IDENTIFYING GENETIC STRUCTURE AND DIVERGENCE BETWEEN *Zonotrichia leucophrys* SUBSPECIES and HABITAT ECOTYPES

BUKOLA GBEMISOLA OGUNTUASE-OSAGIE
Doctor of Philosophy, Federal University of Technology Akure Nigeria, 2017

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

IDENTIFYING GENETIC STRUCTURE AND DIVERGENCE BETWEEN *Zonotrichia leucophrys* SUBSPECIES and HABITAT ECOTYPES

BUKOLA G. OGUNTUASE-OSAGIE

Date of Defense: November 2, 2022

Dr. Theresa Burg Thesis Supervisor	Professor	Ph.D.
Dr. Andrew Iwaniuk Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Steve Wiseman Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Cam Goater Chair, Thesis Examination Committee	Professor	Ph.D.

ABSTRACT

Habitat heterogeneity which includes environmental conditions are known to increase species richness and may lead to speciation or genetic structure. This study used a combination of novel approaches to assess genetic differences in populations of two *Zonotrichia leucophrys* subspecies found in two different habitats: alpine coniferous and riparian deciduous forests. Our study also provided genetic support for the divergence within the *Zonotrichia leucophrys* clade. The application of triple digest restriction-site associated DNA sequencing (3dRADseq) and low-coverage whole genome sequencing (lcWGS) showed for the first time clear genetic divergence among *Z. l. pugetensis*, *Z. l. gambelii* and *Z. l. oriantha*, and between the habitat ecotypes. The use of outlier loci showed that it can detect recent divergence. This study has not only clearly defined genetic structure in the habitat ecotypes and subspecies but has also shown that local conditions defining the habitat play a prominent role in genetic structure of populations.

ACKNOWLEDGEMENTS

First, I attribute the success of this work to the immense support received from my supervisor, Professor Theresa Burg, Department of Biological Sciences, Faculty of Arts & Science, University of Lethbridge. I am very grateful for the wealth of knowledge she exposed me to, her effective style of tutelage. Many a times, I have had to do deeper research into several topics to which I could not provide good answers at first, she only needed to ask one question to assess one's level of understanding. I am grateful for her constant encouragement, her patience in dealing with my repetitive mistakes, her candid and helpful opinions, and suggestions and very importantly, her understanding and support in times of unforeseeable situations and family emergencies that happened too many times. I appreciate the opportunity to work in her laboratory and for providing funding for my program. Without her commitments and overwhelming support, this program would not have been completed hitch-free.

My sincere appreciation goes to all the members on my thesis examination committee: Professor Steve Wiseman from the Department of Biological Sciences, Faculty of Arts & Science, University of Lethbridge, and Professor Andrew Iwaniuk from Neuroscience Department, University of Lethbridge. Their constructive comments and probing questions helped me gain more knowledge and understanding critical to my success in the program. Their applause and encouragement when applicable gave me a sense that I was on the right track, and in a way boosted my confidence. I cannot over emphasize their willingness to hear me speak on my research at committee meetings, their attention during the meetings and their directions.

I would like to appreciate Theresa Burg's lab members, former and current. Samples collected by former members formed the larger part of my research samples for this program. I appreciate former lab member Catherine Welke for collecting majority of the samples in the past and setting a concrete background that has contributed to the success of my program. Former members Kanishka and Ian gave their contributions during my proposal presentation and updates. I am grateful to Brendan Graham, a former postdoctoral researcher in the lab, for his huge support in the field. He showed me many aspects of bird capturing, sample and data collection from birds. He taught me several analyses which I applied to my research. He monitored the progress of my program and was always available to help.

To the current members, specifically Dilini and Danika for providing guidance and support in setting up the computing platform, introducing me to the applicable analyses, responding to my never-ending questions especially when I run into a problem with the scripts, for reviewing my work whenever asked. To Thilini, Aaron, thank you for your time in providing feedback for my write up and your contributions at my presentations. I appreciate you all for the encouragement you gave from my first day in this program till the very last. To the new members of the lab, Bryan, a master's student, and Linda a postdoctoral fellow, though we met towards the end of my program, your input was very helpful.

My profound appreciation goes to the University of Lethbridge for funding received through the international student tuition scholarship, the research stipend received from my supervisor's grant, and funding through graduate teaching assistantship awarded. I am

grateful to Coca-Cola for awarding me Master of Science, Coca-Cola admission scholarship.

My heartfelt appreciation goes to Alberta Conservation Association (ACA) for providing funding for some part of my research, and to the Natural Sciences and Engineering Research Council (NSERC) Discovery grants for funding part of the research.

I appreciate all whose path I crossed during my time at the University of Lethbridge, I am grateful to Associate professors Robert Laird and Dmytro Yevtushenko, my BIOL 5500 instructors, for their contribution to my research proposal presented as part of the BIOL 5500 course. I am grateful to the graduate student coordinators, Jenny Burke and Katrina Mendez for making my teaching experience at the University a memorable one. I equally appreciate my BIOL 1020 teaching mates: Danika, Trinie, and Amy, thank you all for the notable teamwork.

Finally, my appreciation goes to my family and friends in Nigeria, in the United States and in Canada, for their words of encouragement all through my study period, for checking on me and showing their concern for my well-being and my progress. You all contributed to my success. My unreserved gratitude goes to my husband, Roland for his immense support, understanding, and love. Thank you for not expecting too much and thank you for helping at every opportunity. And to my 4 year-old, Mary Jane, thank you for understanding that Mummy had to go to school just like you. You would patiently wait at the entrance to the Water Building for whoever could allow you access to the building so you could fetch me.

Thank you all very much for being part of my success story, God bless.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations, Acronyms, and Symbols

AB	Alberta
BC	British Columbia
BCF tools	Binary variant call format tools
bp	basepair
BRINP1	BMP/retinoic acid-inducible neural-specific protein 1
BSC	Biological species concept
BWA	Burrows-Wheeler Alignment tool
CO	Colorado
DAPC	Discriminant analysis of principal components
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
FDR	False discovery rate
F_{ST}	Wright's fixation index
GWAS	Genome wide association study
K	cluster
kb	Kilobase
lcWGS	Low coverage whole genome sequencing
LEA	R package for landscape and environmental association study
m	meter
MAF	Minor allele frequency
MAP3K15	Mitogen-Activated Protein Kinase Kinase Kinase 15
Mb	Megabase
MSN	Minimum spanning network
*Msp*I	restriction endonuclease isolated from *Moraxella sp.*
NaCl	sodium chloride
ng	nanogram
NJ	neighbour joining
*Nsi*I	restriction endonuclease isolated from *Neisseria sicca*
^{o}C	degrees Celsius
OR	Oregon
PCoA	Principal coordinate analysis
PCR	polymerase chain reaction
PE	paired end
ProK	proteinase K
*Pst*I	restriction endonuclease isolated from *Providencia stuartii*
PTK2	Protein Tyrosine Kinase 2
RNase	ribonuclease
rpm	revolutions per minute
SAM tools	Sequence Alignment Map tools
SDS	sodium dodecyl sulfate
SH3KBP1	SH3 domain containing kinase binding protein 1
ShinyGO	Graphical gene-set enrichment tool
SNPs	Single nucleotide polymorphisms
TE buffer	Tris and EDTA buffer

VCF	Variant call format
Z. l.	*Zonotrichia leucophrys*
µL	microliter
3dRADseq	triple digest restriction-site associated DNA sequencing

Chapter One: General Introduction

1.1 Evolution of Species

The earliest definition of 'evolution' dates back to the 17th century when a Swiss botanist, physiologist, lawyer, and poet, Albercht von Haller (1708-1777) defined evolution as the spreading out of the preformed body plan of an embryo. Later Jan Swammerdam and Marcello Malpighi modified the definition to mean that humans were fully formed, folded up in the ovary and gradually enlarged during growth until reaching the form and size of animals (Haller, 1774, cited from Hall, 2011). In the 19th century, evolution's meaning shifted to the transformation of species or the features of organisms within the context of a population as opposed to an individual. Modern evolution has been built on the theory of transformation and descent with modification following Charles Darwin's publication on the origin of species in 1859. Evolution now can be considered hierarchical which translates to the hierarchical organization of life itself from genes to molecules to organelles to cells to tissues to organs to organisms to populations and lastly to species (Hall, 2011).

Natural selection is central to the theory of evolution. It explains how organisms adapt to their environment. Genetic changes from one generation to the next in response to natural selection can result in changes in observable characters (phenotype) over a short period and can be recorded as a microevolution event (Hairston et al., 2005, Carroll et al., 2007).

Natural selection is most applicable to natural populations of organisms. Variation in environmental conditions or habitat is one of the driving forces for natural selection and subsequently for evolution. Therefore, natural selection is generally, but not the only source

of speciation and divergence as neutral evolution processes including genetic drift, mutation, and demographic events are important factors as well (Schluter 2001; Coyne and Orr, 2004; Lewontin 1974; Kimura 1983; Stern and Orgogozo 2009; Hamilton 2009; Gavrilets and Hastings 2012).

1.2 Habitat heterogeneity and species' diversity

Habitat heterogeneity or variation within a habitat is often positively correlated to species richness and diversity (Bazzaz, 1975; MacArthur and MacArthur, 1961). Habitat heterogeneity has been viewed by some as an ultimate determinant of species richness (Rosenzweig, 1995, Tews *et al*., 2004). Understanding the relationship between species diversity and the environment such as differential response to habitat conditions which may favor selection and consequently divergence between species is fundamental to ecology, and very important to the understanding of speciation by natural selection.

As good as variation in habitat is to the diversity of species, another consideration is the effect on gene flow. For the maintenance of gene flow within a species, habitat connectivity is important and without connectivity populations could be fragmented. In addition to habitat fragmentation, geological features, and varying environmental conditions can also decrease gene flow. It is becoming clear that local adaptation and ecological isolation resulting from these barriers lead to evolutionary divergence and the production of cryptic species i.e. taxa that are morphologically indistinguishable but genetically different. It has also been defined to include non interbreeding, similar looking taxa (Struck et al., 2018; Marchán et al., 2020). This can create challenges when determining evolutionarily significant units (ESU) or defining populations (Crandall et al., 2000). For example, cryptic species are evolutionarily distinct, but morphologically

indistinct, resulting in poor understanding of the species/subspecies which may lead to poor knowledge and understanding of actual population size and range (Neiva et al., 2017).

1.3 Species complex in response to evolution

Many authors and studies today agree that species are independent units emerging from the speciation process. They evolve from a metapopulations (Bock, 2004; Hey 2006; Hey et al., 2003). Emanating from species is the subspecies concept with a lot of controversies dating back to 1950s (e.g., Wilson and Brown, 1953). Although there are several definitions for a subspecies, in general terms they tend to be the aggregate of populations inhabiting a distinct breeding range and are distinct from other populations using characteristics such as morphology and genetic differentiation (Patten, 2009). However, this definition applies to populations that are not reproductively isolated from one another, this criterion separates subspecies from species, species are reproductively isolated (Mayr, 1942). Further, Mayr (1942) cautioned under the consideration of the geographic speciation model that regardless of every biological species going through a subspecies stage, not every subspecies will go on to become a distinct species. Due to the possibility of misidentification, reliable diagnosis of subspecies will require a combination of both genetic and morphological variation (Mousseau and Sikes, 2011).

Despite the many criticisms of the concept of subspecies, studies involving multiple criteria such as the combination of morphological, behavioural, and genetic characters have supported the subspecies concept and that they are evolutionarily definable entities (Gavin et al., 1999; Pasquet 1999, Haig et al., 2004). Mayr's early work in the middle 20th century showed species can evolve similar morphologies (Mayr, 1942). Today, with the use of molecular methods, cryptic species are more clearly defined and are gaining attention in

research (Bickford et al., 2007; Knowlton, 1993; Nygren, 2014; Hawlitschek et al., 2012; Ballentine and Greenberg, 2010). It is now established that cryptic species are widespread across many clades (Perez-ponce de Leon and Poulin, 2016; Pfenninger and Schwenk, 2007). Though from the earlier descriptions of cryptic species, most authors use the term interchangeably with sibling species (Saez, 2005). While some relate to groups with a more recent common ancestry (Knowlton, 1986), not all do. There are several explanations for the occurrence of cryptic species, including recent divergence, where morphological differences may be shallow (Egea et al., 2016). A second explanation refers to phylogenetic niche conservatism where niche evolution and therefore, differences in morphology across generations of species is limited by selection (morphological stasis) (Bravo et al., 2014; Egea et al., 2016; Smith et al., 2011). Lastly there is morphological convergence theory which supports morphological similarities evolving independently among distantly related species in response to similar selection pressure (Bravo et al., 2014; Trontelj et al., 2009). Understanding the concept of cryptic species is important and can be applied to diverging populations of a species, as these may remain cryptic until attaining complete divergence.

1.4 Local adaptation

Studying adaptation is important to both traditional and modern evolution. Local adaptation requires populations evolve to better fit local conditions due to selection. Local adaptation has supported Darwin's view on speciation as a result of adaptation to different niches or habitats (Schilthuizan, 2000; Jiggins and Mallet, 2000; Via, 2001; Schluter, 2001). Genetic variation within and among individuals allows for the potential of a species to adapt to particular environmental conditions (Aitken et al., 2008; Kawecki and Ebert, 2004) including response to predation in the environment. In a study by Møller & Nielsen (2015),

vulnerability of common avian prey species to predation by the sparrowhawk *Accipiter nisus* and goshawk *Accipiter gentilis* increases with reduced genetic variation. Many examples exist of the importance of variation to quick and successful adaptation of species to a novel environment, e.g., in beach mice *Peromyscus polionotus*, typical Santa Rosa Island mice that have successfully adapted to mainland, and three spine stickleback *Gasterosteus aculeatus*, a typical marine species that has successfully colonized freshwater (Barret & Schluter, 2008).

1.5 Molecular methods in population genetics

Molecular methods are important to reveal divergence and differentiation between and within populations. Molecular markers are short fragments of DNA used to provide information about individuals or populations including genetic diversity, divergence, and migration (Avise, 2000). With advances in next-generation sequencing (NGS), there are now alternative sequencing approaches such as restriction-site associated DNA sequencing (RADseq), and whole genome sequencing (WGS) including low coverage whole genome sequencing (lcWGS).

1.5.1 Restriction site associated DNA sequencing (RADseq)

A number of variations in RADseq exist including RAD-Tag (Baird et al., 2008), ddRAD-seq (Peterson et al., 2012), and 2b-RAD (Wang et al., 2012). They all work in similar ways, but with minor modifications to the number and type of enzymes used to fragment the DNA. RADseq, also known as genotyping-by-sequencing (GBS), can be used on non-model organisms (Andrews et al., 2016), and it works by targeting sequences adjacent to common restriction sites found throughout the genome. This approach can generate

massive amounts of data and thousands of loci which can then be used to identify sequence variants in the form of single nucleotide polymorphisms (SNPs). SNPs are useful in investigating research questions such as population structure, genetic diversity, and signatures of selection (Andrews et al., 2016). SNPs generally have low mutation rates making them less informative for population divergence compared to microsatellite markers. However, the large amount of data generated with RADseq can compensate for this shortcoming (e.g., Hodel et al., 2017; Morgan et al., 2017; Bohling et al., 2019; Jeffries et al., 2016). However, some of the difficulties associated with RADseq include construction of libraries which require high molecular weight DNA that may not be available in some cases. Highly degraded DNA results in low numbers of reads, poor quality of the reads and high error rates (Graham et al., 2015). Large amounts of starting material, about 50-100 ng of DNA, are required if there is no DNA amplification step (Andrews et al., 2016; Toonen et al., 2013).

1.5.2 Whole genome sequencing (WGS)

As opposed to the less expensive RADseq approach mentioned earlier where only regions adjacent to common restriction sites are sequenced across the genome, whole-genome sequencing as the name suggests, sequences the entire genome. The major limitation of RADseq is that a very large proportion of the genome is not sampled, and therefore the approach may miss regions under selection or with localized adaptive divergence in the genome (Lowry et al., 2017; Tiffin and Ross-Ibarra, 2014).

With whole genome sequencing you can see well-defined peaks of differentiation that were unclear or undetected with RADseq data (Aguillon, et al., 2021; Aguillon et al., 2018; Campagna et al., 2017; Campagna et al., 2015; Lou et al., 2019; Szarmach et al.,

2021). Genome-scale data are not limited to just neutral genetic variation, they also allow for complex genetic screening (Hedrick, 1999; Primmer, 2009). Further, genome-wide scans for selection and quantitative trait loci are possible with the application of whole-genome sequencing and can identify potential loci of interest for local adaptation (Steiner et al., 2013) and other population analyses (Figure 1.1). A major setback to whole-genome sequencing is cost. For many projects, sequencing at a depth, sufficient to confidently call individual genotypes is still very expensive, plus the need for a reference genome, and large amount of high-quality DNA (Lou et al., 2021).

Low coverage whole genome sequencing (lcWGS) is an alternative to WGS that is emerging as a more cost-effective method. To infer individual genotypes at low coverage is not reliable (Nielsen et al., 2011; 2012). But for some population genetic questions, the center of focus is the population characteristics obtained from the population genotype. Questions on genetic structure and relatedness of individuals are not based on the genotype at any SNP, but on genome-wide variation across the SNPs (Lou et al., 2021). Therefore, lcWGS maximizes the amount of sequence data, spread across the entire genome of many individuals. Sampling many individuals with sample size of 20 at low read depth of 2x provides an accurate result for estimating population parameters (Buerkle & Gompert, 2013; Fumagalli, 2013; Nevado et al., 2014). Many recent successful applications of lcWGS exist to answer population genetics questions across different species e.g. evidence for killer whale *Orcinus orca* divergence (Foote et al., 2016); identification of genes associated with rapid adaptation in Atlantic silversides *Menidia menidia* (Therkildsen et al., 2019); mapping hybrid incompatibility genes in swordtail fish *Xiphophorus hellerii* (Powell et al., 2020) and soft sweeps scanning for white-nose syndrome in bats *Myotis lucifugus* (Gignoux-Wolfsohn et al., 2021).

1.6 Study species

White-crowned sparrow (*Zonotrichia leucophrys*)

White-crowned sparrow, along with its sister species, Golden-crowned sparrow *Zonotrichia atricapilla* are being promoted as an example of rapid speciation with the two splitting as recent as 50,000 years ago (Johnson and Cicero, 2004). Divergence within white-crowned sparrow has been linked to the Pleistocene period about 18,000 years ago, giving rise to five subspecies (Rand, 1948). The rapid divergence within white-crowned sparrow, along with its wide range, different migration behavior and song dialect make this species an interesting one to study for signatures of evolution. Though subspecies have been defined largely based on morphological and geographical distribution, but genetic divergence has not been sufficiently described for these subspecies.

1.6.1 Description

White-crowned sparrows are small songbirds (25-30 g) with five recognized subspecies that vary in their bill color and lores (short feathers between the eyes and the bill) (Table 1.1) as well as range (figure 1.2). Allozyme and mitochondrial DNA studies propose that these subspecies diverged during the Pleistocene (Zink 1982; Zink et al., 1991). *Zonotrichia leucophrys nuttalli* and *Z. l. pugetensis* both have yellow bills and pale grey lores and occur in different geographical locations (Table 1.1): coastal California and coastal Pacific Northwest respectively. *Z. l. oriantha* is found in the Rocky Mountains of the western U.S., while *Z. l. leucophrys* occurs in northeastern Canada. They both have pink bills and black lores. Lastly, *Z. l. gambelii* with an orange bill and pale lores occurs in the northern Rocky Mountains, Alaska, and northern Canada west of Hudson Bay (Dunn

et al., 1995). White-crowned sparrows are generally sexually monomorphic, with the female having a slightly smaller body (Dunn et al., 1995; Morton, 2002).

In addition to genetic and morphometric data, there is support for the subdivision within *Zonotrichia* from the variation in vocal dialects. Subspecies *Z. l. oriantha*, *nuttalli*, and *pugetensis* all exhibit strikingly different dialects, but *Z. l. gambelii* does not seem to have a distinct dialect (Nelson, 1998). It generally lacks the clear trills and paired whistles, and though not distinguishable from *Z. l. leucophrys* by song, the only difference lies in the slight shift in notes pattern (Dunn et al., 1995). There have been different views on the relationship between song dialects and population structure in *Z. leucophrys* subspecies, some with strong support and others with poor support for a link between dialect and population structure (Baker et al., 1982; Chilton et al., 1990; Morton, 2002; Soha et al., 2004).

1.6.2 Distribution

White-crowned sparrows are widespread. They prefer a habitat with thick shrubby cover interspersed with open ground for wintering and breeding. They are generally found to breed in boreal scrub, forest edges, and thickets along the upper parts of Canada to the southern United States (Chilton et al., 1995). White-crowned sparrows are primarily omnivores surviving on a variety of flora species eating seeds, flowers, buds, fruits, and grass. They at times feed on terrestrial arthropods mainly when feeding young. Subspecies of *Z. leucophrys* occur at varying elevations, some selecting their nesting in subalpine meadows, 3,500 m high in the Sierra Nevada and the Rocky Mountains; and others, down to sea level on the Pacific Coast (Dunn et al., 1995; Morton, 2002).

In southwestern Alberta, two distinct subspecies of white-crowned sparrows come into contact: western taiga *Z. l. gambelii* with pale lores and interior west *Z. l. oriantha* with dark lores, however, the genetic differences found by Welke et al. (2021) do not correspond to the two subspecies rather habitat. The distribution and seasonal range of this species combined with morphological and behavioral variation make it an ideal species to study local adaptation because is found in different habitat types. In addition, regions where *Z. leucophrys* are found were once covered by ice sheets, and correspond to four areas of refugia in North America during the Pleistocene namely: south of the ice sheets near the West Coast and supporting populations of *Z. l. nuttalli* and *pugetensis*; southern Rocky Mountain, supporting the *oriantha* populations; northwest by the Yukon-Bering Sea supporting *gambelii* populations and lastly, the northeastern U.S. supporting the *Z. l. leucophrys* populations (Rand, 1948). Morphological variation especially the bill color and geographical distribution do correspond to subspecies delineation within this clade, but genetic background leading to their delineation has not been studied in-depth. Also, a preliminary genetic differentiation suggestive of local adaptation to habitat types created a platform for studying this species in detail with respect to regions driving this divergence.

1.7. Thesis Organization

The second chapter is the data chapter detailing sample selection within the habitat types under consideration, sample collection and genetic analyses. This chapter reports the application of single nucleotide polymorphisms (SNP) generated with triple digest restriction-site associated DNA (3dRAD) sequencing to detect genetic differences among the populations of white-crowned sparrow subspecies: *Zonotrichia leucophrys gambelii, Z. l. oriantha* and *Z. l. pugetensis,* and variation within the subspecies. The same SNP dataset

was also used to detect evidence of local adaptation. The chapter focuses on four research questions: 1) are populations of white-crowned sparrow subspecies occurring in alpine coniferous forest genetically different from those occurring in riparian deciduous forest?, 2) is there evidence that populations found in each forest type are adapting locally to their different habitats?, 3) what genes are associated with the forest types?, and 4) are the subspecies of *Z. l. leucophrys* genetically distinct. This study hypothesized that, divergence within *Z. leucophrys* subspecies have genetic underlining in form of genetic structure and variation in the genome. Also, if there were no habitat influence and subsequent local adaptation, subspecies co-occurring in different contiguous habitats should group according to their subspecies rather than their habitat types.

