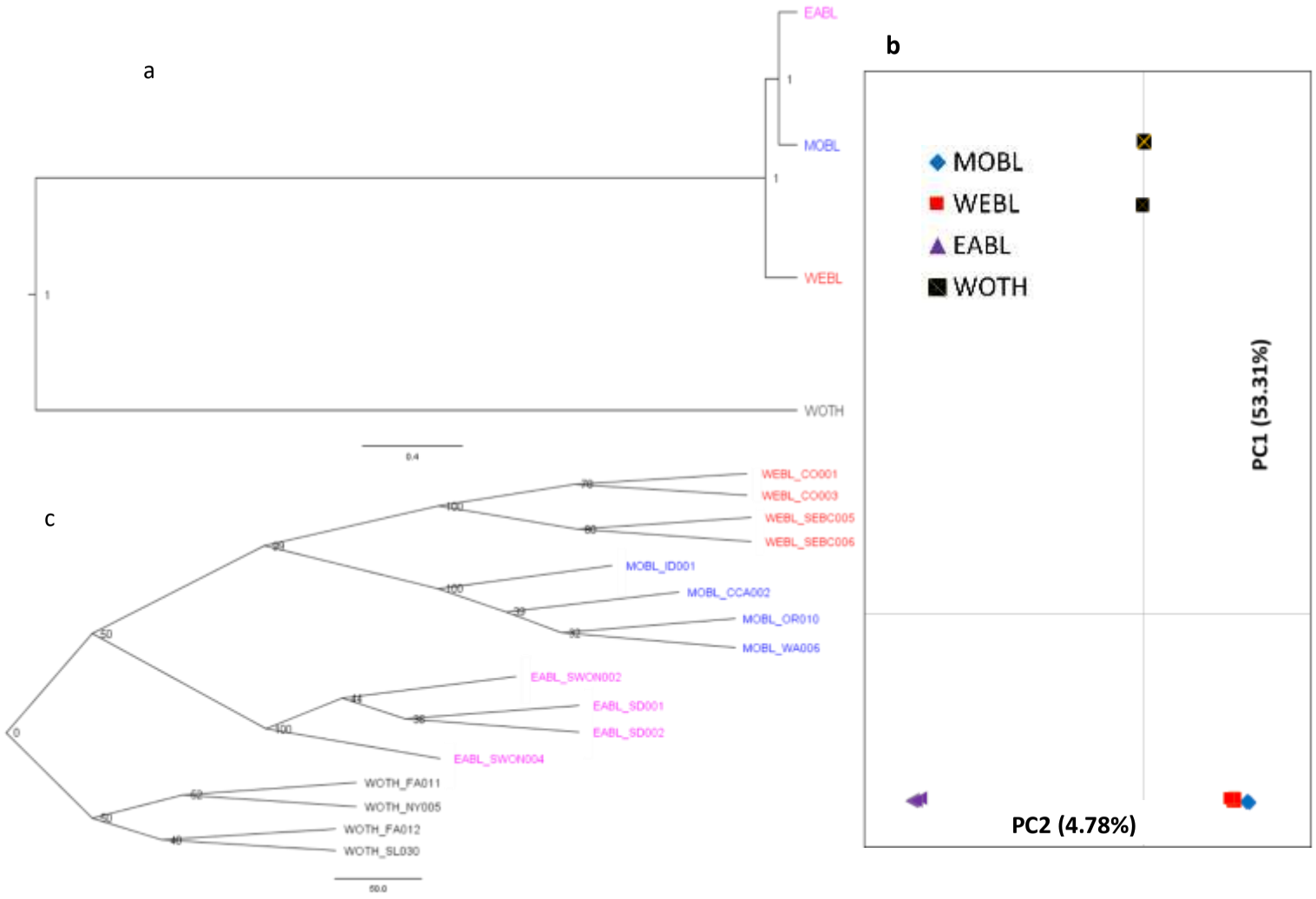

Figure 3.6 **a)** A rooted maximum likelihood tree generated in SNP using a reduced dataset of individuals (n=16) filtered for >10% missing data and 114, 000 unlinked SNPs also filtered at >10% missing data. Posterior probabilities are displayed at each node. **b)** PCoA of genetic distance between all bluebirds and wood thrush **c)**

