

Genomic architecture of speciation in a warbler species complex

by

Silu Wang

M.A., University of Texas, Austin, 2014

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

January 2020

© Silu Wang, 2020

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

Genomic architecture of speciation in a warbler species complex

submitted by Silu Wang in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

in Zoology

Examining Committee:

Darren Irwin

Supervisor

Loren Rieseberg

Supervisory Committee Member

Dolph Schluter

Supervisory Committee Member

Jeanette Whitton

University Examiner

Sally Aitken

University Examiner

Additional Supervisory Committee Members:

Sally Otto

Supervisory Committee Member

Abstract

My PhD research investigated the mechanism of speciation in natural populations. Despite a great progress made since the ‘modern synthesis’ on the genetic basis of speciation, no consensus has been reached in terms of the genomic architecture of speciation: whether speciation is initiated by many regions (scattered across the genome), each of small effects on traits under weak selection (divergent selection or reinforcement), or a few regions that have strong effects on traits under strong selection. Understanding the genomic architecture of speciation is important because it determines whether reproductive isolation could effectively evolve when the diverging lineages still hybridize. I investigated this question in a *Setophaga* warbler species pair at the early stage of divergence: Townsend’s (*Setophaga townsendi*) and Hermit warblers (*S. occidentalis*). These two species hybridize extensively in the hybrid zone in Washington Cascades, demonstrating stable genomic clines over decades. This species complex demonstrates clear pattern of ‘selection with gene flow’, in which a few regions of large effects under strong divergent selection are highly differentiated between lineages, although the rest of the genomes are homogenized by gene flow. These nuclear genomic regions under selection are related to plumage coloration and mitochondrial fatty acid metabolism and are tightly associated with climatic variation among sites. Altogether, mitonuclear adaptation and selection on pigment patterning are prevailing evolutionary forces that counteract gene flow and allow speciation in this warbler system to proceed, despite ancient admixture and ongoing hybridization.

Lay Summary

One of the fundamental questions in biology is: how can one species become more than one. The key to this process, speciation, is hidden in the genomes of individuals in the populations. As the populations differentiate, certain regions in the genome become different between populations. If such differentiated populations meet again and interbreed, the previously accumulated differentiation can be washed away by interbreeding. To become distinct species, there has to be evolutionary forces on certain parts of the genome that maintain differentiation despite interbreeding between populations. My PhD research examined such evolutionary forces in Townsend's and Hermit warblers at the early stage of speciation. It turns out that strong selection is very important in maintaining differentiation at a few genetic regions that contain genes related to plumage coloration and mitochondrial physiology. This inference from natural populations sheds light on speciation mechanism that generates the diversity of life on earth.

Preface

A version of Chapter 2 is now published: **Wang S.**, Rohwer S., Delmore K., Irwin D. E. 2019. Cross-decades stability of an avian hybrid zone. *Journal of Evolutionary Biology*. DOI: <https://doi.org/10.1111/jeb.13524>. I designed the study with advice from Darren Irwin and Sievert Rohwer. I collected the 2015-16 samples. Sievert Rohwer provided the 1987-94 samples. I conducted the laboratory and analytical work advised by Darren Irwin. Kira Delmore designed the GBS pipeline. I wrote the initial draft and worked closely with Darren Irwin on revisions.

A version of Chapter 3 is in review. This project was conducted in collaboration with Sievert Rohwer, Devin de Zwaan, Jacqueline Mackenzie and Darren Irwin. I designed the study, collected 2015-16 DNA samples along with the photos, conducted lab work and genomic analysis with advice from Darren Irwin. Devin designed protocol for pigment quantification. Jacqueline R. Mackenzie conducted this protocol with supervision from Devin de Zwaan and me. I wrote the initial draft of the chapter. Darren Irwin worked with me on earlier versions. Devin de Zwaan and Sievert Rohwer helped edit a later version of the writing.

Chapter 4 was completed in collaboration with Madelyn Ore, Else Mikklesen, Julie Lee-Yaw, Sievert Rohwer, and Darren E. Irwin. I designed the study with advice from Darren Irwin. I requested tissue samples from Sievert Rohwer (Burke Museum) and extracted DNA from the first set of the samples. Then I supervised Else Mikklesen to complete the rest of DNA extraction. Madelyn Ore and I prepared libraries for sequencing. I conducted the genomic analyses with advice from Darren Irwin. I conducted climate analysis with help from Julie Lee-Yaw. I wrote the initial draft, then Darren Irwin helped on revisions. Julie Lee-Yaw and Sievert Rohwer contributed ideas on writing a later version of this chapter.

All research in this thesis is ethically approved by UBC Animal Care Committee (A17-0049).

Table of Contents

Abstract.....	iii
Lay Summary.....	iv
Preface.....	v
Table of Contents.....	vi
List of Tables.....	viii
List of Figures.....	ix
Acknowledgements.....	xi
Dedication.....	xiii
Chapter 1 general introduction.....	1
1.1 Speciation genomics.....	2
1.1.1 Complex speciation in nature: divergence with some level of gene flow.....	2
1.1.2 Two important genetic targets of selection.....	3
1.1.2.1 Mitonuclear coadaptation.....	4
1.1.2.2 Plumage pigmentation genes.....	4
1.2 Ancient and current hybrid zone.....	5
1.3 Setophaga warbler system.....	5
1.4 Research in this dissertation.....	8
Chapter 2 Cross-decades stability of an avian hybrid zone.....	9
2.2 Introduction.....	9
2.2 Methods.....	13
2.2.1 Sampling.....	13
2.2.2 Plumage Hybrid Index.....	13
2.2.3 Genotyping by Sequencing (GBS) and Genomic HI.....	15
2.2.4 Mitochondrial genotyping.....	18
2.2.5 Reduction of geographical dimensions.....	19
2.2.6 Site-level change in HI.....	20
2.2.7 Cline analysis.....	20
2.3 Results.....	22
2.4 Discussion.....	27
Chapter 3 Selection on a pleiotropic plumage gene region underpins early divergence between a warbler species pair.....	32
3.1 Introduction.....	32
3.2 Materials and Methods.....	36
3.2.1 Sampling.....	36
3.2.2 Plumage measurements.....	36
3.2.3 GBS pipeline.....	38
3.2.4 Whole genome sequencing (WGS).....	39
3.2.5 Genomic differentiation.....	39
3.2.6 Admixture mapping.....	40
3.2.7 Geographical Cline analysis on candidate loci.....	41

3.3 Results.....	42
3.4 Discussion.....	51
3.4.1 Pleiotropy.....	52
3.4.2 Dominance.....	53
3.4.3 RALY SNP.....	54
Chapter 4 The correlated evolution of mito-nuclear genes underlying climate adaptation of a warbler species complex.....	56
4.1 Introduction.....	56
4.2 Methods.....	61
4.2.1 Museum samples, mtDNA sequences, and nDNA sequencing.....	61
4.2.2 GBS pipeline.....	62
4.2.3 Population structure and genomic differentiation.....	63
4.2.4 Candidate genetic regions.....	64
4.2.5 mt-nDNA association.....	64
4.2.6 Climate analysis.....	68
4.3 Results.....	68
4.3.1 Population structure.....	70
4.3.2 F_{ST} distribution.....	70
4.4 Discussion.....	77
4.4.1 Biogeography and semi-parallel introgression.....	77
4.4.2 mt-nDNA association.....	79
4.4.3 Genomic Architecture of differentiation.....	82
4.4.4 Caveats and future directions.....	82
4.5 Conclusion.....	84
Chapter 5 Conclusion.....	85
5.1 Genetic architecture of complex speciation.....	85
5.2 Genetic basis of mitonuclear coadaptation and its role in speciation.....	86
5.3 Genetic basis of plumage and social signaling and its role in divergence.....	87
5.4 The relation of plumage and mitochondrial loci.....	88
5.5 Broad implication.....	90
References.....	93
Appendices.....	102
Appendix A Supplementary material for Chapter 2.....	102
Appendix B Supplementary material for Chapter 3.....	106
Appendix C Supplementary material for Chapter 4.....	107

List of Tables

Table 2.1 The least-squares estimates and confidence intervals (CI) of the center (c) and width (w) of plumage, genomic, and mtDNA clines in each sampling period.....	24
---	----

List of Figures

Figure 1.1 Illustration of ‘divergence with gene flow’ model.....	3
Figure 1.2 Range map of the <i>townsendi</i> and <i>occidentalis</i> complex.	7
Figure 2.1 Range map of <i>Setophaga townsendi</i> (top, in magenta) and <i>S. occidentalis</i>	11
Figure 2.2 Genomic principle component analysis of individuals from <i>occidentalis</i> zone(turquoise), the Cascades hybrid zone (yellow), and <i>townsendi</i> zone (magenta).....	18
Figure 2.3 Maps showing mean plumage (A-B) and genomic (C-D) hybrid indices of each of the sampling sites around the Cascades hybrid zone in each sampling period.....	23
Figure 2.4 The plumage (A, D), nuclear genomic (B, E), and mitochondrial (C, F) clines over decades (blue: 1987-94, yellow: 2015-16) do not reveal hybrid zone movement (cline center did not shift)... ..	25
Figure 2.5 Bootstrap distribution of $w_{2015-16}^2 - w_{1987-94}^2$	27
Figure 3.1 Illustration highlighting the key plumage difference between <i>occidentalis</i> (left, turquoise), hybrids (center, yellow), and <i>townsendi</i> (right, magenta).....	35
Figure 3.2 F_{ST} (between inland <i>townsendi</i> and <i>occidentalis</i>) scan is consistent with the isolation with gene flow model.....	43
Figure 3.3 Admixture mapping revealed a SNP inside the RALY gene in chromosome 20.....	44
Figure 3.4 Association of the RALY SNP on crown, Cheek, and breast coloration.....	46
Figure 3.5 Evidence of selection on RALY locus or its linked region.....	48
Figure 3.6 There is signature of hitchhiking around the RALY peak (elevated F_{ST} around RALY), while the effect on plumage coloration is more tightly associated with RALY.	50

Figure 4.1 Geographical distribution of mtDNA ancestry with coastal *townsendi* (banding code: “TOWA”) site names labeled and haplotype network of the mtDNA sequences from Krosby and Rohwer (2009) study.59

Figure 4.2 Principle component analysis of 19083 high quality SNPs in the genome.....66

Figure 4.3 The Weir & Cockerham weighted F_{ST} among *occidentalis*, inland and coastal *townsendi*.....69

Figure 4.4 F_{ST} scan between *occidentalis* and inland *townsendi*, other coastal and inland *townsendi*, between Valdez and Haida Gwaii *townsendi*, as well as *occidentalis* and coastal *townsendi*70

Figure 4.5 Coastal *townsendi* (black) and *occidentalis* (grey) exhibit concordant genetic differentiation from inland *townsendi* at regions in chr 5 and Z that are associated with genes involved in mitochondrial fatty acid metabolism.....73

Figure 4.6 Spatial distribution of ancestry proportion at mitochondrial fatty acid metabolic markers at chr 5 and Z, and chr20 RALY SNP.75

Figure 4.7 Climate principal component analysis of 26 climate variables from ClimateWNA.....76

Acknowledgements

I want to first thank my advisor, Darren Irwin, who is so dedicated for my scientific development towards great rigor. He showed me field photos and research notes from his PhD time and motivated me to think deeper and endure over challenges. I am especially grateful for his respects and cultivation of my research originality and guided me with rigor and questions.

I was so fortunate to have many wonderful mentors for this scientific endeavour. I have tremendous gratitude for Sarah Otto, who initially inspired me to work with Darren for my interest in speciation in birds when I was finishing up my master's study, and has been such a thoughtful, intelligent, and influential committee member. Thank Loren Rieseberg for generously sharing deep knowledge and insights about speciation genomics and all the enlightening conversations. I am very thankful for Dolph Schluter for his always so refreshing, stimulating, and profound statements about evolution, and letting me join his interesting weekly lab meetings. I want to thank Jeannette Whitton who triggered me to think about glacial cycles and climatic history as the context of population differentiation, and has been so supportive for my research and teaching practices. Thank Mike Whitlock for precious discussions about genomic scans, linkage disequilibrium, and population expansion. Wayne Maddison and Derek Tan provided inspiring guidance on scientific illustration. Thank Armando Geraldés for figuring out the genotyping-by-sequencing protocol with me and sharing helpful insights. Thank Geoffrey E. Hill for inspiring conversations that stimulated Chapter 4.

I am very grateful for many Irwin lab members, although we only overlapped briefly. Thank Dave Toews for his sincere friendship, advices, and insights. Thank Kira Delmore for helping me get started with genomic pipelines. Julie Lee-Yaw for teaching me climate analysis. Alison Porter for showing me bird banding and DNA extractions. Haley Kenyon for allowing me

to play with her data. Thank Madelyn Ore for contributing sequence data for Chapter 4. Thank my friend Ailene McPherson, Remi Doret, Andrea Thomaz, Tom Booker, Ken Thompson for interesting discussions. Thank Gil Henriques for providing warbler digital illustration.

My PhD research would not be possible without the helpful people from the Burke Museum: Chris Wood, Sievert Rohwer, and Sharon Birks. Especially thank Chris for sharing field notes and maps from 1980s, so that I could revisit their historical sites. I am very thankful for my wonderful field assistants Ruth Midgley (2015) and Else Mikklesen (2016) sampling. Thank Devin de Zwaan and Jackeline Mackenzie for pigment analysis for Chapter 3.

Research funding was provided by the Natural Sciences and Engineering Research Council of Canada (grants 311931-2012, RGPIN-2017-03919 and RGPAS-2017-507830 to DEI; and PGS D 331015731 to SW), a Werner and Hildegard Hesse Research Award in Ornithology and a UBC Four Year Doctoral Fellowship to SW. Finally I appreciate Environment Canada, U. S. Geological Survey, Departments of Fish & Wildlife of Washington, Idaho, and Montana, and the UBC Animal Care Committee for providing research permits.

Thank my grandfather, Tongsheng Wang, who ignited my passion for biodiversity when I was growing up in his biodiverse farm. Thank my kind mentors Marla Sokolowski and Mark Kirkpatrick for enormous guidance and supports over many years. I am in enormous debt to my grandmother Shaoying Tu, grandfather Daode Min, and father Yi Wang who provided phenomenal support for my early education, but I did not get to see them when they passed away in China during my graduate study. Great thanks to my mother Guixiang Min, sister Siyu Wang, brother-in-law Alex Sheyerman, uncle Hong Wang, and uncle Hongbin Min for their constant love and supports. Lastly, I am extremely thankful for my husband Dahong Chen for the unconditional supports and intellectual synergy.

Dedicated to my grandmother 涂少英

红尘满面凄凉月，何时乡还孝故里？

黄鹤一去玉影碎，白云悠忘生死茫。

细雨凄凄芳草栖，碧露莹莹晴川历。

松鸟苔虫轮回生，竹筐手影然悟空。

Chapter 1 general introduction

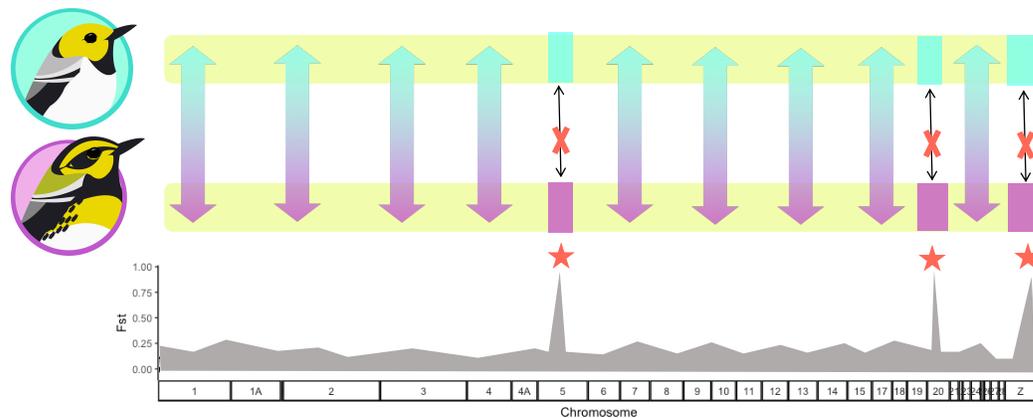
There are at least 8.7 million species on earth (Mora et al. 2011). What is the mechanism generating the diversity of life? Darwin (1859) proposed that natural selection was the driver for speciation. However the exact mechanism by which variation within populations become independent units of evolution was unclear until the modern synthesis between Darwin's theory of evolution and Mendelian genetics (Mendel 1869) in the 20th century (Fisher 1919; Wright 1931; Dobzhansky 1937a). Long term geographical isolation was emphasized because of the difficulty of establishing complete reproductive isolation in the absence of geographical isolation (Mayr 1942). However there is complex speciation in natural populations, in which various amounts of gene flow/hybridization occurred at different time in the speciation history (Bush 1994; Seehausen et al. 2014; Payseur and Rieseberg 2016). Such gene flow contributed to the "grey zones of speciation" (De Queiroz 2007; Roux et al. 2016), where different species concepts and species delineations tend to disagree, and even whether speciation would proceed is uncertain. Thus the core for understanding speciation becomes understanding the signature of gene flow among diverging/diverged lineages subject to divergent selection or selection against hybrids. The "footprints" of these evolutionary forces are registered in the genome, which can be more readily revealed with the advancement in genomic technology. Investigating at these "footprints" in hybrids zones is important, because these zones are places where diverging/diverged lineages hybridize, thus the ideal places to examine the interplay between selection and gene flow in shaping the evolutionary trajectories of the lineages. My thesis examined the genomic signatures of selection and gene flow during the complex speciation in natural populations of New World warblers in which both ancient and ongoing hybrid zones exist.

1.1 Speciation genomics

1.1.1 Complex speciation in nature: divergence with some level of gene flow

As lineages diverge, genomic differences build up among lineages, generating a genomic landscape of differentiation. Speciation in natural populations is often more complex than previously thought due to gene flow between diverging/diverged lineages (Bush 1994; Wu 2001; Seehausen et al. 2014; Payseur and Rieseberg 2016). Such gene flow at different points in time reshapes the genomic landscapes of differentiation. The regions in the genome that are under selection (i.e. genomic targets of selection) can be distinct between diverging/diverged lineages, representing the genomic ‘islands of divergence’, while the rest of the genome becomes homogenized (Wu 2001; Nosil et al. 2009a; Via 2009) (Figure 1.1). However the targets of selection can become co-localized due to selection on recombination or to avoid recombination (Felsenstein 1974). This process can elevate genome-wide level of divergence through genomic hitchhiking or divergent hitchhiking (Maynard Smith and Haigh 1974; Nosil and Feder 2012; Feder et al. 2013), revealing widespread genomic divergence when the species pair transit from one end to the other end of the ‘grey zone’ (Roux et al. 2016), becoming independent units of evolution.

Figure 1.1 Illustration of ‘selection with gene flow’ model in which the genetic targets of selection (stars) remain distinct between species (turquoise versus magenta blocks), while the rest of the genome becomes similar between species due to gene flow.



The genomic underpinning of the transition from one side to the other side of the ‘grey zone’ can be investigated in natural populations/species that still experience extensive gene flow, thus in the early stage of speciation. Genome-wide differentiation may increase due to increasing numbers of targets of selection as well as physical linkage of different parts of the genome with those targets (Hartl 1977; Feder et al. 2012b). Understanding these targets of selection is central to understanding the topping point as speciation unfolds.

1.1.2 Two important genetic targets of selection

In the avian speciation literature, two genomic targets of selection have been increasingly recognized as important targets of selection in recent years: they are nuclear genomic regions related to plumage pigmentation (Gray and McKinnon 2007; Hugall and Stuart-Fox 2012; Poelstra et al. 2013; Toews et al. 2016b; Kim et al. 2019; Knief et al. 2019) and mitochondrial

function (Hill 2017, 2019). My thesis examined the role of such genetic targets in the genomics of divergence.

1.1.2.1 Mitonuclear coadaptation

Mitochondria are the organelles in eukaryotic cells that carry protein coding genes that modulate mitochondrial functions such as cellular respiration (Henze and Martin 2003). The mitochondrial genome co-functions with the nuclear genomes for important cellular functions such as energy production that are critical for individual survival (Rand et al. 2004; Dowling et al. 2008). There are more than 1000 genes involved in animal mitochondrial function (Pagliarini et al. 2008), but only 37 genes are coded in mitochondrial genomes of vertebrates (Rand et al. 2004; Kühlbrandt 2015). Many nuclear-encoded gene products need to physically interact or co-function with the mitochondria-encoded molecules; thus compatibility is critical (Rand et al. 2004; Hill 2019; Hill et al. 2019). When divergently co-adapted mitonuclear complexes mix under secondary contact, suboptimal mitonuclear combinations can be generated through hybridization, which can be selected against, leading to recovery of compatible mitonuclear combinations (Rand et al. 2004). This form of incompatibility has been increasingly recognized as an important intrinsic barrier allowing speciation to unfold (Rand et al. 2004; Sambatti et al. 2008; Gagnaire et al. 2012; Bar-Yaacov et al. 2015; Kühlbrandt 2015; Morales et al. 2018; Hill 2019).

1.1.2.2 Plumage pigmentation genes

Recently in the avian speciation literature, there has been increasing recognition of plumage genes contributing to genomic divergence among populations/species (Sætre et al.

1997; Hill and McGraw 2004; Poelstra et al. 2013, 2014; Toews et al. 2016b; Knief et al. 2019). In some cases, there is limited differentiation between species throughout the genome, except a few plumage genes that contain almost fixed differences between species (Toews et al. 2016b). One reason that plumage-related genetic differentiation may be important is the legacy of plumage-centered species delineation in birds, so that we arbitrarily consider populations with different plumage as distinct species (Sharpe 1909; Mayr 1970; Hill 2006). Alternatively, plumage genes can play an important role in reproductive isolation (Sætre et al. 1997), whether evolved at secondary contact (via reinforcement) (Hill 2018) or in allopatry, thus keeping the populations apart even in the face of gene flow (Wu 2001; Via 2009; Feder et al. 2012a; Poelstra et al. 2013; Knief et al. 2019). The two possibilities can be parsed out by careful examination of gene flow in the hybrid zone: if the plumage genes are an important barrier for gene flow, they should demonstrate steep clines in the hybrid zone relative to the rest of the genome (Poelstra et al. 2013, 2014). I thus looked into these possibilities in the *Setophaga* warblers for my PhD.