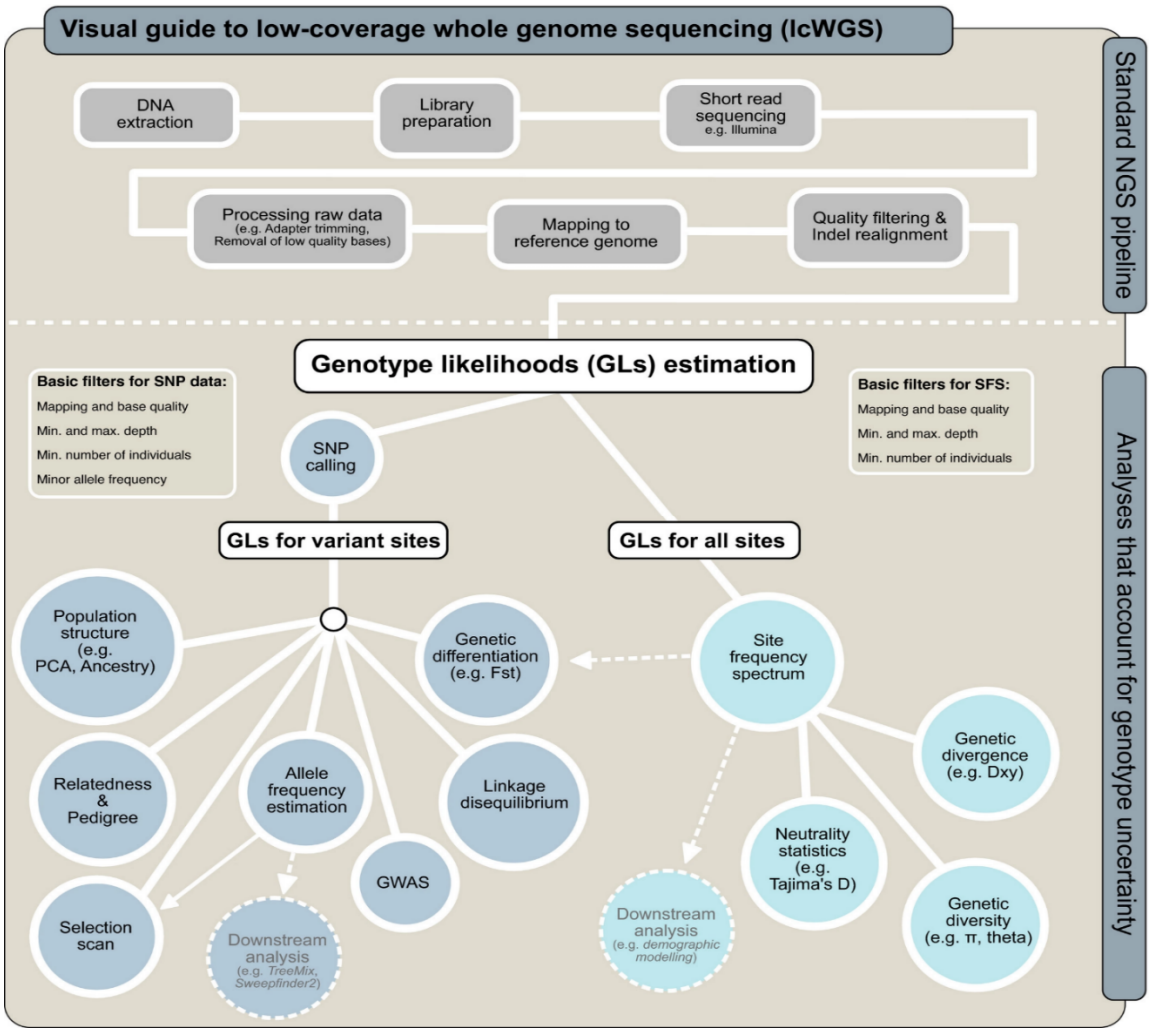


Figure 1.1: A typical computational pipeline for lcWGS data. The preprocessing part of the pipeline is shown at the top, and the bottom is data analysis based on genotype likelihood to account for gaps within the genome based on low coverage sequencing pattern (Lou et al., 2021).

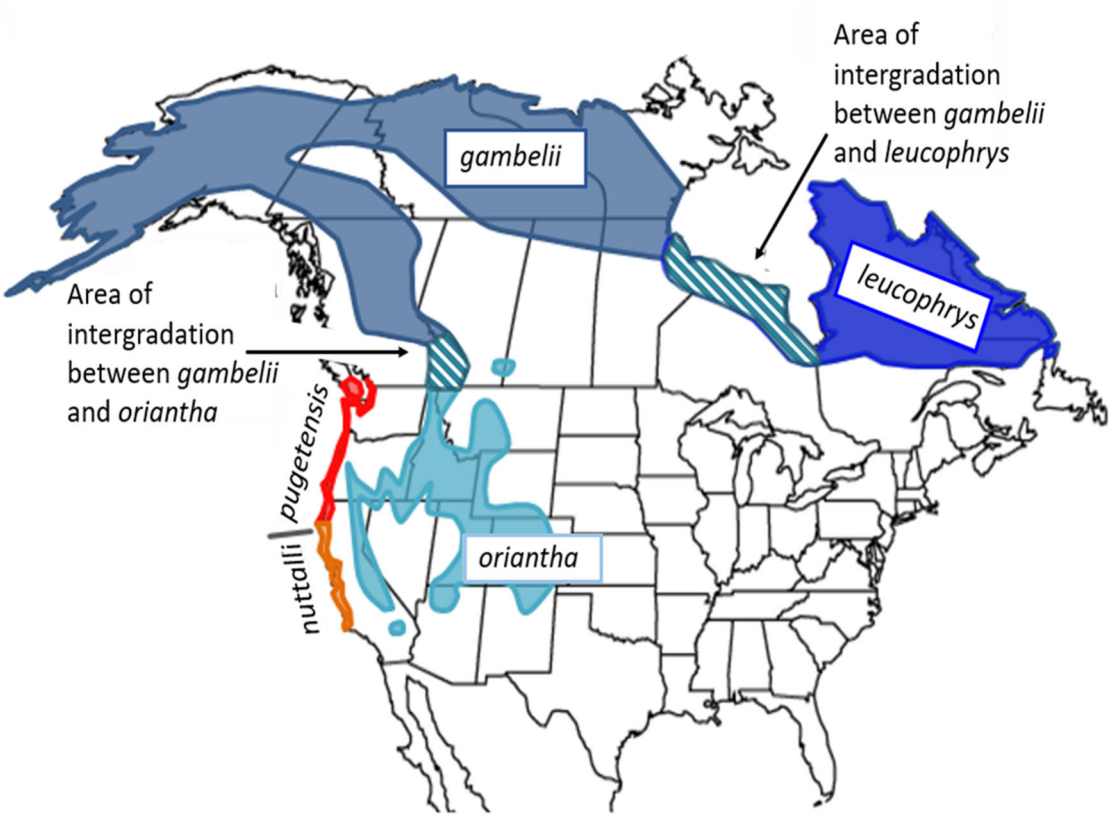


Figure 1.2: Breeding season range map of white-crowned sparrow subspecies, indicating distribution and intergradation. The two Pacific groups: *Z. l. nuttalli* and *pugetensis* have an overlapping range, while *Z. l. pugetensis* are short distance migrant, *Z. l. nuttalli* are resident birds. From Welke et al. (2021).

Table 1.1: Differentiating characteristics of *Zonotrichia* subspecies (Dunn et al., 1995; Pyle, 1997). Measurements did not distinguish between male and female individuals for weight and bill. Photo credits: *gambelii*; *leucophrys* © Don Robinson; *oriantha* Planet of Birds © 2020; *pugetensis* © Connor Cochrane; and *nuttalli* © Mike Baird

Characteristics/ subspecies	*Z. l. gambelii*	*Z. l. leucophrys*	*Z. l. oriantha*	*Z. l. pugetensis*	*Z. l. nuttalli*
Location	Alaska, northern Canada, British Columbia	eastern Canada	western Montane	northern Pacific Coast	southern Pacific Coast
Weight (g)	21.0 – 28.5 (n50)	21.6 – 38.5 (n162)	23.3 – 33.7 (n50)	21.4 – 29.1 (n50)	27.0 – 35.5 (n50)
Wing chord length (mm)	♀69 – 79 (n100) ♂ 74 – 84 (n100)	♀70 – 80 (n100) ♂ 73 – 83 (n100)	♀69 – 78 (n100) ♂ 73 – 82 (n100)	♀64 – 72 (n100) ♂ 67 – 75 (n100)	♀63 – 71 (n100) ♂ 67 – 75 (n100)
Tail length (mm)	♀64 – 73 (n20) ♂ 74 – 84 (n20)	♀63 – 72 (n12) ♂ 66 – 75 (n24)	♀67 – 74 (n12) ♂ 70 – 78 (n10)	♀62 – 70 (n10) ♂ 64 – 74 (n10)	♀62 – 71 (n20) ♂ 65 – 74 (n20)
Bill (nares to tip (mm)	7.0 – 8.3	7.3 – 8.6	7.4 – 8.8	7.0 – 8.3	7.4 – 8.8
Supraloral area	pale	dark	dark	pale	pale
General bill color	Orange	Pink to reddish pink	Dark reddish pink	Dull yellowish	Dull yellowish
Call note	Sharp *pink*	Sharp *pink*	Metallic *pink*	Flatter *pink*	Flatter *pink*

1.8 References

Aguillon, S. M., Walsh, J., & Lovette, I. J. (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society: Biological Sciences.* 288: 20201805.

Aguillon, S. M., Campagna, L., Harrison, R. G., & Lovette, I. J. (2018). A flicker of hope: Genomic data distinguish northern flicker taxa despite low levels of divergence. *The Auk*. 135: 748 – 766.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration, or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*. 1: 95 – 111.

American Ornithologists Union (1957). Checklist of North American Birds. Fifth edition, American Ornithologists Union, Washington, DC.

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*. 17: 81 – 92.

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

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 3: e3376

Baker, C. M., Thompson, D. B., Sherman, G. L., Cunningham, M. A., & Tomback, D. F. (1982). Allozyme frequencies in a linear series of song dialect populations. *Evolution*. 36: 1020 – 1029.

Ball, R. M., & Avise, J. C. (1992). Mitochondrial DNA phylogeographic differentiation among avian populations and the evolutionary significance of subspecies. *The Auk*. 109: 626 – 636.

Ballentine, B., & Greenberg, R. (2010). Common garden experiment reveals genetic control of phenotypic divergence between swamp sparrow subspecies that lack divergence in neutral genotypes. *PLoS One*. 5: e10229.

Barrett, R. D., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution.* 23: 38 – 44.

Bazzaz, F.A. (1975). Plant species diversity in old field successional ecosystems in southern Illinois. *Ecology.* 56: 485 – 488.

Bickford, D., Lohman, D., Sodhi, N., Peter, K. L. Ng., Meier, R., Winker, K., & Ingram, K. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution.* 22: 148 – 55.

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*. 16: 1195 – 1205.

Bock, W. J. (2004). Species: the concept, category, and taxon. *Journal of Zoological Systematics and Evolutionary Research*. 42: 178 – 190.

Bohling, J., Small, M., Von Bargen, J., Louden, A., & DeHaan, P. (2019). Comparing inferences derived from microsatellite and RADseq datasets: a case study involving threatened bull trout. *Conservation Genetics.* 20: 329 – 342.

Bravo, G. A., Remsen, J. V., Jr. & Brumfield, R. T. (2014). Adaptive processes drive ecomorphological convergent evolution in antwrens (*Thamnophilidae*). *Evolution*. 68: 2757 – 2774.

Buerkle, C., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*. 22: 3028 – 35.

Byars, S.G., Ewbank, D., Govindaraju, D. R., & Stearns, S. C. (2010). Colloquium papers: Natural selection in a contemporary human population. *Proceedings of the National Academy of Sciences of the United States of America*. 107: 1787 – 1792.

Campagna, L., Gronau, I., Silveira, L. F., Siepel, A., & Lovette, I. J. (2015). Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*. 24: 4238 – 4251.

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*. 3: e1602404.

Carrasco, L., Giam, X., Papeş, M., & Sheldon, K. (2019). Metrics of Lidar-derived 3D vegetation structure reveal contrasting effects of horizontal and vertical forest heterogeneity on bird species richness. *Remote Sensing.* 11: 1 – 19.

Carroll, S. P., Hendry, A. P., Reznick, D. N., & Fox, C. W. (2007). Evolution on ecological time-scales. *Functional Ecology.* 21: 387 – 393.

Cheek, R. G., Forester, B. R., Salerno, P. E., Trumbo, D. R., Langin, K. M., Chen, N., Scott Sillett, T., Morrison, S. A., Ghalambor, C. K., & Chris F. W. (2022). Habitat-linked genetic variation supports microgeographic adaptive divergence in an island-endemic bird species. *Molecular Ecology.* 31: 2830 – 2846.

Chilton, G., Lein, M. R., & Baptista, L. F. (1990). Mate choice by female white-crowned sparrows in a mixed-dialect population. *Behavioral Ecology and Sociobiology*. 27: 223 – 227.

Clausen, J., Keck, D. D., & Hiesey, W. M. (1948). Experimental studies on the nature of species. III. Environmental responses of climatic races of Achillea. *Carnegie Institution of Washington Publication*. 581: 1 – 189.

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

Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*. 12: 1971 – 1987.

Coyne, J. A., & Orr, H.A. (2004). Speciation. Sunderland: Sinauer Associates. 545 p.

Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*. 15: 290 – 295.

Dunn, J. L., Garrett, K. L., & Alderfer, J. (1995). White-crowned sparrow subspecies: identification and distribution. *Birding.* 27: 182 – 201.

Egea, E., David, B., Choné, T., Laurin, B., Féral, J. P., & Chenuil, A. (2016). Morphological and genetic analyses reveal a cryptic species complex in the echinoid *Echinocardium cordatum* and rule out a stabilizing selection explanation. *Molecular Phylogenetics and Evolution.* 94: 207 – 220.

Foote, A., Vijay, N., Ávila-Arcos, M. *et al.* (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*. 7: 11693

Franks, S. J., Weber, J. J., & Aitken, S. N. (2014). Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications*. 7: 123 – 139.

Fumagalli, M. (2013). Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One*. 8(11): e79667.

Garant, D., Forde, S. E., & Hendry, A. P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology.* 21: 434 – 443.

Gavin, T A., Sherman, P. W., Yensen, E., & May, B. (1999). Population genetic structure of the northern Idaho ground squirrel (*Spermophilus brunneus brunneus*). *Journal of Mammalogy.* 80: 156 – 165.

Gavrilets, S., & Hastings, A. (2012). Founder effect speciation: a theoretical reassessment. *American Naturalist*. 147: 466 – 491.

Gee, J. M., & Warwick, R. M. (1994). Metazoan community structure in relation to the fractal dimensions of marine macroalgae. *Marine Ecology Progress Series*. 103: 141 – 150.

Gignoux-Wolfsohn, S. A., Pinsky, M. L., Kerwin, K., Herzog, C., Hall, M., Bennett, A. B., Fefferman., N. H. & Maslo, B. (2021). Genomic signatures of selection in bats surviving white-nose syndrome. *Molecular Ecology*. 30: 5643 – 5657.

Goetz, S., Steinberg, J. D., Betts, M. G., Holmes, R. T., Doran, P. J., Dubayah, R., & Hofton, M. (2010). Lidar remote sensing variables predict breeding habitat of a Neotropical migrant bird. *Ecology.* 91: 1569 – 1576.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., et al. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources.* 15: 1304 – 1315.

Haig, S. M., Mullins, T. D., & Forsman, E. D. (2004). Subspecific relationships and genetic structure in the spotted owl. *Conservation Genetics*. 5: 683 – 705.

Hairston, N. G., Jr, Ellner, S. P., Geber, M. A., Yoshida, T. & Fox, J. A. (2005). Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters.* 8: 1114 – 1127.

Hamilton, M.B. (2009). Population Genetics. Chichester: John Wiley and Sons. 424 p.

Hawlitschek, O., Nagy, Z., & Glaw, F. (2012). Island evolution and systematic revision of Comoran snakes: why and when subspecies still make sense. *PLoS ONE*. 7: e42970.

Heck, K. L. J., & Crowder, B. (1991). Habitat structure and predator-prey interactions in vegetated aquatic systems. Pages 281–299 in S. S. Bell, E. D. McCoy, and H. R. Mushinsky, editors. Habitat structure: the physical arrangement of objects in space. Chapman and Hall, London, UK.

Hedrick, P. W. (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution.* 53: 313 – 318.

Hey, J. (2006). On the failure of modern species concepts. *Trends in Ecology and Evolution*. 21: 447 – 450.

Hey, J., Waples, R. S., Arnold, M. L., Butlin, R. K., & Harrison, R. G. (2003). Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution* 18: 597 – 603.

Hill, R. A., & Thomson, A. G. (2005). Mapping woodland species composition and structure using airborne spectral and LiDAR data. *International Journal of Remote Sensing.* 26: 3763 – 3779.

Hinsley, S. A., Hill, R. A., Bellamy, P. E., & Balzter, H. (2006). The application of Lidar in woodland bird ecology. *Photogrammetric Engineering and Remote Sensing*. 72: 1399 – 1406.

Hodel, R. G. J., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improve phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports.* 7:17598.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in three spine stickleback using sequenced RAD tags. *PLoS Genetics.* 6: e1000862.

Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology.* 25: 2997 – 3018.

Jeffries, M. (1993). Invertebrate colonization of artificial pondweeds of differing fractal dimension. *Oikos*. 67: 142.

Jiggins, C., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution.* 15: 250 – 255.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters.* 7: 1225 – 1241.

Kelaher, B. P., & Castilla, J. C. (2005). Habitat characteristics influence macrofaunal communities in coralline turf more than mesoscale upwelling on the coast of northern Chile Estuarine. *Coastal and Shelf Science*. 63: 155 – 165.

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge: Cambridge University Press. 367 p.

Knowlton, N. (1986) Cryptic and sibling species among the decapod Crustacea. *Journal of Crustacean Biology*. 6: 356 – 363.

Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology and Systematics*. 24: 189 – 216.

Kosova, G., Abney, M., & Ober, C. (2010). Heritability of reproductive fitness traits in a human population. *Proceedings of the National Academy of Sciences of the United States of America*. 107: 1772 – 1778.

Lewontin, R. (1974). The genetic basis of evolutionary change. New York: Columbia University Press. 346 p.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*. 30: 5966 – 5993.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*. 17: 142 – 152.

MacArthur, R. H., & MacArthur, J. W. (1961). On bird species diversity. *Ecology.* 42: 594 -598.

Marchán, D. F., Novo, M., Sánchez, N., Domínguez, J., Darío, J., Cosín, D., & Fernández, R. (2020). Local adaptation fuels cryptic speciation in terrestrial annelids. *Molecular Phylogenetics and Evolution*. 146: 106767.

Mayden, R. L. (1997). A hierarchy of species concept: the denouement in the saga of the species problem. In: Claridge MF, Dawah HA, Wilson MR (eds), Species: The Units of Biodiversity. Chapman & Hall, London, pp 381–424

Mayr, E. (1942) Systematics and the Origin of Species. Columbia Univ. Press, New York.

Mikles, C. S., Aguillon, S. M., Chan, Y. L., Arcese, P., Benham, P. M., Lovette, I. J., & Walsh, J. (2020). Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird. *Molecular Ecology.* 29: 4295 – 4307.

Møller, A. P., & Nielsen, J. T. (2015). Genetic variation in birds in relation to predation risk by hawks: A comparative analysis, *Current Zoology*. 61: 1 – 9.

Morgan, T. D., Graham, C. F., McArthur, A. G., Raphenya, A. R., Boreham, D. R., Manzon, R. G., et al. (2017). Genetic population structure of the round whitefish (*Prosopium cylindraceum*) in North America: multiple markers reveal glacial refugia and regional subdivision. *Canadian Journal of Fisheries and Aquatic Sciences.* 75: 836 – 849.

Morton, M. (2002). The mountain white-crowned sparrow: migration and reproduction at high altitude. *Studies in Avian Biology*. 24: 1 – 236.

Mousseau, T., & Sikes, D. S. (2011). Almost but not quite a subspecies: A case of genetic but not morphological diagnosability in *Nicrophorus* (*Coleoptera: Silphidae*). *Biological Journal of the Linnean Society.* 102: 311 – 333.

Nacci, D., Proestou, D., Champlin, D., Martinson, J., & Waits, E. R. (2016). Genetic basis for rapidly evolved tolerance in the wild: Adaptation to toxic pollutants by an estuarine fish species. *Molecular Ecology.* 25: 5467 – 5482.

Neiva, J., Serrão, E. A., & Anderson, L. *et al.* (2017). Cryptic diversity, geographical endemism and allopolyploidy in NE Pacific seaweeds. *BMC Evolutionary Biology*. 17: 30.

Nelson, D. A. (1998). External validity and experimental design: the sensitive phase for song learning. *Animal Behaviour.* 56: 487 – 491.

Nevado, B., Ramos-Onsins, S. E., Perez-Enciso, M. (2014). Resequencing studies of nonmodel organisms using closely related reference genomes: optimal experimental designs and bioinformatics approaches for population genomics. *Molecular Ecology.* 23: 1764 – 1779.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PLoS One*. 7: e37558.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*. 12: 443 – 451.

Nygren, A. (2014). Cryptic polychaete diversity: A review. *Zoologica Scripta*. 43: 172 – 183.

Owen, J. C. (2011). Collecting, processing, and storing avian blood: a review. *Journal of Field Ornithology*. 82: 339 – 354.

Pasquet, R. S. (1999). Genetic relationships among subspecies of *Vigna uguiculata* (L.) Walp. based on allozyme variation. *Theoretical and Applied Genetics.* 98:1104 – 1119.

Patten, M. (2009). 'Subspecies' and 'race' should not be used as synonyms. *Nature*. 457: 785.

Patten, M. (2015). Subspecies and the philosophy of science. *The Auk*. 132: 481 – 485.

Pérez-Ponce de León, G., & Poulin, R. (2016). Taxonomic distribution of cryptic diversity among metazoans: Not so homogeneous after all. *Biology Letters.* 12: 2016037.

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: e37135.

Pettay, J. E., Kruuk, L. E. B., Jokela, J., & Lummaa, V. (2005). Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proceedings of the National Academy of Sciences of the United States of America*. *102*: 2838 – 2843.

Pettay, J.E., Helle, S., Jokela, J., & Lummaa, V. (2007) Natural selection on female life-history traits in relation to socio-economic class in pre-industrial human populations. *PLoS ONE.* 2: e606.

Pfenninger, M., & Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology.* 7: 121.

Powell, D. L., GarcíaOlazábal, M., Keegan, M., Reilly, P., Du, K., DíazLoyo, A. P., Banerjee, S., Blakkan, D., Reich, D., Andolfatto, P., Rosenthal, G. G., Schartl, M., & Schumer, M. (2020). Natural hybridization reveals incompatible alleles that cause melanoma in swordtail fish**.** *Science*. 368: 731 – 736.

Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*. 1162: 357 – 368.

Pyle, P. (1997). Identification guide to north American birds, part 1. Slate Creek Press. Bolinas, California. 732p.

Rand, A. L. (1948). Glaciation, an isolating factor in speciation. *Evolution*. 2: 314 – 321.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution.* 29: 165 – 176.

Rosenzweig, M. L. (1995). Species Diversity in Space and Time. Cambridge University Press, New York.

Rosenzweig, M. L. (1995). Energy flow and diversity. Pages 343– 372 *in Species diversity in space and time*. Cambridge University Press, Cambridge, UK.

Sáez, A. G., & Lozano, E. (2005). Body doubles. *Nature*. 433: 111.

Schilthuizen, M. (2000). Dualism and conflicts in understanding speciation. *BioEssays*. 22: 1134 – 41.

Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology and Evolution.* 16: 372 – 380.

Smith, K. L., Harmon, L. J., Shoo, L. P., & Melville, J. (2011). Evidence of constrained phenotypic evolution in a cryptic species complex of agamid lizards. *Evolution; International Journal of Organic Evolution*. 4: 976 – 992.

Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*. 64: 295 – 315

Soha, J. A., Nelson, D. A., & Parker, P. G. (2004). Genetic analysis of song dialect populations in Puget Sound white-crowned sparrows. *Behavioral Ecology*. 15: 636 – 646.