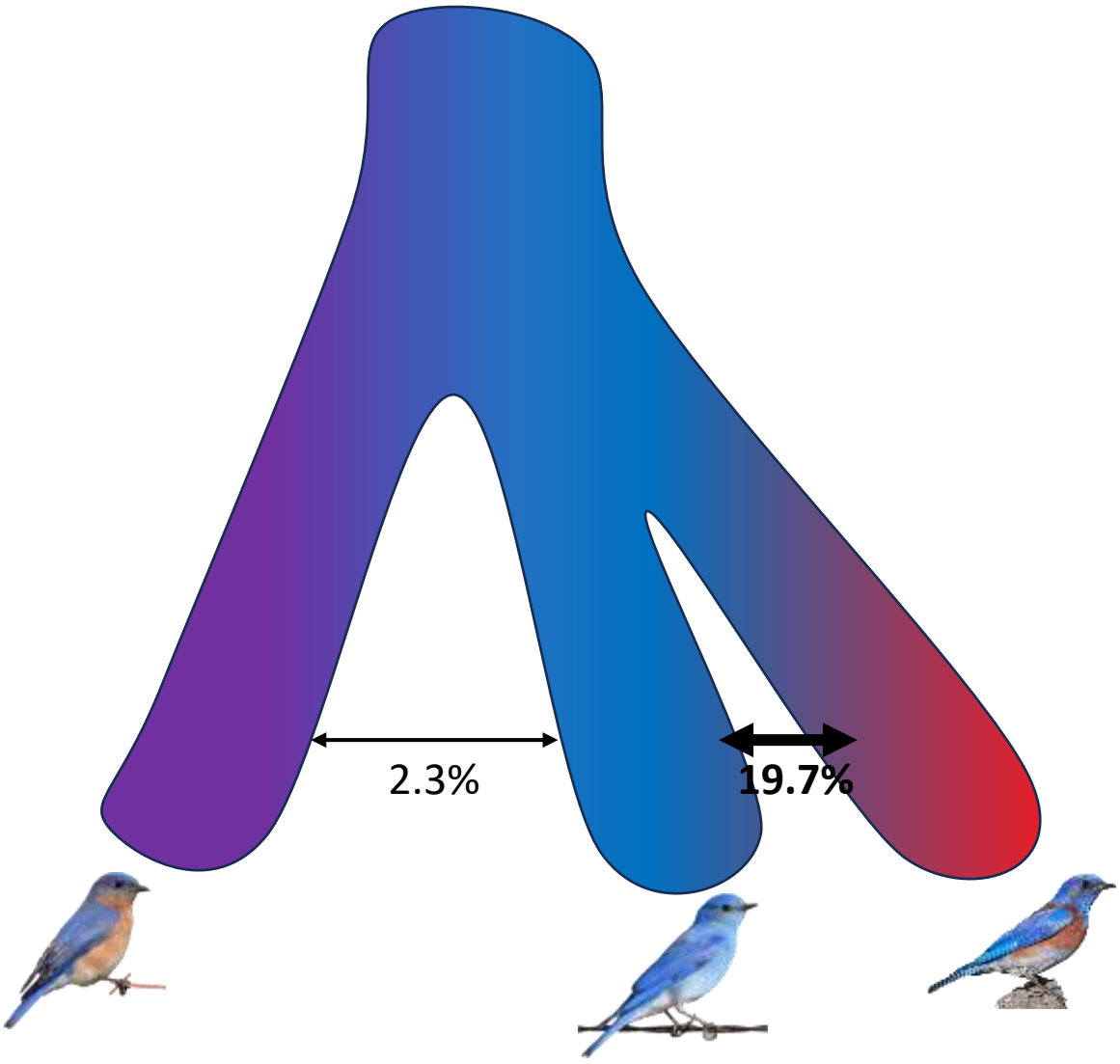

Figure 3.7 A graphical representation of the novel bluebird phylogeny based on the PCoA and SVD quartets with estimated gene flow percentages based on f_4-ratio.

Chapter 4: General Discussion

4.1 General Discussion

Population structure can be difficult to elucidate, especially in highly mobile species with large ranges and nebulous population boundaries (Lah et al., 2016; Nance et al., 2011; Stenzler et al., 2009; Younger et al., 2017). Mountain bluebirds (*Sialia currucoides*) have an expansive range and have adapted to many diverse habitat types provided cavities are available for nesting (Johnson & Dawson, 2020). Mountain bluebirds also exhibit high breeding philopatry. As such, they appear to be excellent candidates for the isolation-by-distance model which posits that increasing geographic distance decreases relatedness of individuals (Wright, 1943). I predicted that mountain bluebirds would demonstrate population structure when sampled across their range—a prediction that was supported using restriction enzyme-associated DNA sequencing (RADseq) data focused on neutral markers. I also predicted that contemporary physical barriers, namely the Rocky Mountains and discontinuous habitat, would inhibit gene flow between isolated populations. Until now, population structure has gone Unknown in mountain bluebirds and my results from distance-based pairwise F_{st}, PCoA and ancestry matrices supported at least four distinct genetic clusters. All results supported a Pacific Northwest complex (composed of Washington, Oregon, Idaho and south-central California, and possibly British Columbia); one in western Alberta (consisting of southwestern Alberta and central Alberta); the Cypress Hills (southeastern Alberta); and one in the southeast Rockies (southeast Montana, Wyoming, and Colorado). The ancestry

matrices also outlined admixture between the Pacific Northwest complex and Nevada, while the PCoA split the Nevada individuals between the Pacific Northwest and the southeastern Rockies. Saskatchewan formed a discrete cluster in the PCoA and ancestry matrices, but a small size precludes deeper analyses.

The separation between the Pacific Northwest complex and populations east of the Rockies supported my prediction that the mountain chain acts as a contemporary barrier to gene flow in mountain bluebirds. The Rockies are a well-documented barrier, inhibiting gene flow in several bird species including mountain chickadees (*Poecile gambeli*) (Spellman et al., 2007); yellow warblers (*Setophaga petechia*) (Milot et al., 2000); red-tailed hawks (*Buteo jamaicensis*) (Hull et al., 2008); and barn owls (*Tyto alba*) (Machado et al., 2018); to name just a few. Although mountain bluebirds have high tolerances for raised elevations (Johnson & Dawson, 2020), the peaks of the Rockies are likely inhibit facilitate gene flow due to the lack of suitable habitat. Fragmented habitat also appeared to inhibit gene flow between the two Alberta population as well. Despite being less than 300 km away from each other (closer than other connected populations within the Pacific Northwest complex), mountain bluebirds in western Alberta showed very little gene flow with the Cypress Hills. The Cypress Hills are an elevated (~1250 m) region home to relic populations of lodgepole pine (*Pinus contorta*), trembling aspen (*Populus tremuloides*), and white spruce (*Picea glauca*) (Sauchyn, 1990). Known as a sky island, the Cypress Hills differ greatly from the semi-arid grasslands that surround them (Sauchyn, 1990). Mountain bluebirds in the west of the province are unlikely to span the gap of unsuitable habitat, inhibiting gene flow between the two populations.