1.2 Ancient and current hybrid zone

Hybrid zones are important places for understanding speciation (Barton 1979b; Barton and Hewitt 1985), because they reveal the evolution of reproductive isolation at precise points in time and space. There are at least 135 animal hybrid zones documented in nature and still more to be described (McEntee et al. 2018). Hybrid zones are important places to understand evolutionary forces at species boundaries (selection, drift, migration) (Endler 1977a,b; Szymura and Barton 1986; Barton and Hewitt 1989; Barton and Gale 1993). In addition, hybrid zones are natural experiments for revealing genetic basis of divergent phenotypes (Buerkle and Lexer 2008; Shriner 2017), uncovering the genetic targets of selection at an early stage along the

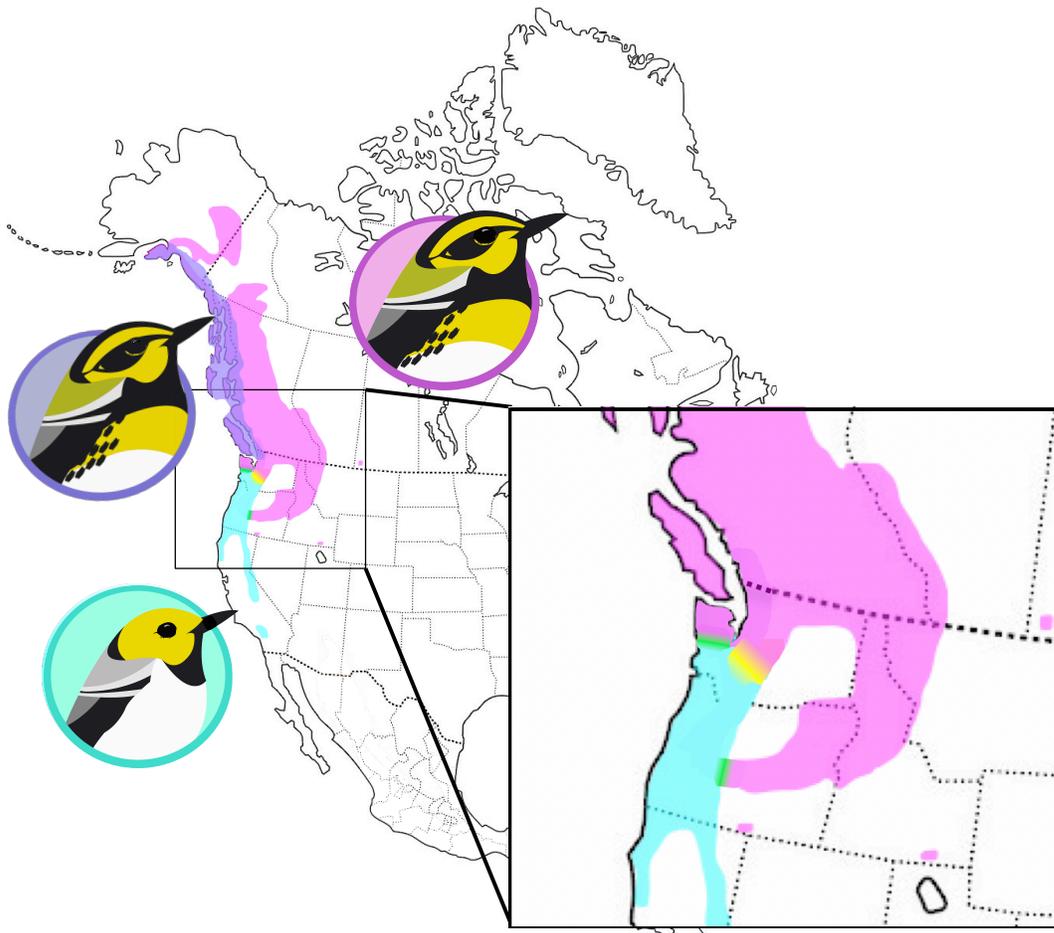
speciation continuum where hybrids can still form. Advanced generation hybrids are particularly useful in revealing these targets, as they have undergone many rounds of recombination, allowing the genetic basis of traits to be more accurately inferred.

1.3 *Setophaga* warbler system

The North American *Setophaga* warbler (Family: Parulidae) clade exhibits recent Pleiocene adaptive radiation (Price et al. 1998, 2000). One of the sister pairs of the *Setophaga* warblers that still experience ongoing gene flow is the Townsend's Warbler, *S. townsendi* (abbreviated as *townsendi*, see range map Figure 1.2), and Hermit Warbler, *S. occidentalis* (abbreviated as *occidentalis*, Figure 1.2) (Rohwer and Wood 1998). The mitochondrial genome is quite distinct between *occidentalis* and inland *townsendi* (~0.5 million years of divergence) (Weir and Schluter 2004), but they still hybridize extensively in three hybrid zones: Olympic mountains, Cascade mountain ranges, and Oregon, USA (Figure 1.2). In the coastal *townsendi* populations (Figure 1.2 range colored in purple), both *occidentalis* and inland *townsendi* mitochondrial DNA (mtDNA) haplotypes can be found, which was thought to be the result of ancient hybridization between the coastal *occidentalis* that was along coastal Canada and Alaska and expanding *townsendi* from the interior after glacial retraction (Krosby and Rohwer 2009). This mtDNA distribution along with the aggressive/territorial behavior asymmetry (*townsendi* is more aggressive than *occidentalis*) (Pearson and Rohwer 2000) led to the hypothesis that the hybrid zones between inland *townsendi* and *occidentalis* are moving into the *occidentalis* range, which may lead to extinction of *occidentalis*. However to date, all inferences of this system have been limited to phenotype and a couple mitochondrial genes. Questions remaining in the system are: (1) would the cross-decades genomic variation over space support hybrid zone movement; and

(2) would the nuclear genomic differentiation be consistent with the mtDNA differentiation; how are coastal *townsendi* differentiate from *occidentalis* and inland *townsendi*?

Figure 1.2 Range map of the *townsendi* and *occidentalis* complex. The *occidentalis* and inland *townsendi* hybridize at three contact zones that are respectively located in: 1) Olympic Mountain and 2) Oregon (colored in jade green), and 3) Cascade Mountain ranges (colored in yellow, the focus of my PhD research). The coastal *townsendi* (colored in purple) harbors mtDNA from both *occidentalis* and inland *townsendi*, and are thus thought to be ancient hybrid populations.



1.4 Research in this dissertation

This was originally a story of two hybridizing warbler species that diverged around the Pleistocene, but it became a story of a species complex with at least 5 distinct genetic clusters as my PhD research progressed. These additional 3 genetic clusters are ancient hybrid populations that represent admixture between the parental species and some unique differentiation. What evolutionary forces have been shaping the differentiation of these populations at such early stages of divergence? With admixture mapping, climate analysis, and across-decades geographical genomic comparisons, I found a clear genomic pattern of ‘selection with gene flow’ in which a few genetic targets of selection became divergent among populations while the rest of the genome was homogenized by gene flow. I found evidence that these genetic targets of selection are related to plumage coloration and mitochondrial fatty acid metabolism.

In chapter 2, I tackle the question whether the hybrid zone has been moving or not and that the hybrid zone has been stable and narrow, what are the evolutionary forces maintaining it? Chapter 3 further investigates the genomic architecture of the early divergence as well as the genetic mechanism of plumage divergence with admixture mapping. Chapter 4 supports the genetic basis of plumage coloration in the ancient hybrid populations along coastal British Columbia and Alaska. In addition, chapter 4 further delves into the other two major regions of differentiation between *occidentalis* and inland *townsendi* and reveals a signature of mitonuclear coadaptation.

Chapter 2 Cross-decades stability of an avian hybrid zone

2.2 Introduction

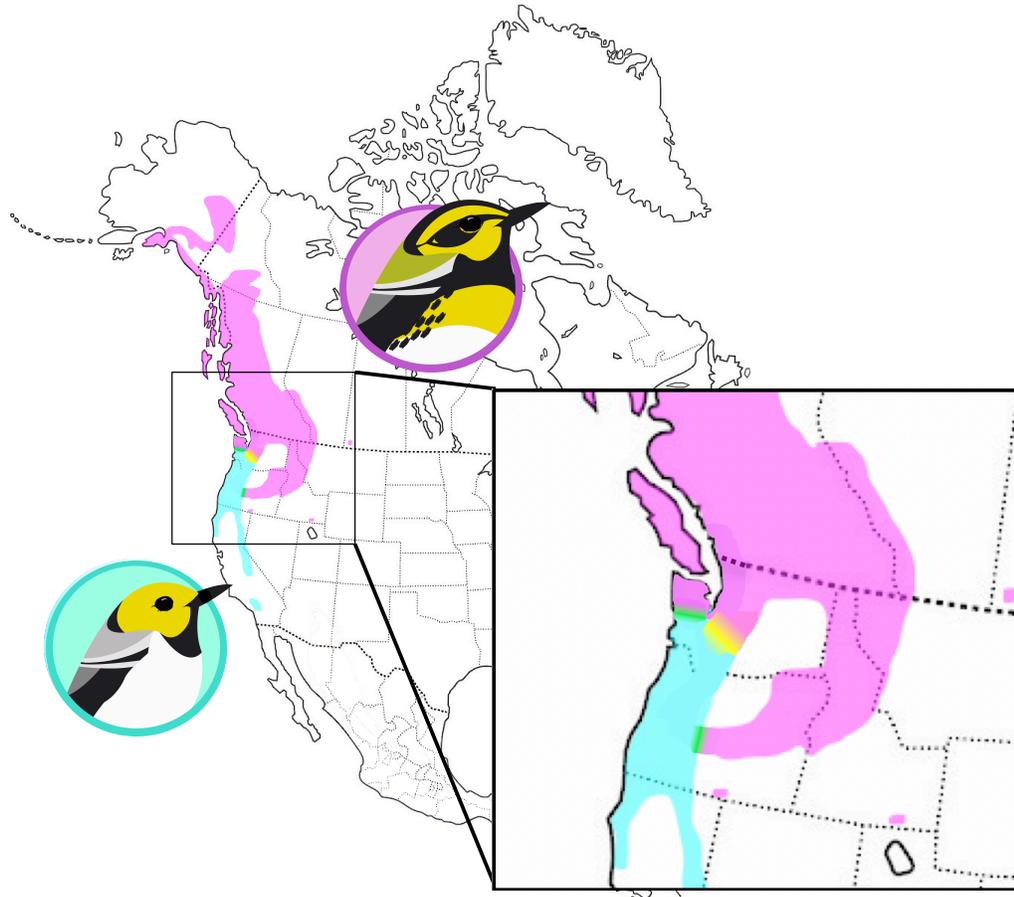
The central process that generates the diversity of life is speciation, in which populations become reproductively isolated and follow different evolutionary trajectories (Mayr 1942). Much of our understanding about the causes of partial reproductive isolation has emerged from the study of hybrid zones, where differentiated lineages interbreed and produce hybrids (Barton 1979b, 2001). Hybrid zones are often tension zones, in which selection against hybrids is balanced by dispersal of parental forms into the zone (i.e., a selection-dispersal equilibrium), leading to a hybrid zone that is stable in width (Slatkin 1973; Barton and Hewitt 1985). In contrast, the location of such a tension zone can move over time, for instance if one of the forms is dominant over the other or if there is a gradient in habitat quality and therefore local population density. If the position of the hybrid zone moves to a density trough, where habitat quality is lower, such tension zones can become stable in position (Barton 1979b).

Buggs (2007) reviewed 25 moving hybrid zones, and since then 11 more hybrid zones have been shown to be moving over time (Buggs 2007; Gay et al. 2008; Wielstra et al. 2017a; Ryan et al. 2018). These can be detected by tracking cline centers over time (Taylor et al. 2015). For instance, the cline center moved 40 km over 32 years in a *Papilio* butterfly (*P. glaucus* and *P. canadensis*) hybrid zone (Ryan et al. 2018) and 11.5 km in 10 years in a hybrid zone between two chickadees (*Parus atricapillus* and *P. carolinensis*) (Taylor et al. 2014b). Hybrid zones can move due to density gradients between parental zones (Barton 1979a), asymmetrical mating (Endler 1977a; Konishi and Takata 2004), competitive displacement (Currat and Excoffier 2005), and differential selection on parental populations (Key 1968). With the accruing reports

of hybrid zone movement, it seems that movement might be more common than stability (Buggs 2007; Wang et al. 2011; Taylor et al. 2014a; Wielstra et al. 2017b; van Riemsdijk et al. 2019).

Here we examine a case that is often cited as an example of hybrid zone movement: the contact zone between Hermit Warblers (family Parulidae, *Setophaga occidentalis*, abbreviated as *occidentalis*), which breed in California, western Oregon, and southwestern Washington; and Townsend's Warblers (*S. townsendi*, abbreviated as *townsendi*), which breed in eastern Oregon, northern and northeastern Washington, British Columbia, Alaska, Idaho, and western Montana (Figure 2.1). These hybridize in three current hybrid zones: 1) the Olympic Mountains and 2) Cascade Mountains of the state of Washington, USA, and 3) the Cascade Mountains of the state of Oregon, USA (Rohwer and Wood 1998) (Figure 2.1). We directly test hybrid zone movement by tracking over time the spatial transition in plumage and genomic hybrid index (HI) across the Cascade Mountains contact zone between these two warbler species.

Figure 2.1 Range map of *Setophaga townsendi* (top, in magenta) and *S. occidentalis* (bottom, in turquoise) and the three known hybrid zones (Rohwer and Wood 1998). The cascade hybrid zone in Washington, USA, is in yellow. The Olympic and Oregon hybrid zones are in green.



This *occidentalis* X *townsendi* hybrid zone is often cited as an example of hybrid zone movement (Buggs 2007; Price 2007; Chunco 2014; Curry 2015; Grether et al. 2017) because of multiple lines of inference. First, mitochondrial haplotypes similar to those of *occidentalis* were found in coastal populations of *townsendi* far north of the current hybrid zone, a pattern interpreted as a 2000 km genetic wake following an original meeting of the two species much further north and movement of the hybrid zone southward to its present location (Rohwer et al. 2001a; Krosby and Rohwer 2009). Second, observations of aggressive behavior suggested an

asymmetry, with *townsendi* males being more aggressive and displacing *occidentalis* males from their territories (Pearson 2000; Pearson et al. 2000). Third, Krosby and Rohwer (2010) found that sites within the Washington Cascades ($N = 11$) and Olympic Mountains ($N = 2$) hybrid zones became more *townsendi*-like in their plumage patterns and colours 10-20 years after initial sampling, although the statistical significance was somewhat marginal ($p = 0.046$).

We extended the study of the Washington Cascades hybrid zone both in terms of time and density of sampling. In 2015-2016 we revisited the 1987-94 sampling sites (Rohwer and Wood 1998) and investigated clines of both plumage and genomic characters across the hybrid zone as well as within much of the *occidentalis* range. We tested whether there has been hybrid zone movement by comparing cline locations across these two sampling periods. Following Rohwer and Wood (1998) and Rohwer et al. (2001a), we used 8 species-specific plumage characters to generate a phenotypic hybrid index (HI) score ranging from *occidentalis* to *townsendi*. We also generated a mtDNA HI (Rohwer et al. 2001a) and a genomic HI using thousands of variable sites across the nuclear genome. We then fit geographic cline models to the variation of phenotypic, mtDNA, and nuclear genomic HI among sampling sites, and we compared whether specific resampled sites at the center of the hybrid zone tended to shift in the same phenotypic direction. If the hybrid zone has been moving in the *townsendi*-to-*occidentalis* direction, we expected geographic cline centers to shift towards the *occidentalis* side over time. We also tested whether cline width was consistent with neutral diffusion, given reasonable estimates of the number of generations between sampling periods as well as individual dispersal; or alternatively whether selection of some form is maintaining a narrow cline and stabilizing the differentiation between *townsendi* and *occidentalis*.

2.2 Methods

2.2.1 Sampling

Previous to the present study, the Washington Cascades hybrid zone was sampled in two time periods: first by Rohwer and Wood (1998) and second by Krosby and Rohwer (2010), both during the breeding season (mid-May to mid-July). The first sampling (34 sites, in which birds were collected within 10 km of each other) was carried out in 1987-1994 (Rohwer and Wood 1998); the second sampling focused on the center of the hybrid zone (11 of the 34 sites plus 2 sites from the Olympic Mountains hybrid zone) during 2005-2008 (Krosby and Rohwer 2010). In the 1987-94 sampling, about 10 birds were sampled at each of the sites scattered around the observable phenotypic center of the *townsendi-occidentalis* transition in the Cascade Mountain range (Rohwer and Wood 1998). In that study, focal birds were attracted using a locally-recorded song playback and whole specimens were collected; these specimens are now stored in the Burke Museum of Natural History and Culture (University of Washington, Seattle, Washington) (Rohwer and Wood 1998; Krosby and Rohwer 2010).

To track this hybrid zone over time, we carried out another round of sampling ($N = 225$) using a catch-and-release approach during the breeding seasons (early May to mid-July) of 2015 and 2016. To test whether the hybrid zone has moved to the *occidentalis* range, we sampled the 34 sites ($N = 191$) that were sampled historically by Rohwer and Wood (1998), as well as 10 more sites ($N = 34$).

2.2.2 Plumage Hybrid Index

We scored the plumage of each bird following the methodology of Rohwer and Wood (1998). Briefly, eight plumage traits characterize species identity: streaks in the mid and lower flanks, gaps at the corner of the bib (a dark area around the neck), darkening on the face, extent

and intensity of yellow on the breast, darkening on the crown, and back color. We divided each trait score by the maximum score possible for each trait, then we averaged the resulting values for each individual to acquire a hybrid index (HI) ranging from 0 to 1, with 0 indicating pure *occidentalis* and 1 indicating pure *townsendi* plumage.

To ensure plumage scores are comparable across time periods, we rescored 152 specimens (in the Burke Museum of Natural History and Culture) from the historical sampling (Rohwer and Wood 1998), while being blind to the originally assigned scores. We then estimated best-fitting relationships (Table S2.1) between trait scores assigned to these individuals by Rohwer and Wood (1998) to plumage scores assigned by the current observer (SW). We applied these transformation functions (Table S2.1) to the other 329 historical individuals (the response variable was rescored plumage value and the predictor was historical score). A single person (SW) conducted all of the rescoring of historical specimens and scoring in the current sampling. The result of this procedure is that the transformed plumage scores of historical sampling are directly comparable to the plumage scores of the current sampling.

The age classes of the focal males need to be considered in plumage color analysis, because otherwise the younger birds tend to demonstrate higher HI than the older birds (Jackson et al. 1992; Rohwer and Wood 1998). To control for this effect, we aged each bird and transformed HI of the young birds by adding a correction factor. To be consistent with aging criteria for the historical sampling, we assigned birds to immature (i.e., prior to the end of their first breeding season, at about one year old) and adult classes (i.e., in their second breeding season or later, at about two years old or older) based on the black shaft streak on the secondary coverts, the size of the white spot at rectrix IV, and the shape of the rectrices (Jackson et al. 1992; Rohwer and Wood 1998). We renamed “immature” and “adult” age classes as “young”

and “old” respectively. The correction factor is the mean difference of each quartile of the old HI and young HI, so that each quartile has a specific correction factor. We did not simply use the mean difference HI between age classes as the correction factor, because the HI distribution of each age class is not normally distributed. All the raw individual plumage HIs from both sampling periods were age-corrected in this way.

We then rescaled age-corrected HI between 0 and 1 so that the HI = 0 represents pure *occidentalis* plumage whereas HI =1 represents pure *townsendi* plumage. The HIs of 60 *occidentalis* individuals (from California and southern Oregon, USA) and 42 *townsendi* individuals (from Idaho and Montana, USA) that are far from the hybrid zone were used to calculate the lower and upper bound of HI in the hybrid zone to represent the plumage variation in the parental populations. Specifically, the 95th percentile of *occidentalis* (0.2548) and 5th percentile *townsendi* (0.8269) HI distribution was respectively considered the lower and upper bound of hybrid zone HI (samples collected in western Washington around the phenotypic cline center (Rohwer and Wood 1998)). Hence the HI was rescaled to vary between 0 and 1 by subtracting 0.2548 and dividing by 0.5721 (0.8269 - 0.2548). Then 0 or 1 was assigned to the individuals with transformed residuals less than zero or greater than 1 respectively. Site mean HI was used to reflect site-level plumage admixture.