Steiner, C. C., Putnam, A. S., Hoeck, P. E. A., & Ryder, O. A. (2013). Conservation genomics of threatened animal species. *Annual Review of Animal Biosciences* 1: 261 – 281.

Stern, D., & Orgogozo, V. (2009). Is genetic evolution predictable? *Science.* 323: 746 – 751.

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., and Gusarov, V. I. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology and Evolution*. 33: 153 – 163.

Szarmach, S., Brelsford, A., Witt, C. C., & Toews, D. (2021). Comparing divergence landscapes from reduced-representation and whole-genome re-sequencing in the yellow-rumped warbler (*Setophaga coronata*) species. *Molecular Ecology*. 30: 5994 – 6005.

Tews, J., Brose, U., Grimm, V., Tielbörger, K., Wichmann, M.C., Schwager, M. & Jeltsch, F. (2004). Animal species diversity driven by habitat heterogeneity/diversity: the importance of keystone structures. *Journal of Biogeography*. 31: 79 – 92.

Therkildsen, N. O., Wilder, A. P., Conover, D. O., Munch, S. B., Baumann, H., & Palumbi, S. R. (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science.* 365: 487 – 490.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution*. 29: 673 – 680.

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., et al. (2013). ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ.* 1: e203.

Trontelj, P., Douady, C. J., Fišer, C., Gibert, J., Gorički, Š., Lefébure, T. & Zakšek, V. (2009). A molecular test for cryptic diversity in ground water: How large are the ranges of macro-stygobionts? *Freshwater Biology*. 54: 727 – 744.

Urban, M. C., Richardson, J. L., Freidenfelds, N. A., Drake, D. L., Fischer, J. F., & Saunders, P. P. (2017). Microgeographic adaptation of wood frog tadpoles to an apex predator. *Copeia.* 105: 451 – 461.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution.* 16: 381 – 390.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods.* 9: 808 – 810.

Welke, C. A., Graham, B., Conover, R. R., Rivers, J. W., & Burg, T. M. (2021). Habitat-linked genetic structure for white-crowned sparrow (*Zonotrichia leucophrys*): Local factors shape population genetic structure. *Ecology and Evolution*. 11: 11700 – 11717.

Whitlock, M. C. (2015). Modern approaches to local adaptation. *American Naturalists.* 186: S1 – 4.

Wiens, J.A. (1982). Forum: Avian subspecies in the 1980's. *The Auk.* 99: 593

Wilson, E. O., & Brown, W. L. Jr. (1953). The subspecies concept and its taxonomic application. *Systematic Biology*. 2: 97 –111.

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*. 65*:* 1897 –1911.

Zink, R. M. (1989). The study of geographic variation. *The Auk.* 106: 157 – 160.

Zink, R. M. (2004). The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London B*. 272: 71 – 78.

Zink, R. M., Barrowclough, G. F., Atwood, J. L., & Blackwell-Rago, R. C. (2000). Genetics, taxonomy, and conservation of the threatened California gnatcatcher. *Conservation Biology*. 14: 1394 – 1405.

Zink, R. M., Dittmann, D. L., & Rootes, W. L. (1991). Mitochondrial DNA variation and the phylogeny of *Zonotrichia. The Auk.* 108: 578 – 584.

Zink, R. M. (1982). Patterns of genic and morphologic variation among sparrows in the genera *Zonotrichia, Melospiza, Junco,* and *Passerella. The Auk.* 99: 632 – 649.

Chapter Two: Identifying genetic structure and divergence between *Zonotrichia leucophrys* subspecies, and habitat ecotypes.

2.1 Introduction

Both reproductive and ecological isolation result in evolution of isolated populations (Yeaman and Whitlock, 2011; Nosil, 2008; Garant et al., 2007). In many cases, it is easier to notice the evidence of evolution by observing differences in the morphology and behavior of the populations, but in some cases involving cryptic species or populations, this may be difficult to notice visually (Struck et al., 2018). Evidence of evolution within a species include observable changes in the morphology of individuals, geographical distributions, elevation range, behavior, vocal variation, and molecular variation (Bravo et al., 2014). Molecular testing has come to be the most reliable tool for detecting evolution and improving our knowledge of cryptic species or populations.

The white-crowned sparrow, *Zonotrichia leucophrys*, is an interesting species to study owning in part to its wide distribution (Chilton et al., 1995; Dunn et al., 1995; Morton, 2002). Five subspecies have been recognised across a wide geographic area, suggesting some level of isolation. Divergence in *Z. leucophrys* is partly associated with populations sheltering in different refugia during the last glacial maximum (LGM) of the Quaternary (Hewitt, 2000; Hewitt, 2001), occurring between 2.5 Ma – 11.5 ka (Gibbard et al., 2010). *Z. l. gambelii* was mostly confined to the northwest near the Yukon-Bering Sea; *Z. l. oriantha* to the south of the ice sheet; *Z. l. nuttalli* and *Z. l. pugetensis* (Pacific group) to the south in the Rocky Mountains; and *Z. l. leucophrys* to the northeastern United States (Rand, 1948). There are several distinguishing features among white-crowned sparrow subspecies: *Z. l. nuttalli* and *Z. l. pugetensis* sometimes referred to as the Pacific Coast race are characterized by pale lore and yellow beak. The back stripes are tan and blackish, the plumage below is much more brown, and the inner greater secondary coverts and outer webs of the tertials are pale rusty brown. The *Z. l. gambelii, Z. l. oriantha* and *Z. l.*

leucophrys are grouped as boreal and montane birds. They are generally grayish below with deep reddish, and pale gray on the back stripes. Specifically, *Z. l. gambelii* from the western boreal has pale lores and an orange bill; *Z. l. oriantha* from the western mountains has dark lores with a dark reddish-pink bill; and *Z. l. leucophrys* from the eastern boreal is very similar to *Z. l. oriantha* and may be difficult to distinguish in the field (Table 1.1). On a very close look, the dorsal and ventral color of *Z. l. oriantha* appear paler compared to *Z. l. leucophrys* (Godfrey, 1965). There are ecological and behavioral variation among white-crowned sparrow subspecies in that the breeding ranges of the boreal and montane subspecies are isolated from one another (Figure 1.2) and different migration patterns. The non-migratory group consisting of *Z. l. nuttalli* (Hafner & Petersen, 1985); and the migratory group consisting of the rest of the subspecies. Additionally, both *Z. l. oriantha* and *Z. l. gambelii* maintain a very strong fidelity to their wintering and breeding sites (Morton, 2002).

Genetic approaches have been applied to further understand this widespread species in terms of genetic divergence due to variation in their morphology, behavior, and habitat. Several studies using mitochondrial, microsatellite and single nucleotide polymorphism datasets have pointed to the divergence of the subspecies with varying degrees of support from total separation of *Z. l. pugetensis* and *Z. l. nuttalli* from the other subspecies to partial differentiation between *Z. l. gambelii, Z. l. oriantha* and *Z. l. leucophrys* (Taylor et al., 2020; Welke et al., 2021). Welke et al. (2021) found some individuals of *Z. l. gambelii* and *Z. l. oriantha* subspecies, irrespective of their subspecies, clustered according to either alpine coniferous habitat or riparian deciduous habitat, suggesting a possible role of habitat as a restriction to gene flow or differential selection pressure resulting in local adaptation.

There has not been a lot of studies with consistent pattern to understand the evolution of the white-crowned sparrows and especially, recent divergence events. This current study therefore assessed populations for patterns of evolution among the subspecies, and within subspecies using high resolution molecular markers (triple digest restriction-site associated DNA sequencing, and low coverage whole genome sequencing) applied to white-crowned sparrow for the first time. The expectation is that single nucleotide polymorphisms assessed across several thousand loci will provide higher resolution in these groups. In addition to establishing support for recent studies where there are indications of genetic differentiation for these groups, this study will more importantly scan the genome looking for loci under selection as these may be driving differentiation between subspecies and ecotypes. The outlier loci may also contain specific genes on which selection is acting. The study will probe further into the identification of regions of differentiation among the subspecies, leading to a clearer understanding of divergence within this clade. Lastly genes found around the outlier SNPs will be identified, along with their functions and roles in the divergence of the habitat ecotypes and subspecies.

This study is in two parts: the first part examined evidence for habitat-linked differentiation within populations of *Z. l. gambelii* and *Z. l. oriantha* occupying contiguous but heterogenous habitats: alpine coniferous forest and riparian deciduous forest. In this first part, several analyses were conducted to understand recent divergence driven by habitat conditions such as microclimatic conditions, the habitat structure and its components. The second part briefly examined speciation of the subspecies *Z. l. gambelii, Z. l. oriantha* and *Z. l. pugetensis* using a combination of 3dRADseq and low coverage whole genome sequence data (lcWGS) as an advanced methodology to previous studies to provide genetic support for the divergence of white-crowned sparrow subspecies since

previous studies have not been able to fully and consistently resolve genetic differentiation for many of the subspecies. This side-by-side comparison of the genome of the three subspecies will allow for an improved understanding of their divergence.

2.2 Materials and Methods

2.2.1 Sample collection

A total of 141 samples were collected from sites across western North America. *Z. l. oriantha* and *Z. l. gambelii* from eight locations in Alberta (Beaver Mines, Waterton, Lethbridge, Jasper, Brule, Banff, Cypress Hills, Crowsnest Pass), three in British Columbia (Mackenzie, Fort St James, Revelstoke), three *Z. l. oriantha* transects from along a 415 m elevational transect north of the Rocky Mountain Biological Laboratory in Gunnison County, Colorado, and three populations of *Z. l. pugetensis* from Oregon. Sample collection occurred during the breeding seasons between June and August. Sampling took place with the use of 12 m mist net with bird song playback specific to white crowned sparrow. Approximately 100 µL of blood was collected from the brachial vein of each bird following Owen (2011), or a tail feather was collected. Blood samples were placed in 99% ethanol and stored at -20ºC upon return to the lab, feather samples placed in tubes were preserved at -20ºC. Before being released, birds were banded with a numbered metal band to avoid resampling.

2.2.2 DNA extraction, library preparation, and sequencing

3dRAD

DNA (deoxyribonucleic acid) extraction was done following a modified method of Aljanabi and Martinez (1997), with a salt extraction method. Blood or feather samples stored in ethanol were incubated for 15 minutes at 57ºC to remove the ethanol. For the lysis step, about 10 µL of blood or 0.5 cm of feather sample (shaft) was put in each tube. To the samples were added 200 µL homogenizing buffer (0.4 M NaCl, 10 mM Tris–HCl pH 8.0, and 2 mM EDTA pH 8.0), 20 µL 20% SDS, 2.5 µL RNase at 20 mg/ml concentration and 10 µL ProK at 20 mg/ml concentration. The tubes with the mixture were inverted to thoroughly mix, and then placed in a rotator for a minimum of 2 hours at 57ºC. For the salt step, 150 µL of 6 M NaCl was added to each tube and inverted or gently vortexed for about 15 seconds. The tubes were then centrifuged for 30 minutes at 14,000 xg. For the For the precipitation step, the supernatant was carefully removed to new labeled tubes where 2 µL glycol blue was added to each sample to enable the binding of DNA to the glycol blue and make the pellet visible. The tubes were gently inverted to ensure a thorough mixing. One millilitre of 95% cold ethanol was added to each tube and the tubes inverted a few times. The tubes were then placed in a -20ºC freezer for few hours or overnight. Next, the tubes were centrifuged for 30 minutes at 14,000 xg, ethanol was carefully removed leaving behind a blue pellet containing the DNA. The next step was the wash step which required adding 1 ml of 70% ethanol at room temperature to the tubes. The tubes were centrifuged for 10 minutes at 14,000 xg, and the ethanol was removed. The wash step was done two times to remove excess salt. After the last wash step and removal of ethanol, the tubes were

left open to evaporate the remaining ethanol. After the ethanol was evaporated, 100 µL 1x TE buffer (10 mM Tris, 1 mM EDTA at pH 8-9) was added to the tubes to resuspend the DNA. The DNA was incubated for 15 minutes at 37ºC and then placed in the fridge. As sequencing requires high quality DNA at a concentration of approximately 20 ng/µL, the purity and concentration of the DNA were checked using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technology Inc. Wilmington, DE, USA), and DNA integrity was checked on agarose gel.

For the 3dRAD sequencing, library preparation was done at the University of Laval Institute of Integrative Biology and Systems. Library was constructed following Parchman et al. (2016), and DNA was digested with *Pst*I, *Msp*I, and *Nsi*I which are known to work well for bird species. Sequencing was performed at Genome Quebec using a medium output, 100 cycles, paired end run on an Illumina NovaSeq 6000 S4 sequencer.

Low coverage Whole Genome Sequencing

Representative samples ranging from 1 to 9 birds from each subspecies and habitat ecotype (Table 2.1) were used for the low-coverage whole genome analyses aimed at obtaining better coverage and comparing findings with the 3dRAD sequencing (Lou et al., 2019). The low-coverage sequencing followed a similar protocol of DNA extraction and purification to obtain high yield DNA (approximately 50 ng/µL). Shotgun PCR free library was prepared with 8 bp unique barcode tagged to each sample. Sequencing was done at Genome Quebec on a paired end run on Illumina Novaseq 6000 S4 PE 150.

2.2.3 Data preparation and genetic analyses

3dRAD

Sabre 1.00, a tool for demultiplexing barcoded reads into separate files, was used to generate fastq files. Cut-adapt program (Martin, 2011) was used to remove adapter sequences added to the restriction cut site during library preparation and to trim the reads to a uniform size of 80 base pairs following the removal of barcodes and adapters from the initial read of about 100 bp. Reads were aligned to the annotated genome of white-throated sparrow *Zonotrichia albicollis* GenBank accession GCA_000385455.1 with BWA program (Li, 2013). Cut-adapt was also applied to the whole genome reads to trim reads to 150 bp, and reads were aligned to the same reference genome as above. Different steps in the STACKS pipeline were then used to call SNPs and genotype individuals at each identified SNP (Catchen et al., 2018). Population statistics such as nucleotide divergence, genetic differentiation (F_{ST}) and genetic structure were computed using a number of tools such as VCftools version 0.1.14 (Auton & Marcketta, 2015), Arlequin version 3.5.2 (Excoffier & Lischer, 2015), DAPC (Jombart et al., 2010), and LEA (Frichot & Francois, 2015).

lcWGS

Reads from the whole genome did not require demultiplexing as it was done by Genome Quebec. Our samples were sequenced at a range of 6.2 to 8.9x coverage. BWA program (Li, 2013) was used to align the reads to the reference genome earlier stated, and to remove duplicate and unmapped reads. SAM and BCF tools were used to estimate the genotype

likelihood using the mpileup option and the actual SNP calling using the bcftools call option. A VCF file was generated for the downstream population analyses.

For both 3dRAD and lcWGS sequencing methods, filtering of datasets included setting the minimum allele frequency, MAF, to 0.05. Filtering of dataset to retain quality sites and individuals was done. 3dRAD data were filtered to remove SNPs missing up to 40% data, and individuals missing up to 10% data. All the lcWGS aligned dataset were retained during filtering as they had no missing data. For the outlier SNPs, it was realized that a good number of outliers were still found in more loose filtering settings, so to include more individuals for the outlier analyses, we filtered at 60% missing SNPs and 20% missing individuals.

2.2.3.1 Evidence of genetic differentiation

With the results from STACKS population pipeline, and with the data filtered to minimize missing data from the populations, DAPC, Discriminant Analysis of Principal Components, program in R (Jombart et al., 2010) was used to infer genetic clusters within and between the habitat types. Other programs for genetic structure analysis such as LEA, an R package for Landscape and Environmental Association Studies (Frichot & Francois, 2015), and PCoA were used to support the structure analysis. These programs possess different algorithms and sensitivity for detecting population structure, and a combination of different programs is often desirable in a study like this (Jones & Wang, 2012; Stift et al., 2019).

2.2.3.2 Genetic distance and divergence

Degree of genetic differentiation or genetic distance between the subspecies and habitat types was assessed with F_{ST} (Wright, 1951), a commonly used method to estimate divergence defined by Wright (1951). The populations were compared using Slatkin (1995) genetic distance computed from F_{ST} in Arlequin version 3.5.2 (Excoffier & Lischer, 2015). Values usually range from 0 to 1 and indicate the proportion of genetic diversity resulting in any observed genetic structure between populations (Holsinger & Weir, 2009). A Neighbour Joining phylogenetic tree (Saitou & Nei, 1987) was constructed using pairwise distance matrix from F_{ST}, with 100 bootstrap replicates.

Nucleotide divergence as a measure of genetic diversity was assessed in the populations. VCFtools (Auton & Marcketta, 2015) was used to estimate per site nucleotide diversity across the three subspecies, *Z. l. pugetensis, Z. l. gambelii* and *Z. l. oriantha* and for the two habitat types, alpine coniferous and riparian deciduous. Divergence along the genome was viewed side by side for the three subspecies, and the two habitat types.

2.2.3.3 Loci under selection

Our study applied a combination of different outlier analysis methods to investigate loci under selection. The choice of at least three methods: BayeScan, PCAdapt and Quantile-based simple outlier was due to the sensitivity, flexibility, and effectiveness of the methods and to make certain that the loci are consistently identified (Pérez-Figueroa, 2010; Dalongeville et al., 2018).

VCFtools, a Variant Call Format program used to process large SNP data and compute essential analyses for variation between and within populations, was used to compute Weir and Cockerham F_{ST} values for the two habitat types. Distribution of loci with significant genetic differentiation (only loci with high F_{ST}) was viewed with Manhattan plot produced in R being a common way of visualizing loci or genes under selection especially for genome wide association studies (GWAS) (Fraslin et al., 2022) and outlier loci were identified by applying quantile function set at 99.9 using Rstudio version 1.3.1093.

BayeScan (Fischer et al., 2011) was another method used to detect outliers in our study. The program uses a pairwise analysis between individuals from the different forest types using a False Discovery Rate (FDR) of 0.05. Outliers were those loci deviating from the detection factor of q value = 0.05.

Lastly, PCAdapt (Luu et al., 2017) was also used to identify outlier SNPs. PCAdapt can partition individuals into groups or clusters and link SNPs to the clusters. The program assigns Bayes factor scores to each SNP, with a larger score indicating more support for local adaptation. Quantile-based, simple outlier method, a traditional method for detecting outliers was implemented in R with the quantile set at 99.9%.

2.2.3.4 Gene identification

I searched for genes close to regions containing our outlier SNPs positions. With a combination of NCBI gene viewer and the ShinyGO v 0.76 program (Ge et al., 2020) for gene annotation and ontology. The ShinyGo tool is an intuitive, graphical web application based on annotation database derived from Ensembl and STRING-db. We identified

common genes from the outlier loci found in the three methods to look for outlier loci. The genes were identified, and their functions investigated, particularly with reference to the habitat conditions.

2.4 Results

2.4.1 Population genetic structure in Zonotrichia leucophrys habitat ecotypes

To account for subtle variation within populations and to establish strong, consistent support for the differentiation, multiple programs were used to assess genetic structure in the habitat ecotypes. Genetic structure detection programs, DAPC and PCoA consistently gave similar results. The populations assessed with SNPs marker using 3dRAD sequencing showed some structuring (Figure 2.2), with an indication of two mixed genetic clusters and a distant cluster (individuals from Cypress Hills) with all the loci (18,481 SNPs) after filtering, however, the plot was inconclusive due to the mixing of individuals between the habitat types. For instance, some members of the coniferous populations clustered with the deciduous cluster. Using only 137 F_{ST} outlier SNPs gave a clearer structure within the habitat types. DAPC membership probability plot showed membership probability of one for all members within each group i.e., members with the value of one exclusively belong to the cluster they are assigned to (Figure 2.3A). Likewise, PCoA using the 137 outlier loci from 3dRADseq showed clear clustering of individuals into either alpine coniferous or riparian deciduous forest, largely supported by the first principal components with genotype data (Figure 2.3B). In addition, 222 outliers from lcWGS provided a similar support for the two ecotypes (Figure 2.3C) separating the individuals into two groups: alpine coniferous and riparian deciduous habitats. The separate clustering of the deciduous

individuals on Figure 2.3C does not correspond to any biologically meaningful groupings with respect to the habitat types. The admixture proportion map showed the approximate distance of sample locations to each other and individual assignment to the forest type, indicating the contiguous habitat, and its heterogeneity when described as a species range (Figure 2.4).

2.4.2 Population genetic structure in the Zonotrichia leucophrys subspecies: pugetensis, gambelii and oriantha

A combination of 3dRADseq and lcWGS was applied to assess genetic differentiation in the subspecies. The PCoA plot with 129 outlier SNPs across 84 individuals showed three genetic groups corresponding to three of the subspecies (Figure 2.5). The PCoA shows an individual from *Z. l. pugetensis* clustering a considerable distance away from its group, this individual is from Vancouver Island in British Columbia, a known contact zone between *Z. l. pugetensis* and *Z. l. gambelii*. Likewise, a mix of individuals between the two clusters of subspecies *gambelii* and *oriantha* are samples collected from Beaver Mines and Crowsnest Pass, within the contact zone between subspecies *oriantha* and *gambelii.*

Outlier SNPs were called from the lcWGS dataset and used for genetic structure analysis. Both PCoA (Figure 2.6A) and LEA structure plot (Figure 2.6B) showed four substantive population groupings with 576 SNPs providing support for the genetic differentiation of subspecies *pugetensis*, *gambelii* and *oriantha* and the divergence of northern *Z. l. oriantha* from southern *Z. l. oriantha*. The first axis of the PCoA explained 41.96% of the variation and separated *pugetensis* from the other two subspecies, the second axis (18.76%) separated *gambelii* and *oriantha*.

2.4.3 Genetic distance in both subspecies and habitat ecotypes

On the neighbour joining tree, constructed with 137 3dRAD outlier SNPs for two habitat types, clusters with strong node support are shown in Figure 2.7. For our subspecies genetic distance analysis, there was a need to control for the effect of within population variation. To minimize the effects of habitat variation, we retained the same individuals for *Z. l. pugetensis* and *Z. l. oriantha* from Colorado (southern *Z. l. oriantha*) but restricted individuals from *Z. l. gambelii* and *Z. l. oriantha* to riparian deciduous forest. The result obtained from our phylogenetic tree (Figure 2.8A) resolved all the subspecies clades with 100 bootstrap support. The tree shows that *Z. l. pugetensis* and *Z. l. gambelii* are genetically closer to the northern *Z. l. oriantha* than the southern *Z. l. oriantha*. Both the phylogenetic tree and the minimum spanning network (Figure 2.8B) provide support for the same from genetic clustering reported earlier.

2.4.4 F_{ST} and nucleotide divergence

Our pairwise comparisons using 3dRAD data pointed to high F_{ST} values between alpine coniferous and riparian deciduous populations with significant F_{ST} values ranging from 0.16 to 0.43 (Figure 2.9 at $p < 0.05$). All the riparian deciduous habitat populations: BM, CH, and WT showed significant genetic difference from the alpine coniferous habitat populations: MK, BR and BA.

Pairwise comparison for all the subspecies using the lcWGS showed significant differentiation ($F_{ST} = 0.47$) between *Z. l. gambelii* and *Z. l. pugetensis*; and $F_{ST} = 0.53$ for differentiation between *Z. l. pugetensis* and the southern *Z. l. oriantha.* There was no

significant difference between *Z. l. pugetensis* and northern *Z. l. oriantha* at $p < 0.05$ (Figure 2.11A) but this may be due to small sample size.

To identify variation within the genome, we examined a single locus NW_005081640.1 based on its many gene functions compared to other loci assumed to be under selection. Regions within 2 Mb of the locus revealed variation between the habitat ecotypes. Prominent divergence was noted around 500 kb, between 1 and 1.5 Mb and around 2.5 Mb (Figure 2.10). Subspecies nucleotide divergence along the same locus spanning across 3 Mb of the genome showed some variation as well, the region around 900 Mb most especially showed some variation between subspecies *gambelii*, northern and southern *oriantha*. While the region around 1.9 Mb clearly shows variation between all the groups (Figure 2.11B).