Although mountain bluebirds form distinct clusters in western Alberta and the Cypress Hills in the east of the province, exactly when these populations diverged remains unclear. Shafer et al. (2010) proposed cryptic refugia within the ice sheets during the LGM (~20 kyBP), including one in southwestern Alberta in the form of an ice-free corridor or a nunatak. However, it is unlikely any northern refugia with tundra-like conditions or heavy coniferous canopy would have been suitable habitat for mountain bluebirds at the time. Despite relying on the availability of nest cavities, mountain bluebirds prefer open woodlands with low canopy cover (Holt & Martin, 1997). Even if sections of boreal forest were thinned with frequent wildfires regimes, bluebirds only tend to nest in burn sites that were severely burned 1-29 years prior (Taylor, 1969). Short-lived bursts of colonization followed by extirpation from the burn sites would be unlikely to create the pattern observed. Therefore, it is more likely that mountain bluebirds spread into these habitats as the aspen parkland spread northwards allowing for greater population densities to accrue. In the Pacific Northwest and Nevada, mountain bluebirds likely utilized a mixture of subalpine meadows and forests along with high elevation parkland which accounted for most of the refugia in western North America (Roberts & Hamann, 2015). In the southeastern Rockies, mountain bluebirds likely took advantage of the areas dominated by trembling aspen such as the southwestern Tablelands and the High Plains (Roberts & Hamann, 2015).

Contrary to my prediction, I did not find any evidence of ongoing hybridization between mountain bluebirds and western (*S. mexicana*) or eastern (*S. sialis*) in the western prairie provinces based on the data set. While a lack of recent hybrids may be due to a small sample size, it more likely represents an insufficient population density of

congeners for hybridization to occur. Hubbs' principle (1955) states that numerically imbalanced species are more likely to mate with heterospecifics due to the absence of potential mates. However, both western and eastern bluebirds are likely still too rare in Alberta to create sustained opportunities for hybridization (the same is likely true in Saskatchewan for eastern bluebirds).

Although I did not detect any early stage or advance-generation hybrids, the D-statistic and f_4-ratio tests revealed support for past introgression between mountain and western bluebirds and mountain and eastern bluebirds. The f_4-ratio results demonstrated unequal levels of introgression which varied greatly between 19% between the mountain and western bluebirds and 2% for the mountain and eastern bluebirds. A greater level of introgression mountain and western bluebirds stands to reason; both species currently share much of their breeding ranges in temperate western North America, often only separated by elevational preference (Johnson & Dawson, 2020) and have likely done so for most of the two species' histories. As the two species' ranges expanded and contracted with the cycles of Pleistocene glaciation, it is possible that they formed repeated ephemeral hybrid zones, such as the one observed in western Montana (Duckworth & Semenov 2017). In contrast, mountain and eastern bluebirds are mostly allopatric throughout much of their breeding ranges and only appear to overlap in a few breeding sites (Johnson & Dawson, 2020). Their much more recent history of peripatry likely corresponds to fewer opportunities to interbreed.

Even small amounts of hybridization may have knock-on effects in terms of phylogenetics. I was unable to resolve the polytomy within *Sialia*, even with the novel combination of mitochondrial and nuclear markers, as suggested by Voelker and

Klicka (2008). I did detect strong signatures of mitonuclear discordance between the two types of markers and incomplete lineage sorting (ILS) within the three mitochondrial genes used. The ILS within the mtDNA may be a result of past hybridization events, a pattern supported by evidence from my D-statistic and f_4-ratio tests. Duckworth and Semenov (2017) also noted that hybrid pairing was one-sided, with female mountain pairing with male western bluebird. This may partially explain why mtDNA, which is maternally inherited and usually quicker to introgress compared to nuDNA (Chan & Levin, 2005), fails to place mountain bluebirds as sister taxa to western bluebirds based on the gene trees while nuDNA does. Alternatively, differences between mtDNA gene trees and nuDNA phylogenies may be due selective pressures acting strongly on mitochondrial genes which act a as single molecule without recombination.

4.2 Future Directions

Identifying how hybridization and gene flow have shaped mountain bluebirds in the past and present allows us to make important predictions on the future of the species. Like all bluebirds, mountain bluebirds are a charismatic species beloved by many for both their beauty and important ecological contributions as insectivores. As climate change and anthropogenic habitat alterations allow their congeners to expand, a better understanding of how the three species interact in places of overlap will be increasingly important. In this study, I provided novel insights into the population structure of mountain bluebirds and added to the existing knowledge of bluebird hybridization along with the nuances of their phylogenetic relationships. However, there are still many unexplored avenues. In terms of sample size, underrepresenting peripheral populations

risks missing out on potentially important trends. For example, the North American Breeding Bird Survey noted that between 1968 to 2015, mountain bluebirds experienced a decline of 4.9% per year in Manitoba and a decline of 2.7% in Saskatchewan (Sauer et al., 2017). These declines appear to be corroborated by anecdotal evidence from long time monitoring studies (Johnson & Dawson, 2020), as well as my own field experiences trying to find individuals in the southeast of Saskatchewan. Additional samples from the northeastern portion of mountain bluebirds' range would provide a better idea as to the loss of genetic diversity through measurements of nucleotide diversity and inbreeding indexes. Improved sampling from other peripheral sites that are not declining (e.g., central British Columbia or southern California) would also give a better understanding of the genetic variation that exists within mountain bluebirds.