2.2.3 Genotyping by Sequencing (GBS) and Genomic HI

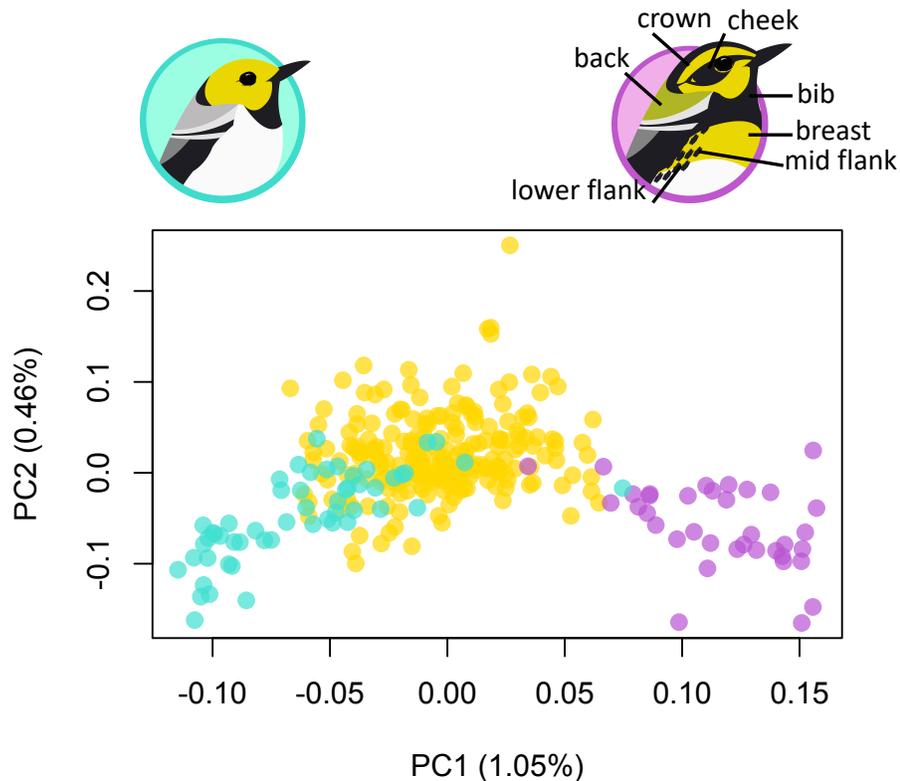
For analysis of genomic variation, we selected a subset of the full sampling (see *Sampling*) that includes 12 sites ($N = 45$) in the 1987-94 sampling (Rohwer and Wood 1998) that together cover the phenotypic cline center of the hybrid zone and 36 sites ($N = 142$) in the 2015-16 sampling (we gathered 155 individuals from 51 sites in total but only included sites with at least 3 individuals for site-level inference). We extracted DNA with a phenol-chloroform

protocol either from tissue samples (from the historical sampling, obtained from the Burke museum) or blood samples in Queen's lysis buffer (Seutin et al. 1991) (collected from the field in 2015-16).

We prepared genotyping-by-sequencing (GBS) libraries for DNA samples of 181 (39 from 1987-94 and 142 from 2015-16) hybrid zone individuals and 90 individuals from the parental ranges (54 *occidentalis* and 36 *townsendi*) (Elshire et al. 2011; Alcaide et al. 2014). GBS allows reduced-representation sampling of variation from many genomes by fragmenting the genomes with a restriction enzyme, ligating the products with adaptors and barcodes (to identify sequences from each individual), and amplifying the sequences with PCR (Elshire et al. 2011). We then sent libraries to Genome Quebec for paired-end DNA sequencing (read length = 125 bp) with an Illumina HiSeq 2500 automated sequencer. Bioinformatic processing of sequencing reads was done using a similar approach as Irwin et al. (2016). In brief, we demultiplexed and then trimmed sequences with Trimmomatic 0.36 (Bolger et al 2014) [TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30] before aligning them to the Zebra Finch *Taeniopygia guttata* reference genome (version 3.2.4) (Warren et al. 2010) using BWA-MEM (default settings) (Li and Durbin 2009). We used HaplotypeCaller and GenotypeGVCFs of GATK (McKenna et al. 2010) to identify single nucleotide polymorphisms (SNPs) and genotype individuals at those SNPs. We acquired 4,097,089 SNPs, which we filtered using VCFtools (Danecek et al. 2011). The filtering regime involved 1) removing indels, 2) keeping SNPs with genotype quality (GQ) > 20, 3) keeping SNPs with minor allele frequency (MAF) ≥ 0.05 , 4) removing SNPs with > 30% missing data, and 4) keeping only biallelic SNPs. This resulted in 21,852 SNPs remaining for downstream analysis.

To calculate individual HI, we conducted principal components analysis using the SNPRelate package (Zheng et al. 2012) in R (R Core Team 2014). The first principal component (PC1) reflects the between-species difference, in which individuals sampled from the parental range were at either ends of the PC1 range, thus PC1 can be taken as a hybrid index for each individual (Figure 2.2). To rescale PC1 so that 0 represents *occidentalis* and 1 represents *townsendi* ancestry, we first found the lower (-0.0071) and upper bound (0.0688), which was respectively the 95th percentile of the PC1 score of pure *occidentalis* ($N = 34$) and the 5th percentile of pure *townsendi* ($N = 36$). Then we rescaled PC1 scores by subtracting the lower bound value and then dividing by the range of PC1 (upper bound - lower bound). We assigned 0 or 1 to the transformed score that was lower than zero or greater than 1 to ensure the rescaled PC1 varies between 0 and 1. This rescaled PC1 was interpreted as a genomic HI, with 0 corresponding to pure *occidentalis* ancestry and 1 to *townsendi* ancestry. In each geographic sampling site, the HI values of individual birds were averaged to represent the mean ancestry at that site. We did not use model-based approach such as Faststructure (Raj et al. 2014) to estimate hybrid index because this species pair is at an early stage of differentiation, has a complex biogeographic history not well suited to the assumptions of some models, and because these programs tend to force the hybrids to either of the parental cluster and the output admixture proportions are heavily affected by the prior distributions (see supplementary method).

Figure 2.2 Genomic principle component analysis of individuals from *occidentalis* zone (turquoise), the Cascades hybrid zone (yellow), and *townsendi* zone (magenta). The key plumage traits are labeled on the *townsendi* illustration (warbler illustration by Gil Henriques).



2.2.4 Mitochondrial genotyping

Rohwer et al. (2001b) showed that allopatric *occidentalis* and inland *townsendi* harbor almost fixed differences in the mitochondrial cytochrome oxidase I (COI) gene, although both haplotypes were widely distributed across the Cascade hybrid zone (N = 283). We scored the *occidentalis* haplotype as 0 and the *townsendi* haplotype as 1. To track changes in frequencies of the mtDNA haplotypes over time, we subsampled 30 individuals (9 sites, at least 3 individuals each) in the 2015-16 sampling and sequenced the NADH dehydrogenase subunit 2 (ND2) gene. We amplified the ND2 gene with the known primer pair (Krosby and Rohwer 2009) and

conducted Sanger sequencing with the forward primer only, which covers ~800bp in the ND2 gene that contain most of the differentiated sites between *occidentalis* and *townsendi*. ND2 and COI should be tightly linked in the mt genome (because mtDNA is matrilineally inherited); to ensure the genotyping with COI is consistent with the ND2, we randomly sequenced the ND2 gene of 33 individuals from the 1987-94 sampling and confirmed a 100% match between COI and NADH genotyping. We calculated the site mean HI by taking the mean ancestry score of each site.

2.2.5 Reduction of geographical dimensions

To investigate the spatial variation in hybrid index (HI), we transformed the two-dimensional sampling into a one-dimensional transect by finding the shortest distance of each site to the plumage HI 50% isocline in the historical sampling (Rohwer and Wood 1998). We first constructed Local Polynomial Regression Fitting (LOESS) models in R (R Core Team 2014), which locally fit the polynomial surface of HI such that the fit at each point was based on its neighboring points weighted by their distance to the focal point. In the LOESS models, we specified the degree of polynomials to be 2 and the span (the proportion of points included in the neighborhood used for fitting at each point) to be 0.6. The shortest distance of each site to the plumage HI = 0.5 isocline in the 1987-94 sampling (Rohwer and Wood 1998) was calculated using the *sp* package (Pebesma and Bivand 2005) in R. The sites east of the isocline were assigned positive values while western sites were assigned negative values. We used these location scores of each site for analysis of the spatial variation in both plumage and genomic HI. We added 1200 km to the location score of each site to generate location scores that were all above zero while still being indicative of the location of each site in relation to the HI = 0.5 isocline of the historical sampling (which was therefore located at 1200 km).

2.2.6 Site-level change in HI

To detect if there was any shift over time in mean HI at sites near the center of the hybrid zone (the sites within 25 km from the 50% plumage HI isocline in 1987-94 sampling), following Krosby and Rohwer (2010) we conducted a one-tailed Wilcoxon signed-rank test, in which the mean HI of each site was compared between time periods.

2.2.7 Cline analysis

To examine the location and width of the geographical clines, we used the `nls` function in R to fit an equilibrium sigmoidal cline model (Szymura and Barton 1986; Gay et al. 2008) to the relationship between site location (x) and site mean HI (y). The fit of this model was used to estimate cline center (c) and width (w):

$$y = \frac{1}{1 + e^{-\frac{4(x-c)}{w}}} \quad [1]$$

This was done for each sampling period. The confidence intervals of c and w were acquired by assuming asymptotic normality, using the `confint` function in R.

Under the null model of no hybrid zone movement, differences between time periods in the estimated locations of cline centers is due only to uncertainty caused by sampling error. We therefore used AIC (Akaike 1973) to compare cline models of 2015-16 to the ‘fixed-center cline model’ that has cline centers fixed to the estimates of cline centers in 1987-94, $c_{1987-94}$ (widths were allowed to differ). If there was no difference in cline center between time periods, the fixed-center cline models would be expected to demonstrate lower AIC than the full model (with different cline centers in each time period).

To test if there is selection maintaining the cline, we took advantage of the across-decades data in which the number of generations between the sampling periods can be estimated

with reasonable confidence. We tested the change of cline width against the neutral diffusion model (Barton and Hewitt 1985),

$$w = \sqrt{2\pi\sigma^2 t} \quad [2]$$

, where w is the cline width, σ is the standard deviation of parental-offspring distance, and t is the number of generations since the formation of hybrid zone. Conventionally, the expected w under neutrality is approximated and compared to the observed w (Gay et al. 2008; Brelsford and Irwin 2009). However, such an approach requires quality estimation of t , which is usually unavailable in natural hybrid zones. Alternatively, t can be calculated based on w estimated from equation [2], and if t is drastically smaller than the age of the initial documentation of the hybrid zone, neutrality can be rejected. Such approach is insufficient for hybrid zones that lack historical records, and/or have moderate selection that is not strong enough to cause drastic deviations from neutrality.

Given these challenges, we develop a new approach that uses sampling from two time periods, in which equation [2] can be used to express $w_{t_2}^2 - w_{t_1}^2$, the difference in cline width squares between the time periods:

$$w_{t_2}^2 - w_{t_1}^2 = 2\pi\sigma^2 \Delta t \quad [3]$$

, where Δt is the number of generations between the two time points.

We used a generation time of 1.8 years (Milá et al. 2007) and the more conservative σ of 10 km/generation^{0.5} used for another closely-related *Setophaga* (Brelsford and Irwin 2009), although the σ estimate of the currently studied *Setophaga* species pair was 30 km/generation^{0.5} (Rohwer et al. 2001b). Thus the Δt between 1987-94 sampling and 2015-16 sampling is 21 years / 1.8 years/generation \approx 12 generations. Inserting these estimates in the equation [3] results in a value of 7540 km² for $w_{2015-16}^2 - w_{1987-94}^2$ under the neutral diffusion model.

2.3 Results

If the hybrid zone had been moving from the range of *townsendi* towards that of *occidentalis*, we expect phenotypic and/or genetic characteristics at sites at the center of the zone to have shifted toward *townsendi* in the more recent time period. We did not detect a significant *townsendi*-biased shift from 1987-94 to 2015-16 for site mean plumage HI (one-tailed Wilcoxon signed-rank test: $W = 110$, $p = 0.43$), nor for the nuclear genomic HI ($W = 21$, $p = 0.89$) (Figure S2.2), nor for the mtDNA HI ($W = 4$, $p = 0.22$) (Figure S2.2).

The mean plumage (Figure 2.3A,B) and nuclear genomic (Figure 2.3C,D) hybrid indices at each site were used to fit the geographical clines. We did not detect significant movement of genomic or plumage clines (Table 2.1, Figure 2.4). The 1987-94 cline center fits well to the 2015-16 data (Table 2.1, AIC $c_{1987-94} < \text{AIC full model}$ in 2015-16 sampling for both plumage and genomic data), confirming that the center did not significantly shift in the recent time period. The mtDNA cline was poorly fit with large confidence interval around the w estimates, consistent with earlier studies that showed lack of a narrow transition in mtDNA frequency (Rohwer et al. 2001b).

Figure 2.3 Maps showing mean plumage (A-B) and genomic (C-D) hybrid indices of each of the sampling sites around the Cascades hybrid zone in each sampling period (A, C, 1987-94; and B, D, 2015-16).

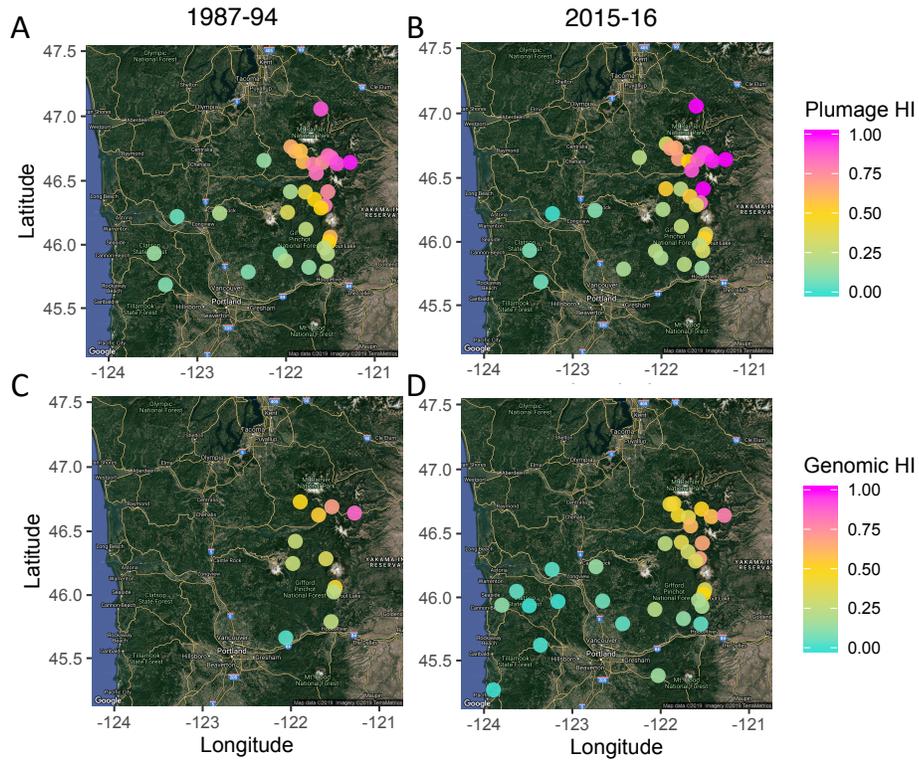
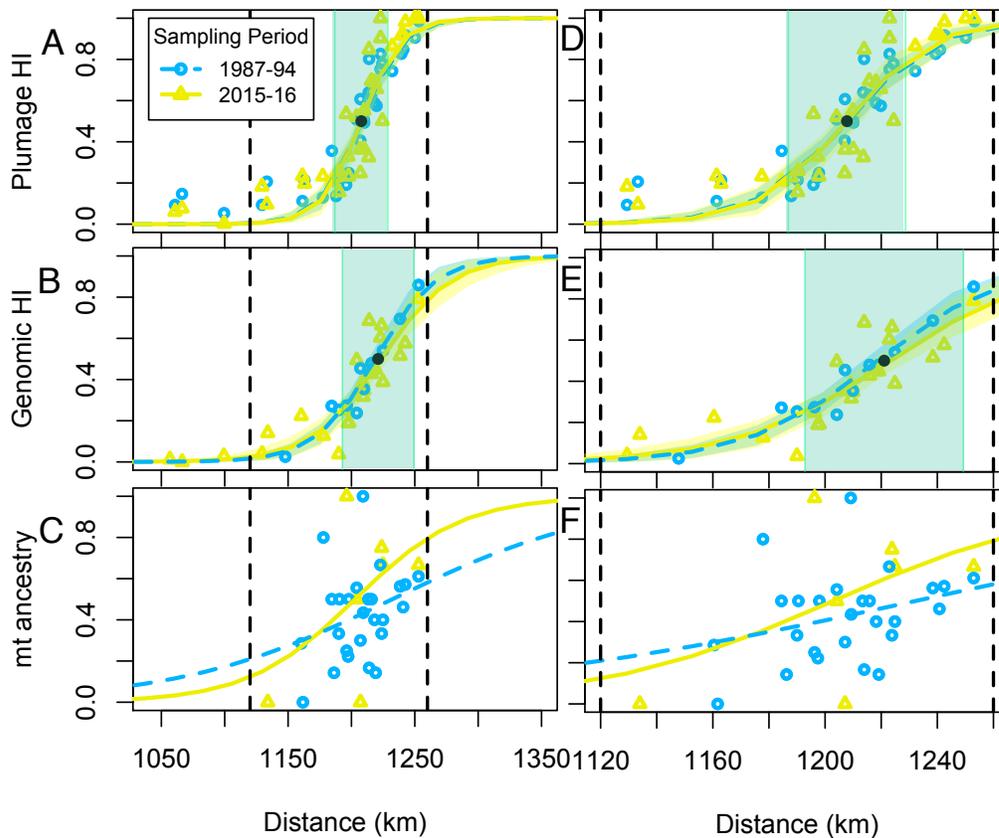


Table 2.1 The least-squares estimates and confidence intervals (CI) of the center (c) and width (w) of plumage, genomic, and mtDNA clines in each sampling period. The cline movement can be inferred from the comparison of the AICs of the cline model versus the models of which the cline center was fixed the previous time period. If the cline been moving, the AIC of fitted center models should be lower ($\Delta\text{AIC} > 2$) than the fixed-cline-center models ($c_{1987-94}$). For each cline trait, the lower AIC is bolded.

	Period	AIC	AIC ($c_{1987-94}$)	c	95% CI c	w	95% CI w
Plumage	1987-94			1207.93	(1204.92, 1210.93)	64.90	(50.77, 79.03)
	2015-16	-33.92	-37.90	1207.65	(1203.48, 1211.83)	62.06	(42.703, 81.42)
Genomics	1987-94			1218.84	(1214.48, 1223.20)	94.35	(72.11, 116.58)
	2015-16	-39.52	-41.3	1223.28	(1216.88, 1229.68)	112.28	(73.45, 151.10)
mt	1987-94			1232.30	(1191.53, 1273.07)	337.10	(-94.88, 769.08)
	2015-16	8.91	6.07	1202.89	(1151.31, 1254.41)	167.89	(-174.15, 509.93)

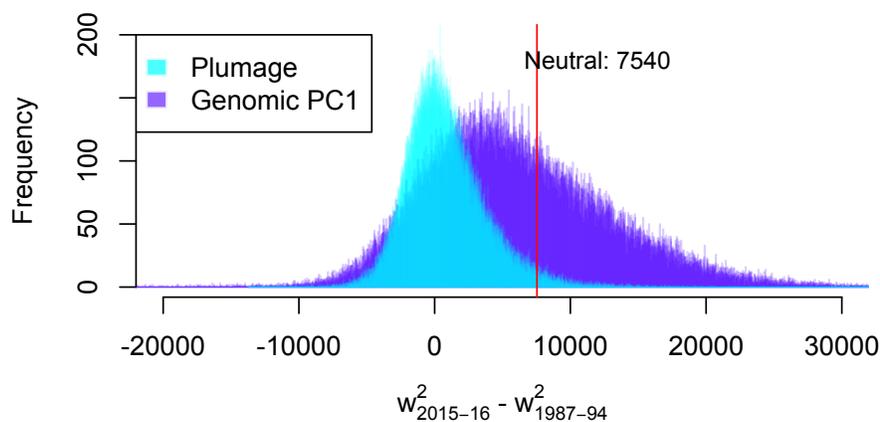
Figure 2.4 The plumage (A, D), nuclear genomic (B, E), and mitochondrial (C, F) clines over decades (blue: 1987-94, yellow: 2015-16) do not reveal hybrid zone movement (cline center did not shift). The shades (blue: 1987-94, yellow: 2015-16) around the clines in A, B, D, E represent 95% bootstrap confidence intervals of the cline models. Graphs in the right column (D-F) are respectively zoomed-in cline plots of those in the left column (A-C) towards the cline center (1120 to 1260 km). The vertical green shade is the standard error around mean cline center (black dot) position (between time periods).



To test if the plumage and genomic clines are widening as expected under neutral diffusion, we calculated the bootstrap distribution of the observed genomic and plumage $w_{2015-16}^2 - w_{1987-94}^2$. The bootstrap distribution (Figure 2.5) was acquired by resampling the sites with replacement, fitting clines within each time period, and computing $w_{2015-16}^2 - w_{1987-94}^2$ at each

iteration, over a total of 100,000 iterations. For plumage, the bootstrap distribution of $w_{2015-16}^2 - w_{1987-94}^2$ (point estimate: -676.36; 95% CI: -3978.35, 7536.59) was significantly less than that predicted by the neutral diffusion model ($w_{2015-16}^2 - w_{1987-94}^2 = 7540 \text{ km}^2$; see Methods), suggesting selection has maintained the narrow clines (Figure 2.5). In contrast, the $w_{2015-16}^2 - w_{1987-94}^2$ of the nuclear genomic PC1 cline (point estimate: 5245.268; 95% CI: -5125.625, 21056.125) did not significantly deviate from neutral expectation. However, applying the neutral diffusion equation [2] to the current width of the nuclear genomic cline (112.28 km) with a reasonable estimate of dispersal ($\sigma = 10 \text{ km/generation}^{0.5}$), the estimated time since secondary contact is just 20 generations ago (or just 36 years, if generation time is 1.8 years). This then suggests that the zone formed around 1979, however there has been documentation of *occidentalis* and *townsendi* hybrids in this hybrid zone since 1944 (Jewett 1944), suggesting there might have been selection maintaining the genomic cline narrow as well.

Figure 2.5 Bootstrap distribution of $w_{2015-16}^2 - w_{1987-94}^2$ of the plumage cline is significantly different than the expected $w_{2015-16}^2 - w_{1987-94}^2$ under the neutral diffusion model (Barton and Hewitt 1985), suggesting the plumage cline have become narrower than predicted by neutral diffusion over two decades. The purple and cyan lines represent the estimated genomic and plumage $w_{2015-16}^2 - w_{1987-94}^2$ values.