2.4.5 Loci under selection

Different tests were run to detect possible loci under selection within the habitat ecotypes. Figure 2.12A shows the per site genetic differentiation analysis using the Weir and Cockerham F_{ST}. Outliers ranged from 0.3 to 0.5 with a quantile cut off set at 99.9. PCAdapt and BayeScan (Figure 2.12B) tests were run as additional tests to identify the outliers. Figure 2.13 shows the comparison of the three methods used, with PCAdapt identifying a total of 88 unique SNPs as outliers, BayeScan identified 24 and the quantile-based simple outlier analysis identified 18 unique outliers. The quantile method shared the largest number of outliers SNPs with PCAdapt (seven loci), followed by PCAdapt and BayeScan (two loci), and BayeScan and the quantile method (two loci). Only one locus NW_005081556.1 was identified using all three methods.

2.4.6 Gene identification

Nine outlier SNPs common to at least two methods were selected for gene identification. These SNPs were isolated from the per-site F_{ST} or nucleotide divergence analysis for their positions on the scaffold. Eight of the nine outlier loci identified by two or more methods for detecting loci under selection contained one or more known functional genes. A total of 10 genes: AGO2, PTK2, MAP3K15, SH3KBP1, LSAMP, PCLO, CTIF, BRINP1, PPIL1, and BARX2 were mapped on the loci, and three of the loci (Figure 2.14) contained genes that are related to stress in animals most often triggered by environmental factors. Table 2.2 summarizes the genes that may be linked to the divergence observed in the habitat ecotypes along with their function. Figure 2.15 gives a general representation of the functional network in the most diverse gene identified, BRINP1.

2.5 Discussion

2.5.1 Population genetic structure in habitat ecotypes

A previous study by Welke et al. (2021) proposed the concept of ecotype for some populations within white-crowned sparrow clustering according to two habitat types: alpine coniferous and riparian deciduous habitat populations, not all individuals were resolved into the habitat ecotypes in their study, and that called for consistent support using other high-resolution markers to support their preliminary findings. The major focus of this study therefore is first to establish the genetic differences between populations of white-crowned sparrow found in alpine coniferous from those found in riparian deciduous forest. Our study of genetic structure in two *Zonotrichia* subspecies: *gambelii* and *oriantha* was

accomplished with a combination of two clustering methods: DAPC and LEA with both 3dRADseq and low coverage whole genome sequencing (lcWGS). The forest populations assessed with outlier SNPs generally showed two population clusters with individuals having between 0.8 to 1 membership probability (Figure 2.3) to either riparian deciduous or alpine coniferous forest ancestry. The microsatellite markers earlier used by Welke et al. (2021) reported the same differentiation between the two ecotypes. Many of the results from our study were based on outlier SNPs analyses, a fine scale SNPs dataset assumed here to be effective in identifying recent divergence as complete SNPs with large amount of data with thousands of both neutral and significant SNPs were not able to give clear genetic structure (Figure 2.2, Appendix 2). However, an interesting finding with the complete SNPs data showed the divergence of Cypress Hills population from all others. Cypress Hills has been described as a sky island due to its significantly high elevation compared to the surrounding lowland, and it is known to harbor genetically distinct and isolated populations of warbling vireos *Vireo gilvus* (Carpenter et al., 2021; Dempsey et al., 2020). This may provide some support for the differentiation of populations due to habitat structure and conditions. Using outlier loci, our LEA and PCoA analyses, provided strong support for habitat related differentiation for forest populations of *Zonotrichia leucophrys*.

Our study has shown genetic divergence corresponding to the heterogenous habitat within the home range of *Zonotrichia leucophrys* indicating limited gene flow between alpine coniferous forest and riparian deciduous forest populations. Because variation in the habitat can lead to population fragmentation if low gene flow persists, the populations can diverge both morphologically and genetically in response to the local condition of each population (Barley et al., 2015, Marchán et al., 2020). Therefore the different habitat

conditions may explain the genetic differentiation reported for our populations. We generally do not expect large variation between these two populations since these habitats are contiguous and are largely defined by the dominant tree species at each sampling site. Though Kawecki & Ebert, (2004) noted that variation in habitat as seen in our study locations, can be characterized by several factors including small to large climatic variation, and can consequently result in local adaptation for the ecotypes, most of the habitat ecotypes in our study are geographically very close to each other (Figure 2.1).

Forest vegetation has been inferred to influence predation risk. LaManna et al. (2015) reported that in the presence of increased coniferous vegetation, predation risk reduces for chipping sparrow but not for other sparrow species. Other studies have shown that in hard wood forest habitat, ground nest predation was lower compared to understory nest (Martin, 1993), and others reported that nest predation was higher in coniferous forests than in deciduous forest (Seitz and Zegers, 1993). The implication of this for our white-crowned sparrow populations could be that the different habitat conditions created by the two habitat types are influencing predation density differently, and the white-crowned sparrow populations are responding to this variation. This may then result in adaptation of the populations to these local conditions since predation rate differs among different habitats and forests (Belammy et al., 2018). In addition, these habitat ecotypes may have developed different songs further enhancing their adaptation to their habitat as seen in willow flycatcher where geographically adjacent populations have distinct vocal identity (Sedgwick, 2001). However, there is a need to study this in detail. Our study showed regions of divergence within locus NW_005081640.1 and nucleotide variation corresponding to the divergence (Figure 2.10), it is suspected that these regions are associated with the genetic differences recorded for these ecotypes.

2.5.2 Genetic differentiation in the Zonotrichia leucophrys subspecies: pugetensis, gambelii and oriantha

Our results from the 3dRAD and lcWGS showed that the three subspecies are genetically distinct. Previous studies have pointed to the divergence within the *Zonotrichia leucophrys* clade linked to the historical processes resulting in their speciation in the Pleistocene (Zink et al., 1991; Morton, 2002). More recent studies from Taylor et al. (2021) and Welke et al. (2021) both pointed to divergence especially for *nuttalli* and *pugetensis* separating from *gambelii, oriantha* and *leucophrys* with a combination of microsatellite and nuclear SNP markers. Our structure plots did not show evidence of intergradation between *pugetensis* and either of the two other subspecies (*gambelii* and *oriantha*). This could be related to the geographic distance between subspecies *pugetensis* and the other subspecies, the non-overlapping breeding range of subspecies *pugetensis* with the other two, and the high fidelity of both subspecies *gambelii* and *oriantha* to their breeding sites (Morton, 2002).

Our results with the low coverage whole genome data showed divergence within the *Z. l. oriantha* populations. Populations from the northern part of the range (Alberta and British Columbia) clustered separately from those in the south (Colorado). Welke et al.'s study (2021) had previously documented this separation with microsatellite data. As opined in the earlier study mentioned above, geographic distance, and the southern *oriantha* having the genetics of the pure parental form of *Z. l. oriantha* are factors implicated here. This finding however requires detailed studies in future to understand separation within this *Z. l. oriantha* populations which will provide better understanding to studies of both ancestral populations and their diverged populations with opportunity to unravel factors that have led to their divergence.

Zonotrichia subspecies delineations have in the past been supported by distribution, migratory and phenotypic characteristics (Cortopassi & Mewaldt, 1965; Morton, 2002; Rand, 1948), while recent studies with different markers have reported weak to moderate genetic differences for the *gambelii, oriantha* and *leucophrys* subspecies (Taylor et al., 2021; Welke et al., 2021), our study has clearly shown that subspecies *gambelii* and *oriantha* are closely related, but genetically distinct. Though Taylor et al. (2021) had their *Z. l. gambelii* from Alaska and Manitoba, outside our sampling sites, and some of their *Z. l. oriantha* came from within our sampling sites: Colorado and Beaver Mines, they generally had very few *oriantha* samples compared to ours. While samples from Welke et al. (2021) are same sampling sites as ours, though not all of their samples and sites were used in our study and not all of ours appeared in their study. Our approach has also equated the power of microsatellite markers in detecting recent divergence to that of outlier SNPs.

We assessed nucleotide and F_{ST} divergence in the three subspecies with 576 SNPs from lcWGS data, pairwise F_{ST} revealed the level of differentiation with values from 0.47 for comparison between *Z. l. gambelii* and *Z. l. pugetensis* to 0.53 between *Z. l. pugetensis* and the southern *Z. l. oriantha.* There was no significant difference between *Z. l. pugetensis* and northern *Z. l. oriantha*. This implies that subspecies *pugetensis* is closer to the northern *Z. l. oriantha* than to the southern *Z. l. oriantha*. This observation is quite interesting as this may mean that one of the *Z. l. oriantha* groups is of the ancestral form, and both genetic groups may be helpful in studies relating to recent divergence and evolution.

The nucleotide divergence reported for the first time from our study along the locus NW_005081640.1 spanning across 3 Mb of the genome showed region of divergence around 900 kb between subspecies *gambelii*, northern and southern *oriantha*. While the region around 1.9 Mb clearly showed variation between all the groups. These regions of

divergence may have different alleles which may have contributed to the genetic differentiation shown for these subspecies. Our study has confirmed that *gambelii* and *oriantha* are genetically distinct. We were able to find differences between *Z. l. gambelii* and *Z. l. oriantha* with outlier SNPs from lcWGS and 3dRADseq data.

2.5.3 Detecting loci under selection for the habitat ecotypes, and identifying associated genes

A combination of approaches for detecting outlier loci that may be driving selection was used in our study. Multiple genes: PTK2, MAP3K15, SH3KBP1 and BRINP1 were identified within three outlier loci. All have functions relating to cellular stress. Some studies on gene-environment association have linked cellular stress to environmental factors such as ultraviolet radiation and DNA damage, reduced or excessive nutrients, and presence of some environmental compounds (Llanos et al., 2009; Shiozaki, 2009; Ongusaha et al., 2008).

BRINP1 is a very diverse gene with many of its functions overlapping with other identified genes which are linked to defense response behaviour, learning or memory, vocalization behaviour, fear response, short term memory, and cognition (Figure 2.15). These genes are suspected to be one of the factors driving differentiation between the two ecotypes. As reported by earlier studies, predation is one of the important selection forces driving the evolution of prey species (Gliwicz, 1986; Lima, 1998). Many species avoid becoming prey by living out of sight of their predators e.g., horseshoe bat emerging at different period to reduce predation risk (Duverge et al., 2000). In some, it may lead to species gradually moving towards habitats with low predator densities. It is likely that this is the case for either of our habitats, preferring habitat that either conceal their presence or

is unconducive to their predators. This is further supported in a related sparrow species, LaManna et al. (2015) found that increased coniferous vegetation led to an increase in chipping sparrow population, conversely, deciduous forest supported Lincoln's sparrow better. For our study, the stress related genes are further implicated in their link to behavioural defense response and cognition following the work of Berkowicz et al. (2016) where it was reported that mice with BRINP1 knock out, one of the stress related genes identified in our study, displayed reduced sociability, changes in vocalization capacity in addition to autism-like behaviour. These behavioural characteristics may contribute to species susceptibility to predation. It is important to note that some of the genes identified in our study such as: MAP3K15, CTIF, and BRINP1 with roles related to vocalization may be linked in function to very similar genes related t HVC (High Vocal Center), a nucleus in the brain of the songbirds (Mello et al., 1997). Other genes (BARX2, PPIL1, and CTIF) with link to brain development, neural adhesion, olfactory signalling, and sensory processing of sound together with previously mentioned genes may all be playing a role in cognition. Cognition is a very complex trait that may affect ability of animal to detect and respond to changes in its environment. Cognition may affect how animals discern threat: predators or competitors, food availability, mate choice, nesting, migratory time and pattern. Differential expression of genes regulating cognition may therefore by an important factor in the divergence of populations. However, more study is needed to establish a link between white-crowned sparrow and predation risk in both alpine coniferous and riparian deciduous forest where we have found divergence between these populations. Based on the supporting previous studies and our findings, the genes identified in our study may correspond to habitat ecotypes and both microclimate conditions and

predation risk in the forest types are largely suspected to be driving the genetic divergence observed.

2.6 Conclusions

Our study has established for the first time, clear genetic divergence between subspecies *gambelii* and *oriantha* using outlier SNPs. Another interesting finding is that our study found that subspecies *pugetensis* is genetically closer to *northern Z. l. oriantha* than the southern *Z. l. oriantha*, an indication that both *pugetensis* and northern *Z. l. oriantha* may share a larger proportion of their ancestral genes. Further, microclimatic conditions may have a large effect on genetic structure even with contiguous habitat as our study clearly showed genetic differentiation between populations of *Zonotrichia leucophrys* in alpine coniferous and riparian deciduous forests regardless of subspecies. However, there is a need to include other subspecies especially, *Z. l. leucophrys* in future study to determine if this finding applies to all subspecies of white-crowned sparrow known to utilize both deciduous and coniferous habitats. Predation was also suspected to be one of the driving forces for the divergence of the ecotypes, as we found a gene linked to defence behaviour in the ecotypes. This also requires additional study to establish relationship between the ecotypes and predation risk in each habitat.

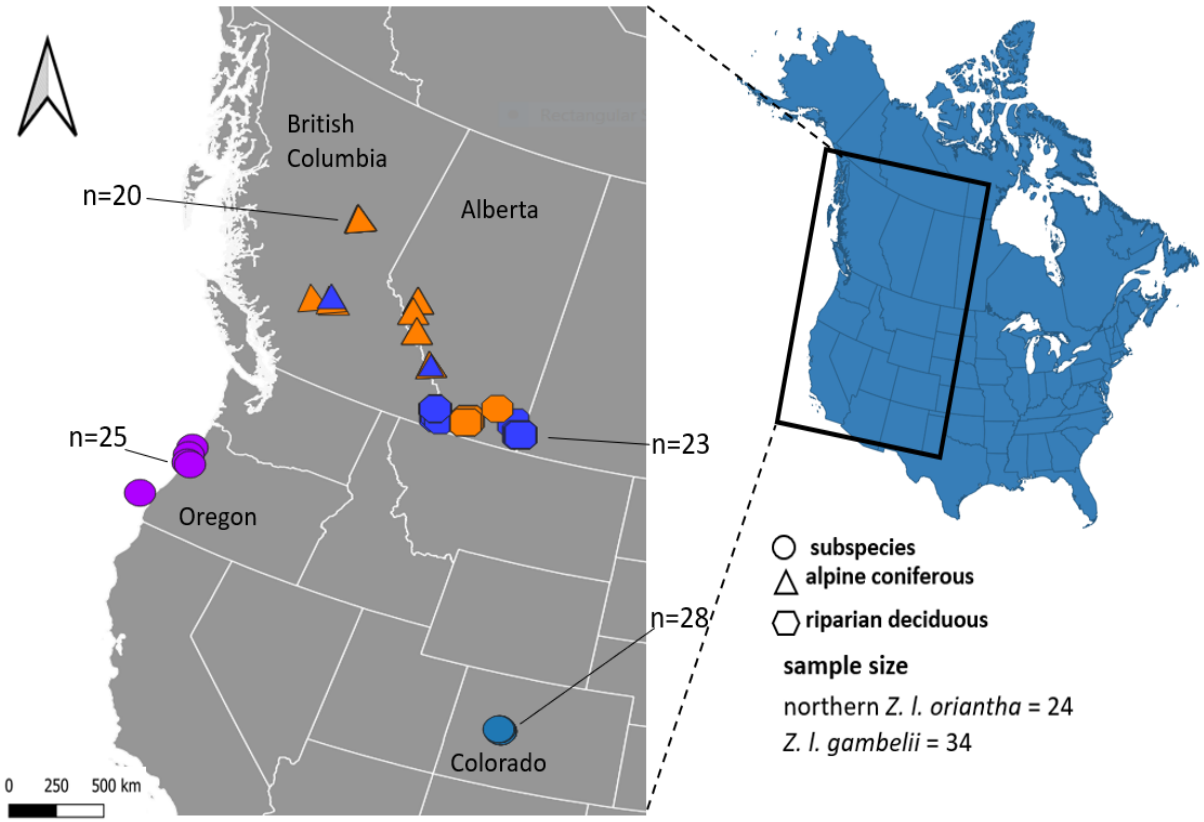


Figure 2.1: Overview of the samples used in this study. Oregon (OR) samples were subspecies *pugetensis* (purple), Colorado (CO) has the southern *Z. l. oriantha* population (dark blue), and both British Columbia (BC) and Alberta (AB) have a mix of northern *Z. l. oriantha* (blue) and *Z. l. gambelii* (orange). Alpine coniferous (triangle shape) and riparian deciduous (octagon shape) ecotypes are also indicated on the map.

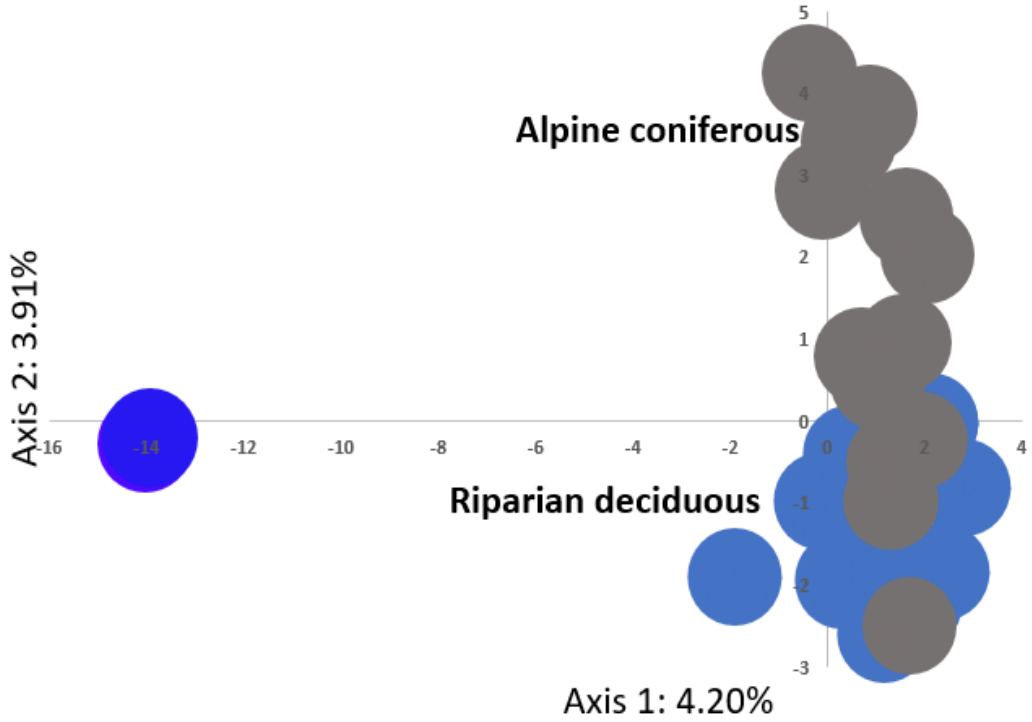


Figure 2.2: Scatter plot for populations found in alpine coniferous and riparian deciduous forests with 28 individuals after filtering using 3dRAD data showed two undifferentiated clusters. The cluster to the left consists of deciduous habitat individuals from Cypress Hills, southern Alberta (filtering option: 40% missing data and 10% missing individuals, 18,481 SNPs).

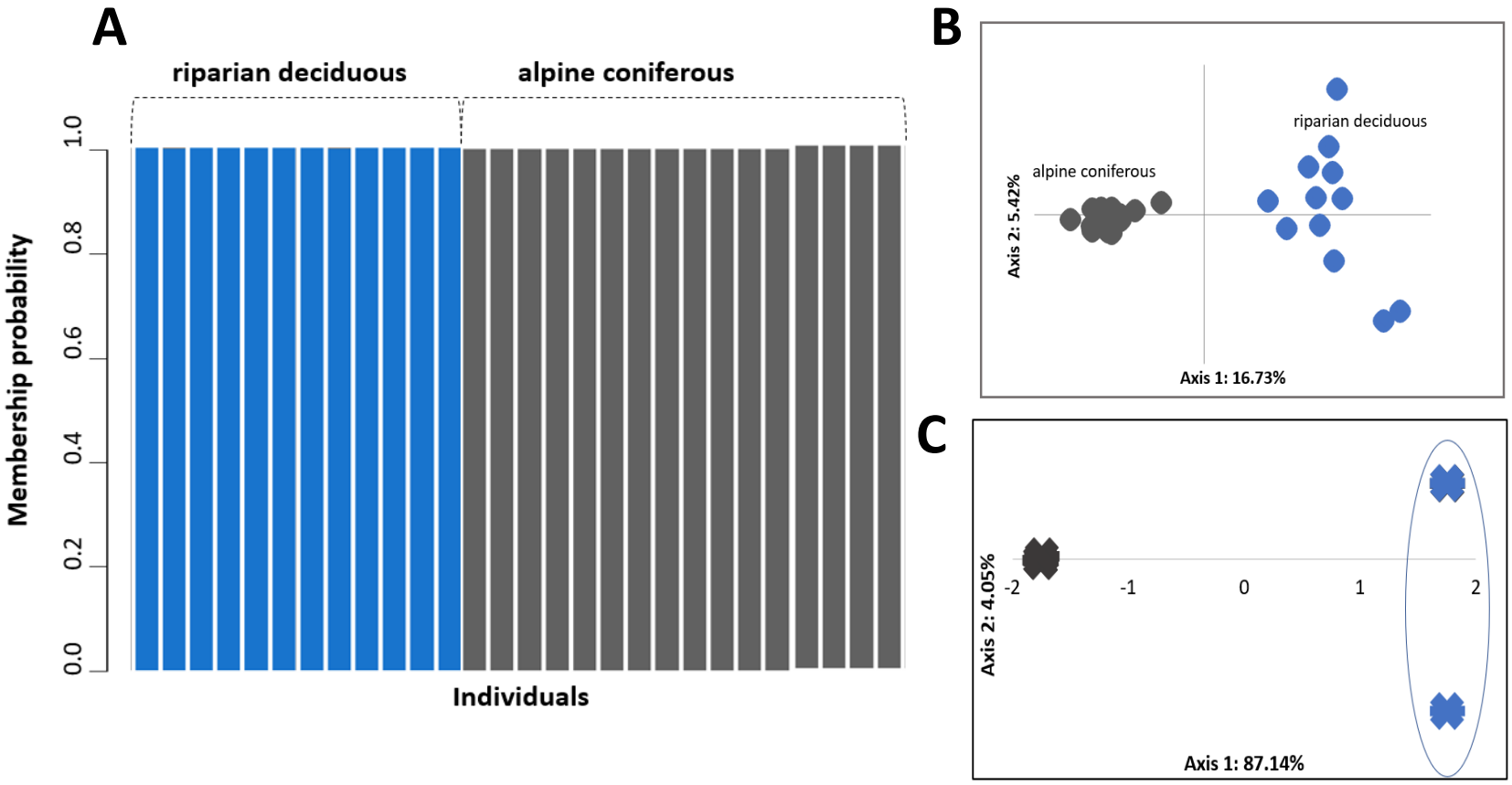


Figure 2.3: DAPC (A) 137 outlier SNPs from 3dRAD, and PCoA (B & C) plots with 137 outlier SNPs and 222 outliers from 3dRADseq and lcWGS dataset respectively, for the two forest populations. The colours represent the habitat types in the bar plots and PCoA, riparian deciduous (blue) and alpine coniferous (dark gray). C. The PCoA on the right is from the lcWGS and the circled individuals are from deciduous habitat, Lethbridge, southern Alberta.