Peripheral populations also serve as an excellent contact point between the other bluebird species as well. At the current time, western and eastern bluebirds are still too rare in Alberta to use as robust sample sites for either species. Researchers focused on hybridization should include more western bluebird samples from locations with known hybrids, such as western Montana, where hybrids have been the subject of an existing long term study (see Duckworth & Semenov, 2017). For eastern bluebirds, existing studies on hybridization are lacking especially from the northwestern range where they overlap with mountain bluebirds. More samples from Saskatchewan and Manitoba would provide a better idea of hybridization dynamics in central Canada.

Contemporary hybridization is only part of the picture, however. Understanding past hybrid events can indicate whether introgression contributes significantly to the ILS among mitochondrial genes and the overall mitonuclear discordance. The history of

hybridization is likely further confounded by a rapid speciation event (with the proto-*Sialia* ancestor emerging ~3 myBP) (Klicka et al., 2005). Despite containing only three extant species, the bluebird polytomy has proven to be challenging, but the genus also has the potential to be an excellent model system for better understanding hybridization and mitonuclear discordance in birds as a whole. My study was limited by the lack of continuity amongst the genes used and the individuals, especially in the case of the COI gene where all sequences were obtained from public databases. Using sequences from different individuals for each gene bolsters samples size, but it also makes concatenation unfeasible with each gene tree being independent of each other. Independent trees based off single genes are unlikely to provide a wholistic and accurate representation of the entire genome's history. A possible solution is to use whole genome sequencing (WGS). WGS would also allow for the concatenation of all mitochondrial genes in concert with nuDNA for the same individuals. It would also provide additional mtDNA markers as well as the remainder of the nuclear genome to determine where selection may be acting upon and the overall impact of hybridization.

4.3 General Conclusions

In this study, I used RADseq data to analyze the population structure within a widespread North American songbird, mountain bluebirds, using an array of putatively neutral markers. From that population structure, I then suggested possible glacial refugia that the species may have used to survive in during the LGM. In chapter three, I used RADseq data again to determine the presence or absence of early and late-stage hybrid bluebirds in the western prairie provinces. Despite a lack of contemporary hybrids in the

dataset, I did find evidence of past admixture events between mountain bluebirds and its two congeners. Unsurprisingly, admixture appeared greatest between the two westernmost species, (western and mountain bluebirds) indicating multiple ephemeral hybrid zones arose over time while in isolation from their eastern counterpart. I also attempted to address the polytomy within the bluebird genus using nuclear markers in concert with mitochondrial genes. I found ample evidence of ILS amongst bluebird mitochondrial genes and nuclear discordance between the mtDNA and the nuDNA. Overall, mountain bluebirds appeared to exhibit previously unknown population structure at least in part due to IBD and physical barriers to gene flow. Additionally, there is evidence that they survived in multiple refugia during the last LGM. Although bluebird species do not appear to be hybridizing in the western prairie provinces (at least at widespread, detectable levels), evidence points to multiple hybridization events amongst the three species in the past. Ancient hybridization events previously unknown to researchers may also account for the clear ILS occurring within mitochondrial genes, as well as the mitonuclear discordance that contributes to the current polytomy. Ultimately, the history and impact of hybridization within *Sialia* appears complex and opens numerous avenues for future research into the subject.

4.4 Literature Cited

Chan, K. M., & Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution, 59(4), 720-729.

Duckworth, R. A., & Semenov, G. A. (2017). Hybridization associated with cycles of ecological succession in a passerine bird. The American Naturalist, 190(4), E94-E105.

Holt, R. F., and K. Martin (1997). Landscape modification and patch selection: The demography of two secondary cavity nesters colonizing clearcuts. Auk 114:443-455.

Hubbs, C. L. (1955). Hybridization between fish species in nature. Systematic Zoology, 4(1), 1-20.

Hull, J. M., Ernest, H. B., Harley, J. A., Fish, A. M., & Hull, A. C. (2009). Differential migration between discrete populations of juvenile Red-tailed Hawks (*Buteo jamaicensis*). The Auk, 126(2), 389-396.

Johnson, L. S. and R. D. Dawson (2020). Mountain bluebird (*Sialia currucoides*), version 1.0. In Birds of the World (P. G. Rodewald, Editor). Cornell Lab of Ornithology.

Klicka, J., Voelker, G., & Spellman, G. M. (2005). A molecular phylogenetic analysis of the "true thrushes" (Aves: Turdinae). Molecular Phylogenetics and Evolution, 34(3), 486-500.

Lah, L., Trense, D., Benke, H., Berggren, P., Gunnlaugsson, Þ., Lockyer, C., Öztürk, A., Öztürk, B., Pawliczka, I., Roos, A. and Siebert, U. (2016). Spatially explicit analysis of genome-wide SNPs detects subtle population structure in a mobile marine mammal, the harbor porpoise. PloS One, 11(10), p.e0162792.