2.4 Discussion

The *Setophaga townsendi* / *occidentalis* hybrid zone has been treated as one of the most well-known examples of hybrid zone movement (Buggs 2007; Price 2007), based on range-wide patterns in mitochondrial DNA (Krosby & Rohwer 2009), evidence for competitive dominance of *townsendi* over *occidentalis* (Pearson 2000; Pearson et al. 2000), and some evidence ($p = 0.046$) in a shift in the hybrid zone over a short period of time (Krosby & Rohwer 2010). However, our examination of plumage and genomic variation across this hybrid zone over a broader temporal and geographical span suggests little if any movement of the zone.

The presence of *occidentalis* mitochondrial (mt) haplotypes in the coastal portion of the current range of *townsendi* was interpreted to be a result of hybrid zone movement (Krosby and Rohwer 2009), a remnant of a hypothesized *occidentalis* population that was further north along the coast before being replaced by the advancing *townsendi*. Such inference agrees with knowledge of glacial cycles, which suggested that the hybrid zone formed further north (coastal British Columbia) when interior *townsendi* came into contact with coastal *occidentalis* by dispersing westward through the Skeena River valley around 5000 years ago (roughly 2700 generations), and then spread northward and southward over time (Rohwer et al. 2001a), eliminating coastal populations of *occidentalis* and generating the abundance of *occidentalis* haplotypes in coastal populations of *townsendi* from Washington to Alaska. The cline stasis revealed in the current study does not rule out the possibility of very long-term Washington Cascade hybrid zone movement towards the *occidentalis* direction or a potential movement of the coastal Olympic Mountains hybrid zone (Figure 2.1). The Washington Cascade hybrid zone could have been moving before reaching stability in the recent past, or it could still be moving but at a rate that is not detectable within 25 years. In contrast, with a shorter time interval (10 years) and lower dispersal (up to 2.4 km/generation^{0.5} as opposed to 30 km/generation^{0.5} in this *Setophaga* sister pair (Rohwer et al. 2001b)) (Weise and Meyer 1979), the black-capped / Carolina chickadee hybrid zone moved 11.5 km (Taylor et al. 2014a). If there is any force driving the movement of this *Setophaga* hybrid zone, it would be a weaker force than the climate change driving the chickadee hybrid zone movement, or it could have been dampened in the recent past.

An alternate explanation (in addition to hybrid zone movement) for the presence of *occidentalis* mtDNA in coastal *townsendi* populations was considered by Krosby and Rohwer

(2009). In this scenario, the mtDNA distribution may have resulted from *occidentalis* mtDNA haplotypes adaptively introgressing (reviewed by Toews and Brelsford 2012) into the coastal *townsendi* range, leaving most of the nuclear genomic cline behind. Krosby and Rohwer (2009) made a compelling argument that geographic structure in the mtDNA haplotype network was not consistent with a selective sweep of the *occidentalis* mtDNA along the coastal distribution of *townsendi* (Krosby and Rohwer 2009): rather, the *occidentalis* mtDNA in the *townsendi* populations shows geographic structure, suggesting those populations were well established and geographically structured prior to hybridization and introgression of plumage alleles from *townsendi*.

Our results prompt consideration of a different biogeographic scenario (other than long-distance hybrid zone movement or adaptive introgression) for this *Setophaga* species pair. If the *occidentalis* populations resided in various glacial refugia along the coast (from California to Alaska) during last glacial maxima (LGM) (Krosby and Rohwer 2009), they could have interbred with inland *townsendi* that had undergone post-LGM expansion (from Idaho or Montana, U.S.A) at various contact zones along the west coast. The ancient hybridization between inland *townsendi* and northern *occidentalis* could have initially occurred along a long front paralleling the coastal mountains, just inland from the coastline. Then, the hybrid zones could have moved southwest, at each location just a short distance from the mountains toward the coast, causing the phenotype of coastal populations to change from *occidentalis* to *townsendi*. This scenario requires only slow movement of the hybrid zone at each coastal area and is more consistent with the stability we have observed in the present Cascades hybrid zone. According to this scenario, coastal *townsendi* are ancient hybrid populations between north-coastal

occidentalis and expanded inland *townsendi*, a product of admixture with only little hybrid zone movement.

This scenario assumes that the presumably earlier contact through the broad and low Skeena River valley (Rohwer et al. 2001) had little effect on the north- and south-ward capture of *occidentalis* haplotypes by *townsendi*. By this new hypothesis, the zone moved slowly towards the coast. The later retreat of glaciers from east-west corridors that were much higher in elevation than the Skeena River corridor led to *townsendi* moving through these corridors followed by multiple local replacements of *occidentalis* phenotypes along the coast from Alaska through British Columbia.

Male *townsendi* are more dominant at heterospecific territorial interactions in the hybrid zone (Pearson 2000; Pearson and Rohwer 2000), which might favor *townsendi*-like plumage signals. If so, the plumage cline might tend to move further towards the *occidentalis* zone in comparison with the rest of the genomic clines. The surprisingly stable plumage, mtDNA, and nuclear genomic clines over two decades suggest that any movement driven by competitive displacement might have paused in the recent past or have been too slow to be detected over decades. Future work could investigate clines of individual genetic markers to further dissect such stability. An additional possibility is that any advantage that *townsendi* may have in competitive interactions may come with costs that result in little or no net fitness benefits. More importantly, the narrowness of the hybrid zone implies selection against hybrids has been shaping this hybrid zone, and the presently observed lack of movement suggests long-term stability of these two species and the hybrid zone between them. Future study of associations among plumage, mtDNA, and genomic characters may reveal the mechanisms of selection against hybrids and its role in the evolution of reproductive isolation in this system.

We tested whether the *occidentalis X townsendi* hybrid zone has moved by tracking plumage, mtDNA, and nuclear genomic variation over decades, but did not detect significant movement. Although there is still a possibility of long-term hybrid zone movement in this hybrid zone, the hybrid zone has been stable in location over the recent 2-3 decades. The plumage cline is narrower than predicted under the neutral diffusion model, implicating selection in the maintenance of this narrow hybrid zone.

Chapter 3 Selection on a pleiotropic plumage gene region underpins early divergence between a warbler species pair

3.1 Introduction

Examining the genomic distribution of differentiation between two populations can reveal targets of divergent selection, advancing our understanding of the speciation process (Rice and Hostert 1993; Nosil et al. 2009b; Feder et al. 2012b; Via 2012; Wolf and Ellegren 2017; Irwin et al. 2018). At the onset of divergence between sister taxa, differentiation at narrow regions of chromosomes, or “islands of differentiation”, is expected (Wu 2001; Bradshaw and Schemske 2003; Turner et al. 2005; Nosil et al. 2009a; Schluter and Conte 2009; Via 2009). These differentiations are maintained by strong selections on a few genes of large phenotypic effects, which through physical linkage results in elevated divergence of the nearby neutral regions, a phenomenon known as divergent hitchhiking (Nosil et al. 2009a; Via 2009, 2012). More differentiation can accumulate as speciation progresses, thus extending regions of high differentiation and forming “continents of divergence” (Nosil et al. 2009b). Alternatively, speciation can be initiated via many regions of small effects on traits under multifarious selection (one selection targeting multiple traits) and gradually accumulate genome-wide differentiation via a ‘correlated evolutionary response’ (Orr 1995; Johnson and Porter 2000; Orr and Turelli 2001; Matsubayashi and Katakura 2009; Lawniczak et al. 2010; Williams and Oleksiak 2011). A central debate in speciation research has been whether speciation occurs through gradual differentiation throughout the genome or more rapidly at a few key regions. Genomic analyses of sister species in the early stages of speciation are needed to address this debate.

Islands of divergence tend to be associated with divergent (and often diagnostic) traits that are involved in reproductive isolation (Seehausen and Schluter 2004; Kronforst et al. 2006; Whibley et al. 2006). Divergent traits frequently play important roles in social interactions at

geographic boundaries between closely-related species (West-Eberhard 1984; Anderson and Grether 2010). These features are commonly involved in mate-choice (Kronforst et al. 2006; Whibley et al. 2006) or aggressive signaling (Mikami et al. 2004; Seehausen and Schluter 2004) that are divergently selected (Uy et al. 2009; Poelstra 2013; Toews et al. 2016b,a; Barrera-Guzmán et al. 2018; Knief et al. 2019) and facilitate species recognition and reproductive isolation (Sætre et al. 1997; Hill and McGraw 2004). We are just beginning to understand the genetic underpinnings of traits related to species divergence (Crawford and Nielsen 2013; Toews et al. 2016b,a; Shriener 2017; Knief et al. 2019). Integrating the genetic basis underlying plumage signals with corresponding evolutionary forces will deepen our understanding of the formation and maintenance of species barriers.

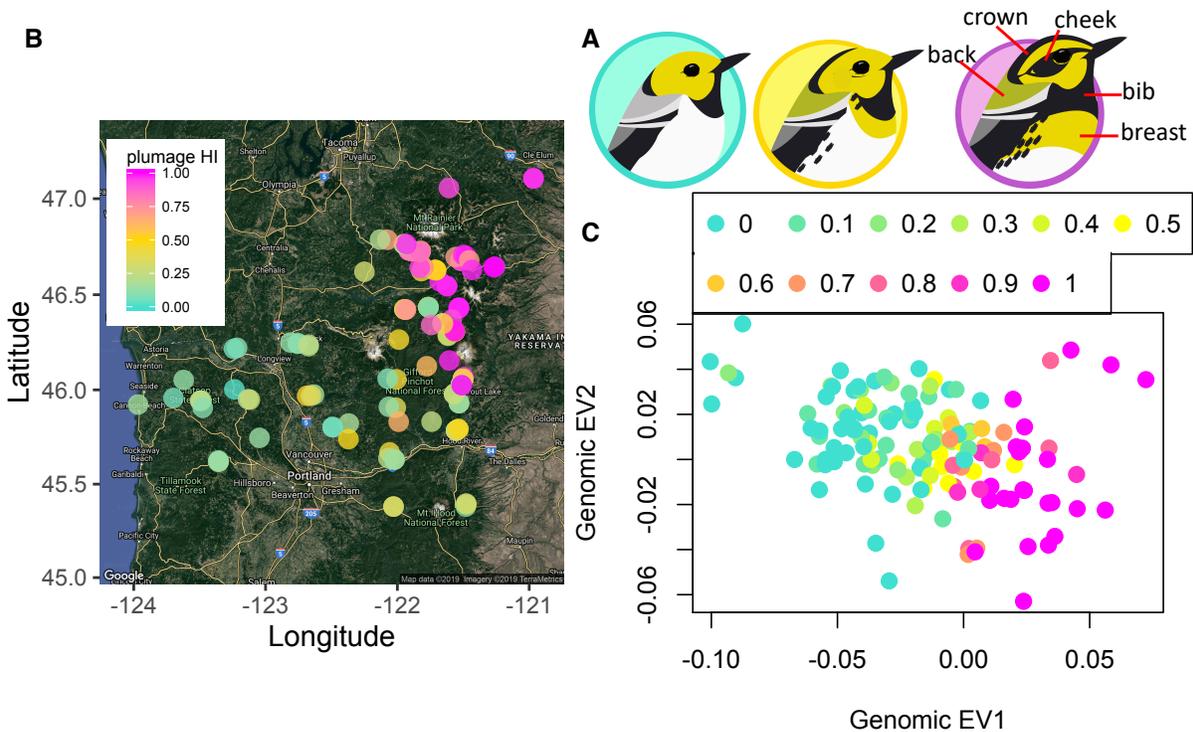
Carotenoid and melanin pigmentation commonly act as species-diagnostic traits (Sætre et al. 1997; Kronforst et al. 2006; Uy et al. 2009; Hill 2015; Toews et al. 2016b,a; Barrera-Guzmán et al. 2018). At the boundaries between closely-related species, the coloration per se might not differ, but the patterning of the pigments are employed as signals reflecting species identity (Hill 2015). Therefore, the regulatory genes of melanin and carotenoid pathways are expected to differ between species with similar coloration but distinct patterning.

Hybrid zones, locations where divergent lineages interbreed, provide opportunities to understand the association of species-diagnostic features and differentiation. Social signals can be crucial for conspecific mate recognition and competition for mating opportunities, hence they can play a role in premating reproductive isolation (van Doorn et al. 2004; Kronforst et al. 2006; Whibley et al. 2006; Pfennig and Pfennig 2009; Dijkstra and Border 2018). When the plumage signals established in isolated populations come into secondary contact, hybrids can be selected

against due to their intermediate or mismatched parental signals interfering with mate recognition and competition (Pfennig and Pfennig 2009; Dijkstra and Border 2018).

The hybrid zone between *Setophaga townsendi* (abbreviated as *townsendi*) and *S. occidentalis* (abbreviated as *occidentalis*) along the Cascade mountains experiences extensive gene flow between closely related species that diverged an estimated 400,000 years ago (Krosby & Rohwer, 2010; Pearson & Rohwer, 2000; Rohwer & Wood, 1998; Weir & Schluter, 2004). The males of this species pair differ in several plumage features related to carotenoid and melanin patterning on the crown, cheek, breast, and back. For example, *townsendi* has a black cheek patch (Figure 3.1A right), whereas *occidentalis* displays a completely yellow cheek (Figure 3.1A, left). Because apparent hybrids predominantly resemble *occidentalis* in crown and cheek coloration (Figure 3.1A center), with intermediate breast (the extent of yellow plumage) and back colours (green plumage in the mantle), Rohwer & Wood (Rohwer and Wood 1998) predicted that face coloration of *occidentalis* and hybrids would be controlled by a single-locus dominant allele. In addition, whether the other carotenoid and melanin patterning differences between species are underpinned by the same genetic mechanism needs to be investigated. These signals can be important in male-male competition for territories, as individuals in the hybrid zone demonstrate an aggression bias towards different plumage types (Pearson & Rohwer, 2000; Pearson, 2000).

Figure 3.1 **A**, illustration highlighting the key plumage difference between *occidentalis* (left, turquoise), hybrids (center, yellow), and *townsendi* (right, magenta) (illustration by Gil Jorge Barros Henriques). **B**, a map showing individual samples in this study and plumage hybrid index of each based on 8 plumage traits (0 for pure *occidentalis*, in turquoise; 1 for pure *townsendi*, in magenta). **C**, Genomic eigenvector 1 (EV1) and eigenvector 2 (EV2) with individual datapoints colored by plumage hybrid index. The genomic EV1 reflects the variation among individuals that are *occidentalis*-like (low genomic eigenvector EV1) versus *townsendi*-like (high genomic EV1).



We investigated whether speciation in this young species pair occurred through gradual widespread differentiation across the genome or via a small number of restricted regions. If selection has targeted a small number of specific genomic regions, genomic islands of differentiation should be narrow and dispersed (Rice and Hostert 1993; Turner et al. 2005; Nosil et al. 2009a). In addition, we tested the following hypotheses regarding the genetic basis of

species-diagnostic traits: 1) one of these islands of divergence pleiotropically underpins species diagnostic features (crown and cheek darkening, breast yellow, bib size, greenish back); 2) cheek darkening is influenced largely by an allele of dominant effect, consistent with the prediction made two decades ago by Rohwer & Wood (1998); and 3) there is stable selection against hybrids at the locus underlying key species diagnostic features.

3.2 Materials and Methods

3.2.1 Sampling

Whole specimens of *Setophaga* warblers were collected over two historical sampling sessions in the Washington Cascade hybrid zone (Rohwer and Wood 1998; Krosby and Rohwer 2010). The first sampling ($N = 314$ individuals; 35 sites) was carried out in 1987-1994 (Rohwer and Wood 1998), while the second ($N = 127$; 11 sites) was in 2005-2008 (Krosby and Rohwer 2010) and covered a subset of the sites from the original sampling. We accessed these specimens at the Burke Museum of Natural History and Culture (University of Washington, Seattle, Washington).

We carried out a third round of sampling using a catch-and-release approach during the breeding season (early May to mid-July) in 2015-16 (Figure 3.1B). Upon locating a territorial male by song, a mist net with a playback (of a locally-recorded song) at the bottom was set up nearby. After capturing an adult, photographs and a blood sample were taken for further analysis. We re-sampled the sites that were sampled in 1987-94 (Rohwer and Wood 1998). For details, see Chapter 2.

3.2.2 Plumage measurements

Melanin- and carotenoid-based plumage traits allow identification of the two species (Figure 3.1B), but there is also variation within each species (Rohwer and Wood 1998; Owen-Ashley and Butler 2004). To quantify plumage variation within and between populations, we focused on five distinct plumage traits in males: the relative amount of black (melanin) and yellow (carotenoid) in the 1) cheek, 2) crown, and 3) breast and the 4) throat bib, as well as, 5) the intensity of green chroma on the back (Figure 3.1B).

For each warbler captured in the field, we took three pictures from different angles: 1) frontal with head tilted up (for bib and breast measurements), 2) profile (cheek), and 3) from above (crown) (Figure S3.1). Plumage colour metrics were measured using Adobe Photoshop CC in CIE (*Commission Internationale de l'Eclairage*) LAB colour space. LAB colour space is a 3-dimensional space consisting of 3 distinct, perpendicular axes: 1) Luminosity (L) ranging from 0 (black) to 100 (white), 2) 'a' ranging from green (negative) to red (positive), and 3) 'b' ranging from blue (negative) to yellow (positive; Adobe 2017). We chose this colour space because it linearizes the variables of interest along three distinct axes: black ('L'), yellow (along the 'b' axis), and green ('a').

For the cheek and mantle, we selected a standardized area and averaged the pixels to record the 'b' and 'a' values, respectively. For the cheek, we selected and averaged the entire area from above the eye (but excluding the eye) to the throat badge, and from the base of the bill to the mantle using the profile photos. Differences in ambient light conditions at the time a picture is taken can confound comparison of colour metrics among individuals. To address this, we used the white-balance feature in Photoshop, using the white plumage of each individual's belly as a standard, to correct for differences in ambient light among photos and standardize the colour metrics. We acknowledge that without spectral analysis, we do not incorporate UV

reflectance which is a ubiquitous aspect of signaling in avian systems (Eaton and Lanyon 2003). However, our methods allow us to estimate the relative intensity of melanin- and carotenoid-based plumage traits during the breeding season.

To measure the size of the black bib, we used the program Analyzing Digital Images (ADI; Bull and Israel 2015). A scale was included in all photos to standardize size measurements among individuals. We measured bib size (Figure 3.1A) by creating a polygon around the bib and calculating the area ($\pm 0.1 \text{ mm}^2$).

3.2.3 GBS pipeline

Following Alcaide et al. (2014), we prepared genotyping-by-sequencing (GBS) (Elshire et al. 2011) libraries from 352 individual DNA samples. In brief, genomes were digested with restriction enzyme and ligated with barcode and adaptors, amplified with PCR, and size selected (fragment length of 300 - 400 bp) for sequencing. Libraries were sequenced at Genome Quebec with paired-end sequencing (read length = 125 bp) on an Illumina HiSeq 2500 automated sequencer. The resulting sequences were processed following a GBS pipeline (Irwin et al. 2016). All the sequence can be acquired through GenBank (accession number: PRJNA573930, ID: 573930). We demultiplexed the reads with a custom script and trimmed them using Trimmomatic 0.36 (Bolger et al 2014) [TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30], then we aligned reads to a *Taeniopygia guttata* reference version 3.2.4 (Warren et al. 2010) using bwa (Li et al 2009) (default settings). We assume synteny of *Setophaga* and *T. guttata* genomes given the limited rearrangement in avian genomes (Ellegren 2010; Zhang et al. 2014), but the conclusions of this study would be unlikely to be affected by a moderate number of rearrangements. We conducted SNP calling with GATK (McKenna et al. 2010), which produced 4,097,089 SNPs. The SNP filtering was done with VCFtools (Danecek et al. 2011), which

includes removing indels, genotype quality (GQ) > 20, minor allele frequency (MAF) \geq 0.05, removing loci with >30% missing data, and only including biallelic SNPs, resulting in 21,852 SNPs remaining.

3.2.4 Whole Genome Sequencing (WGS)

In addition, we selected 5 samples from *occidentalis* in Pinehurst, CA, U.S.A. (UWBM 66152, 66153, 66148-66150) and 5 samples from inland *townsendi* in Tok, AK, U.S.A. (UWBM 84816-84819, 84860) from the Burke museum for whole genome re-sequencing. For DNA extraction, we used 2 mm³ of tissue digested in Qiagen buffers—following the tissue extraction procedure—and separated DNA using UPrep spin columns (Genesee). We standardized DNA concentrations with after quantifying concentrations with a Qubit fluorometer, and then generated sequencing libraries with the Illumina TruSeq Nano kit—which includes an 8-cycle PCR enrichment—selecting 350 bp insert sizes. We individually indexed each sample, and sequenced the combined libraries across a single lane of an Illumina NextSeq using the paired-end 150 bp sequencing chemistry. We combined these 10 samples with other wood warblers from other projects, but consistently included 24 individuals per sequencing lane in order to generate comparable coverage across samples.

3.2.5 Genomic differentiation

To quantify the level of differentiation throughout the genome, we calculated F_{ST} (Weir and Cockerham 1984) with VCFtools (Danecek et al. 2011) between allopatric *townsendi* (Montana and Idaho, USA; $N = 38$) and *occidentalis* (Oregon and California, USA; $N = 23$) for each of the filtered SNPs.

The genomic architecture of differentiation from the GBS was compared to that of the WGS data. The reads from the WGS data were aligned to the same *Taeniopygia guttata* reference (version 3.2.4) (Warren et al. 2010) with bwa (Li et al 2009) (default settings). ANGSD (Korneliussen et al. 2014) was employed to calculate genotype frequencies accounting for genotyping uncertainty before estimating F_{ST} for each of the non-overlapping (step size = 10kb) 10kb windows.