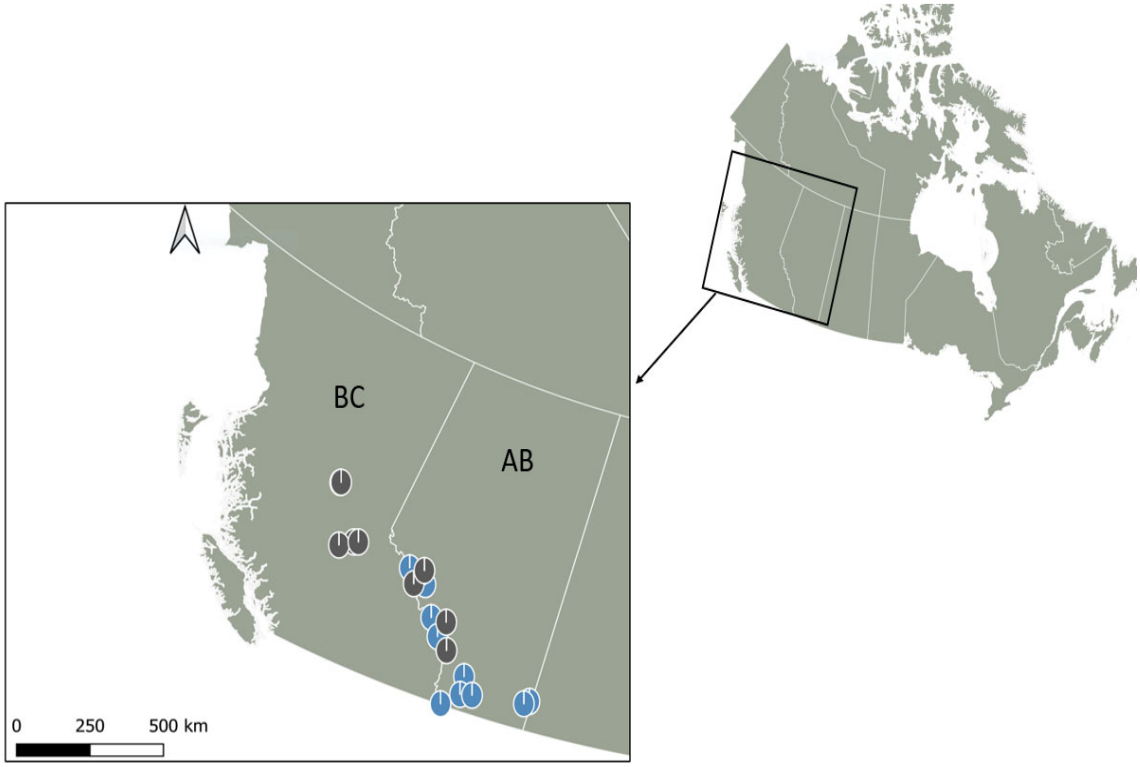


Figure 2.4: Admixture map with ancestral proportion of individuals assessed across the two contiguous forest types. Habitat types are colour coded, riparian deciduous (blue) and alpine coniferous (dark gray).

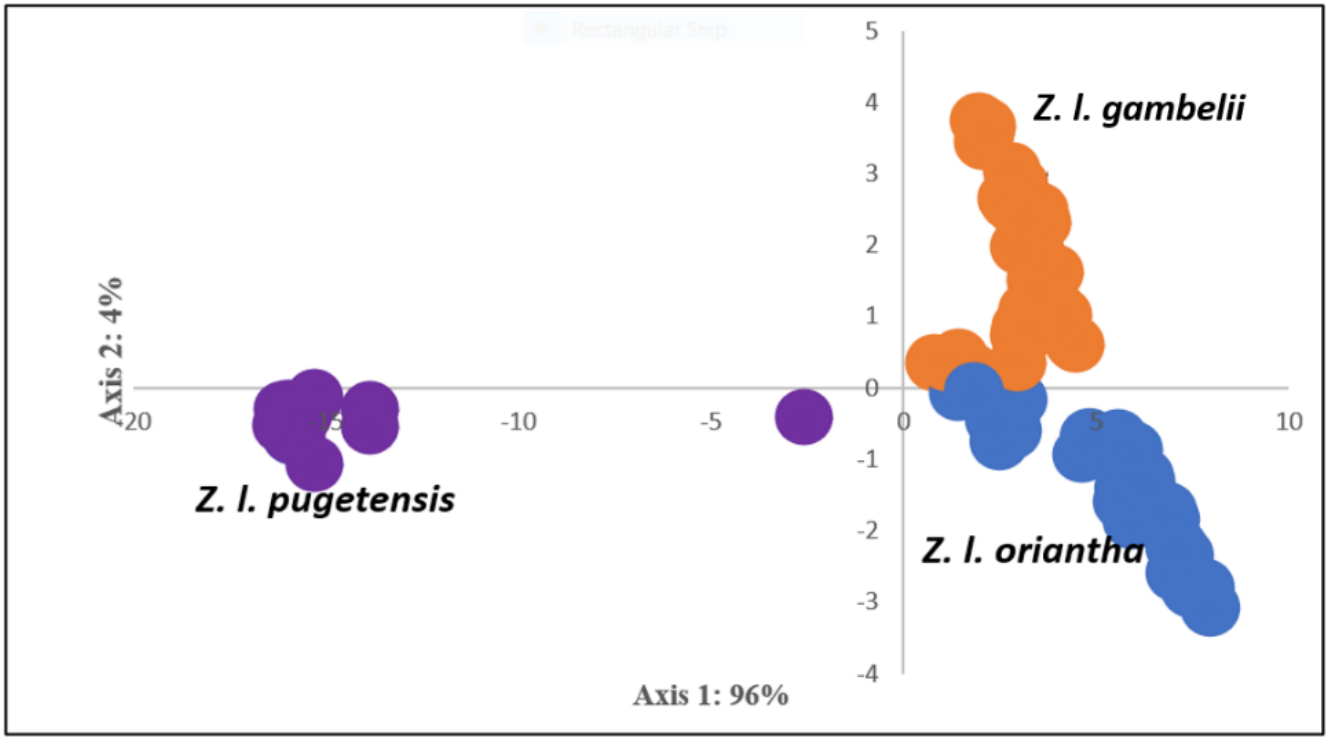


Figure 2.5: PCoA plot of the 3 subspecies using 129 outlier SNPs from 3dRAD datasets. Subspecies clusters are color coded as: *Z. l. pugetensis* (purple), *Z. l. oriantha* (blue), *Z. l. gambelii* (orange). The *pugetensis* individual located away from its group is a sample collected from Vancouver Island in British Columbia, a contact zone between *gambelii* and *pugetensis*, and a mix of individuals between the *oriantha* and *gambelii* clusters are samples from Beaver Mines and Crowsnest Pass in Alberta corresponding to the intergradation zone between *gambelii* and *oriantha*.

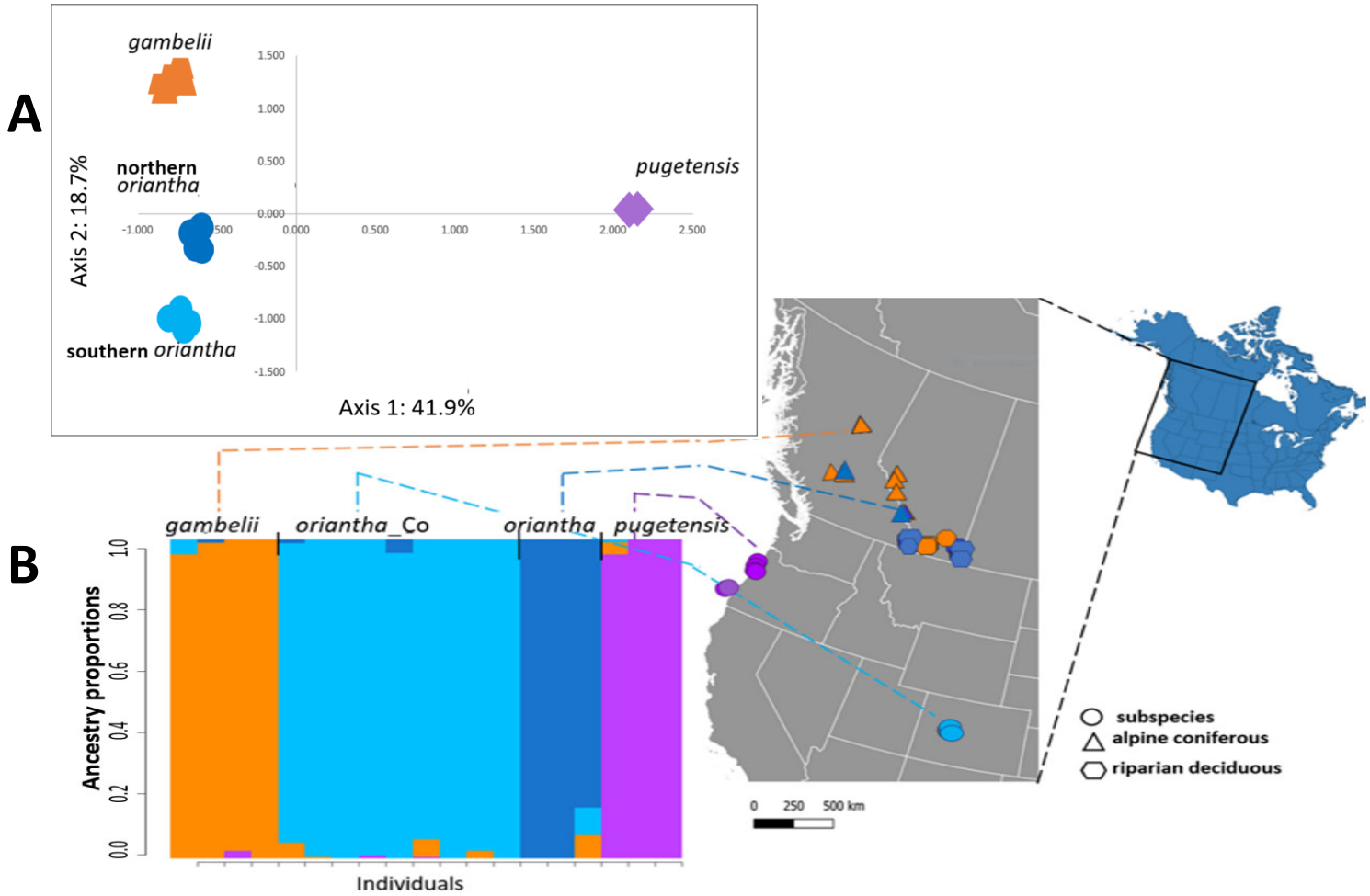


Figure 2.6: A. PCoA indicates four genetic clusters color coded as in the structure plot (B), orange represents *Z. l. gambelii*, light blue, Z. *l. oriantha* (southern), dark blue, *Z. l. oriantha* (northern), and purple is *Z. l. pugetensis*. **B** shows the origin of the genetic cluster on the map above it. (Dataset from 576 lcWGS SNPs after controlling for within group variation).

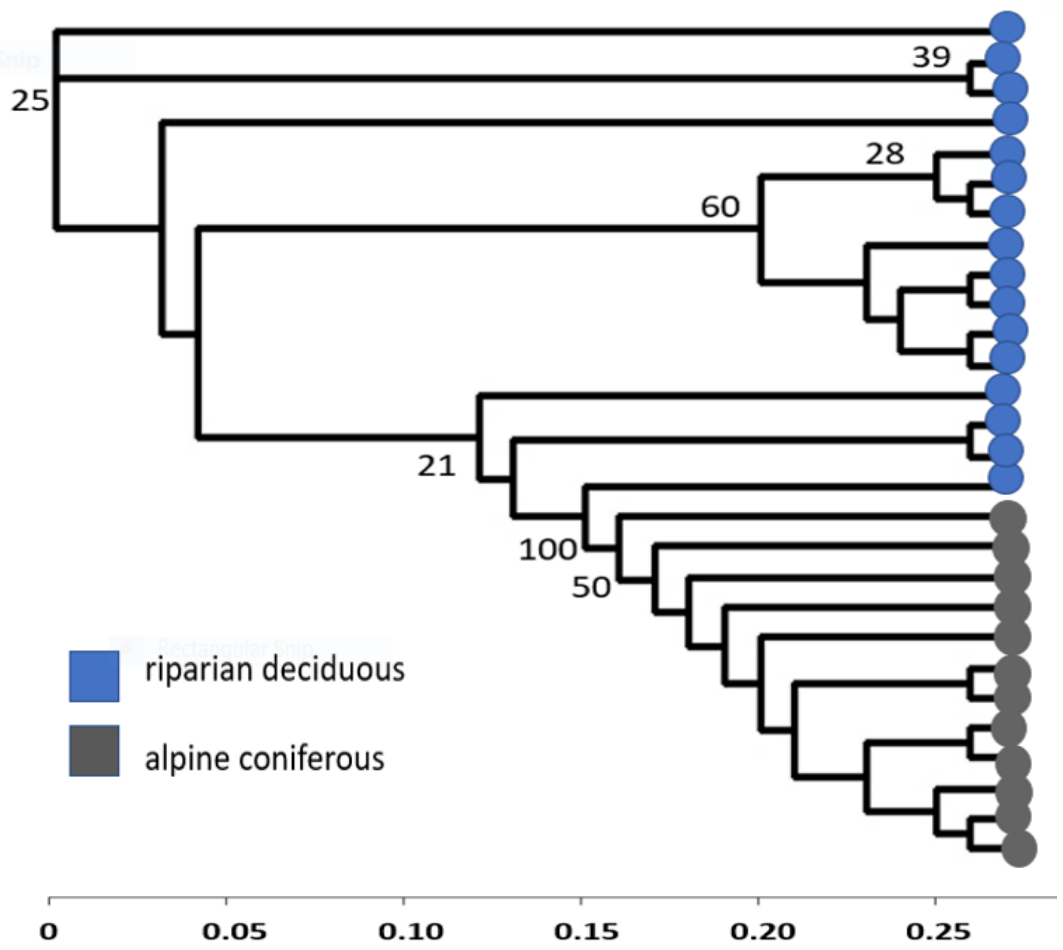


Figure 2.7: Neighbour joining tree with 137 outlier SNPs for the habitat types using 3dRAD dataset. Individuals are color coded: blue represents the deciduous populations and dark gray, the coniferous populations.

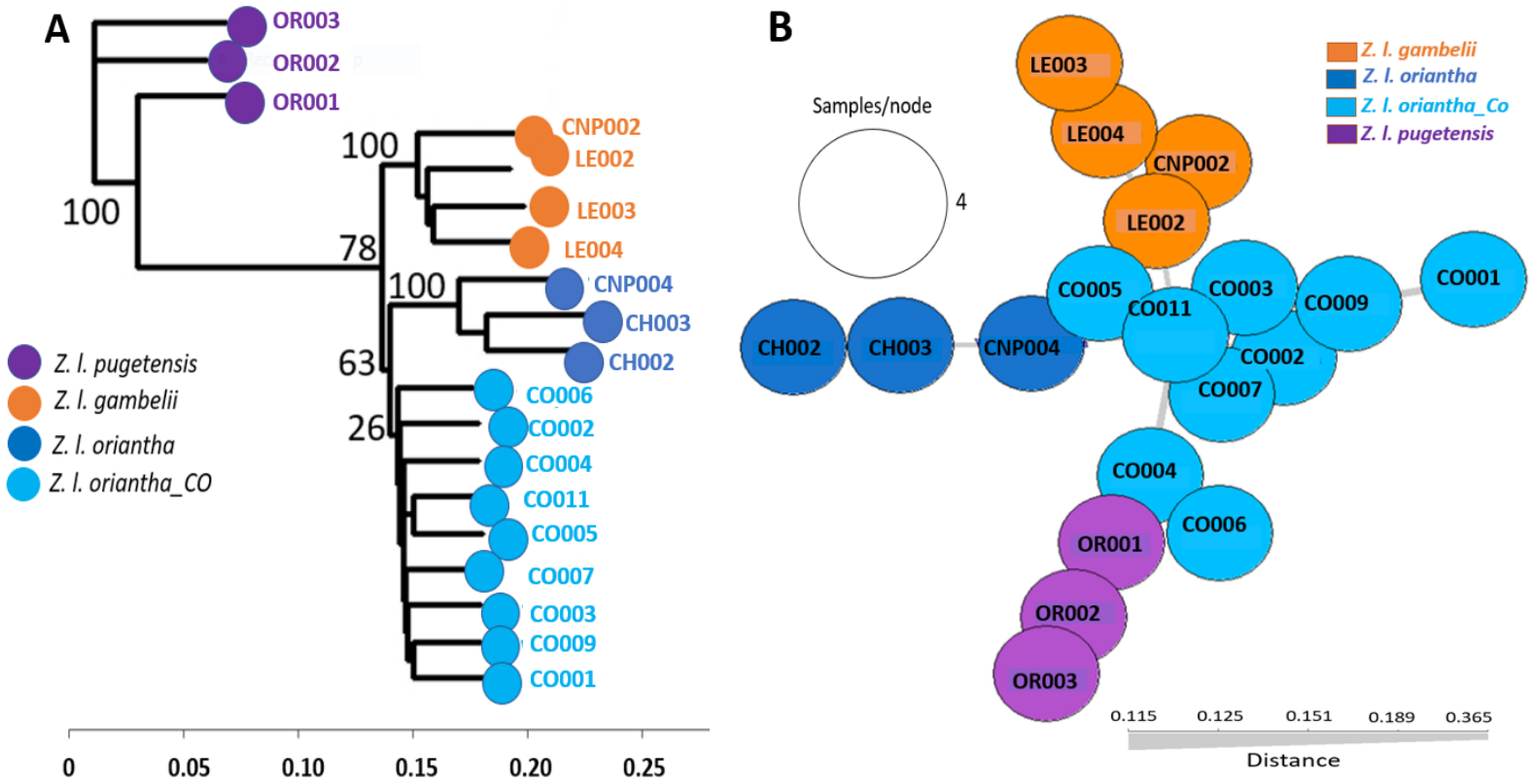


Figure 2.8: Neighbour joining phylogenetic tree (A) and minimum spanning network (B) with lcWGS outlier SNPs for the subspecies. Individuals are color coded on the figures: orange represents *Z. l. gambelii*, light blue: *Z. l. oriantha*, dark blue: *Z. l. oriantha* (southern) and purple: *Z. l. pugetensis*.

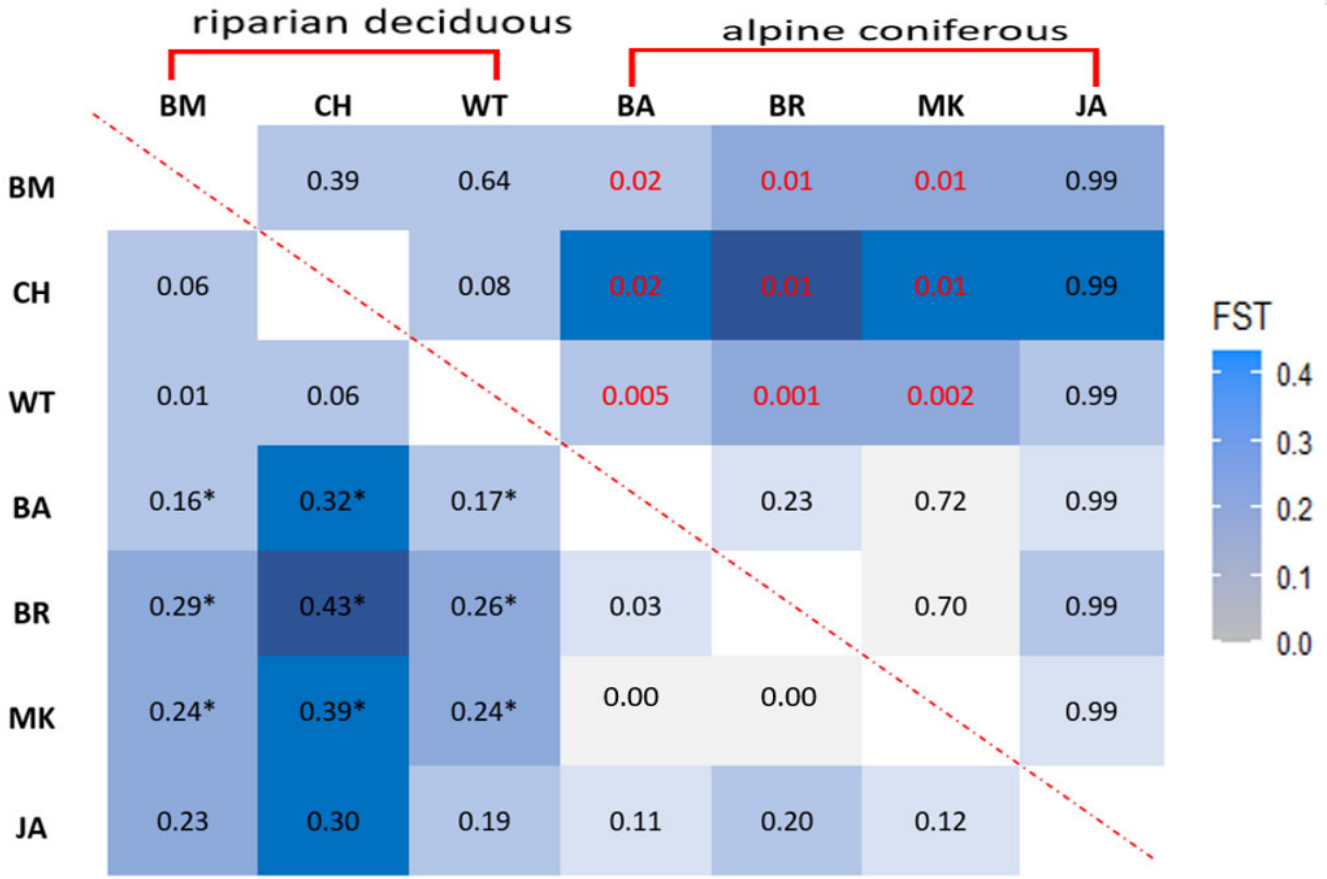


Figure 2.9: Pairwise F_{ST} values (below the diagonal) between individuals of the two habitat clusters using 137 SNPs from 3dRADseq data. Individuals from Mackenzie (MK), Brule (BR), Jasper (JA) and Banff (BA) are alpine coniferous populations, while those from Beaver Mines (BM), Cypress Hills (CH) and Waterton (WT) are the riparian deciduous populations. P values appear above the diagonal and are significant at $P < 0.05$. Significant values are in red.