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

Milot, E., Gibbs, H. L., & Hobson, K. A. (2000). Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Molecular Ecology, 9(6), 667-681.

Nance, H. A., Klimley, P., Galván-Magaña, F., Martínez-Ortíz, J., & Marko, P. B. (2011). Demographic processes underlying subtle patterns of population structure in the scalloped hammerhead shark, *Sphyrna lewini*. PLoS One, 6(7), e21459.

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. Proceedings of the Royal Society B: Biological Sciences, 282(1804), 20142903.

Sauer, J. R., Niven, D. K., Hines, J. E., Ziolkowski Jr, D. J., Pardieck, K. L., Fallon, J. E., & Link, W. A. (2017). The North American breeding bird survey, results and analysis 1966–2015. Version 2.07. 2017. Laurel (MD): USGS Patuxent Wildlife Research Center.

Sauchyn, D. J. (1990). A reconstruction of Holocene geomorphology and climate, western Cypress Hills, Alberta and Saskatchewan. Canadian Journal of Earth Sciences, 27(11), 1504–1510. doi:10.1139/e90-160.

Shafer, A. B., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology, 19(21), 4589-4621.

Spellman, G. M., Riddle, B., & Klicka, J. (2007). Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to Quaternary climate change. Molecular Ecology, 16(5), 1055-1068.

Stenzler, L. M., Makarewich, C. A., Coulon, A., Ardia, D. R., Lovette, I. J., & Winkler, D. W. (2009). Subtle edge-of-range genetic structuring in transcontinentally distributed North American Tree Swallows. The Condor, 111(3), 470-478.

Taylor, D. L. (1969). Biotic succession of lodgepole pine forests in Yellowstone National Park. Ph.D. dissertation, University of Wyoming, Laramie, WY, USA.

Voelker, G., & Klicka, J. (2008). Systematics of *Zoothera* thrushes, and a synthesis of true thrush molecular systematic relationships. Molecular Phylogenetics and Evolution, 49(1), 377-381.

Wright, S. (1943). Isolation-by-distance. Genetics, 28(2), 114.

Younger, J. L., Clucas, G. V., Kao, D., Rogers, A. D., Gharbi, K., Hart, T., & Miller, K. J. (2017). The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. Molecular Ecology, 26(15), 3883-3897.

Appendix 1: Supplementary Information for Chapter 2

Neutral markers reveal subtle population structure in a widespread songbird

Appendix 1.1 Sample ID, location, coordinates, and band/museum ID for each individual used in Chapter 2, listed by population. Museum samples from the Burke Museum of Natural History and Culture (UWBM); Denver Museum of Nature and Science (DMNS); and the Museum of Southwestern Biology

ID	Location	Lat(° N)	Long(° W)	Source	Band/Museum ID
MOBL_CAB001	RR25, north of Black Diamond, AB	50.739231	-114.278847	WILD	1461-24182
MOBL_CAB002	Hwy 546 by Bar n' Ranch, Turner Valley, AB	50.659507	-114.350579	WILD	1461-24183
MOBL_CAB003	RR34, Millarville, AB	50.724286	-114.371092	WILD	1461-24186
MOBL_CAB004	RR34, Millarville, AB	50.724286	-114.371092	WILD	1461-24187
MOBL_CAB005	RR32 off Twn 194, Turner Valley, AB	50.622485	-114.324948	WILD	2951-01928
MOBL_CAB006	RR32 off Twn 194, Turner Valley, AB	50.622485	-114.324948	WILD	1461-24190
MOBL_CAB007	Township Rd 194, Turner Valley, AB	50.630146	-114.314822	WILD	1461-24193
MOBL_CAB008	Township Rd 194 (part 2), Turner Valley, AB	50.63018	-114.278893	WILD	1461-24194
MOBL_CAB009	Township Rd 194 (part 2), Turner Valley, AB	50.63018	-114.278893	WILD	1461-24195
MOBL_CAB010	Township Rd 194 (part 2), Turner Valley, AB	50.63018	-114.278893	WILD	2951-19805
MOBL_CAB011	Township Rd 210, Millarville, AB	50.753855	-114.463482	WILD	1461-24198
MOBL_CAB012	RR25, north of Black Diamond, AB	50.739231	-114.278847	WILD	1461-24180
MOBL_CAB013	RR25, north of Black Diamond, AB	50.739231	-114.278847	WILD	1461-24181
MOBL_CAB014	Hwy 546 by Bar n' Ranch, Turner Valley, AB	50.659507	-114.350579	WILD	1461-24184
MOBL_CAB015	RR34, Millarville, AB	50.724286	-114.371092	WILD	1461-24185
MOBL_CAB016	Township Rd 195, Turner Valley, AB	50.644708	-114.331262	WILD	1461-24188
MOBL_CAB017	Township Rd 195, Turner Valley, AB	50.644708	-114.331262	WILD	1461-24189
MOBL_CAB018	Township Rd 194, Turner Valley, AB	50.630146	-114.314822	WILD	1461-24191
MOBL_CAB019	Township Rd 194, Turner Valley, AB	50.630146	-114.314822	WILD	1461-24192
MOBL_CAB020	Township Rd 194 (part 2), Turner Valley, AB	50.63018	-114.278893	WILD	1461-24196
MOBL_CAB021	Township Rd 210, Millarville, AB	50.753855	-114.463482	WILD	1461-24197
MOBL_CAB022	Township Rd 210, Millarville, AB	50.753855	-114.463482	WILD	2981-10506
MOBL_CAB023	Township Rd 204A, Millarville, AB	50.724604	-114.437862	WILD	1461-24199
MOBL_CAB024	Township Rd 204A, Millarville, AB	50.724604	-114.437862	WILD	1461-24200
MOBL_CAB025	RR42, Millarville, AB	50.72565	-114.463245	WILD	1271-84204