3.2.6 Admixture mapping

To identify SNPs that are associated with variation in the five plumage traits, admixture mapping was conducted with GenABEL package in R (Aulchenko et al. 2007) using 189 birds captured from hybrid zone (around the center of the hybrid zone in Washington, U.S.A.), 46 from the *occidentalis* zone (Southern Oregon and California, U.S.A.), and 7 from the *townsendi* zone (Northeast Washington & Montana, USA). The GenABEL function *ibs*, which calculates identity-by-state relatedness among individuals, was used to control for population structure. Genomic control of inflation factor λ (which represents the effect of genetic structure and sample size) was conducted (Aulchenko et al. 2007). Briefly, λ was estimated from the genomic data assuming that randomly-selected markers from the genome are not associated with the trait (after controlling for population substructure). λ was then used to correct the test statistic χ^2 of each association test, so that the test is relative to the null hypothesis of no phenotype-genotype association (Aulchenko et al. 2007). Because the cheek darkening and breast yellow intensity were left-skewed, rank order transformation of the phenotypes was used.

To find the number of independent hypotheses for multiple hypothesis correction, we estimated the total number of independent linkage blocks (a proxy for the number of independent

hypotheses) in the dataset. Principle Component Analysis (Gao, Starmer, & Martin, 2008; Johnson et al., 2010) was run until 99.5% of the variance (to cover the significant proportion of variance, with 0.05 as the critical threshold) was explained with the R package SNPRelate (Zheng et al. 2012), producing 210 PCs. We applied a Bonferroni correction to the independent blocks, such that significance level $\alpha = 0.05/210$. Because we conducted admixture mapping on five plumage traits, we further applied Bonferroni correction across the 5 plumage traits, which gave rise to the final alpha value of $0.05/(210 \times 5) \approx 4.76 \times 10^{-5}$. Candidate SNPs (i.e., those with p -values below this alpha value) were annotated in Ensembl Zebra Finch (taeGut3.2.4;(Zerbino et al. 2018)).

3.2.7 Geographical Cline analysis on candidate loci

As one candidate gene, RALY (see Results), stood out in the above analysis as particularly strongly associated with plumage variation, we investigated the spatial and temporal variation in this locus relative to the plumage hybrid index and rest of the genome (using samples described in Chapter 2). We fit the relationship between RALY allele frequency and location using an equilibrium geographic cline model for each SNP (Szymura and Barton 1986; Gay et al. 2008).

Geographical cline analysis was done following Chapter 2. Briefly, we collapsed the two-dimensional sampling into a one-dimensional transect by measuring the location of each site to the 0.5 isocline of the plumage hybrid index (HI) in the historical sampling (Rohwer and Wood 1998), as follows. First, a Local Polynomial Regression Fitting (LOESS) model in R (R Core team 2015) was used to fit variation in HI across the hybrid zone, and this model was used to estimate the HI = 0.5 isocline in the 1987-94 sampling (Rohwer and Wood 1998). Then for each site, the shortest distance to the 0.5 isocline was calculated with ‘sp’ package (Pebesma and

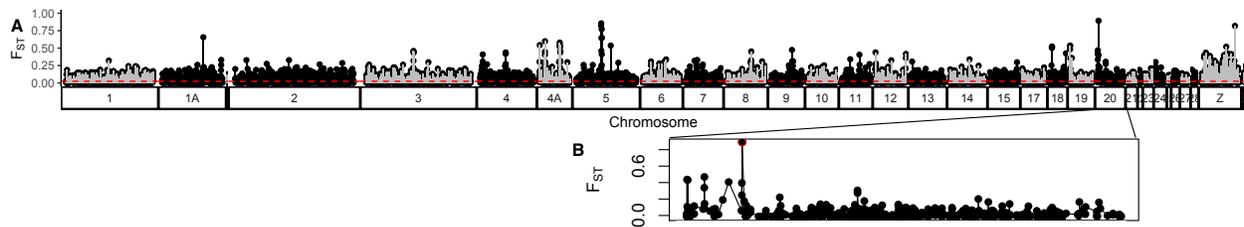
Bivand 2005). The sites east or west of the isocline were specified as having positive versus negative distance values, respectively. We added 1200 km to the distance score of each site so that all distance values are above zero, while the relative distance of each site to the isocline is preserved. Then the data was fit to the equilibrium sigmoidal cline model (Szymura and Barton 1986; Gay et al. 2008) $y = \frac{1}{1+e^{-\frac{4(x-c)}{w}}}$, in which cline center (c) and width (w) was estimated (where y is HI and x is location with respect to the HI = 0.5 isocline). To examine whether selection (i.e., divergent selection and/or selection against hybrids) is acting on the candidate loci, we followed (Chapter 2) and tested whether the increase in cline width ($w^2_{2015-16} - w^2_{1987-94}$) is significantly less than expected under the neutral diffusion model (Barton & Hewitt, 1985). The change of w^2 (between sampling periods) as opposed to w allows comparison of the observed change in cline width to the neutral expectation (Chapter 2). To understand the RALY cline width change relative to the plumage and genomic cline, we compared $w^2_{2015-16} - w^2_{1987-94}$ relative to the plumage and genomic HI cline (Chapter 2), which was respectively based on the scores of the 8 plumage landmarks (with 0 representing pure *occidentalis* and 1 being pure *townsendi*), and the scaled genomic PC1 (with 0 representing pure *occidentalis* and 1 being pure *townsendi*).

3.3 Results

There was low genome-wide weighted average F_{ST} between allopatric *occidentalis* and *townsendi* populations (Weir and Cockerham's $F_{ST} = 0.03$; Figure 3.2 A). However, four high regions of differentiation ($F_{ST} > 0.6$) were found (Figure 3.2 A, Figure S3.2, S3.3, gene association and functions summarized in Table S2) that map to Zebra Finch chromosome (chr) 5 (nucleotide position 25064223-25875302 in the Zebra Finch genome, mean $F_{ST} = 0.75$), chr 20 (1981369, mean $F_{ST} = 0.90$), and chr Z (66226657, $F_{ST} = 0.82$). The SNP on chr 20 is within an

intron of the gene RALY that encodes heterogeneous nuclear ribonucleoprotein, the cofactor for cholesterol biosynthetic genes (Sallam et al. 2016). We hereafter refer to this SNP as the RALY SNP. This RALY SNP demonstrated the highest F_{ST} (0.90) inside an “island” of relatively high F_{ST} on chr 20 between *occidentalis* and *townsendi* parental populations across the genome (Figure 3.2B). The pure *occidentalis* population mostly contains GG homozygotes (refer as OO), while the *townsendi* population contains CC homozygotes (refer as TT).

Figure 3.2 F_{ST} (between inland *townsendi* and *occidentalis*) scan is consistent with the isolation with gene flow model in which a few regions are under selection (divergent selection or selection against hybrids) while the rest of the genome is similar due to gene flow. A, F_{ST} genomic scan of the genome revealed an island of differentiation around the highest peak in chromosome 20 (B, zoom in chromosome 20).



Admixture mapping of some traits showed an even stronger pattern of a small number of loci standing out, now specifically in terms of their association with divergent plumage traits in the hybrid zone. In particular, the colours of the crown, cheek, and the breast were each very strongly associated with the same standout SNP, which remarkably is the same RALY SNP (on chr 20) that stood out in the F_{ST} analysis (crown: $\chi^2 = 57.89$, $r^2 = 0.57$, $p = 5.95 \times 10^{-15}$, Figure 3.3A, Figure 3.4A, D; cheek: $\chi^2 = 44.76$, $r^2 = 0.55$, $p = 2.23 \times 10^{-11}$, Figure 3.3C, Figure 3.4B,E;

breast: $\chi^2 = 28.47$, $r^2 = 0.57$, $p = 9.53 \times 10^{-8}$, Figure 3.3D, Figure 3.4C,F). This RALY SNP shows a partial dominance pattern of the G allele for the three plumage traits, with G/C heterozygotes tending to have similar phenotypes as G/G homozygotes, although there appears to be some additivity as well (Figure 3.4). The RALY gene (Figure 3.3B) is associated with yellow skin pigmentation in mice and quail (Michaud et al. 1994; Nadeau et al. 2008), and is adjacent to two other skin color genes -- ASIP (115941 bases away) and EIF2S2 (30223 bases away) (Figure 3.3B).

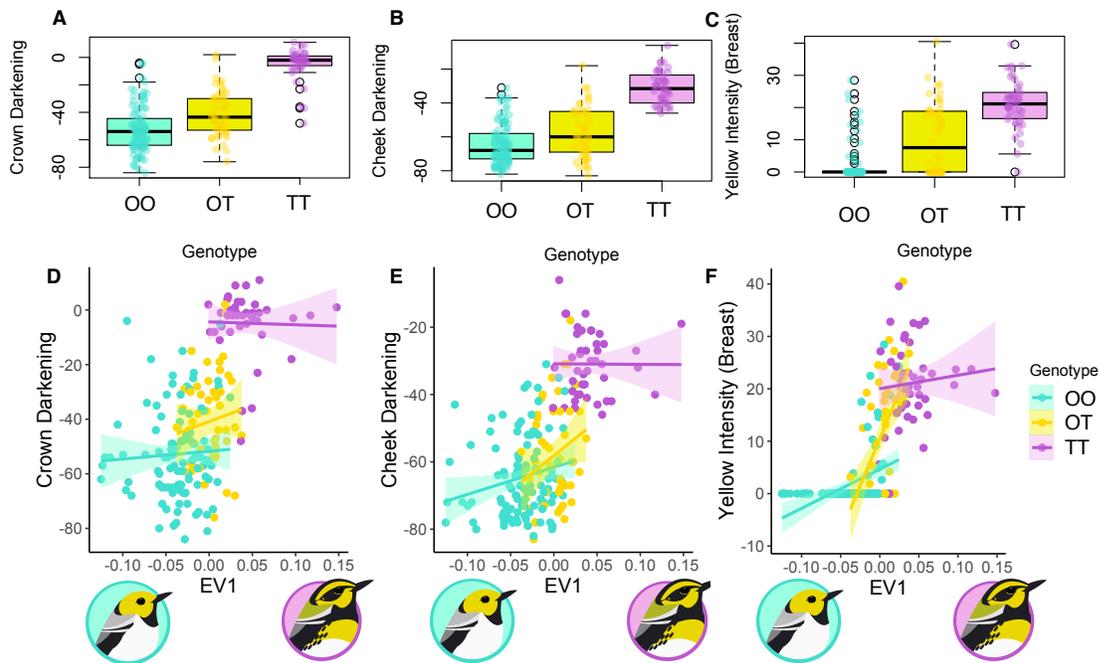
Figure 3.3 Admixture mapping revealed a SNP inside the RALY gene in chromosome 20 at position 1981369 that was significantly associated with crown (**A**, **B**) and cheek darkening (**C**), as well as the intensity of yellow on the breast (**D**), but not with bib size (**E**) nor with the color of the back (**F**). **A-F**, genomic scan of inflation factor-corrected $-\log(p\text{-value})$ of genotype-phenotype association tests. The horizontal red lines represent the Bonferroni-corrected critical threshold. The plumage trait tested in each scan is indicated by the red triangle in the cartoons. **A-C**, there was a strong peak at chromosome 20 position 1981369 inside the RALY gene that indicates a pleiotropic effect on crown, cheek, and breast coloration. **B**, genetic map of around position 1981369 on chromosome 20 and its nearby SNPs in this study. The green boxes represent the stretches of the protein-coding genes, EIF2S2 and RALY, with the exons depicted as the red vertical bars connected by introns (red lines). **E-F**, no significant SNP was detected that is associated with bib size or greenish back.

A low-resolution sequencing approach such as GBS, which sequences less than 1% of the genome, would be unlikely to detect a SNP that is directly causal for phenotypic variation. It is more likely that the RALY SNP is closely physically linked to the casual DNA variant (SNP or some other type of variant) for the phenotypic differences. It is also possible that multiple linked SNPs in this region are responsible for the variation in different phenotypic traits; such that the apparent pleiotropic effect (i.e., strong association with three plumage traits) of the RALY SNP might just be pleiotropy through close linkage. Either way, the genetic region near the RALY SNP appears to have an effect on multiple species-diagnostic coloration traits. The RALY SNP is in physical proximity to the two other pigmentation genes mentioned above, although 3 other SNPs (1955244, 1972476, and 1972481) that are 8888-26125 bp away from RALY SNP (Supplementary Table 2; Figure 3.3B), physically closer to ASIP and EIF2S2, did not show association with phenotype (Figure 3.3B). Two of these SNPs showed low minor allele frequencies (0.05-0.09) in the hybrid zone, and thus are not expected to be highly associated with trait variation. However, the fact that SNP 1973476 (closer to the other genes than the candidate RALY SNP) demonstrated similar minor allele frequency as the significant RALY SNP and was not significantly associated with plumage coloration highlights the importance of the region around position 1981369 inside the RALY gene.

Three additional SNPs were found to be significantly associated with the intensity of yellow on the breast (Figure 3.3D). Two are on chr 4A at nucleotide location 5588139 and 5588235, both inside an ortholog of mammal Immunoglobulin Binding Protein 1 (IGBP1) gene. One is at nucleotide location 6623798 on chr 13, inside gene Annexin A6, which codes for a phospholipid binding protein, Annexin VI (Chang et al. 2007). We did not detect SNPs that significantly explained variation in either bib size or green coloration on the back (Figure 3.3E,

F). The RALY SNP explains 57% of variation in crown (Figure 3.4A, D), 52% of the variation in cheek (Figure 3.4B, E), and 43% variation in breast coloration (Figure 3.4C, F).

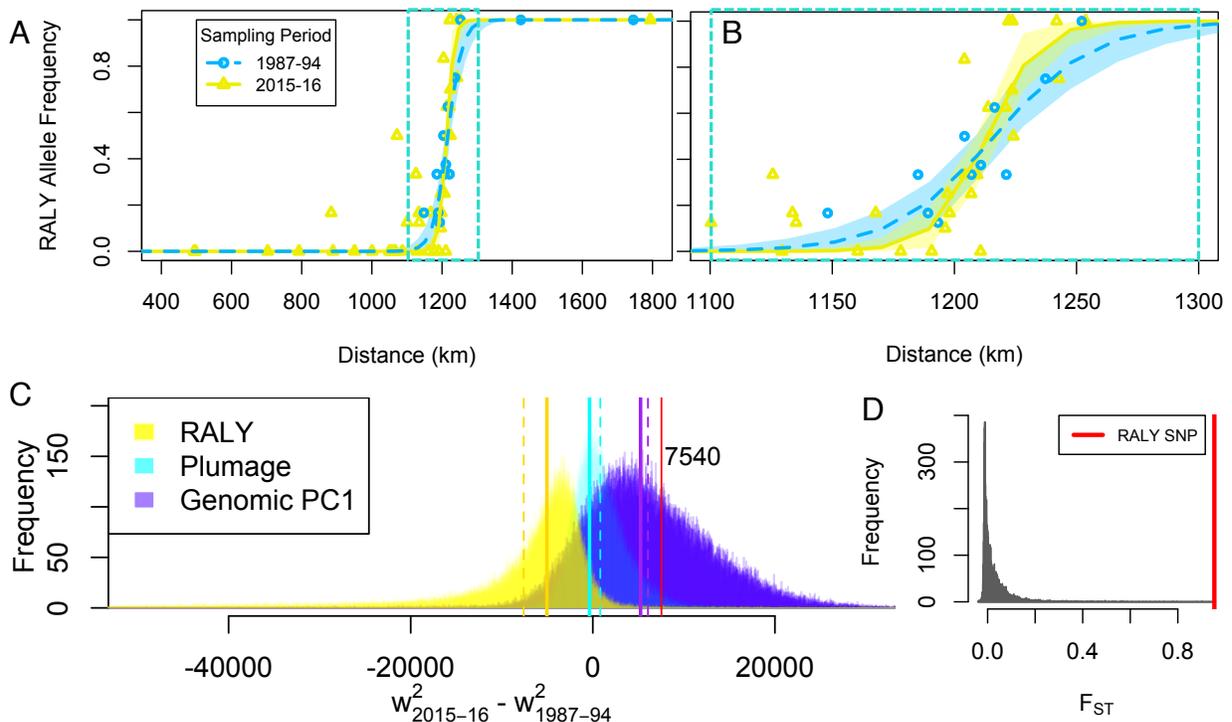
Figure 3.4 Association of the RALY SNP on crown (A, D), cheek (B, E), and breast (C, F) coloration, in which the pure *occidentalis* genotypes is denoted as “OO”, pure *townsendi* genotype is denoted as “TT”, and heterozygotes as “OT”. D-F, The associations of the RALY genotype and cheek darkening among individuals are significant ($p < 10^{-7}$) after accounting for the underlying genomic ancestry. The data are consistent with a combination of dominance and additive effects on the three plumage traits.



We found evidence that the RALY region is under divergent selection or reinforcement, as it is the most extreme F_{ST} outlier, while most of the genome is not very differentiated (average = 0.03) (Figure 3.5D). In addition, the RALY SNP demonstrated a spatial cline that was stable in

location over two decades (Figure 3.5A). The cline center was at 1216.35 ± 4.0 (SE) km in 1987-94 (Figure 3.5A, blue curve), did not significantly shift in 2015-16 sampling (Figure 3.5A, yellow curve), occurring at 1213.140 ± 2.2 (SE) km. Furthermore, the width of the RALY cline became narrower (95% CI of $w^2_{2015-16} - w^2_{1987-94}$: -39004.70-1035.13 km²) than neutral expectation ($w^2_{2015-16} - w^2_{1987-94} = 7540$ km²) (Figure 3.5B, see Chapter 2 for explanation of method).

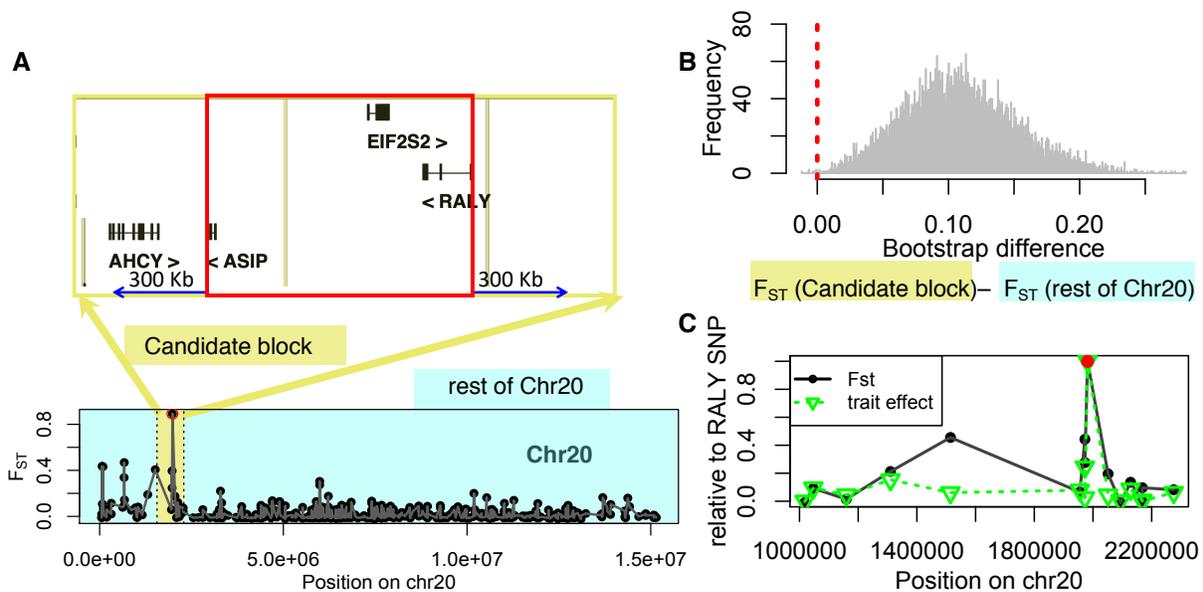
Figure 3.5 Evidence of selection on RALY locus or its linked region. The RALY SNP shows stable clines (sampled in 1987-94 and 2015-16) across the hybrid zone, which are extremely narrow (**A, B**, zoom in around 1100 to 1300 km), and have become significantly narrower in the recent sampling (**C**). The bootstrap distribution of $w^2_{2015-16} - w^2_{1987-94}$ (the change of squared cline width between sampling periods) of the RALY cline is less than expected under neutral diffusion, suggesting selection has been maintaining narrow clines of genomic PC1, plumage, and RALY SNP (**C**). The selection at RALY (greater deviation from neutral expectation than plumage and genomic PC1) might indirectly maintain the narrow plumage and genomic cline (**C**). The vertical lines are $w^2_{2015-16} - w^2_{1987-94}$ values (solid line: sample estimates; dotted line: bootstrap means) respectively for RALY cline (yellow), plumage cline (turquoise), and genomic PC1 cline (purple). (**D**), RALY is the extreme outlier in the distribution of F_{ST} between inland *townsendi* and *occidentalis* across the genome (RALY SNP F_{ST} value depicted by the red line).



We then investigated whether selection at the RALY locus or linked region caused divergent hitchhiking, which would result in greater genomic divergence allowing speciation to unfold. Divergent hitchhiking occurs when the targets of divergent selection indirectly cause the elevation of differentiation of its linked regions (Via, 2009). If there is divergent hitchhiking around the ASIP-RALY gene block, the flanking regions of this candidate gene block should demonstrate greater differentiation between species than the rest of the chromosome 20. We tested whether the bootstrap difference of F_{ST} between the candidate genetic block (300kb flanking region of the ASIP-RALY genetic block, excluding the RALY locus itself, colored in yellow Figure 3.6A) versus the rest of chr20 (colored in light blue Figure 3.6A) is significantly greater than 0. Indeed, we found signature of divergent hitchhiking, as ASIP-RALY candidate genetic block demonstrated significantly greater F_{ST} than the rest of the regions in chr20 (Figure 3.6A, B, grey distribution of difference is greater than 0, 95% CI: 0.107-0.109, $t_{9999} = 254.31$, $p < 10^{-15}$).