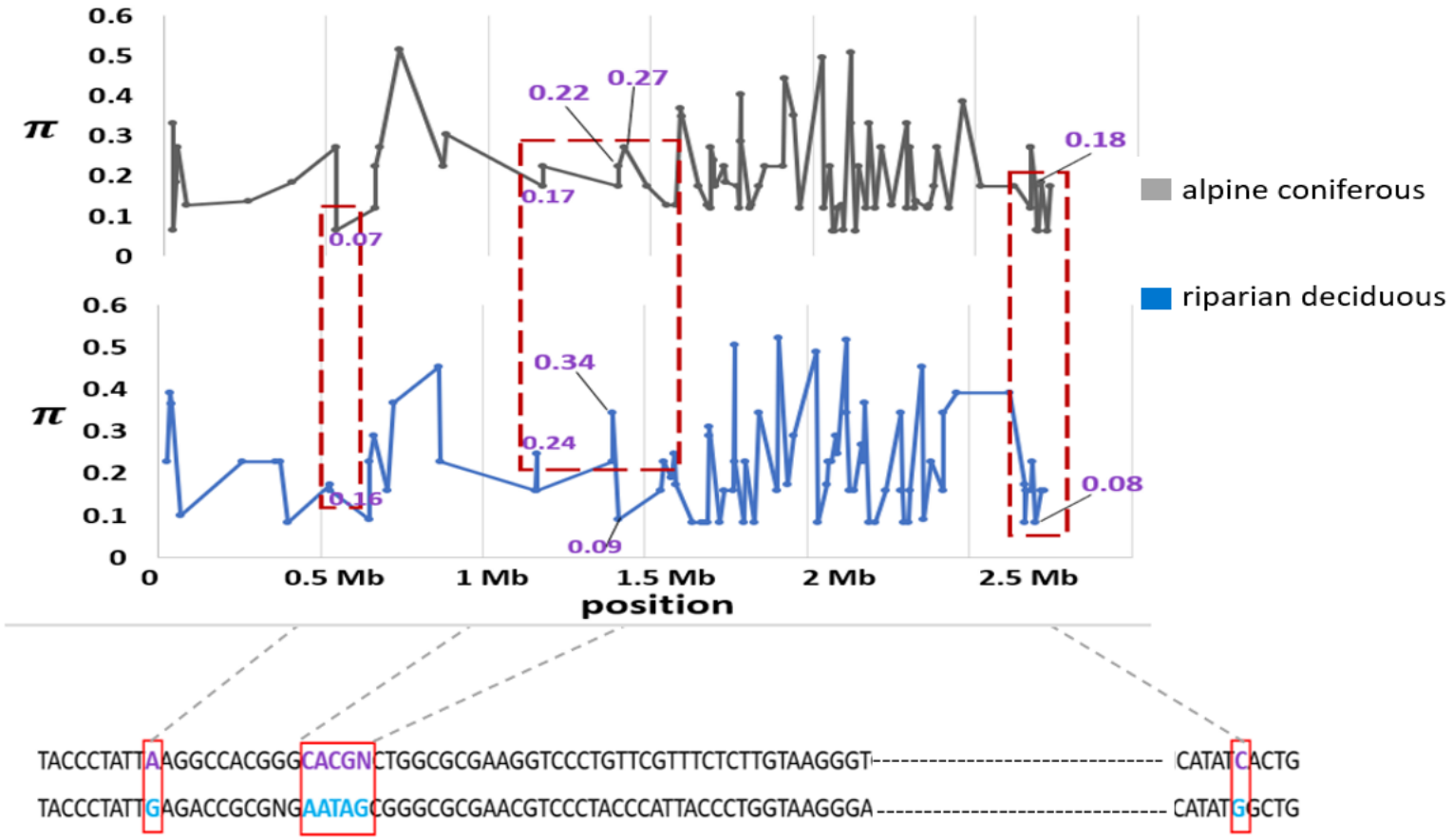


Figure 2.10: Nucleotide divergence is compared between the two habitat ecotypes at locus NW_005081640.1 from the 18,481 3dRAD SNPs, color coded as coniferous (dark gray) and deciduous (blue). Below the nucleotide plot are the sequences. Lines correspond to regions of divergence marked on the plot above it.

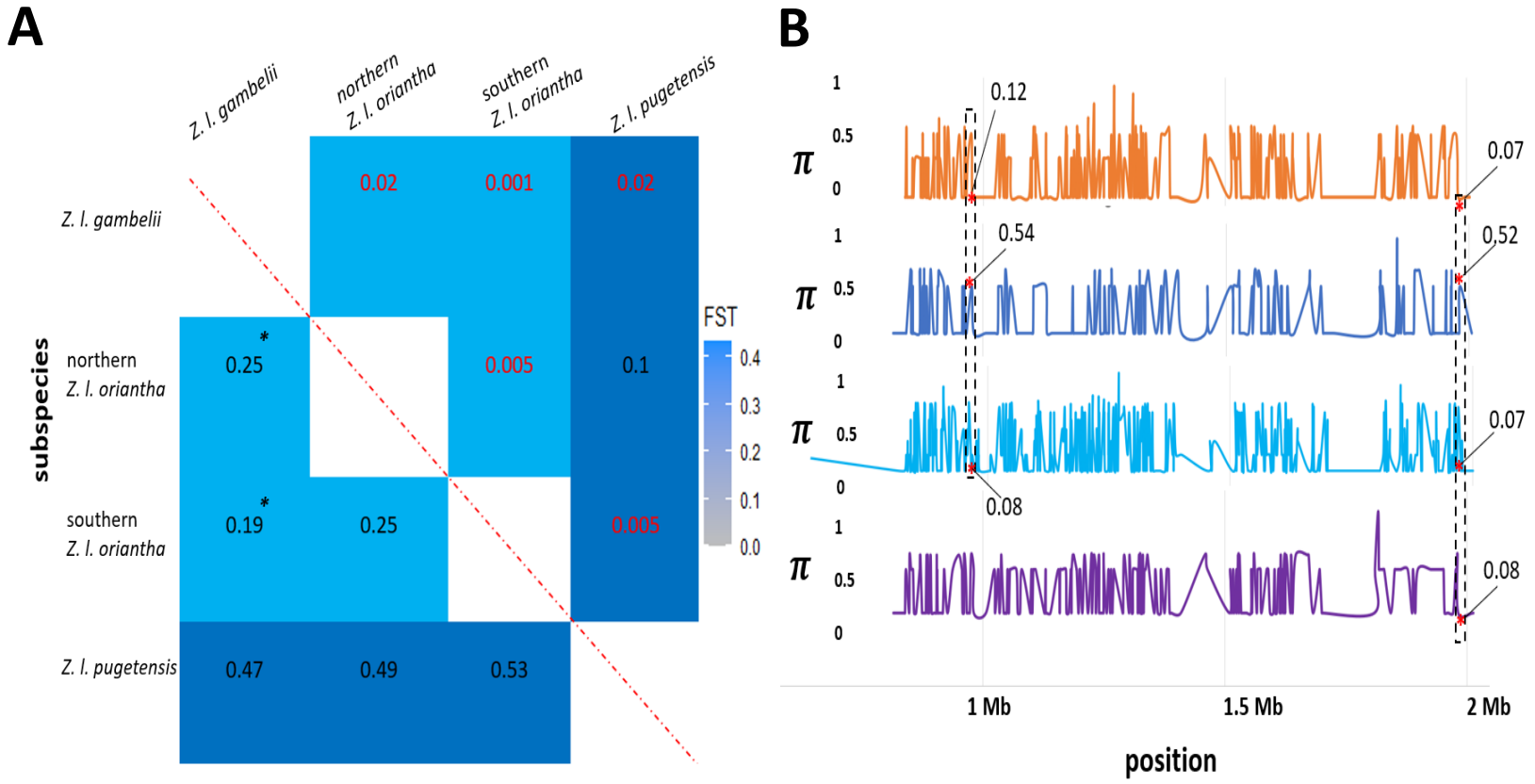


Figure 2.11: Pairwise F_{ST} (A) indicating the distance between the 3 subspecies and the previous genetic cluster corresponding to southern oriantha were each compared. F_{ST} values appear below the diagonal and P values above the diagonal. Significant values are in red. Nucleotide divergence (B) for the four groups, pugetensis (purple), gambelii (orange), northern oriantha (dark blue), southern oriantha (light blue) shows regions of divergence depicted with broken lines for locus NW_005081640.1 with the most diverse gene functions from the thinned lcWGS SNPs.

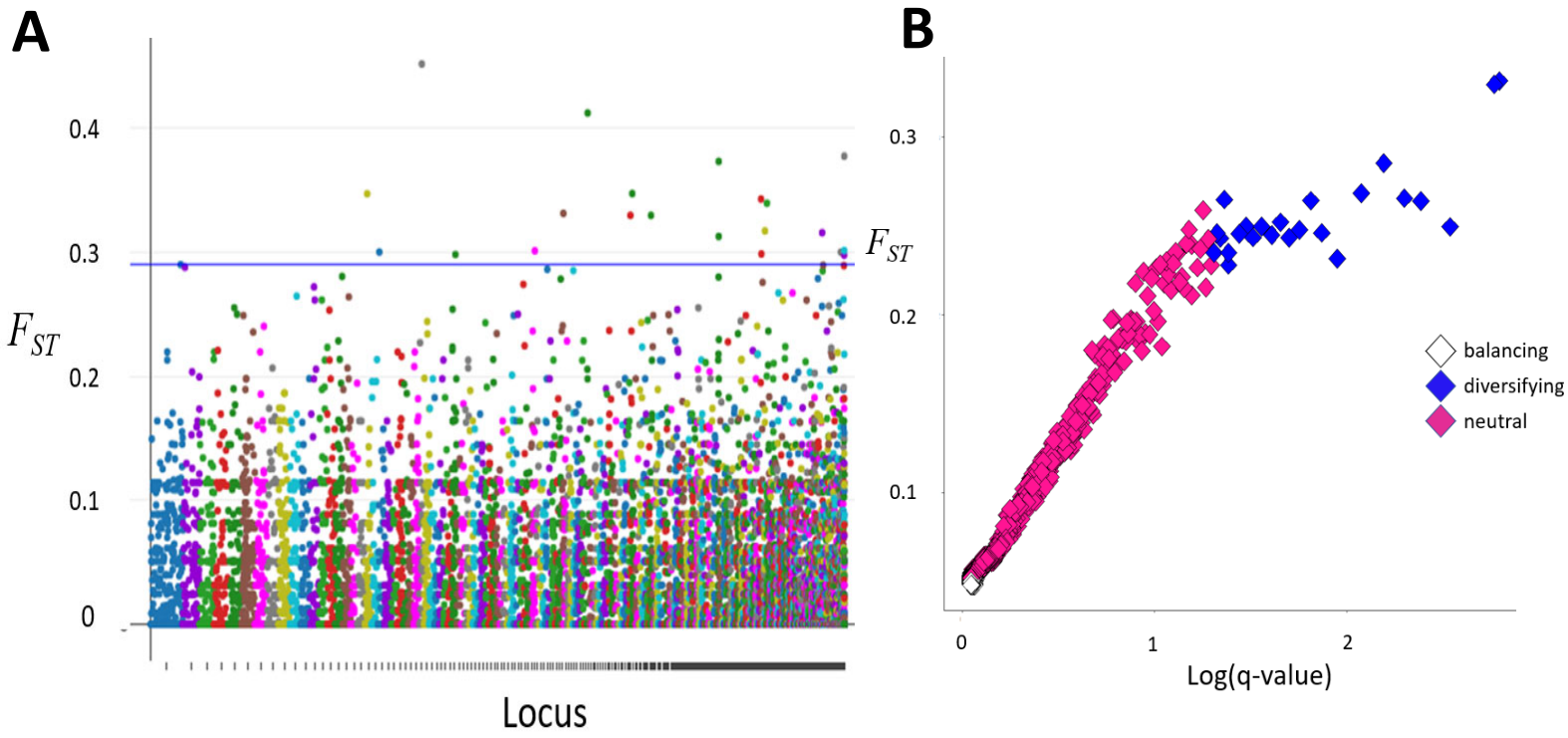


Figure 2.12: Outlier detection methods applied to the 3dRAD seq data (habitat types) show the result of a Manhattan plot (A). Outliers were detected at 99.9 quantile with the simple outlier detection method run with quantile function in R studio. The BayeScan method (B) indicates the presence of outliers. The blue diamonds represent the loci suspected to be under positive selection.

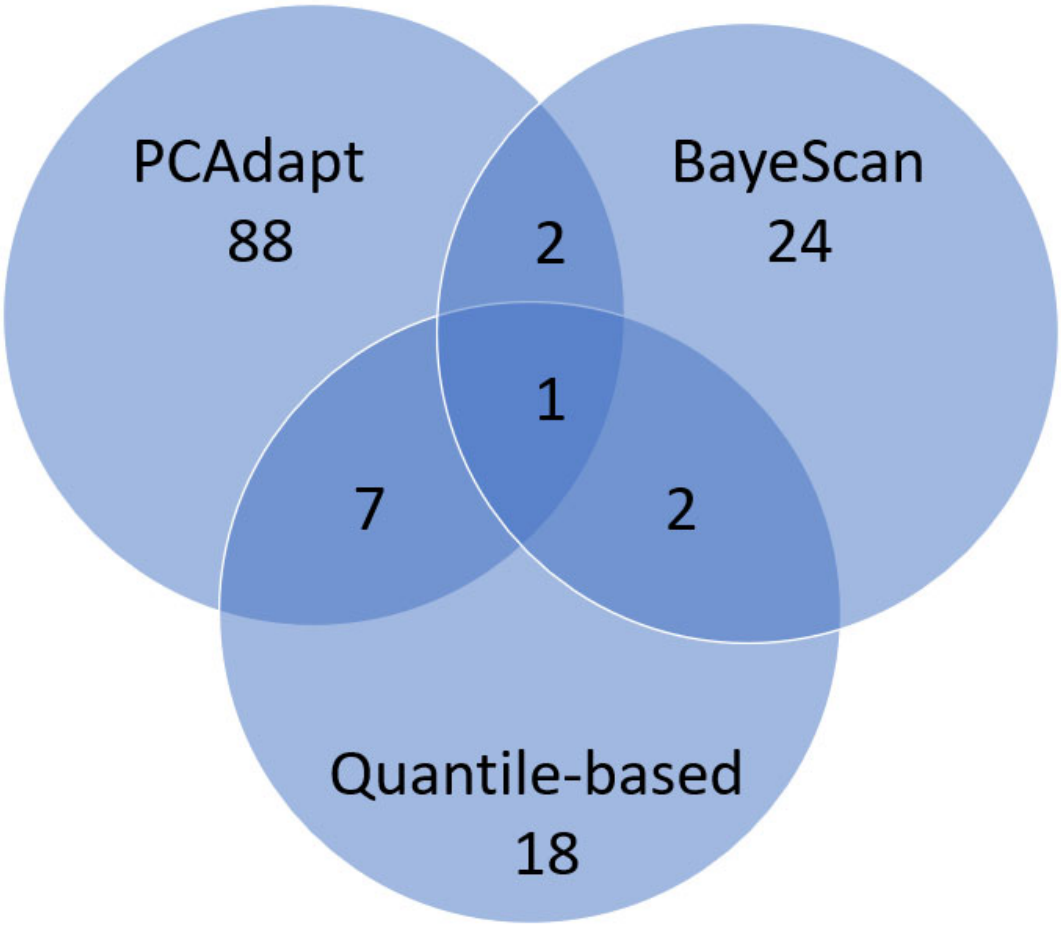


Figure 2.13: Venn diagram comparing the three outlier detection approaches for habitat types. Total number of detected SNPs are indicated in each circle corresponding to the approach used.

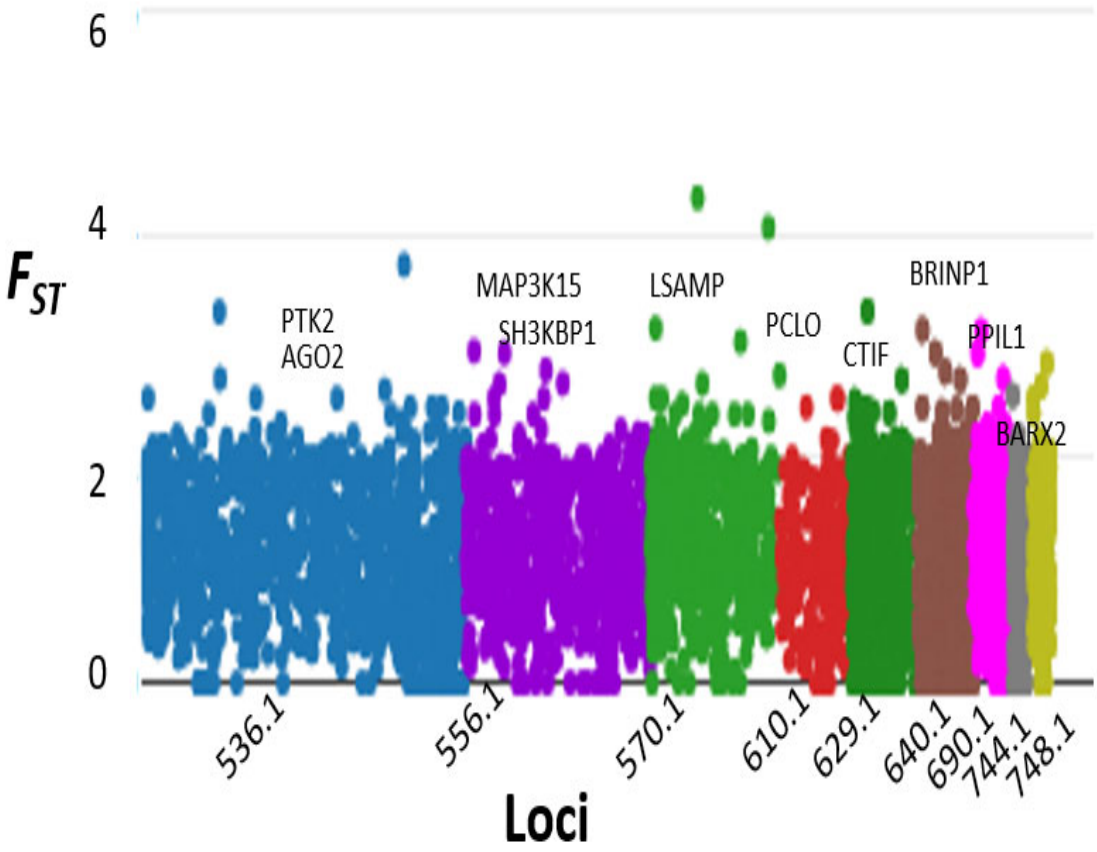


Figure 2.14: Manhattan plot of the F_{ST} values of the selected outliers for the habitat ecotypes using three approaches for detecting loci under selection. Genes linked to stress, suspected to be important to the occurrence of individuals in the forest types are mapped on the scaffold.

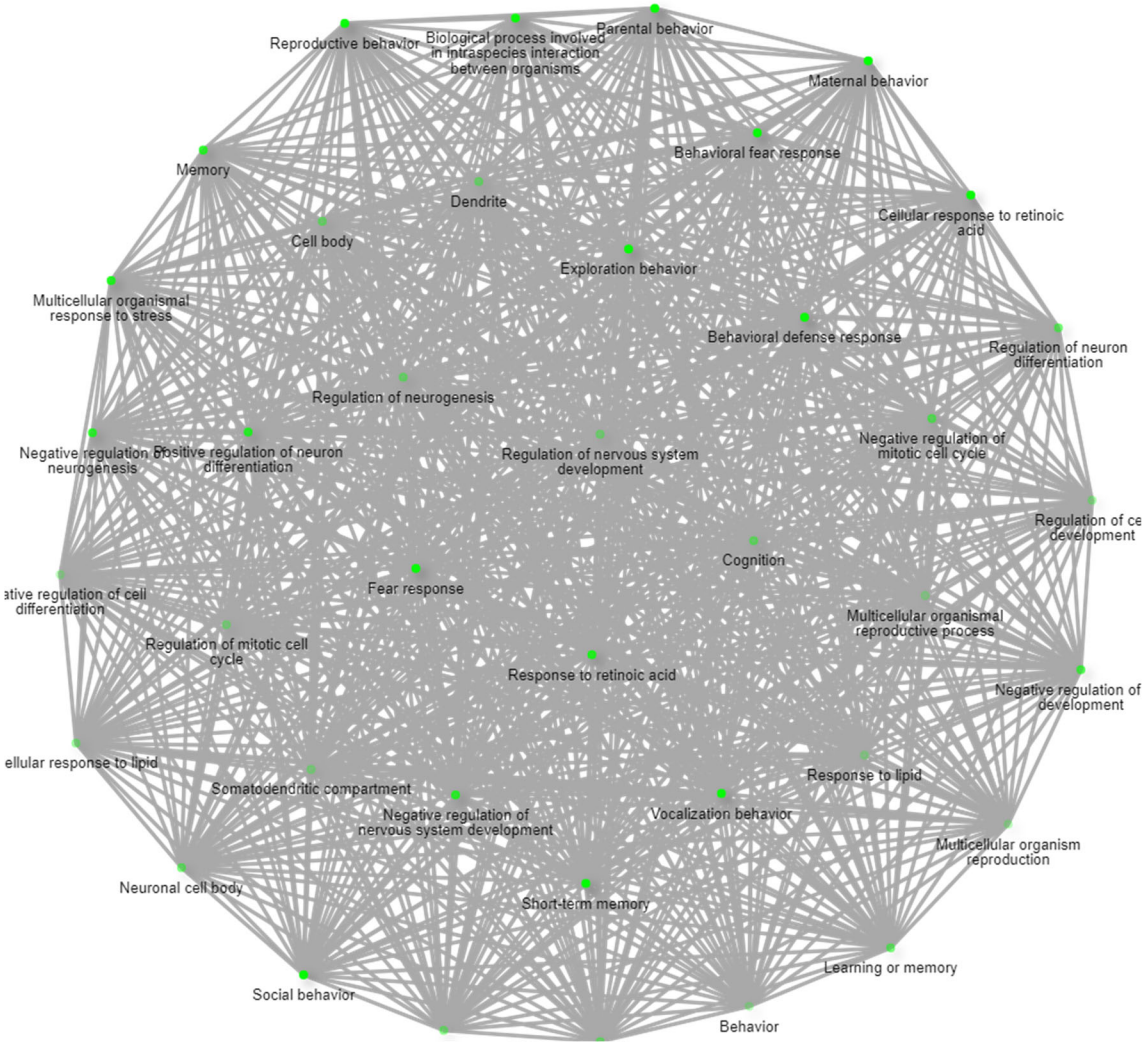


Figure 2.15: Gene function network of the most diverse gene (BRINP1) found on the locus NW_005081640.1.

Table 2.1: Number of samples used in this study with data filtering options applied. Locations are abbreviated as BA Banff, FTSJ Fort St. James, JA Jasper, BR Brule, MK Mackenzie, BM Beaver Mines, CH Cypress Hills, WT Waterton, CNP Crowsnest Pass, LE Lethbridge, RV Revelstoke, CO Colorado, and OR Oregon.

Sequencing method	Initial No of samples/location	No of samples after filtering	Filtering option	No of SNPs retained	Outlier SNPs retained
3dRADseq (subspecies)	BA (n=4) FTSJ (4) JA (1) BR (5) MK (6) BM (3) CH (5) WT (5) CNP (6) LE (6) RV (4) CO (16) OR (19)	84	50_80: Allow SNPs with 50% missing data; Allow individuals with 20% missing data	8,508	129
3dRADseq (habitat types)	BA (4) JA (1) BR (5) MK (6) BM (3) CH (3) WT (6)	28	60_90: Allow SNPs with 40% missing data; Allow individuals with 10% missing data	18,481	137
lcWGS (subspecies)	CNP (2) LE (3) CH (2) CO (9) OR (3)	19	Thinned at 100,000 distances	15,320	576
lcWGS (habitat types)	MK (1) RV (2) LE (4)	7	Thinned at 100,000 distances	15,320	222

Table 2.2: Outlier loci with genes found around their positions in the genome. The functions of the genes are listed (detected with ShinyGO v 0.76, Ge et al., 2020 bioinformatics.sdstate.edu/go74/).

LOCUS	POSITION	GENE	FUNCTION
NW_005081536.1	42083893	AGO2	cell identity maintenance, functions in early development, regulates nuclear-transcribed mRNA catabolic process and deadenylation-dependent decay.
		PTK2	maintains response to growth hormone, muscle stretch and pH.
NW_005081556.1*	1601131	MAP3K15	plays a role in apoptotic cell death triggered by cellular stresses.
		SH3KBP1	possible involvement in the regulation of cellular stress response through its interaction with MAP3K.
NW_005081570.1	4341360	LSAMP	controls component of membrane.
NW_005081610.1	939201	PCLO	includes synaptic function and involvement in body growth and size.
NW_005081629.1	750061	CTIF	establishing active synaptic zones and in synaptic vesicle trafficking. Related pathways include olfactory signaling pathway and sensory processing of sound.
NW_005081640.1	866386	BRINP1	controls vocalization behavior, behavioral defense response, biological process in intraspecies interaction, response to stress, memory, and cognition.
NW_005081690.1	20019	PPIL1	influences isomerase activity, protein folding, spliceosome, and plays a role in embryonic brain development.
NW_005081744.1	968151	BARX2	encodes a member of the homeobox transcription factor family and may control the expression of neural adhesion molecules.
NW_005081748.1	528383	LOC102068172	no gene function found

*Detected with all three methods.

2.7 References

Aljanabi, S., & Martinez, I. (1997). Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*. 25: 4692 – 4693.