MOBL_CAB026	Township Rd 204A, Millarville, AB	50.724604	-114.437862	WILD	2791-66420
MOBL_CAB027	Ranch Rd, Seebe, AB	51.078094	-115.044972	WILD	1271-84205
MOBL_CBC001	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79745
MOBL_CBC002	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79746
MOBL_CBC003	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79750
MOBL_CBC004	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79754
MOBL_CBC005	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79755
MOBL_CBC006	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79757
MOBL_CBC007	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79758
MOBL_CBC008	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79759
MOBL_CBC009	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79762
MOBL_CBC010	Alkali Lake, BC	51.787552	-122.228681	UNBC	2741-79764
MOBL_CCA001	Wyman Creek Road, east slope of White Mountains, CA	37.368205	-118.400431	UWBM	122376
MOBL_CCA002	Bishop, CA	37.430469	-118.078315	UWBM	117763
MOBL_CO002	Chatfield Reservoir, 500 feet east of marina, CO	39.542711	-105.061291	DMNS	43255
MOBL_CO003	Highlands Ranch, 4800 McArthur Ranch Road, CO	39.523614	-104.934161	DMNS	45969
MOBL_CO005	Lyons, 702 Ponderosa Hill Road, CO	40.290286	-105.27956	DMNS	48333
MOBL_CO006	Parker, 10958 Touchstone Loop, CO	39.517342	-104.828582	DMNS	48681
MOBL_CO007	Nathrop, Elk Run, CO	38.675947	-106.11956	DMNS	52292
MOBL_CO008	Loveland, 1251 Turkey Walk trail	40.375507	-105.263089	DMNS	54422
MOBL_CO009	37°25.77'N, 107°45.04'W, CO	37.429571	-107.75071	UWBM	109294
MOBL_CO010	37°25.77'N, 107°45.04'W, CO	37.429571	-107.750653	UWBM	101774
MOBL_CYP001	Cypress Hills, AB	49.667915	-110.267791	WILD	1461-24125
MOBL_CYP002	Cypress Hills, AB	49.66487	-110.15219	WILD	1461-24126
MOBL_CYP003	Cypress Hills, AB	49.66487	-110.15219	WILD	1461-24127
MOBL_CYP004	Cypress Hills, AB	49.66856	-110.14986	WILD	1461-24128
MOBL_CYP005	Cypress Hills, AB	49.66856	-110.14986	WILD	1461-24129
MOBL_CYP006	Cypress Hills, AB	49.66178	-110.16735	WILD	1461-24130
MOBL_CYP007	Cypress Hills, AB	49.66772	-110.26599	WILD	1461-24131
MOBL_CYP008	Cypress Hills, AB	49.655321	-110.457639	WILD	1461-24132
MOBL_CYP009	Cypress Hills, AB	49.648027	-110.448721	WILD	1461-24133
MOBL_CYP010	Cypress Hills, AB	49.604692	-110.441782	WILD	1461-24134
MOBL_CYP011	Cypress Hills, AB	49.59826	-110.470626	WILD	1461-24135
MOBL_CYP012	Cypress Hills, AB	49.64286	-110.439802	WILD	MOBL 1 (egg)
MOBL_ID001	Elk Meadow, ID	44.118368	-114.885108	UWBM	121932
MOBL_ID002	Elk Meadow, ID	44.118368	-114.885108	UWBM	122006

MOBL_ID003	Tyndall Meadows, ID	44.563234	-115.5509	UWBM	122051
MOBL_MT001	45°13'N, 110°40'W, MT	45.216682	-110.666656	UWBM	100267
MOBL_MT002	45°13'N, 110°40'W, MT	45.216682	-110.666656	UWBM	100268
MOBL_MT003	45°13'N, 110°40'W, MT	45.216682	-110.666656	UWBM	100273
MOBL_NV001	39°20.5'N,114°13.0'W, NV	39.341667	-114.216667	UWBM	104812
MOBL_NV002	39°20.5'N,114°13.0'W, NV	39.341667	-114.216667	UWBM	104813
MOBL_NV003	39°20.5'N/114°13.0'W, NV	39.341667	-114.216667	UWBM	104853
MOBL_NV004	36.25.5'N/115.45'W, NV	36.425	-115.75	UWBM	109692
MOBL_NV005	36.25.5'N,115.45'W, NV	36.425	-115.75	UWBM	109693
MOBL_NV006	36.25.5'N,115.45'W, NV	36.425	-115.75	UWBM	109694
MOBL_NV007	Spring Mountains; Deer Creek at north end, NV	36.340433	-115.652415	UWBM	112035
MOBL_NV008	36°24.62'N, 12, NV	36.340433	-115.652415	UWBM	112036
MOBL_NV009	38°51.598'N, 116°36.628'W, NV	38.859967	-116.610467	UWBM	115347
MOBL_NV010	38°46.667'N, 116°37.827', NV	38.777783	-116.63045	UWBM	115356
MOBL_OR001	Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	122390
MOBL_OR002	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	116080
MOBL_OR003	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	100469
MOBL_OR004	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	101914
MOBL_OR005	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	101912
MOBL_OR006	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	101913
MOBL_OR007	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	101911
MOBL_OR008	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	101910
MOBL_OR009	forest road 39, Wallowa-Whitman National Forest, OR	44.363253	-120.314385	UWBM	100466
MOBL_OR010	forest road 39, Wallowa-Whitman	44.363253	-120.314385	UWBM	100470