Although there is signature of divergent hitchhiking, it seems that the plumage pigmentation effect is more narrowly underpinned by ASIP-RALY region (Figure 3.6C). In the hybrid zone, the recombination over multiple generations can break down the hitchhiking gene blocks (that tend to co-segregate with the trait-determining loci in allopatric populations), revealing specific genetic underpinnings of the differentiated phenotypes. The genetic underpinning of the plumage trait should be very close to the narrow region around the RALY SNP.

Figure 3.6 There is signature of hitchhiking around the RALY peak (elevated F_{ST} around RALY), while the effect on plumage coloration is more tightly associated with RALY. Thus the potential causal locus of this trait should be very close to RALY. **A, B** signature of divergent hitchhiking: the candidate gene block (the pigmentation gene block and its 300kb flanking region excluding the RALY locus) has higher F_{ST} than regions in the rest of the chr20 (**B**, the bootstrap difference of F_{ST} is significantly greater than 0, indicated by the vertical dotted line, 95% CI: 0.107-0.109). **C**, in comparison to F_{ST} , the trait effect is more specific to the RALY peak: there is a steeper decay of scaled trait effect size (relative to RALY peak) than the scaled differentiation (relative F_{ST} to RALY peak).



3.4 Discussion

We observed clear ‘islands of differentiation’ on chr 1A, 5, 20, and Z (Figure 3.2), suggesting that selection on specific regions of large effect underlies speciation in the face of gene flow in this system. This agrees with the predicted pattern of selection with gene flow: as neutral regions homogenize through gene flow, restricted targets of divergent selection have been maintained (Nosil & Feder, 2012; Nosil, Funk, et al., 2009; Rice & Hostert, 1993; Via, 2009).

Indeed, the RALY SNP (chr 20) that resides on one of the ‘islands of differentiation’ has strong effects on key divergent traits: cheek, crown, and breast coloration, and demonstrate signatures of selection (divergent selection or reinforcement). The consistently narrow geographical RALY cline in the hybrid zone over several decades despite ongoing gene flow (Chapter 2) further revealed the key role of selection (divergent selection and/or selection against hybrids) around the RALY SNP in driving this young sister species apart. Indeed, the significantly negative $w^2_{2015-16} - w^2_{1987-94}$ deviates from the expectation under neutral diffusion, is consistent with strong selection that has been directly/indirectly shaping the spatial-temporal distribution of RALY genotypes. The $w^2_{2015-16} - w^2_{1987-94}$ estimate at the RALY cline was less than the plumage cline and the genomic PC1 cline (Figure 3.5B), suggesting that a narrow region centred at the RALY locus might be the direct target of selection, which indirectly maintains the stable plumage and genomic clines (Chapter 2). The signature of divergent hitchhiking of the flanking regions of RALY (Figure 3.6) suggests that such selection may further extend genomic differentiation by elevating the differentiation at linked sites.

The ‘selection with gene flow’ pattern along with evidence of narrow genetic region under strong selection that has large effect on divergent traits suggests a simple genomic

architecture of divergence. Such a strong target of selection could effectively counteract gene flow and maintain species boundary at secondary contact, allowing speciation to progress (Endler, 1977; Gavrillets, 2004; Nosil, et al., 2009). There may still be potential weak divergent selections that have polygenic architecture that complement the strong selection and facilitate genome-wide divergence with correlated evolutionary response. Future study should look for genetic signatures of any potential weak multifarious selections on top of the simple genetic architecture contoured by strong selection. By understanding the genetic architecture underlying species-delineating traits, we gain insight on genomic evolution and selection in the early stages of divergence.

3.4.1 Pleiotropy

This RALY locus underlying multiple plumage patches could provide a strong pleiotropic mechanism (one region that affects multiple phenotypes (Fisher 1930)), linking divergent selection and reproductive isolation. The RALY SNP is associated with a suite of melanin- and carotenoid-related plumage patches (crown, cheek, and breast coloration) that differ diagnostically between species (Figure 3.3) and are likely under some form of selection. Melanin and carotenoid plumage traits are often involved in mate-choice and male-male competition (Kingma et al., 2008; McGraw, 2000; Pryke & Andersson, 2003; Tarof, Dunn, & Whittingham, 2005). If carotenoid-related plumage traits are targets of female mate-choice or male-male competition for mating opportunities in *Setophaga* warblers, pleiotropy can play a powerful role in limiting mating between species and maintaining sympatric speciation by strong selection. Pleiotropy is a more effective mechanism linking selection to reproductive isolation than linkage or polygenic inheritance, although genetic evidence for pleiotropic genes that

influence both selected traits and reproductive isolation has been scarce (see review by Nosil, 2012). However among the limited existing examples, the classic cases involve coloration: the wingless gene affecting reproductive isolation and wing coloration in *Heliconius* butterfly (Kronforst et al. 2006), and the YUP locus affects pollinator isolation and flower coloration in monkey flowers (Bradshaw and Schemske 2003). Our results suggest the RALY SNP could be an additional example of a speciation gene. The stable and narrow cline of RALY SNP and extreme differentiation at the RALY locus imply divergent selection and/or selection against hybrids at this locus, which suggests that RALY affects reproductive isolation in this system. Future study is needed to validate the possibility by examining the role of RALY-regulated plumage signals in mate preference or male-male competition.

3.4.2 Dominance

Our results support the prediction that Rohwer and Wood (1998) made two decades ago that the cheek coloration in hybrids is underpinned by a single dominant locus. Indeed, heterozygotes of the RALY SNP exhibit similar yellow cheeks as the homozygous GG individuals (i.e., yellow cheek), which were found predominantly in the *occidentalis* population (Figure 3.4D-F). Dominance in the signal trait might reduce gene flow if it is in the opposite direction of dominance in the receiver trait (i.e., heterozygous receivers are more sensitive to *townsendi* signals), due to ‘opposing genetic dominance’ in receiver and signaler traits (Forister 2005; Nosil 2012). This is because the F1 hybrids (heterozygous for both signaling and receiving) demonstrate incompatible preference and signal (Nosil 2012), thus reduced mating opportunity. Therefore the *occidentalis* RALY dominance effect on the cheek signal could suppress gene flow if it is opposed by a *townsendi* dominant effect on receiving trait.

3.4.3 *RALY* SNP

The *RALY* SNP is at the peak of divergence (Figure 3.2B), however these nearby regions are: (1) not as high in F_{ST} as the *RALY* SNP (Figure 3.6 A), and (2) not significantly associated with plumage traits after controlling for population substructure (Figure 3.3, 3.6C). *RALY* might be the speciation gene dominating the island of differentiation due to its effect on plumage signaling.

The *RALY* gene encodes for heterogeneous nuclear ribonucleoprotein, the RNA binding protein that is known to regulate the expression of its downstream gene that codes for agouti signaling protein (ASIP) (Nadeau et al. 2008). Deletion of *RALY* in Japanese quails and mice leads to novel transcripts of ASIP, which cause a yellow skin phenotype called “lethal yellow” (Michaud et al. 1994; Nadeau et al. 2008). In mice, a lethal yellow mutation (*Ay*) in ASIP exhibits dominant pleiotropic effects on other traits in addition to skin coloration, which includes obesity, diabetic condition, etc., that are unrelated to the ASIP gene (Michaud et al. 1994). ASIP is well known for its influence on skin pigment by binding to melanocortin receptors and competitively excluding its agonists, preventing black/brown pigmentation (Lovett et al. 1987; Wolf Horrell et al. 2016). Genetic variants of *RALY* in warblers could result in differential expression ASIP leading to variations in carotenoid and melanin patterning. The pleiotropic dominance effect of the *RALY* SNP we identified in this natural hybrid zone is consistent with known *RALY* effect on mutants in the lab.

On top of the strong pleiotropic ASIP-*RALY* modulation, there could be other regions of the genomes contributing to further refinement of species-specific coloration in this system. Specifically, other significant SNPs on chr 4A and 13 are associated with breast coloration. In addition, there might be important genetic regions that this GBS study did not cover.

Conclusion

We identified a few outstanding regions of divergence between two closely related *Setophaga* warblers, revealing selection in the face of gene flow. The most highly differentiated RALY SNP was strongly associated with three distinct plumage traits, revealing a major-effect region related to speciation. In contrast, we did not find genetic regions significantly associated with the other two species-diagnostic traits, implying many genes of small effects.

The extensive similarity across the genome except a few regions of divergence raises a question regarding the distinctiveness of *townsendi* and *occidentalis* and their future. Key to this question is whether these regions of differentiation represent sufficient reproductive isolation that will maintain separation sufficient for further differentiation to build up. The RALY F_{ST} peak, the stable narrow RALY cline, and the signature of divergent hitchhiking around it suggest that there is either divergent selection or reinforcement associated with this RALY genetic region, generating reproductive isolation associated with its differentiation. Future studies should track the genomic landscape of differentiation around the ASIP-RALY region and investigate the mechanism of apparent selection (divergent selection or selection against hybrids). This integrative study revealed the underlying selective forces in hybridizing lineages and shed light on the genetic architecture of speciation and the future of young species pairs.

Chapter 4 The correlated evolution of mito-nuclear genes underlying climate adaptation of a warbler species complex

4.1 Introduction

Mitochondrial (mtDNA) and nuclear (nDNA) genomes co-function to a great extent in maintaining critical functions that influence fitness in all eukaryotes (Ballard and Whitlock 2004; Calvo and Mootha 2010; Lane 2011; Bar-Yaacov et al. 2015; Hill 2019). Different ecological contexts can select for distinct variants of mtDNA, and because many nuclear genes encode proteins that function within mitochondria, the two sets of DNA are expected to co-evolve, each being the target of selection favoring compatibility with the other (Morales et al. 2018; Hill 2019; Hill et al. 2019). Sub-optimal mito-nuclear combinations may arise at species boundaries as a by product of hybridization. Combinations of nDNA from one species and mtDNA from the other species can cause lowered fitness of hybrids, keeping hybrid zones narrow and preventing the two species from merging together (Burton et al. 2013; Hill 2017). Hence coadaptation of mtDNA and nDNA is increasingly recognized as being relevant to speciation (Burton and Barreto 2012; Burton et al. 2013; Hill 2017).

While secondary contact between differentiated populations sometimes leads to narrow hybrid zones (Barton and Hewitt 1989), another possible outcome is the formation of a hybrid or mixed population over a broad region (Rieseberg 1997; Schumer et al. 2016; Elgvin et al. 2017). Strong selection can be revealed in these sustained ancient hybrid populations when gene flow is no longer prevalent, while the suboptimal mito-nuclear combinations can still arise at every generation. Despite the increasing interest in mito-nuclear interactions at species boundaries of natural populations with complex population histories (Sambatti et al. 2008; Gagnaire et al. 2012; Bar-Yaacov et al. 2015; Boratyński et al. 2016; Baris et al. 2017; Morales et al. 2018),

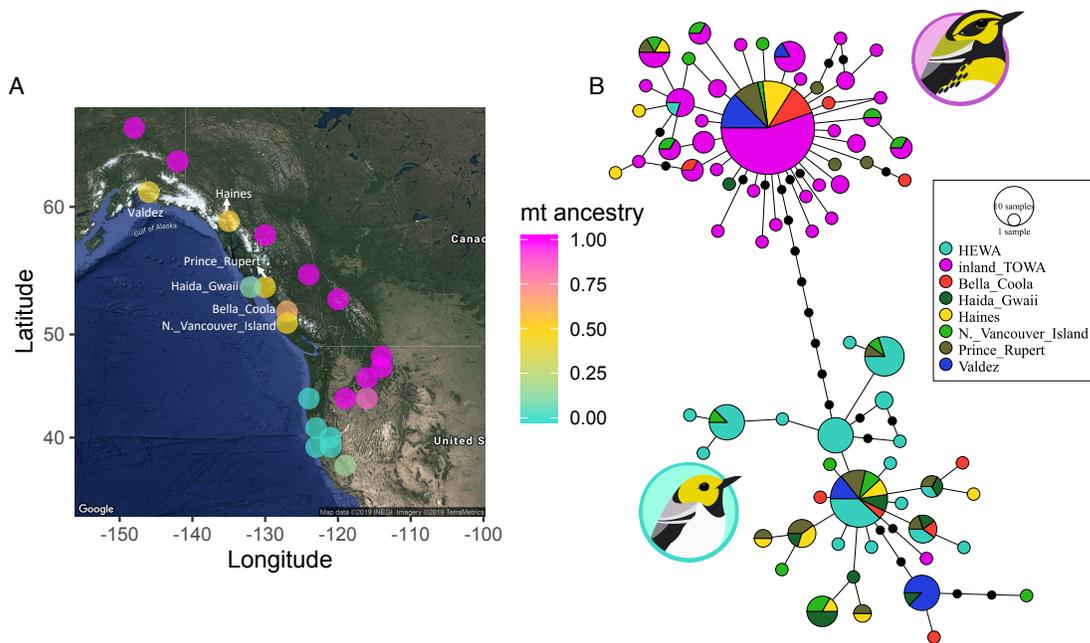
whether mito-nuclear interactions are important early on in the speciation process is not well understood.

Here we examine the relationship between mtDNA and nDNA variation in a warbler species complex with a complex history of hybridization that involves ancient and ongoing hybridizing populations. In particular, we tested whether nuclear regions of high genetic differentiation between the ancient hybrid population and either of the parental populations are related to mitochondrial physiology. We also ask whether there is mito-nuclear genotype association within- and among-populations and whether mito-nuclear genotypes are associated with climatic variation among populations.

Townsend's warblers *Setophaga townsendi* (referred to as *townsendi*), which inhabit coastal and inland regions of northwestern North America, and Hermit warblers *S. occidentalis* (referred to as *occidentalis*), which inhabit coastal regions south of *townsendi*, diverged around 0.5 million years ago (Weir and Schluter 2004). They now hybridize in several regions within Washington and Oregon, USA (Rohwer and Wood 1998). The nDNA differentiation between inland *townsendi* and *occidentalis* is pronounced at a few small regions, whereas the rest of the genome shows very little differentiation (chapter 3). The inland (interior British Columbia, Washington, Idaho, Montana) *townsendi* population contains an mtDNA haplotype cluster that is distinctive from that of *occidentalis* (Krosby and Rohwer 2009). Such distinct mtDNA ancestry could be a response to divergent selection to habitats of different climatic conditions, given the climatic differences between coastal and inland regions of western North America. The differentiation between *occidentalis* and inland *townsendi* is reminiscent of the coastal-inland climate-related mtDNA differentiation in Australian yellow robins (Morales et al. 2017). In contrast, the coastal *townsendi* population harbors both of those mtDNA haplotype groups

(Figure 4.1), suggesting this population is the product of ancient hybridization between *occidentalis* and inland *townsendi* (Krosby and Rohwer 2009). The *occidentalis* mtDNA haplotype cluster is structured suggesting differentiation (potentially due to isolation in different refugia) within *occidentalis*, whereas the inland *townsendi* cluster is more star-like consistent with post glacial rapid population expansion (Figure 4.1B) (Krosby and Rohwer 2009). Even though there is ~0.8% sequence divergence between the inland *townsendi* and *occidentalis* clusters, both of the haplotypes occur within sites of the coastal *townsendi* populations. Altogether, this mtDNA haplotype distribution (Figure 4.1) suggests that coastal *townsendi* populations are hybrid populations of inland *townsendi* and *occidentalis* that formed when the structured *occidentalis* populations (that were along coastal Canada and Alaska) hybridized with the *townsendi* that expanded from the interior when the glaciers receded (Krosby and Rohwer 2009) (Figure S4.1).

Figure 4.1 Geographical distribution of mtDNA ancestry with coastal *townsendi* (banding code: “TOWA”) site names labeled (**A**) and haplotype network (**B**) of the mtDNA sequences from Krosby and Rohwer (2009) study. **A**, 0 and 1 respectively represents pure *occidentalis* (turquoise, lower cluster in B) and inland *townsendi* mtDNA ancestry (magenta, upper cluster in B). **B**, each circle represents a haplotype and sizes of the circles are proportional to the number of individuals carrying each haplotype. The lines (regardless of their lengths) between the circles represent one mutation between haplotypes, the black dots on the lines represent additional mutations among haplotypes. The coastal *townsendi* (Bella Coola: orange, Haida Gwaii: dark green; Haines: yellow; North Vancouver Island: light green; Prince Rupert: brown; Valdez: royal blue) populations harbor admixed mtDNA haplotype (some mtDNA haplotypes nested within the turquoise *occidentalis* (banding code: “HEWA”) cluster whereas some are in the magenta colored inland *townsendi* cluster).



Krosby and Rohwer (2009) proposed that the initial contact zone between *occidentalis* and *townsendi* was far north of their current hybrid zone in Washington, and that the zone moved south to its present position, leaving a wake of *occidentalis* mtDNA behind (Krosby and Rohwer 2009). However, recent tracking of genomic and plumage clines in the Cascade hybrid zone over several decades did not reveal ongoing movement (chapter 2), casting doubts on a history of rapid and long-distance southward movement of the hybrid zone. An alternative hypothesis is that inland *townsendi* came into contact with *occidentalis* along a broad inland-to-coastal front parallel to the British Columbia and Alaska coast, resulting in hybrid zone movement just a short distance from the coast mountains toward the coastline and generating a pattern of coastal populations with admixed mtDNA ancestry while plumage and the bulk of nuclear DNA resemble those of inland *townsendi* (Figure S4.1).

To date, the taxonomy of this species group and hypotheses regarding its biogeographic history are based largely on phenotypic appearance, supplemented by limited genetic data from just a few independent parts of the genome (especially mtDNA, which is maternally inherited). Here, we survey variation at tens of thousands of single nucleotide polymorphisms (SNPs) throughout the nuclear genome, inferring levels of genetic relatedness among sampling regions, how that relatedness is structured across the genome, and whether there are certain SNPs that are highly associated with geographic variation in mtDNA. In particular, we ask 1) whether the nuclear genomic data is consistent with the mtDNA inference that coastal *townsendi* resulted from admixture, producing population sub-structure within coastal *townsendi*; 2) which genomic regions differ between coastal and inland *townsendi* and whether these regions are known to be related to mtDNA function; 3) whether the ancestries at these divergent nuclear regions are associated with mtDNA ancestries within and among sites; and 4) whether the spatial

distribution of mito-nuclear genotypes is associated with climatic variation. In addition, we asked which genomic regions differ between birds with *occidentalis* plumage and those with *townsendi* plumage. In particular, do the regions that were found to be associated with plumage in the *occidentalis* and *townsendi* hybrid zone (e.g. the RALY SNP) (chapter 3) also show an association with plumage over a broad geographic scale? The genomic architecture of differentiation in this species complex with ancient and ongoing hybridization could reveal prevalent evolutionary forces maintaining population/species boundaries in the face of gene flow.

4.2 Methods

4.2.1 Museum samples, mtDNA sequences, and nDNA sequencing

Sequences of the mtDNA NADH dehydrogenase subunit 2 gene (ND2) for 223 individuals (95 coastal *townsendi*, 81 inland *townsendi*, and 47 *occidentalis*) from the Krosby and Rohwer (2009) study were acquired from GenBank (accession numbers FJ373895-FJ374120), and haplotyped as 0 for the *occidentalis* haplotype cluster and 1 for the *townsendi* cluster (Krosby and Rohwer 2009) (Figure 4.1). To understand the relationships among the mt sequences, we generated a minimum spanning haplotype network (Bandelt et al. 1999) with PopART (Leigh and Bryant 2015).

Among these individuals with previously-sequenced mtDNA (i.e., from Krosby and Rohwer 2009), we selected a subset of tissue samples (64 inland *townsendi*, 58 coastal *townsendi*, and 15 *occidentalis*; obtained from the Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington) for nuclear genomic sequencing. We supplemented this set of genetic samples with 30 blood samples that we obtained directly from

birds caught in the field during the breeding season of 2016 (see chapter 3 for details); these included 25 *occidentalis* from California, USA, and 5 inland *townsendi* from Montana, USA. After sequencing, we removed two of the *occidentalis* samples because of insufficient read depth and a labeling error respectively, thus 23 *occidentalis* remained for further analysis.

4.2.2 GBS pipeline

Following Alcaide et al. (2014), we prepared genotyping-by-sequencing (GBS) (Elshire et al. 2011) libraries from DNA samples of the 165 individuals described above. Briefly, we digested genomes with the restriction enzyme PstI, then ligated fragments with barcode and adaptors, and amplified with PCR. Amplified DNA was then pooled into two libraries which were then paired-end sequenced: the first (80 individuals) was sequenced with an Illumina HiSeq 2500 automated sequencer (read length = 125bp), and the second (85 individuals) was sequenced with an Illumina HiSeq 4000 (read length = 100bp). To control for plate effects, we randomly assigned samples to different plates and included replicates of three samples among plates. We then conducted principal component analysis to visually evaluate whether there is any plate effect, in which the duplicated samples were checked and then removed from further analysis. The sequences were processed as in Irwin et al. (2016) (scripts provided at the Dryad Digital Repository, doi:10.5061/dryad.t951d). Briefly, the reads were demultiplexed with a custom script and then trimmed using Trimmomatic (Bolger et al 2014) [TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30]. Assuming synteny between *Setophaga* and *Taeniopygia guttata* (Zebra Finch) genomes given the limited evidence of limited rearrangement in avian genomes (Ellegren 2010; Zhang et al. 2014), we aligned reads to a *T. guttata* reference (Warren et al. 2010) with bwa-mem (Li 2010) (default settings). Variable sites were identified with GATK (McKenna et al. 2010), which resulted in 3,446,846 variable sites among the 165

individuals in the study. We then filtered the variable sites with VCFtools (Danecek et al. 2011) according to the following criteria: 1) removing indels, 2) keeping sites with genotype quality (GQ) > 20, 3) keeping sites with minor allele frequency (MAF) ≥ 0.05 , 4) removing sites with > 30% missing genotypes among individuals, and 5) keeping biallelic single nucleotide polymorphisms (SNPs) only. Thereafter 19,083 SNPs remained.