Auton, A., Marcketta, A., & Danecek, P. (2015). *VCFtools. (Version 0.1.14) [Software]*. Available from https://vcftools.github.io/index.html

Barley, A. J., Monnahan, P. J., Thomson, R. C., Grismer, L. L., & Brown, R. M. (2015). Sun skink landscape genomics: Assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*. 24: 1696 – 1712.

Berkowicz, S. R., Featherby, T. J., Qu, Z., Giousoh, A., Borg, N. A., Heng, J. I., & Bird, P. I. (2016). Brinp1−/− mice exhibit autism-like behaviour, altered memory, hyperactivity, and increased parvalbumin-positive cortical interneuron density. *Molecular Autism*. 7: 1- 20.

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

Cortopassi, A. J., & Mewaldt, L. R. (1965). The circumannual distribution of white-crowned sparrows. *Bird Banding*. 592: 141 – 169.

Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC genomics*. 19: 1 - 13.

Dunn, J. L., Garrett, K. L., & Alderfer, J. (1995). White-crowned sparrow subspecies: identification and distribution. *Birding.* 27: 182 – 201.

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

Fischer, M. C., Foll, M., Excoffier, L., & Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*. 20: 1450 – 1462.

Fraslin, C., Koskinen, H., Nousianen, A., Houston, R. D., & Kause, A. (2022). Genome-wide association and genomic prediction of resistance to *Flavobacterium columnare* in a farmed rainbow trout population. *Aquaculture*. 557: 738332.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution.* 6: 925 – 929.

Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 36: 2628 – 2629. http://bioinformatics.sdstate.edu/go74/

Gibbard, P. L., Head, M. J., Walker, M. J. C., and the Subcommission on Quaternary Stratigraphy. (2010). Formal ratification of the Quaternary system/period and the Pleistocene series/epoch with a base at 2.58 Ma. *Journal of Quaternary Science.* 25: 96 – 102. ISSN 0267-8179.

Godfrey, W. E. (1965). Review of "Geo-graphic variation in white-crowned sparrow *Zonotrichia leucophrys*" by R. C. Banks. *The Auk.* 82: 510 – 511.

Hafner, D. J., & Petersen, K. E. (1985). Song dialects and gene flow in the white-crowned sparrow, *Zonotrichia leucophrys nuttalli*. *Evolution*. 39: 687 – 694.

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*. 405: 907 – 913.

Hodel, R. G. J., Chen, S., Payton, A. C., McDaniel, S. F., Soltis, P., & Soltis, D. E. (2017). Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. *Scientific Reports.* 7:17598.

Holsinger, K. E., & Weir, B. S. (2009) Genetics in geographically structured populations: defining, estimating, and interpreting F(ST). *Nature Reviews Genetics.* 10: 639 – 650

Jombart, T., Devillard., S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics.* 11: 94.

Jones, O. R., & Wang, J. (2012). A comparison of four methods for detecting weak genetic structure from marker data. *Ecology and Evolution*. 2: 1048 – 1055.

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters.* 7:1225 – 1241.

LaManna, J. A., Hemenway, A. B., Boccadori, V., & Martin, T. E. (2015). Bird species turnover is related to changing predation risk along a vegetation gradient. *Ecology*. 96: 1670 – 1680.

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

Luu, K., Bazin, E., & Blum, M. G. B. (2017). PCAdapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources.* 17: 67 – 77.

Marchán, D. F., Novo, M., Sánchez, N., Domínguez, J., Darío, J., Cosín, D., & Fernández, R. (2020). Local adaptation fuels cryptic speciation in terrestrial annelids. *Molecular Phylogenetics and Evolution*. 146*:* 106767.

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

Martin, T. (1993). Nest predation and nest sites. *BioScience.* 43: 523 – 532.

Mello, C. V., Jarvis, E. D., Denisenko, N., & Rivas, M. (1997) Isolation of song-regulated genes in the brain of songbirds. *Methods in Molecular Biology.* 85: 205 – 217.

Morgan, T. D., Graham, C. F., McArthur, A. G., Raphenya, A. R., Boreham, D. R., & Manzon, R. G. (2017). Genetic population structure of the round whitefish (*Prosopium cylindraceum*) in North America: multiple markers reveal glacial refugia and regional subdivision. *Canadian Journal of Fisheries and Aquatic Sciences.* 75: 836 – 849.

Morton, M. (2002). The mountain white-crowned sparrow: migration and reproduction at high altitude. *Studies in Avian Biology*. 24: 1 – 236.

Owen, J. C. (2011). Collecting, processing, and storing avian blood: a review. *Journal of Field Ornithology*. 82: 339 – 354.

Parchman, T. L., Buerkle, C. A., Soria-Carrasco, V., & Benkman, C. W. (2016). Genome divergence and diversification within a geographic mosaic of coevolution. *Molecular Ecology.* 25: 5705 – 5718.

Pérez-Figueroa, A., García-Pereira, M., Saura, M., Rolán-Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*. 23: 2267 – 2276.

Rand, A. L. (1948). Glaciation, an isolating factor in speciation. *Evolution*. 2: 314 – 321.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 4: 406 – 425.

Sedgwick, J. A. (2001). Geographic Variation in the Song of Willow Flycatchers: Differentiation Between *Empidonax Traillii Adastus* and *E. T. Extimus*. *The Auk*. 118: 366–379.

Slatkin, M., (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 139: 457 – 462.

Stift, M., Kolář, F., & Meirmans, P. G. (2019). Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*. 123: 429 – 441.

Taylor, R. S., Bramwell, A. C., Clemente-Carvalho, R., Cairns, N. A., Bonier, F., Dares, K., & Lougheed, S. C. (2021). Cytonuclear discordance in the crowned-sparrows, *Zonotrichia atricapilla* and *Zonotrichia leucophrys*. *Molecular Phylogenetics and Evolution*. 162162: 107216.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution.* 38:1358 – 1370.

Welke, C. A., Graham, B., Conover, R. R., Rivers, J. W., & Burg, T. M. (2021). Habitat-linked genetic structure for white-crowned sparrow (*Zonotrichia leucophrys*): Local factors shape population genetic structure. *Ecology and Evolution*. 11: 11700 – 11717.

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

Zink, R. M., Dittmann, D. L., & Rootes, W. L. (1991). Mitochondrial DNA variation and the phylogeny of *Zonotrichia. The Auk.* 108: 578 – 584.

Chapter Three: General Discussion

3.1 Major findings

3.1.1. Zonotrichia leucophrys divergence

Zonotrichia leucophrys, and its closely related sister species *Zonotrichia atricapilla* have undergone rapid speciation as recently as 50,000 years ago (Johnson & Cicero, 2004). *Z. leucophrys* has five recognised subspecies with divergence times estimated within the last 18,000 years. Their subspecies ranges correspond to North America refugia during the Pleistocene (Rand, 1984). Our study, comparing three of the five subspecies (*Z. l. gambelii, Z. l. oriantha, Z. l. pugetensis*) found considerable genetic differences. Though evidence of intergradation has been reported between *Z. l. gambelii and Z. l. oriantha* especially along their contact zone (Lein & Corbin, 1990), our study clearly showed complete divergence between the two using outlier SNPs from 3dRADseq and lcWGS data, despite a large number of samples from Alberta, a known contact zone of the two. Our hypothesis is that *Z. l. leucophrys* may be responding to very subtle variations within their populations such as their morphological differences and variation within their habitats e.g. colder regions in the northeast for *gambelii* subspecies. This is further supported by an earlier study where *Z. l. nutalli and Z. l. pugetensis* with overlapping breeding areas, showed distinct genetic divergence corresponding to song variation (Lipshutz et al., 2017). Morphological variation supports subspecies delineations in *Zonotrichia leucophrys*, but few studies have substantiated their genetic divergence. The fine-scale genomic data in our study have clearly provided genetic support for the morphological divergence within this species.

3.1.2. Recent divergence within Zonotrichia leucophrys *populations*

Consistency is important to support evolutionary process as they provide credible support for research findings, and when different methods and markers are involved, hidden variation may be uncovered. This is because when populations diverge, large part of their genomes will still maintain paraphyletic genealogies. Over time, many of the genes will be resolved to monophyly by lineage sorting (Avise and Ball, 1990). However, lineage sorting takes time especially for a large population, and therefore recent divergence is difficult to track unless with a very high-resolution markers in some cases (Knowles and Carstens, 2007). It is interesting then to know that our study revealed an additional genetic cluster with the low coverage whole genome data i.e the separation of the northern and southern *Z. l. oriantha*. Subspecies divergence using 3dRADseq data showed three clear groups corresponding to the subspecies, whereas result with lcWGS data showed four genetic groups corresponding to the three subspecies including divergence between the northern and southern *oriantha* subspecies. The use of whole genome as seen in this study contributed to the detection of genetic clusters within *Z. l. oriantha* with fine scale data, and this may be due to the sequencing pattern of lcWGS where the whole genome is sequenced at low coverage.

Welke et al. (2021) is the first known study pointing to divergence between the *Zonotrichia leucophrys* populations in the alpine coniferous and riparian deciduous forests, cutting across two subspecies, *Z. l. gambelii and Z. l. oriantha.* Our study provides additional support for population divergence according to habitat types with the use of outlier SNPs from both 3dRAdseq and low coverage whole genome sequencing. Our data showed the divergence between the two ecotypes, once again supporting the claim of high

divergence within *Zonotrichia leucophrys* (Johnson & Cicero, 2004). The genetic structure established in our study for *Z. leucophrys* populations in the two forest types can be considered a response to subtle habitat variation in the *Z. leucophrys* habitat, due to the contiguity of the two habitats. These heterogenous habitat when considered as the species' range, appears to form a continuum, as migratory routes are established throughout the range. Another reason for considering the forest type variation a subtle one is that the two subspecies come into contact within these forest types, even though there is evidence of intergradation, our results clearly identified substructure grouping corresponding to the habitat types. This again is in support of the assumption that *Z. leucophrys* may be susceptible to small changes or isolation in their environment. However, this assumption, particularly response to habitat types, has not been tested in the other subspecies, *Z. leucophrys leucophrys*. It will be interesting to know if this assumption holds for the other subspecies.

3.1.3. Evidence of selection in the habitat ecotypes

We detected some loci suspected to be under selection, many of which contain genes related to stress such as PTK2, MAP3K15, SH3KBP1 and BRINP1. We believe microclimatic conditions induced by these habitat types along with other variation within the habitats may be driving the genetic differences we have recorded. We have linked genetic structure to habitat types in this study. Habitat is defined by its components, both biotic and abiotic factors. The alpine coniferous and riparian deciduous forest types in our study are characterised by different plant species, especially, the tree species. For instance, the wider canopy cover in deciduous tree populations can create different microclimatic conditions to which the *Z. leucophrys* populations are responding. This is supported by several studies

that have shown within canopy temperature and relative humidity levels are different from the outside the canopy conditions (Jones 1992; Pau et al., 2018). Over time, this canopy effect results in lower seasonal maximum temperatures, and higher minimum temperatures and relative air humidity (Renaud et al., 2011; Gaudio et al., 2017; Prevosto et al., 2020). Altogether, the variation in canopy conditions and the canopy structure are known to drive species richness in forest (Ozanne et al., 2003). Ultraviolet radiation is another environmental factor that can be influenced by canopy cover. Dense canopy can reduce the amount and intensity of light reaching the forest floor. Contrarily, habitat with less dense canopy will receive significantly more radiation. Ultraviolet radiation is associated with cellular stress, and so are reduced or excessive nutrients (Llanos et al., 2009; Shiozaka 2009; Ongusasha et al., 2008). Further, the importance of habitat includes its use for foraging, mating, nesting, and cover (Litvaitis et al., 1996). Studies have pointed to genetic underlining, innate and learned behavior, and resources within the habitat as factors contributing to habitat selection (Hutto, 1985; Rosenwieg, 1981; Block and Breenan, 1993). Species' tolerance level and response to habitat variation varies, some species may survive and thrive in habitat considered harsh to others. LaManna et al. (2015) found that different species of sparrow responded differently to different forest types. In their study, chipping sparrow (*Spizella passerina*) density and nest density increased with coniferous forest, while Lincoln's sparrow (*Melospiza lincolnii*) density and nest density increased with deciduous forest. Several genes identified in our study (AGO2, PTK2, MAP3K15, CTIF, and BRINP1) are linked to functions such as cognition, sensory and behavior which are all complex traits needed for animal to thrive in its environment. We propose that populations of *Z. leucophrys* subspecies in both alpine coniferous and riparian deciduous habitats may be affected by the differential expression of these genes, and consequently their

independent adaptation to the habitat conditions. As mentioned, cognition is a very complex trait which may be involved in prey detection and avoidance. Many studies have described vulnerability trait as the ability of prey to avoid being preyed upon (Gravel et al., 2016; Green & Cote, 2014; Rossberg et al., 2010). These traits include prey characteristics such as poor ability to avoid detection, escape behavior, defense, and social communication (Sheriff & Thaler, 2014; Hawlena & Schmitz, 2010). Specifically, Berkowicz et al. (2016) have linked the absence of BRINP1 in mice to reduced sociability and autism-like characteristics which may enhance vulnerability in animals. Another study has also shown that similar genes with MAP3K15 and PPIL1 are believed to be under selection for song sparrow (*Melospiza melodia*), swamp sparrow (*Melospiza georgiana*) and Nelson's sparrow (*Ammodramus nelsoni*) (Walsh et al., 2019). In addition to the suggested predation index and microclimatic conditions, there is a need to fully investigate other variation such as food availability and quality, nesting conditions and density, competition and other components of these two habitats.

3.2. Future directions

We have provided genetic support for the morphological differentiation of three of the five *Z. leucophrys* subspecies; our work did not include the northeastern *Z. l. leucophrys* or California *Z. l. nuttalli*. Though previous studies have supported the divergence of these subspecies (Lipshutz et al., 2017 (*Z. l. nuttalli* and *Z. l. pugetensis*); Taylor et al., 2020), there is a need to include them in a future study with the use of fine scale next generation sequencing techniques and analyses like the outlier SNPs used in this study from 3dRAD and lcWGS. As shown in our study, the combined power of different approaches led to the detection of within population variation e.g., the northern and southern *oriantha*, and the

coniferous and deciduous populations. These methods may help reveal substructure within subspecies *nuttalli* and *leucophrys*.

Our study has proposed the susceptibility of *Z. leucophrys* subspecies to subtle variation in their habitat such as forest type. It will be interesting to test this on the remaining subspecies. Importantly, it will be interesting to study the hybrids of subspecies *gambelii* and *oriantha* occurring in the contiguous habitat types defined in this study to test for hybrid superiority, and their response to the habitat variation.

Our study supported the genetic differentiation of northern *Z. l. oriantha* (from British Columbia and Alberta) from the southern *Z. l. oriantha* (from Colorado) with lcWGS data. Our findings seem interesting more so that Welke et al. (2021) earlier opined that the southern *Z. l. oriantha* may be harboring the genetics of the pure parental form of *Z. l. oriantha*. Our study shows support for the southern *Z. l. oriantha* being the ancestral subspecies from the point view of divergence starting within the *Z. l. gambelii,* the subspecies with the widest geographical range from the far north to south. Our phylogenetic tree shows that southern *Z. l. oriantha* is closer to *gambelii* than the northern *oriantha* is. Our minimum spanning network also shows the same support and lastly our pairwise F_{ST} values show that the genetic distance between *gambelii* and southern *oriantha* is shorter than between gambelii and northern oriantha. This may mean that the southern *oriantha* contains more of the genotype of the ancestral form which is shared with the other two subspecies. Studying both the northern and southern *Z. l. oriantha* in details will help answer some evolutionary questions relating to ancestral and recent divergence events.
It will be a great addition, if environmental association study can be done in future to assess the relationship of the microclimatic conditions within the two habitat types to the genotype of *Z. leucophrys* populations.

3.3. Closing statements

Our study has established strong support for the subspecies divergence within the *Z. leucophrys* clade. We have also provided genetic support for the differentiation between the populations of *Z. l. gambelii* and *Z. l. oriantha* (northern oriantha) found in alpine coniferous and riparian deciduous forests. We have shown that genetic divergence within *Z. leucophrys* occurs in the absence of well-established physical barriers to gene flow. Stress-linked genes largely found in *Z. leucophrys* populations suggest that microclimatic conditions in the coniferous and deciduous forests may be influencing the genetic structure detected in our populations.

3.4 References

Avise, J.C., & Ball, M. R. (1990). Principles of genealogical concordance in species concepts and biological taxonomy. Oxford Survey of Evolutionary Biology. Edited by Futuyma, D.J., & Antonovics, J. New York: Oxford University Press, 45-67.

Bellamy, P. E., Burgess, M. D., & Mallord, J. W. (2018). Nest predation and the influence of habitat structure on nest predation of wood warbler *Phylloscopus sibilatrix*, a ground-nesting forest passerine. *Journal of Ornithology*. 159: 493 – 556.

Berkowicz, S. R., Featherby, T. J., Qu, Z., Giousoh, A., Borg, N. A., Heng, J. I., & Bird, P. I. (2016). Brinp1−/− mice exhibit autism-like behaviour, altered memory, hyperactivity and increased parvalbumin-positive cortical interneuron density. *Molecular Autism*. 7: 1- 20.

Block, W. M., & Brennan, L. A. (1993). The habitat concept in ornithology: Theory and applications. P. 35- 91 In: D.M. Power (ed.). *Current Ornithology*. Volume 11. Plenum Press, New York.

Carpenter, A. M., Graham, B. A., Spellman, G. M., Klicka, J., & Burg, T. M. (2021). Genetic, bioacoustic and morphological analyses reveal cryptic speciation in the warbling vireo complex (*Vireo gilvus:* Vireonidae: Passeriformes). *Zoological Journal of the Linnean Society.* 195: 45 - 64.

Dempsey, Z. W., Goater, C. P., and Burg, T. M. (2020). Living on the edge: Comparative phylogeography and phylogenetics of *Oreohelix* land snails at their range edge in Western Canada. *BMC Evolutionary Biology*. 20: 3.

Gaudio, N., Gendre, X., Saudreau, M., Seigner, V., & Balandier, P. (2017). Impact of tree canopy on thermal and radiative microclimates in a mixed temperate forest: a new statistical method to analyse hourly temporal dynamics. *Agriculture and Forest Meteorology*. 23: 71 – 79.

Gravel, D., Albouy, C., & Thuiller, W. (2016). The meaning of functional trait composition of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 371: 20150268.

Green, S. J. and Côté, I. M. (2014). Trait-based diet selection: prey behaviour and morphology predict vulnerability to predation in reef fish communities. *Journal of Animal Ecology*. 83: 1451 – 1460.

Hawlena, D., & Schmitz, O. J. (2010). Physiological stress as a fundamental mechanism linking predation to ecosystem functioning. *The American Naturalist*. 176: 537 – 556.

Hutto, R. L. (1985). Habitat selection by nonbreeding migratory land birds. p. 455-476 In: M.L. Cody (ed.). Habitat Selection in Birds. Academic Press, Orlando, FL.

Johnson, N. K., & Cicero, C. (2004). New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution*. 58: 1122 – 1130.

Jones, H. G. (1992). Plants and Microclimate. Cambridge: Cambridge University Press.

Knowles, L.L., & Carstens, B.C. (2007). Delimiting species without monophyletic gene trees. *Systematic Biology*. 56: 887 – 895.

LaManna, J. A., Hemenway, A. B., Boccadori, V., & Martin, T. E. (2015). Bird species turnover is related to changing predation risk along a vegetation gradient. *Ecology*. 96: 1670 – 1680.

Lein, M. R., & Corbin, K. W. (1990). Song and plumage phenotypes in a contact zone between subspecies of the white-crowned sparrow (*Zonotrichia leucophrys*). *Canadian Journal of Zoology*. 68: 2625 – 2629.

Lipshutz, S. E., Overcast, I. A., Hickerson, M. J., Brumfield, R. T., & Derryberry, E. P. (2017). Behavioral response to song and genetic divergence in two subspecies of white - crowned sparrows (*Zonotrichia leucophrys*). *Molecular Ecology*. 656: 3011 – 3027.

Litvaitis, J. A., Titus, K., & Anderson, E. M. (1994). Measuring vertebrate use of territorial habitats and foods. p. 254-74 In: T. A. Bookhout (ed.). Research and Management Techniques for Wildlife and Habitats. 5th ed. The Wildlife Society, Bethesda, Md

Llanos, S., Cuadrado, A., & Serrano, M. (2009). MSK2 inhibits p53 activity in the absence of stress. *Science Signaling*. 2: 57.

Martin, T. E. (1993). Nest predation among vegetation layers and habitat types: revising the dogmas. *American Naturalist*. 141: 897 – 913.

Ongusaha, P., Qi, H. H., Raj, L., Kim, Y. B., Aaronson, S. A., Davis, R. J., Shi, Y., Liao, J. K., Lee, S. W. (2008). Identification of ROCK1 as an upstream activator of the JIP-3 to JNK signaling axis in response to UVB damage. *Science Signaling.* 1: 14.

Ozanne, C. H. P., Anhuf, D., Boulter, S. L., Keller, H., Kitching, R. L., & Körner, C. (2003). Biodiversity meets the atmosphere: a global view of forest canopies. *Science.* 301: 183 – 186.

Pau, S., Detto, M., Kim, Y., & Still, C. J. (2018). Tropical forest temperature thresholds for gross primary productivity. *Ecosphere*. 9: e02311.

Prévosto, B., Helluy, M., Gavinet, J., Fernandez, C., and Balandier, P. (2020). Microclimate in Mediterranean pine forests: what is the influence of the shrub layer? *Agriculture and Forest Meteorology.* 282 – 283.

Rand, A. L. (1948). Glaciation, an isolating factor in speciation. *Evolution*. 2: 314 – 321.

Renaud, V., Innes, J. L., Dobbertin, M., & Rebetez, M. (2011). Comparison between open-site and below-canopy climatic conditions in Switzerland for different types of forests over 10 years (1998-2007). *Theoretical and Applied Climatology.* 105: 119 – 127.

Rosenweig, M. L. (1981). A theory of habitat selection. *Ecology*. 62: 327 - 335.

Rossberg, A., Brännström, A. & Dieckmann, U. (2010). How trophic interaction strength depends on traits – a conceptual framework for representing multidimensional trophic niche spaces. *Theoretical Ecology*. 3: 13 – 24.

Seitz, L. C., & Zegers, D. A. (1993). An experimental study of nest predation in adjacent deciduous, coniferous, and successional habitats. *Condor*. 95: 297.

Sheriff, M. J. & Thaler, J. S. (2014). Ecophysiological effects of predation risk; an integration across disciplines. *Oecologia.* 176: 607 – 611.

Shiozaki, K. (2009). Nutrition-minded cell cycle. *Science Signaling.* 2: 74.

Taylor, R. S., Bramwell, A. C., Clemente-Carvalho, R., Cairns, N. A., Bonier, F., Dares, K., & Lougheed, S. C. (2021). Cytonuclear discordance in the crowned-sparrows, *Zonotrichia atricapilla* and *Zonotrichia leucophrys*. *Molecular Phylogenetics and Evolution*. 162162: 107216.

Wachowiak, W., Zaborowska, J., Łabiszak, B. et al. (2018). Molecular signatures of divergence and selection in closely related pine taxa. *Tree Genetics & Genomes.* 14: 83.

Walsh, J., Benham, P. M., Deane-Coe, P. E., Arcese, P., Butcher, B. G., Chan, Y. L., Cheviron, Z. A., Elphick, C. S., Kovach, A. I., Olsen, B. J., Shriver, W. G., Winder, V. L., & Lovette, I. J. (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows, *Evolution Letters*. 3: 324 – 338.

Welke, C. A., Graham, B., Conover, R. R., Rivers, J. W., & Burg, T. M. (2021). Habitat-linked genetic structure for white-crowned sparrow (*Zonotrichia leucophrys*): Local factors shape population genetic structure. *Ecology and Evolution*. 11: 11700 – 11717.