	National Forest, OR				
MOBL_SAB001	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24143
MOBL_SAB002	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24144
MOBL_SAB003	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24145
MOBL_SAB004	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24146
MOBL_SAB005	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24147
MOBL_SAB006	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24148
MOBL_SAB007	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24149
MOBL_SAB008	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24150
MOBL_SAB009	RR25A, S of Beaver Mines, AB	49.3961	-114.257422	WILD	1461-24153
MOBL_SAB010	RR25A, S of Beaver Mines, AB	49.3961	-114.257422	WILD	1461-24154
MOBL_SAB011	Township Rd 55B, Beaver Mines, AB	49.426222	-114.180194	WILD	1461-24155
MOBL_SAB012	Township Rd 55B, Beaver Mines, AB	49.426222	-114.180194	WILD	1461-24156
MOBL_SAB013	Township Rd 55B, Beaver Mines, AB	49.426222	-114.180194	WILD	1461-24157
MOBL_SAB014	Township Rd 55B, Beaver Mines, AB	49.426222	-114.180194	WILD	1461-24158
MOBL_SAB015	Township Rd 55B, Beaver Mines, AB	49.426222	-114.180194	WILD	MOBL 2
MOBL_SAB016	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24159
MOBL_SAB017	RR30, Burmis, AB	49.582157	-114.272346	WILD	1461-24160
MOBL_SAB018	Nikkel Tempest (N), Lethbridge, AB	49.8578	-112.51552	MBTS	MOBL 10
MOBL_SAB019	RR25, east of Burmis, AB	49.594078	-114.243035	WILD	1461-24162
MOBL_SAB020	RR25, east of Burmis, AB	49.594078	-114.243035	WILD	1461-24163
MOBL_SAB021	RR25, east of Burmis, AB	49.594078	-114.243035	WILD	1461-24164
MOBL_SAB022	RR23A, north of Lundbreck, AB	49.611402	-114.21314	WILD	1461-24165
MOBL_SAB023	RR24B, north of Lundbreck, AB	49.571372	-114.227283	WILD	1461-24166
MOBL_SAB024	RR24B, north of Lundbreck, AB	49.571372	-114.227283	WILD	1461-24167
MOBL_SAB025	Spread Eagle Rd, AB	49.25329	-114.00112	MBTS	MOBL 9
MOBL_SAB026	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24169
MOBL_SAB027	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24170
MOBL_SAB028	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24171
MOBL_SAB029	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	MOBL 3
MOBL_SAB030	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24172
MOBL_SAB031	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24173

MOBL_SAB032	RR22 of HWY 507, near Burmis, AB	49.527066	-114.186116	WILD	1461-24174
MOBL_SAB033	RR30, Burmis, AB	49.582157	-114.272346	WILD	MOBL 4
MOBL_SAB034	Township Rd 64A, south of Burmis, AB	49.50161	-114.152732	WILD	1461-24175
MOBL_SAB035	Township Rd 64A, south of Burmis, AB	49.50161	-114.152732	WILD	1461-24176
MOBL_SAB036	Township Rd 63A, south of Burmis, AB	49.485275	-114.135172	WILD	1461-24177
MOBL_SAB037	Township Rd 63A, south of Burmis, AB	49.485275	-114.135172	WILD	1461-24178
mobl_DD-01-19N1	Kamloops, BC	50.748442	-120.556584	TRU	8001-07034
mobl_DD-07-19N1	Kamloops, BC	50.748442	-120.556584	TRU	8001-07007
mobl_DD-18-19N1	Kamloops, BC	50.748442	-120.556584	TRU	8001-07001
mobl_DD-23-19N1	Kamloops, BC	50.748442	-120.556584	TRU	8001-07022
mobl_EL-30-19N1	Kamloops, BC	50.557808	-120.347142	TRU	2741-86746
mobl_J-02-19N1	Kamloops, BC	50.654852	-120.278928	TRU	2741-86757
mobl_J-08-19N1	Kamloops, BC	50.654852	-120.278928	TRU	2741-86762
mobl_LL-06-19N1	Kamloops, BC	50.503467	-120.338701	TRU	2741-87610
mobl_LL-10-19N1	Kamloops, BC	50.503467	-120.338701	TRU	2741-86720
mobl_LL-14-19N3	Kamloops, BC	50.503467	-120.338701	TRU	2741-87616
MOBL_SK001	Estevan, SK	49.04432	-102.45071	WILD	1461-24136
MOBL_SK002	10 km SW of Whitewood, SK	50.25243	-102.51045	WILD	1461-24138
MOBL_SK003	Whitewood, SK	50.27047	-102.52159	WILD	1461-24139
MOBL_SK004	10 km SW of Whitewood, SK	50.25212	-102.51027	WILD	1461-24140
MOBL_SK005	Whitewood, SK	50.735753	-104.719713	WILD	1461-24141
MOBL_SK006	7 km NE of Craven, SK	50.769681	-104.679601	WILD	1461-24142
MOBL_WA001	West Branch LeClerc Creek, WA	48.688515	-117.51603	UWBM	121316
MOBL_WA002	Misery Drift Inn, WA	46.122631	-117.514364	UWBM	122458
MOBL_WA003	Oroville, WA	48.9101	-119.5235	UWBM	72619
MOBL_WA004	Naneum Creek, WA	47.128533	-120.481401	UWBM	121334
MOBL_WA005	Pomeroy, WA	46.46524	-117.592029	UWBM	122661
MOBL_WA006	Stewart Meadow, WA	48.679288	-117.518769	UWBM	122493
MOBL_WA007	Stewart Meadow, WA	48.679288	-117.518769	UWBM	122338
MOBL_WA008	Cox Meadow, WA	48.489972	-119.056661	UWBM	122634
MOBL_WA009	near FR 46 (Kendall Skyline Drive), WA	46.050604	-117.909241	UWBM	119898
MOBL_WA010	Darland Mountain, WA	46.484703	-121.234776	UWBM	119522
MOBL_WY001	12 miles north of	44.20397	-107.40025	DMNS	45088