4.2.3 Population structure and genomic differentiation

Population structure was examined with principle component analysis (PCA) in the SNPRelate (Zheng et al. 2012) package in R and ancestry assignments in Faststructure with a uniform prior, 10^8 iterations, and K values from 1 to 6 (Raj et al. 2014). We focused initially on assessing the differences between *occidentalis*, coastal *townsendi*, and inland *townsendi*. We then noticed that three obvious nuclear genetic clusters were found within our samples of coastal *townsendi*: Valdez, AK (USA), Haida Gwaii, BC (Canada), and the rest of the coastal *townsendi* populations. We further compared these three coastal *townsendi* populations to the inland *townsendi* and *occidentalis* groups. We used the SNPRelate (Zheng et al. 2012) package in R to examine which SNPs were highly correlated with principal component axes.

To quantify population differentiation across the genome, for each of the 19,083 filtered SNPs we calculated F_{ST} (Weir and Cockerham 1984) with VCFtools (Danecek et al. 2011) between 1) inland *townsendi* ($N = 69$) and *occidentalis* ($N = 38$); 2) coastal *townsendi* ($N = 58$: 10 Haida Gwaii, 15 Valdez, 33 others) and inland *townsendi*; and 3) *occidentalis* and each of the three coastal *townsendi* clusters.

4.2.4 Candidate genetic regions

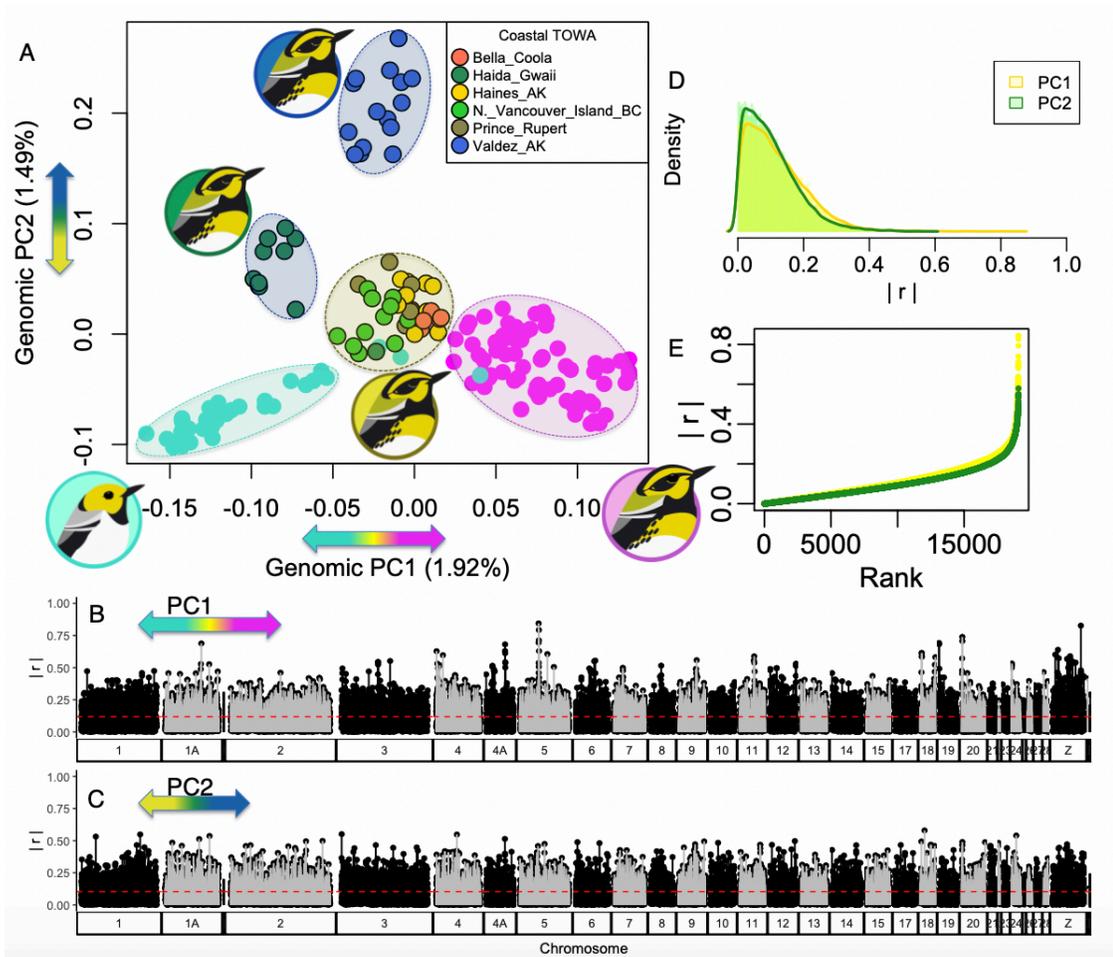
The SNPs at F_{ST} peaks between inland and coastal *townsendi* that are also consistent with the peaks between *occidentalis* and inland *townsendi* were considered candidate loci that may be influenced by selection associated with coastal versus inland populations. One possibility is that these loci are linked to genes that have a mitochondrial function, such that selection shows geographic variation similar to the pattern of mtDNA variation. To examine whether these loci are known to be associated with mitochondrial function, we examined what is known about the protein-coding genes in vicinity of the candidate SNPs, using Ensembl (Hunt et al. 2018) and the zebra finch reference genome. If a large region of elevated F_{ST} is involved, Zebrafinch Gene Ontology analysis (Wu and Watson 2009) was conducted to test regional functional enrichment relative to the rest of the genome. While *occidentalis* and inland *townsendi* differ at the RALY locus that is associated with plumage difference (chapter 3), we did not expect this region to differ between coastal and inland *townsendi* due to their identical plumage features.

4.2.5 mt-nDNA association

If there is selection for mt-nDNA coadaptation, there should be ancestry association between mtDNA and nDNA within and among sites. If individuals with mismatched mt-nDNA ancestries are selected against, there should be a significant association between mtDNA and nDNA ancestries within each population. Such a force is potentially counteracted by random mating at every generation, which breaks down the mt-nDNA combinations, thus strong selection would be inferred to maintain adaptive mt-nDNA combinations within a single randomly mating population. Over time however, specific geographic regions may tend to favor a particular mtDNA ancestry and compatible nDNA ancestry, increasing mt-nDNA concordance among sampling sites. To examine within-population association between mtDNA and nDNA ancestry, we conducted permutation test of independence with the *coin* package (Hothorn et al.

2008) in R to examine if there is association between mtDNA group (0 or 1) and nuclear candidate SNP genotype (0 as homozygous *occidentalis*, 0.5 as heterozygous, or 1 as homozygous *townsendi*) within Valdez and the northern Vancouver Island populations (coastal *townsendi* sites with $N \geq 10$). For example, the Haida Gwaii population is almost fixed for *occidentalis* mt haplotype (Figure 4.1), thus we expect the nuclear genotype to be predominantly *occidentalis* in this population, if there is mt-nDNA coadaptation. To examine between-sites mito-nuclear ancestry, we employed a partial mantel test (Legendre 1998) with the *vegan* package in R to quantify the association between the distance matrices of mtDNA and chr 5 (or chr Z maker respectively) controlling for admixture of *occidentalis* and inland *townsendi* ancestry. In particular, the partial mantel test examined correlation between the distance matrix of mtDNA and nDNA conditioned on the distance matrix of admixture index by permuting the distance matrix of the mtDNA. The admixture index is represented by PC1 of the genomic PCA with the candidate SNPs (all the SNPs in the 700kb differentiation block on chr5 and the SNP at the peak on chrZ) removed. We employed the genomic PC1 instead of a model-based admixture proportion because the complex admixture history between the parental populations of the coastal *townsendi* and the model-based approaches (Raj et al. 2014) tend to force admixed individuals into either of the parental clusters, and the output admixture indices for each individual largely depend on the prior distribution (see Chapter 2). In contrast, genomic PC1 naturally represents admixture between inland *townsendi* and *occidentalis* (Figure 4.2).

Figure 4.2 Principle component analysis of 19083 high quality SNPs in the genome. **A**, the coastal *townsendi* is intermediate in PC1 but distinct from inland *townsendi* and *occidentalis* in PC2. PC1 represents admixture between *townsendi* and *occidentalis*, and PC2 represents unique differentiation of coastal *townsendi*. **B**, **C**, Absolute correlation coefficient between SNPs and PC1 (**B**) and PC2 (**C**). The horizontal red dash clines are the mean. **B**, certain regions in chromosome 1A, 5, 18, 20, and Z are highly correlated with PC1. **C**, SNPs are correlated with PC2 similarly across the genome (no obvious peak). **D**, density distribution of absolute correlation coefficient of each SNP with PC1 (yellow shade) and PC2 (green shade). **E**, there were more SNPs that are high-correlated with PC1 (yellow line) than PC2 (green line).



4.2.6 Climate analysis

To investigate whether there might be selection on mt-nDNA related to climate, we tested true association of site-level mt-nDNA ancestry (the averaged site ancestry score of mtDNA, chr 5, and chr Z marker ancestry) and climate variation. To effectively capture the climate variation among sites, we extracted annual climate data from ClimateWNA (Wang et al. 2012) for all the sites, which includes 26 climate variables, and conducted PCA. To test association between key climate variation and mt-nDNA ancestry among sites while accounting for spatial autocorrelation, we performed a partial mantel test. Specifically, we first computed site-based distance matrices of the site coordinates, climate PC1, climate PC2, as well as mt-nDNA ancestry respectively. Then we tested whether there is a correlation between the mt-nDNA distance matrix and the distance matrix of climate PC1 (and PC2 respectively) by permutating the distance matrix of mito-nuclear ancestry, while controlling for site location (represented by the distance matrix of site coordinates).

4.3 Results

4.3.1 Population structure

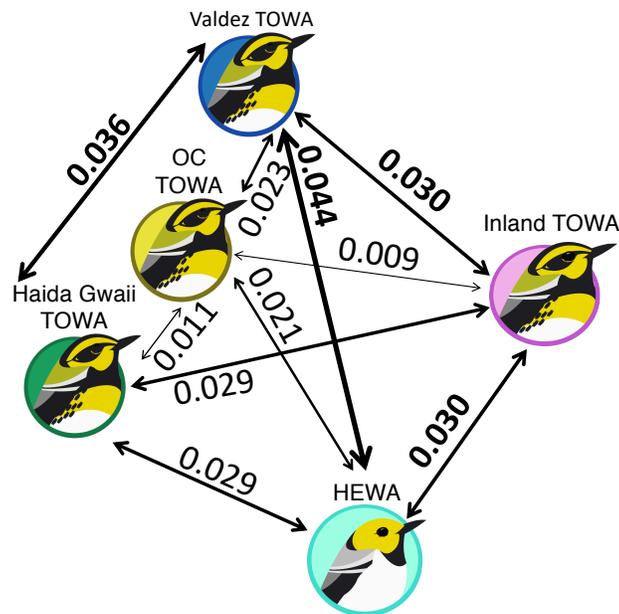
The mtDNA haplotypes are distinct between inland *townsendi* and *occidentalis*, with 0.8% sequence divergence (Krosby and Rohwer 2009; Figure 4.1). Within various coastal *townsendi* populations, there are individuals with both *townsendi* haplotypes and *occidentalis* haplotypes (Figure 4.1), suggesting that these coastal *townsendi* populations are hybrid populations between *townsendi* and *occidentalis* (Krosby and Rohwer 2009). Nuclear genomic variation as assessed through variation in the 19,083 SNPs reveals a pattern broadly consistent with the variation in mtDNA. Inland *townsendi* and *occidentalis* form two clearly differentiated clusters differing in the first principal component (PC1) of a PCA (Figure 4.2A), and most individuals from coastal *townsendi* have a somewhat intermediate position. Two coastal *townsendi* sampling regions, Valdez and Haida Gwaii, are noticeably differentiated from the rest, forming distinct clusters (Figure 4.2A). Valdez differs primarily along the second principal component (PC2), whereas Haida Gwaii differs by a combination of PC1 and PC2. While PC1 is highly correlated with a few strong outlier SNPs, PC2 shows only modest correlations with particular SNPs (Figure 4.2B).

4.3.2 F_{ST} distribution

Genome-wide levels of differentiation show that inland *townsendi* and *occidentalis* are very similar (Weir and Cockerham's $F_{ST} = 0.030$) except for a few peaks of differentiation (chapter 2) (Figure 4.3, 4.4A). As in the PCA, F_{ST} analysis indicates differentiation of the Valdez and Haida Gwaii *townsendi* populations (F_{ST} between Valdez and *occidentalis*: 0.044; Valdez and inland *townsendi*: 0.030; Haida Gwaii and *occidentalis*: 0.029; Haida Gwaii and inland

townsendi: 0.029). The rest of the coastal *townsendi* are more similar to inland *townsendi* (0.009) than to *occidentalis* (0.021).

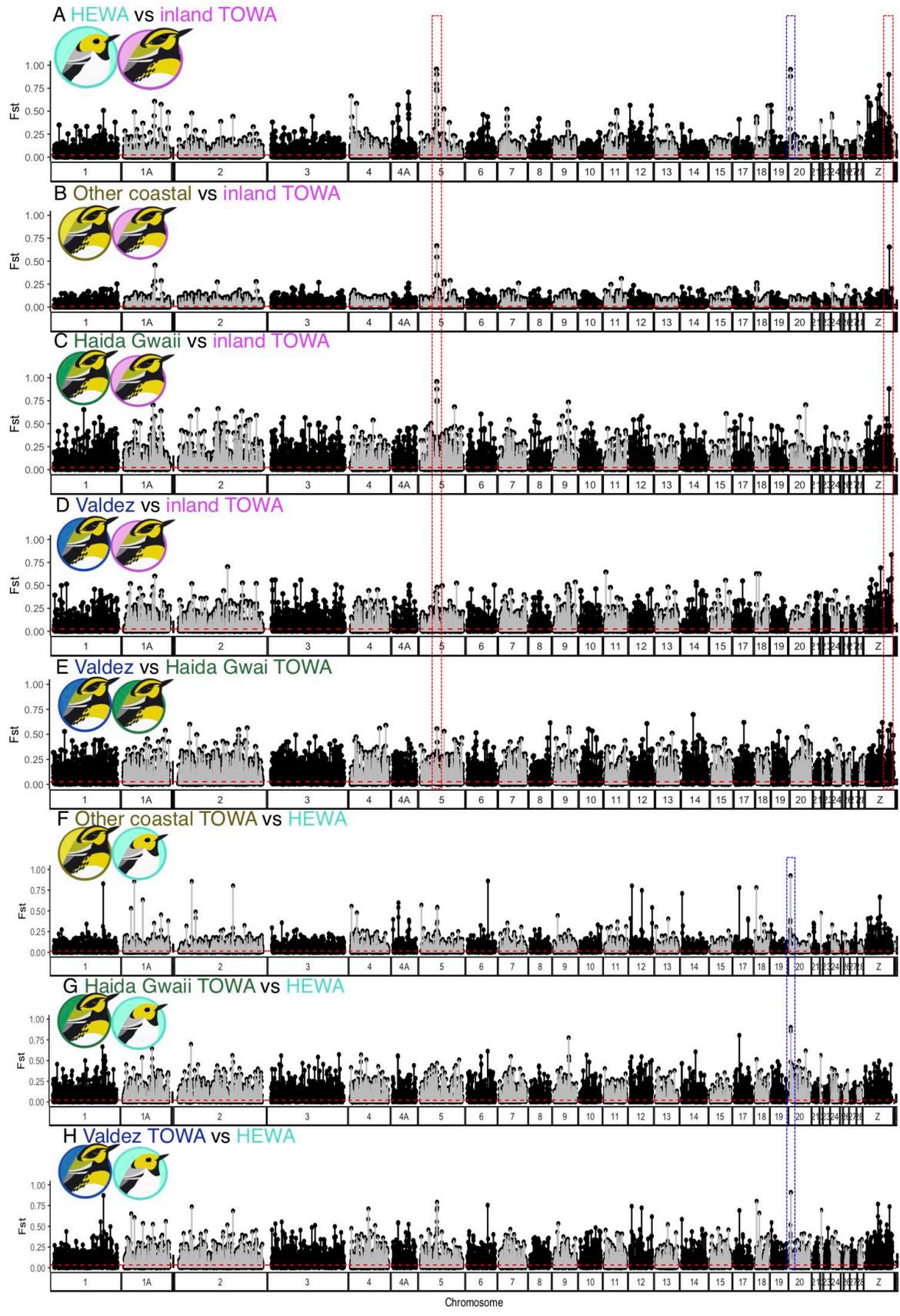
Figure 4.3 The Weir & Cockerham weighted F_{ST} among *occidentalis*, inland and coastal *townsendi* (green: Haida Gwaii, blue: Valdez, dark yellow: other coastal (OC) *townsendi*), which demonstrates a gradient of differentiation from the parental populations (inland *townsendi*, abbreviated as inland TOWA, colored as magenta; *occidentalis*, abbreviated as HEWA, color in turquoise). Each double-head arrow represents a pairwise comparison among the populations. The populations are oriented as their geographical location. The widths of the arrows are weighted by the F_{ST} between each pair of populations. Surprisingly some coastal *townsendi* populations demonstrates significantly greater differentiation (paired t-test, $p < 0.001$) from the parental populations than between the parental populations ($F_{ST} = 0.03$).



Inland *townsendi* and *occidentalis* have a number of peaks of differentiation, with the three highest standing out in particular (Figure 4.4A) and mapping to chromosomes (chr) 5, 20,

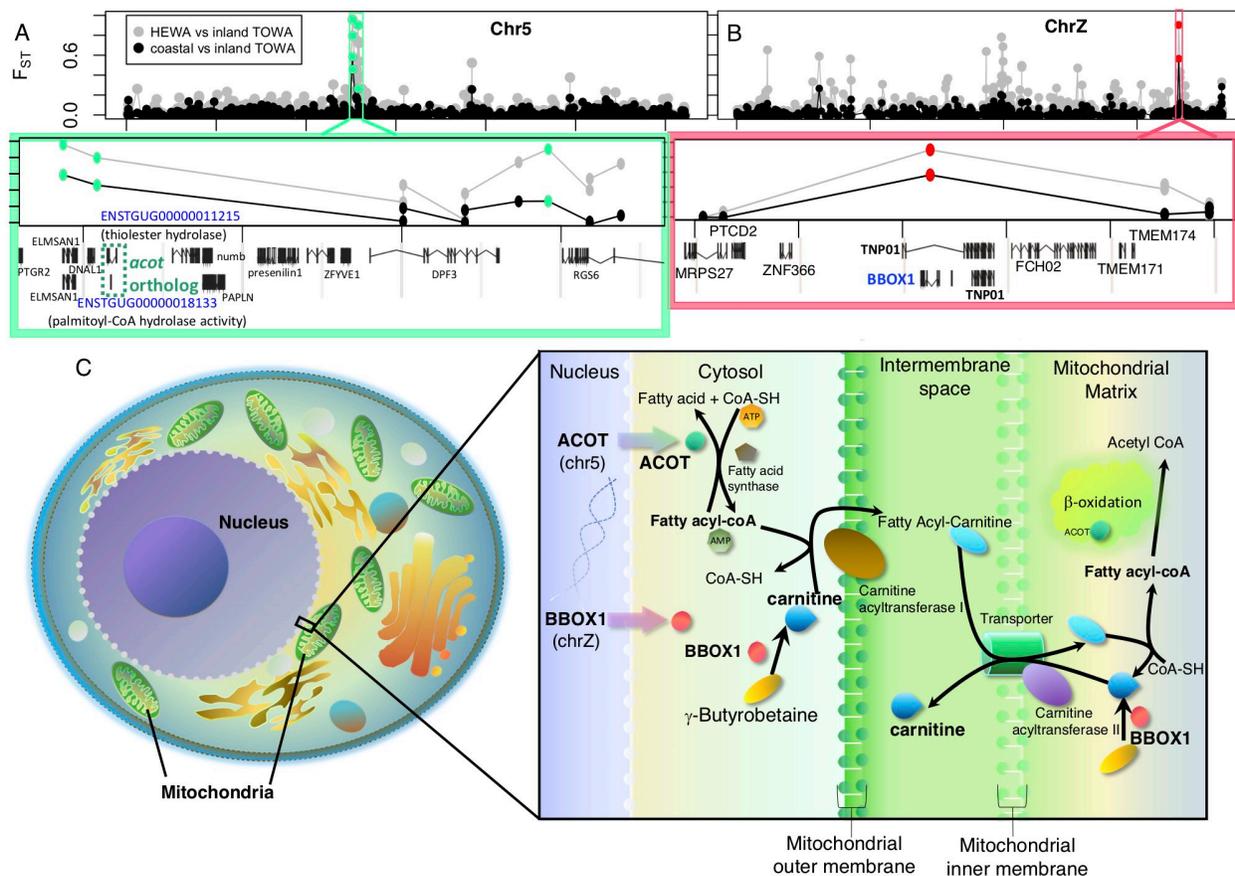
and Z, and a moderate one on chr1A in the *T. guttata* reference. One of these (on chr 20) is in the intron region of the RALY gene (chapter 3), which is known to regulate pigmentation in quail and mice (Michaud et al. 1994; Nadeau et al. 2008). Our earlier study of admixture mapping in the ongoing hybrid zone between inland *townsendi* and *occidentalis* in Washington Cascades (chapter 3) revealed that this locus is highly associated with plumage colour patterns within that zone. As predicted by that finding, the present survey of genomic variation over a much broader geographic region shows high differentiation at the RALY SNP between sampling regions that differ in plumage (i.e., between *occidentalis* and *townsendi*, Figure 4.4A, F-H) and low differentiation between regions with similar plumage (i.e., between coastal and inland *townsendi*, Figure 4.4B-D).