Appendix 1: White crowned sparrow samples used in this study with information on sample ID, location corresponding to the sample ID, band or museum ID, their subspecies, type of sequencing, and their geographical coordinates. Latitudes and longitude are recorded in decimal degrees for all samples except the OR samples that are recorded in UTMs. Band number prefixes are in bold.

Sample ID	Location	Band/Museum	Subspecies/Habitat	Latitude	Longitude	3dRAD	lcWGS
		1451-	***Z. l.***				
WT002*	Haybarn, Waterton, AB	65771	*oriantha*/deciduous	49.07961	-113.85930	yes	
WT003*	Haybarn, Waterton, AB	65773	*oriantha*/deciduous	49.07961	-113.85930	yes	
WT004*	Stables & Camp, Waterton, AB	65779	*gambelii*/deciduous	49.06286	-113.887278	yes	
WT005**	Stables & Camp, Waterton, AB	65783	*oriantha*/deciduous	49.06286	-113.887278	yes	
WT006*	Haybarn II, Waterton, AB	65776	*oriantha*/deciduous	49.07881	-113.8615	yes	
WT008**	Stables & Camp, Waterton, AB	65780	*oriantha*/deciduous	49.06286	-113.887278	yes	
WT007**	Stables & Camp, Waterton, AB	65781	*oriantha*/deciduous	49.06286	-113.887278	yes	
BA005*	Cave & Basin, Banff, AB	65760	*gambelii*/coniferous	51.17058	-115.587472	yes	yes
BA001*	Cave & Basin, Banff, AB	65761	*oriantha*/coniferous	51.17058	-115.587472	yes	yes
BA003*	Cave & Basin, Banff, AB	65762	*oriantha*/coniferous	51.17058	-115.587472	yes	yes
BA004*	Cave & Basin, Banff, AB	65763	*oriantha*/coniferous	51.17058	-115.587472	yes	
		1391-					
BA007*	MAPS station, Bow Valley Parkway Banff, AB	86780	*oriantha*/coniferous	51.20358	-115.750028	yes	
BA008*	MAPS station, Bow Valley Parkway Banff, AB	86781	*gambelii*/coniferous	51.20358	-115.750028	yes	yes
		1501-					
CH001*	Elkwater, Cypress Hills, AB	42950	*oriantha*/deciduous	49.66184	-110.31291	yes	

CH002*	Elkwater, Cypress Hills, AB	42953	*oriantha*/deciduous	49.66351	-110.30308	yes
CH003*	Elkwater, Cypress Hills, AB	WCSP 3	*oriantha*/deciduous	49.66184	-110.31291	yes
CH004*	Elkwater, Cypress Hills, AB	42949	*oriantha*/deciduous	49.66184	-110.31291	yes
CH005*	Elkwater, Cypress Hills, AB	42948	*oriantha*/deciduous	49.66184	-110.31291	yes
CH006*	Elkwater, Cypress Hills, AB	WCSP 73	*oriantha*/deciduous	49.66184	-110.31291	yes
		-				
JA001*	Royal AB Museum, AB	14954	*gambelii*/coniferous	52.1000	-121.93333	yes
BR002*	Brule, AB	Z07.1.8	*gambelii*/coniferous	53.315	-117.869	yes
BR003*	Brule, AB	Z07.9.1	*gambelii*/coniferous	53.315	-117.869	yes
BR004*	Brule, AB	Z07.9.4	*gambelii*/coniferous	53.315	-117.869	yes
BR005*	Brule, AB	Z07.9.4	*gambelii*/coniferous	53.315	-117.869	yes
BR006*	Brule, AB	Z07.9.4	*gambelii*/coniferous	53.315	-117.869	yes
		1451-				
BV001*	West Castle Wetlands, Castle Mountain and Beaver Mines, AB	65765	*oriantha*/deciduous	49.37661	-114.378389	yes
BV002*	Lynx Creek Rd, Beaver Mines AB	65766	*oriantha*/deciduous	49.45925	-114.37211	yes
BV003*	Mill Creek Road, Beaver Mines, AB	65770	*oriantha*/deciduous	49.34667	-114.147306	yes
BV004*	Mill Creek Road, Beaver Mines, AB	65769	*oriantha*/deciduous	49.34667	-114.147306	yes

BV005*	Mill Creek Road, Beaver Mines, AB	65768	*oriantha*/deciduous	49.34667	-114.147306	yes
BV006*	Mill Creek Road, Beaver Mines, AB	65767	*oriantha*/deciduous	49.34667	-114.147306	yes
OK001*	Okanagan Falls	3593	*gambelii*/coniferous	49.35	-119.56666	yes
OK002*	Okanagan landing	3604	*gambelii*/coniferous	51.91667	-123.03333	yes
OK003*	Okanagan landing	3605	*gambelii*/coniferous	51.91667	-123.03333	yes
OK004*	Okanagan landing	3607	*gambelii*/coniferous	51.91667	-123.03333	yes
OK005*	Okanagan landing	3609	*gambelii*/coniferous	51.91667	-123.03333	yes
MK001*	Mackenzie, Municipal RV Park, Clearcut, BC	35775	*gambelii*/coniferous	55.32492	-123.095861	yes
MK002*	Mackenzie, Municipal RV Park, Clearcut, BC	35776	*gambelii*/coniferous	55.32492	-123.095861	yes
MK003*	Mackenzie, Municipal RV Park, Clearcut, BC	35777	*gambelii*/coniferous	55.32492	-123.095861	yes
MK004*	Mackenzie, Municipal RV Park, Clearcut, BC	35778	*gambelii*/coniferous	55.32492	-123.095861	yes
MK005*	Mackenzie, Municipal RV Park, Clearcut, BC	35779	*gambelii*/coniferous	55.32492	-123.095861	yes
MK006*	Clearcut off highway 39, Mackenzie, BC	35780	*gambelii*/coniferous	55.32492	-123.095861	yes
FTSJ001**	Necoslie Road, Fort St James, BC	CBC 248	*gambelii*/coniferous	54.41603	-124.22	yes
FTSJ002*	Necoslie Road, Fort St James, BC	CBC 249	*gambelii*/coniferous	54.41603	-124.22	yes

FTSJ003**	Necoslie Road, Fort St James, BC	CBC 250	*gambelii*/coniferous	54.41603	-124.22	yes
FTSJ004**	Necoslie Road, Fort St James, BC	CBC 260	*gambelii*/coniferous	54.41603	-124.22	yes
LETH001**	Popson Park, Lethbridge, AB	LETH 37	*gambelii*/deciduous	49.556338	-112.871861	yes
LETH002**	Popson Park, Lethbridge, AB	LETH 38	*gambelii*/deciduous	49.556338	-112.871861	yes
LETH003**	Popson Park, Lethbridge, AB	LETH 39	*gambelii*/deciduous	49.556338	-112.871861	yes
LETH004**	Popson Park, Lethbridge, AB	LETH 40	*gambelii*/deciduous	49.556338	-112.871861	yes
LETH005**	Popson Park, Lethbridge, AB	LETH 41	*gambelii*/deciduous	49.556338	-112.871861	yes
LETH006**	Popson Park, Lethbridge, AB	LETH 42	*gambelii*/deciduous	49.556338	-112.871861	yes
RV001*	Blanket Creek Forestry Road, BC	65757	*gambelii*/deciduous	50.82672	-118.121722	yes
RV002*	Blanket Creek Forestry Road, BC	65758	*gambelii*/deciduous	50.82672	-118.121722	yes
RV003*	Blanket Creek Forestry Road, BC	65759	*gambelii*/deciduous	50.82672	-118.121722	yes
VI 002*	Saanich VI, BC	023 WCSP2	*gambelii*/coniferous	48.559416	-123.703055	yes
CO002	Base of slope, CO	01498	*oriantha*	38.98	-106.98	yes
CO003	Paradise basin, CO	50810	*oriantha*	38.99355	-107.05327	yes
CO004	Paradise basin, CO	50888	*oriantha*	38.99355	-107.05327	yes
CO005	Paradise basin, CO	74921	*oriantha*	38.9945	-107.05358	yes
CO006	Site 4, high, CO	50863	*oriantha*	38.99085	-107.01211	yes
CO007	Paradise basin, CO	74916	*oriantha*	38.99389	-107.05367	yes
CO012	Site 4c North pole basin, CO	74928	*oriantha*	38.99409	-107.01385	yes
CO013	Site 2-3 Amigo, CO	74995	*oriantha*	38.97649	-106.999	yes
CO014	Site 4, CO	01405	*oriantha*	39.0000	-107.1	yes
CO015	Site 4_box, CO	01226	*oriantha*	38.99594	-107.01431	yes
CO016	Site 2, CO	60009	*oriantha*	38.97594	-106.99929	yes
CO017	Site 4c, CO	60018	*oriantha*	38.99346	-107.01339	yes
CO018	Site 4c, CO	60020	*oriantha*	38.99289	-107.01421	yes

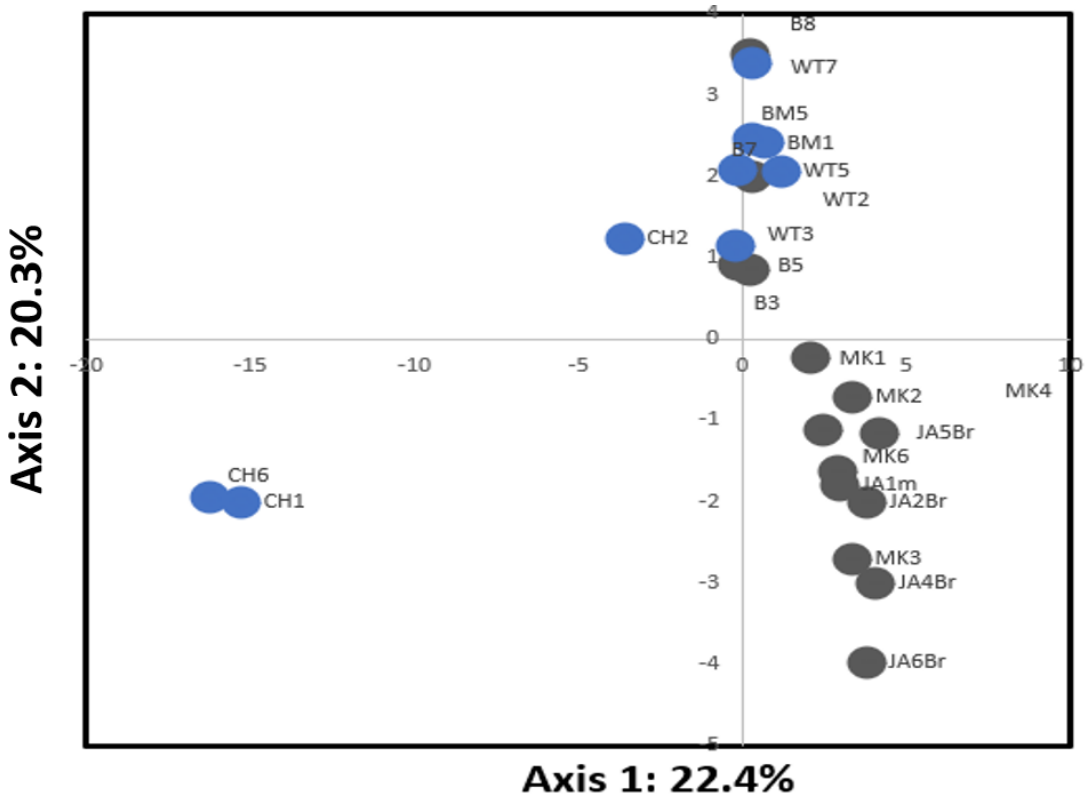
CO019	site SW-Uno, CO	60022	*oriantha*	39.02885	-107.05412	yes
CO020	Site SE, CO	60016	*oriantha*	39.02856	-107.05242	yes
CO021	Schofield high perch, CO	50900	*oriantha*	39.02856	-107.05242	yes
CO022	SE streamside, CO	74945	*oriantha*	39.02856	-107.05242	yes
CO023	Schofield high perch, CO	50824	*oriantha*	39.02819	-107.05202	yes
CO024	SW, CO	60028	*oriantha*	39.02629	-107.05093	yes
		1501-		**UTM**		
OR002	Trask, OR	30802	*pugetensis*	467899	5029621	yes
OR003	Blackrock, OR	30894	*pugetensis*	467031	4979666	yes
OR004	Blackrock, OR	30895	*pugetensis*	467031	4979666	yes
		1131-				
OR005	Blackrock Rubo, OR	88336	*pugetensis*	468438	4976926	yes
OR006	Blackrock Rubo, OR	88335	*pugetensis*	468438	4976926	yes
OR007	Blackrock Rubo, OR	88334	*pugetensis*	468438	4976926	yes
OR008	Trask fair, OR	88341	*pugetensis*	470013	5028446	yes
OR009	Trask TOL, OR	88310	*pugetensis*	467899	5029621	yes
OR010	Willamina TSUN, OR	88337	*pugetensis*	461730	5004440	yes
OR011	Trask, OR	88323	*pugetensis*	467925	5027734	yes
OR012	Willamina TSUN, OR	88338	*pugetensis*	461730	5004440	yes
OR013	Willamina TSUN, OR	88342	*pugetensis*	461547	5004392	yes
OR014	Trask_WEFL, OR	88385	*pugetensis*	468534	5028924	yes
OR015	Trask_WEFL, OR	88319	*pugetensis*	468544	5028888	yes
OR016	Trask_WEFL, OR	88320	*pugetensis*	468544	5028888	yes
		1421-				
OR017	Luckiamute_BRON, OR	12725	*pugetensis*	461578	4962017	yes

OR018	Luckiamute_BRON, OR	12727	*pugetensis*	461578	4962017	yes
OR019	Luckiamute_NWRO, OR	12724	*pugetensis*	460698	4962009	yes
OR020	Luckiamute_NWRO, OR	12723	*pugetensis*	460613	4962029	yes
		2541-				
OR021	Blackrock_SAPA, OR	57218	*pugetensis*	467007	4979702	yes
Low coverage whole genome sequencing						
CNP001	Allison Creek Road 5 Crowsnest Pass, AB	35769	*oriantha*/deciduous	49.73077	-114.60774	yes
CNP002	Allison Creek Road 5 Crowsnest Pass, AB	35770	*oriantha*/deciduous	49.73077	-114.60774	yes
LE001	Popson Park, Lethbridge, AB	43	*gambelii*/deciduous	49.55638	-112.8718611	yes
LE002	Popson Park, Lethbridge, AB	44	*gambelii*/deciduous	49.55638	-112.8718611	yes
LE003	Popson Park, Lethbridge, AB	45	*gambelii*/deciduous	49.55638	-112.8718611	yes
CH001	Graburn Road, Cypress Hills, AB	24216	*oriantha*/deciduous	49.65731	-110.1021196	yes
CH002	Graburn Road, Cypress Hills, AB	24217	*oriantha*/deciduous	49.65731	-110.1021196	yes
				UTM		
OR001	Willamina, OR	30507	*pugetensis*	462413	5003443	yes
OR002	Luckiamute, OR	12917	*pugetensis*	60628	4961894	yes
OR003	Luckiamute, OR	88906	*pugetensis*	460715	4961840	yes
CO001	Site 2, CO	96077	*oriantha*	38.98	-107.01	yes
CO002	Site 2, CO	96091	*oriantha*	38.98	-107	yes
CO003	Site 2, CO	96080	*oriantha*	38.98	-107	yes
CO004	Site 2, CO	1340	*oriantha*	38.98	-107	yes
CO005	Site 2, CO	1436	*oriantha*	38.98	-107	yes

		1301-				
CO006	Schofield, CO	50815	*oriantha*	39.02713	-107.0523	yes
CO007	Paradise slope, CO	50809	*oriantha*	38.99	-107.05	yes
		2331-				
CO009	Site 4, CO	96075	*oriantha*	38.99	-107.01	yes
		2661-				
CO010	Site 4, CO	01411	*oriantha*	39	-107.01	yes

* Sample identified by mitochondrial DNA

** Samples identified by photo



Appendix 2: Scatter plot from DAPC for habitat types clustering using 3dRADseq data. The deciduous habitat type cluster is a mix of populations from both habitat types. (No of SNPs = 450,000).

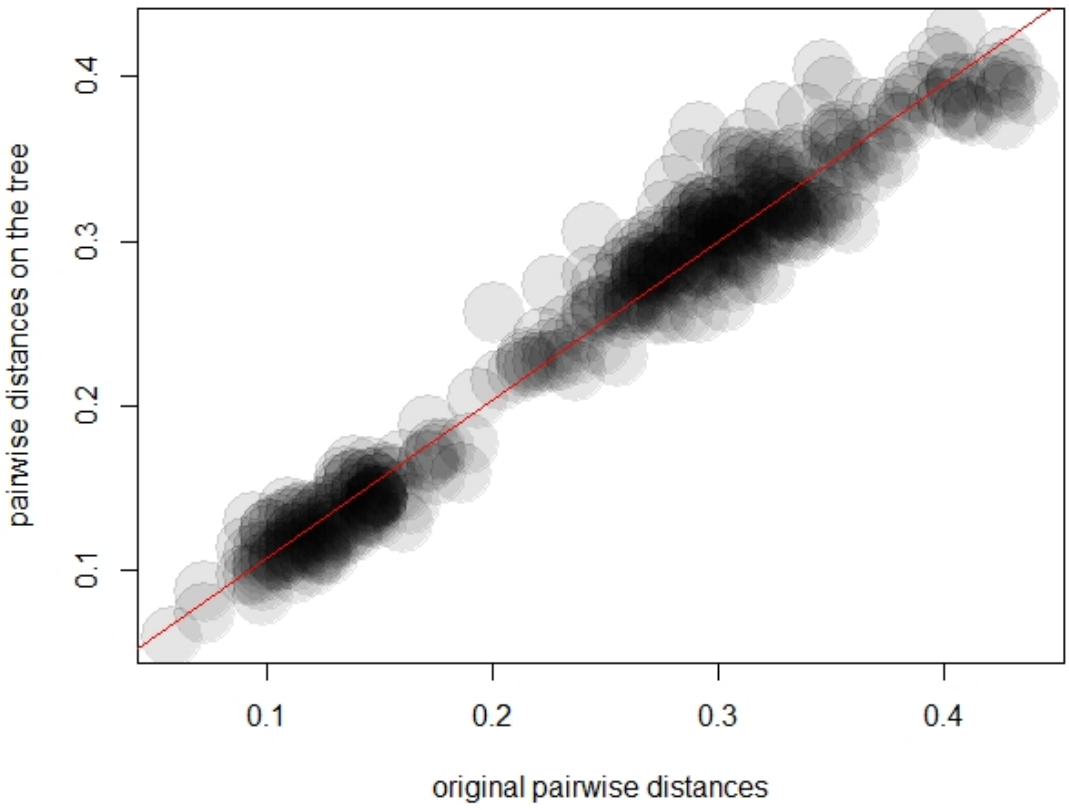
Appendix 3: Outlier SNPs from PCAdapt's analysis of loci under selection using 3dRADseq data for habitat ecotypes. P values for all the loci are less than 0.0001.

LCOUS

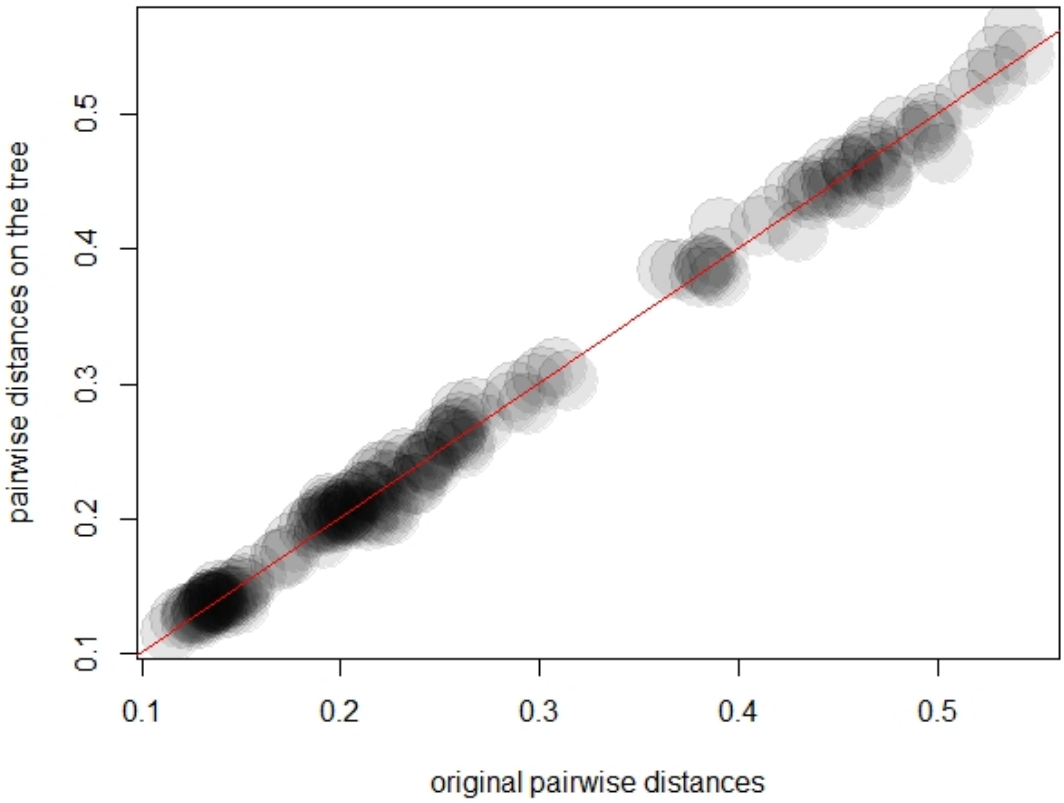
NW_005081536.1
NW_005081538.1
NW_005081540.1
NW_005081543.1
NW_005081543.1
NW_005081545.1
NW_005081545.1
NW_005081545.1
NW_005081551.1
NW_005081552.1
NW_005081556.1
NW_005081556.1
NW_005081557.1
NW_005081559.1
NW_005081559.1
NW_005081560.1
NW_005081569.1
NW_005081569.1
NW_005081570.1
NW_005081570.1
NW_005081576.1
NW_005081581.1
NW_005081583.1
NW_005081598.1
NW_005081598.1
NW_005081600.1
NW_005081603.1
NW_005081604.1
NW_005081606.1
NW_005081607.1
NW_005081608.1
NW_005081608.1
NW_005081610.1
NW_005081612.1
NW_005081612.1
NW_005081613.1
NW_005081613.1
NW_005081613.1
NW_005081613.1
NW_005081613.1
NW_005081617.1

NW_005081617.1
NW_005081618.1
NW_005081629.1
NW_005081639.1
NW_005081640.1
NW_005081649.1
NW_005081650.1
NW_005081654.1
NW_005081656.1
NW_005081663.1
NW_005081675.1
NW_005081683.1
NW_005081686.1
NW_005081686.1
NW_005081690.1
NW_005081696.1
NW_005081696.1
NW_005081714.1
NW_005081725.1
NW_005081731.1
NW_005081733.1
NW_005081740.1
NW_005081748.1
NW_005081748.1
NW_005081754.1
NW_005081757.1
NW_005081758.1
NW_005081758.1
NW_005081763.1
NW_005081777.1
NW_005081779.1
NW_005081779.1
NW_005081779.1
NW_005081779.1
NW_005081781.1
NW_005081782.1
NW_005081782.1
NW_005081782.1
NW_005081796.1
NW_005081808.1
NW_005081819.1
NW_005081952.1
NW_005081955.1
NW_005081973.1
NW_005081999.1

NW_005082111.1
NW_005084522.1



Appendix 4: Reliability test for neighbour joining tree constructed for the habitat ecotypes using 3dRADseq data. Pairwise distance of the original data correlates positively with pairwise distance on the tree at r = 0.95.



Appendix 5: Reliability test for neighbour joining tree constructed for the subspecies groups using lcWGS dataset. Pairwise distance of the original data correlates positively with pairwise distance on the tree at r = 0.99.