	Tensleep, WY\				
MOBL_WY002	Green Mountains, WY	42.33115	-107.692317	DMNS	45983
MOBL_WY003	Green Mountains, WY	42.355433	-107.714367	DMNS	45987
MOBL_WY004	Ferris Mountains, WY	42.28561667	-107.311567	DMNS	46023
MOBL_WY005	Powder Mountain Wyoming, Camp 2, WY	41.04946	-108.245689	DMNS	50307
MOBL_WY006	Pine Mountain, WY	41.059028	-108.96669	DMNS	50364
MOBL_WY007	Medicine Bow National Forest, WY	41.107813	-107.302244	DMNS	50366

Appendix 2: Supplementary Information for Chapter 3

RADseq and mitochondrial sequencing reveal clear mitonuclear discordance and signals of historic introgression within the genus *Sialia*

Appendix 2.1 Sample ID, location, coordinates, and band/museum ID for each individual used in Chapter 3, listed by population. Museum samples from Denver Museum of Nature and Science (DMNS), with remainder provided by Queen's University (QU) and Environment and Climate Change Canada (ECCC).

ID	Location	Lat(° N)	Long(° W)	Source	Band/Mus ID
EABL_ND001	Sandhill, J.Clark Salyer NWR, ND	48.6162	-100.728886	WILD	EABL1
EABL_SD001	26 miles W-NW of Pickstown	43.1690021	-98.80748198	DMNS	44758
EABL_SD002	1 mile E-NE of Murchison	45.391505	-103.253161	DMNS	45192
EABL_SD003	no locality recorded	45.37446	-103.7247	DMNS	45231
EABL_SK001	Estevan, SK	49.10217	-102.6519	WILD	1461-24137
EABL_SWON001	London, ON	42.877521	-81.214293	ECCC	EABL80
EABL_SWON002	London, ON	42.877521	-81.214293	ECCC	EABL85
EABL_SWON003	London, ON	42.877521	-81.214293	ECCC	EABL87
EABL_SWON004	London, ON	42.877521	-81.214293	ECCC	EABL88
EABL_SWON005	London, ON	42.877521	-81.214293	ECCC	EABL121
EABL_SWON006	London, ON	42.877521	-81.214293	ECCC	EABL122
EABL_SWON007	London, ON	42.877521	-81.214293	ECCC	EABL123
EABL_SWON008	London, ON	42.877521	-81.214293	ECCC	EABL126
EABL_SWON009	London, ON	42.877521	-81.214293	ECCC	EABL131
EABL_SWON010	London, ON	42.877521	-81.214293	ECCC	EABL135
WEBL_CO001	Castle Pines, Near Mirage Dr. and Prospect Dr., CO	39.42833	-104.88722	DMNS	34889
WEBL_CO002	Golden, 1298 Northridge Ct	39.688916	-105.294584	DMNS	43277
WEBL_CO003	Castle Pines Village	39.447118	-104.88525	DMNS	43499
WEBL_CO004	N of Cottonwood, CO	39.77475	-105.3755278	WILD	CO117
WEBL_SEBC001	Okanagan, BC	49.682391	-119.723529	QU	2261-23901
WEBL_SEBC002	Okanagan, BC	49.682391	-119.723529	QU	2261-23902
WEBL_SEBC003	Okanagan, BC	49.682391	-119.723529	QU	2261-23903
WEBL_SEBC004	Okanagan, BC	49.682391	-119.723529	QU	2261-23905
WEBL_SEBC005	Okanagan, BC	49.682391	-119.723529	QU	2261-23906
WEBL_SEBC006	Okanagan, BC	49.682391	-119.723529	QU	2261-23907
WEBL_SEBC007	Okanagan, BC	49.682391	-119.723529	QU	2261-23908
WEBL_SEBC008	Okanagan, BC	49.682391	-119.723529	QU	2261-23910
WEBL_SEBC009	Okanagan, BC	49.682391	-119.723529	QU	2261-23911
WEBL_SEBC010	Okanagan, BC	49.682391	-119.723529	QU	2261-23912
WEBL_MOBL_CO001	County Road 74E, 7 miles E of Red Feather Lakes, CO	40.749232	-105.422898	DMNS	34818

WEBL_MOBL_CO004	Conifer, Jubilee Trail, CO	39.459332	-105.380534	DMNS	47180