Figure 4.4 F_{ST} scan between *occidentalis* and inland *townsendi* (A), other coastal and inland *townsendi* (B-D), between Valdez and Haida Gwaii *townsendi* (E), as well as *occidentalis* and coastal *townsendi* (F-G). Three distinctive differentiation peaks were found between inland and coastal *townsendi* that reside in chromosome 1A, 5 and Z (red boxes, A-C). The RALY locus demonstrates consistent differentiation between *occidentalis* and various *townsendi* (blue boxes, F-G). The red horizontal dotted lines represent the genome-wide mean F_{ST} .



Similar to the chr 20 RALY peak, the chr5 and chrZ regions also showed extreme differentiation in the comparison of inland *townsendi* and *occidentalis* (on chr 5 and Z), but opposite to RALY region, these regions stand out in the comparison between coastal and inland *townsendi* as the two highest regions of differentiation between those groups (Figure 4.4, A-B). The chr 5 differentiation (Figure 4.5, A) involves a ~ 700kb region that is significantly enriched for lipid metabolism ($p = 0.0013$, $p_{adjusted} = 0.021$) related to mitochondrial function with particular relevance to acyl-CoA metabolic process ($p = 0.0027$, $p_{adjusted} = 0.021$), thiolester hydrolase ($p = 0.002$, $p_{adjusted} = 0.021$), and palmitoyl-coA hydrolase activity ($p = 3.7 \times 10^{-6}$, $p_{adjusted} = 0.0001$), due to the genes ENSTGUG00000011215 and ENSTGUG00000018133 (orthologs of ACOT, acyl-CoA thioesterase). The chr Z SNP (position 66226657 in the *T. guttata* reference) is within the intron of the BBOX1 gene (gamma-butyrobetaine hydroxylase 1) (Figure 4.5B), which codes for a biosynthesis enzyme of carnitine. Carnitine is the central player in the carnitine shuttle of mitochondria, which activates and transports fatty acid into mitochondria for beta-oxidation (Figure 4.5C) (Calvo and Mootha 2010; Tars et al. 2014; Longo et al. 2016). The other gene associated with this chrZ differentiation is a cytoplasmic-related gene TNP01 that encodes nuclear-cytoplasmic signaling protein, transportin1 (Brelstaff et al. 2011) (Figure 4.5B). This chrZ region of differentiation could be narrow, as SNPs flanking this chrZ peak do not demonstrate high F_{ST} (Figure 4.5B). The chr 5 and Z regions are functionally linked to each other as well through the carnitine shuttle of mitochondria (Figure 4.5C). An additional peak was found in the F_{ST} scan between the inland versus other coastal *townsendi* at chr1A (54442413) (Figure 4.4B), which is in the inter-genic region between golgi gene CHST11 (Carbohydrate Sulfotransferase 11) and the cytoplasmic-functioning gene TXNRD1 (Thioredoxin Reductase 1) (Figure S4.3).

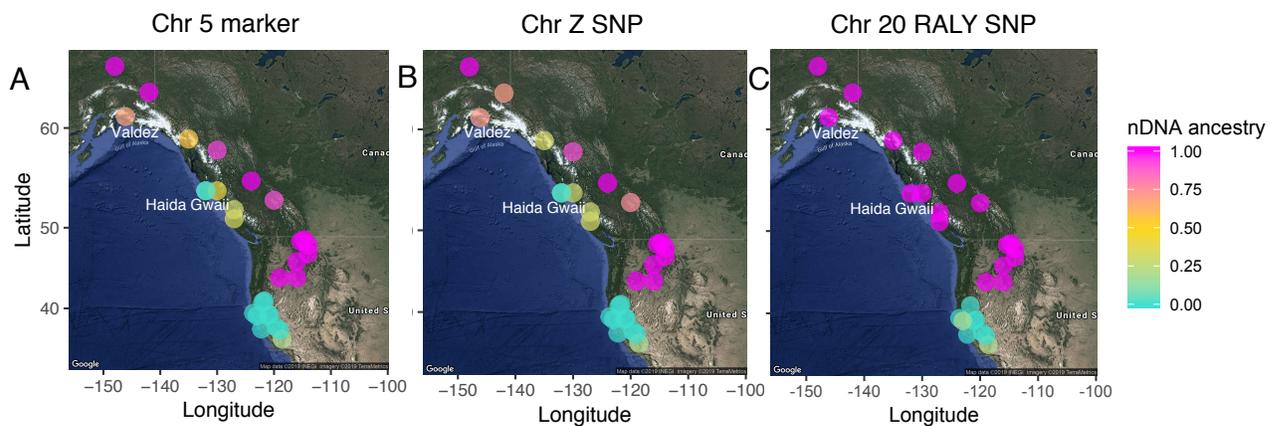
Figure 4.5 Coastal *townsendi* (black) and *occidentalis* (grey) exhibit concordant genetic differentiation from inland *townsendi* at regions in chr 5 (**A**) and Z (**B**) that are associated with genes involved in mitochondrial fatty acid metabolism (**C**). **A-B**, F_{ST} scan on chr 5 (**A**) and Z (**B**) with the zoom-in views around the F_{ST} peaks. There are vertical blue lines every 10,000 bases. zooming in around BBOX1 gene on the Z-chromosome. **A**, the region of differentiation delineated by the jade green F_{ST} peaks are significantly enriched for acyl-CoA metabolism, because of the two orthologs of ACOT (sided the jade green box). **B**, the violet red F_{ST} peak is localized at the Z-chromosome within the intron of gene BBOX1 (involved in fatty acid transportation across mitochondria membranes) and a cyto-nuclear signaling gene TNP01. **C**, Illustration of the mitochondrial carnitine shuttle in which the nuclear genes associated with chr 5 (ACOT) and Z (BBOX1) differentiation were bolded. BBOX1 synthesizes carnitine (bolded), which is essential to transport fatty acyl-coA (bolded) into mitochondrial matrix for beta-oxidation. If not transported into mitochondria, the fatty acyl-coA can be converted back to fatty acid catalyzed by ACOT. This illustration is a synthesized existing illustrations about carnitine shuttle (Mehta 2013; Beaudet 2017).



If there is strong selection maintaining mt-nDNA coadaptation, then within each admixed coastal *townsendi* population, individual mtDNA (Figure 4.1 A) and nDNA ancestries (Figure 4.6 A, B) should be correlated. Among populations, the population-level mtDNA and nDNA ancestries should be correlated as well. In Valdez, there was a significant association of the mtDNA ancestry and chr 5 marker ancestry ($Z = 2.44$, $p = 0.015$, $p_{(FDR-corrected)} = 0.03$; Figure S4.4A), but only a marginal association with Z marker ancestries ($Z = 2.14$, $p = 0.033$, $p_{(FDR-corrected)} = 0.065$; Figure S4.4, B). In the North Vancouver Island population, neither the chr 5 ($Z = -1.41$, $p = 0.157$) nor the chr Z ($Z = -0.57$, $p = 0.5723$) marker was significantly associated with mtDNA ancestry (Figure S4.4, A-B). Consistent with the mt-nDNA coadaptation prediction, the

townsendi homozygotes for the chr5 and Z markers are missing in the Haida Gwaii *townsendi* population (Figure 4.6), in which the *occidentalis* mt haplotypes are almost fixed (Figure 4.1, A). Among sites, both the chr5 (partial mantel pearson's product-moment $r = 0.736$, $p < 10^{-4}$) and chrZ marker (partial mantel pearson's product-moment $r = 0.270$, $p = 0.03$) were correlated with the mtDNA ancestry after controlling for the effect of admixture represented by the distance matrix of genomic PC1 (see method).

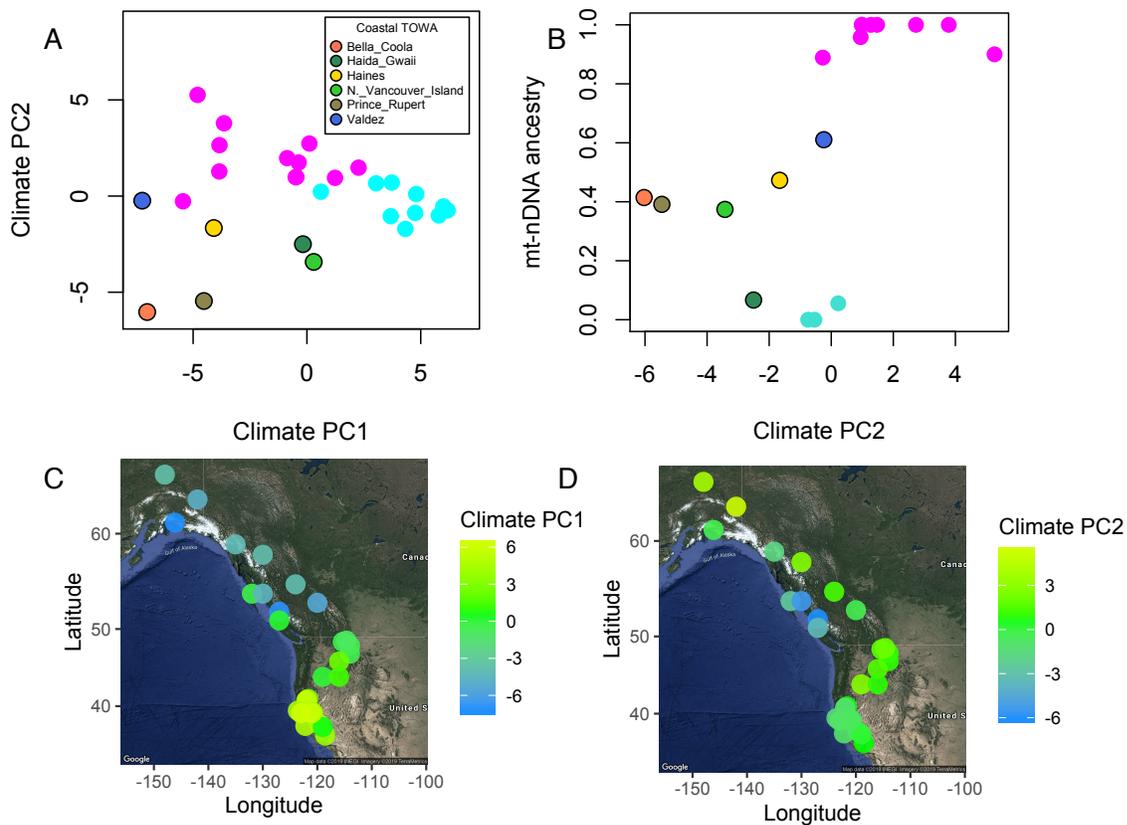
Figure 4.6 Spatial distribution of ancestry proportion at mitochondrial fatty acid metabolic markers at chr 5 (A) and Z (B), and chr20 RALY SNP (C).



Climate PC1 (Figure 4.7A, S4.4, S4.5) explains 64.4% of the variation in climate among sites; this PC was not particularly explained by one or a few climate variables (Figure S4.5AB). Climate PC2 explains 23.5% of the variation and was predominantly explained by four climate variables (Figure S4.5, C): Temperature Difference (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP), Winter Precipitation (PPT_wt). The climate in coastal *townsendi* habitat is similar to inland *townsendi* habitat along PC1, but more similar to *occidentalis* habitat along PC2 (Figure 4.7), although there is great climate variation among various coastal

townsendi populations. Overall, the coastal *townsendi* habitat is more moist and stable in temperature, which is consistent with the coastal-inland humidity gradient (captured by PC2), and the distribution of mt-nDNA ancestry is related to this geographical variation in climate. The mt-nDNA ancestry is significantly correlated with climate PC1 (partial mantel test, $r = 0.194$, $p = 0.040$) as well as climate PC2 (partial mantel test, $r = 0.221$, $p = 0.025$).

Figure 4.7 Climate principal component analysis of 26 climate variables from ClimateWNA. A, climate PC1 and PC2, in which *occidentalis* (turquoise), inland (magenta) and coastal *townsendi* habitats are different. B, site mean mtDNA ancestry is associated with local climate PC2. C-D, spatial distribution of climate PC1 (C) and PC2 (D).



4.4 Discussion

Inland *townsendi* and *occidentalis* are distinct in mtDNA (Krosby and Rohwer 2009) and exhibit three strong regions of differentiation in the nuclear genome (chapter 3). The coastal *townsendi* harbor admixed mtDNA and nDNA ancestry from *occidentalis* and inland *townsendi*. Two of the three regions of strong differentiation between *occidentalis* and inland *townsendi*, on chr 5 and Z, differentiate coastal and inland *townsendi* as well. Both of these nDNA regions contain genes that are strong candidates for coadaptation with mtDNA, as they are both involved in the mitochondrial carnitine shuttle for fatty acid metabolism. Coadaptation was further supported by mt-nDNA association within Valdez population and among populations. This potential mito-nuclear coadaptation is likely associated with climatic adaptation, or with habitat difference predisposed by climate difference, because the site-level mito-nuclear ancestry covaries with the site climate conditions. Altogether, the fatty-acid metabolic mito-nuclear combinations are likely divergently-selected, potentially in response to different climate conditions.

4.4.1 Biogeography and semi-parallel introgression

Our genomic evidence is consistent with Krosby and Rohwer's conclusion, based on mtDNA, that coastal British Columbia and Alaska was inhabited by geographically structured *occidentalis* populations before *townsendi* expanded from inland areas and mixed with them (Krosby and Rohwer 2009). The *occidentalis* and inland *townsendi* mtDNA haplotypes demonstrate many differences (~0.8%) while within a coastal *townsendi* site, both haplotypes can be found. It is unlikely that the polymorphisms in mtDNA and nDNA in the coastal *townsendi* were caused by incomplete lineage sorting, as opposed to hybridization (Figure S4.1). If such pattern is due to incomplete lineage sorting, it implies that over many generations, both

inland *townsendi* and *occidentalis* lost the alternative haplotype, while the coastal *townsendi* maintained both. If so, over such time course, more differentiation is expected between the *occidentalis* haplotypes found in *occidentalis* versus coastal *townsendi*, as well as between inland *townsendi* haplotypes found in coastal versus inland *townsendi*. However, we found no support for such a prediction, since within coastal *townsendi*, some mtDNA haplotypes are identical to *occidentalis* and while some are identical to inland *townsendi* haplotypes.

The higher genome-wide differentiation of the Haida Gwaii and Valdez populations (Figure 4.2-5) is consistent with at least partially isolated cryptic refugia of *occidentalis* in coastal Alaska and Haida Gwaii during the last glacial maximum (LGM) (Shafer et al. 2010). Following expansion of *townsendi* from the inland area, presumably after the last glacial period, hybridization between *townsendi* and *occidentalis* apparently led to populations of mixed ancestry along the coast of BC and Alaska (Figure S4.1). These coastal populations have the plumage patterns and colors of *townsendi*, which is why they have been classified as members of that species. This uniform *townsendi* appearance has concealed a more complex history of hybridization with ancient and geographically differentiated populations of *occidentalis*.

Following expansion of *townsendi* from the interior, gene flow into Haida Gwaii may have been weak due to the expanse of water separating it from the mainland, explaining why that population is more similar to *occidentalis* than other coastal *townsendi* are. Gene flow into Valdez could have also been impeded by geographical barriers, as Valdez is surrounded by mountain ranges (Chugach mountains, Wrangell mountains, and St. Elias mountains). However, both nuclear and genomic data indicate that Valdez has substantial ancestry from both *townsendi* and *occidentalis*. Despite genome-wide differentiation among these three coastal *townsendi* genetic clusters, there is an interesting parallelism: all the three populations exhibit the inland

townsendi-like RALY marker that is associated with plumage (chapter 3), and predominantly *occidentalis*-like mitochondria-related markers. Such parallelism might be driven by parallel adaptation to the coastal climate after the introgression over the extensive contact zone along the west coast of Canada after the westward expansion of *townsendi*.

4.4.2 mt-nDNA association

We found the key coastal-inland *townsendi* nDNA difference resides at a chr 5 and chr Z associated with mitochondrial fatty acid metabolism (Figure 4.5), an intriguing result given that coastal and inland *townsendi* differ so strongly in their mitochondrial haplotype frequencies. This finding points to the possibility of selection based on mito-nuclear interactions (Burton and Barreto 2012; Hill 2017). The BBOX1 gene encodes Gamma-butyrobetaine dioxygenase (Vaz et al. 1998), the enzyme that catalyses L-carnitine synthesis (Paul et al. 1992), which is critical for transporting fatty acids across mitochondrial membranes during beta oxidation (Longo et al. 2016). Carnitine co-functions with mtDNA and the chr5 region that is enriched for mitochondrial fatty acid metabolism and form a ‘carnitine shuttle’ (Figure 4.5C). The *occidentalis* nDNA may be partially incompatible with the *townsendi* mtDNA in jointly forming the functional carnitine shuttle leading to selection against mismatched mito-nuclear ancestries. Such selection maintaining mito-nuclear concordance can be counteracted by random mating in admixed populations at each generation and is thus difficult to detect in samples of individuals from a single population. However, mito-nuclear ancestry concordance can be more easily detected through comparison of many populations. The site-level mito-nuclear ancestry concordance (Figure 4.6) reveals the potential selection maintaining functionally compatible mito-nuclear ‘carnitine shuttle’ over a large temporal and spatial scale.

The pattern of genomic differentiation between coastal and inland *townsendi* is consistent with a divergence with gene flow scenario, where the targets of selection (e.g. the differentiation peaks on chr 5 and Z) are maintained while the rest of the genome is homogenized between populations by gene flow (Wu 2001; Via 2009; Nosil and Feder 2012). However the underlying process is slightly different. Instead of gradually accumulating genomic differentiation in sympatry or parapatry, this system accumulated differentiation at allopatry when *occidentalis* and inland *townsendi* were isolated by ice sheets (Weir and Schluter 2004; Krosby and Rohwer 2009).

The source of such selection is climate-correlated. In other word, the inland-coastal *townsendi* mt-nDNA difference might be shaped in part by climate adaptation. The climate in the coastal *townsendi* is similar to inland *townsendi* habitat along PC1, but similar to *occidentalis* habitat along PC2. The mito-nuclear ancestry is significantly correlated with climate PC1 and PC2, suggesting potential selection on the mt-nDNA combinations for climate adaptation. Admittedly, correlations between any two traits that have large-scale geographic variation are expected. However, the climate differences between *occidentalis*, inland and coastal *townsendi* are very strong and suggest a strong causal relationship. This pattern is consistent with the *Eopsaltria australis* (Eastern Yellow Robin) system in which distinct mt-nDNA combinations are maintained between inland and coastal habitat (Morales et al. 2018). Fatty acid metabolic genes have also been shown to be targets of climatic adaptation in humans, within Siberian (Clemente et al. 2014) and Greenlandic Inuit populations (Fumagalli et al. 2015). Temperature (Zoladz et al. 2017) and humidity (Atkin and Macherel 2009) both influence mitochondrial fatty acid metabolism during beta oxidation, which highly depends on carnitine (Atkin and Macherel 2009; Zoladz et al. 2017). BBOX1-ACOT-mtDNA genotypes might result in functional

difference in fatty acid metabolism that is adapted to specific climate (moist and stable versus dry and variable) in the breeding habitat of these warblers. However, we can not rule out the possibility that such mitonuclear genetic combinations were selected by other habitat variables that are shaped by climatic conditions. Either directly or indirectly, climatic effect on mitonuclear adaptation should not be overlooked.

Because *occidentalis* has apparently inhabited coastal areas for a long period of time, the *occidentalis* mt-nDNA gene combination may be more suited for coastal habitats compared to those of *townsendi*. If the *occidentalis* mt-nDNA genotype is favored in the coastal habitats, the frequency of *occidentalis* mt-nDNA gene combinations would tend to increase in coastal *townsendi* populations over time. However, ongoing inland-coastal *townsendi* gene flow would slow or prevent such an increase. The Haida Gwaii island and Valdez population could have escaped from such a balance between selection and gene flow due to their isolation from the rest of the populations respectively by the sea and mountain ranges. Another possibility is that frequency-dependent selection is maintaining long-term mt-nDNA polymorphism in the coastal *townsendi*. Future investigation on the spatial and temporal variation of mtDNA-BBOX1-ACOT co-segregation would shed light on the evolutionary forces shaping the present and future of coastal *townsendi* population.

A remarkable similarity between the *townsendi* / *occidentalis* system and the *E. australis* system (Morales et al. 2018) is the overlapping peak of differentiation within chromosome 1A between inland and coastal populations. In *Eopsaltria* system the large 15.4-Mb linkage block on chromosome 1A that tightly co-segregated with mtDNA ancestry within an admixture zone (Morales et al. 2018) includes the genomic region of high differentiation between inland-coastal *townsendi* on chromosome 1A (Figure 4.4B). The remarkable similarity between the two systems

that are 47.3 million years divergent suggests that these regions might be a hotspot for mtDNA co-function and/or tend to be the targets of divergent selection between coastal and inland habitats in this avian clade. There was around 2 million years divergence between the north versus south *E. australis* (Morales et al. 2017), while only 0.5 million years between *townsendi* and *occidentalis* (Weir and Schluter 2004). The comparison between *E. australis* and *Setophaga* systems is consistent with the current understanding of the importance of the genomic landscape in the speciation continuum which the “islands of differentiation” are small at early stages of divergence (Rice and Hostert 1993; Turner et al. 2005; Nosil et al. 2009a) and expand into “continents of differentiation” as divergence progresses (Nosil et al. 2009b).

4.4.3 Genomic Architecture of differentiation

The distribution of F_{ST} across the genome comparing various coastal *townsendi* to either *occidentalis* and inland *townsendi* is consistent with the “genic” view of differentiation (Chapter 3), in which peaks of differentiation represent genetic targets of selection (divergent selection or selection against hybrids) that are highly distinct between populations despite the rest of the genome being homogenized by gene flow (Wu 2001; Via 2009; Nosil 2012). The chr1A, chr5, and chrZ differentiation are likely targets of selection for mito-nuclear climatic adaptation. The other highly differentiated region of differentiation (the RALY marker at chr20) is under selection related to plumage signaling that underlies the narrow Cascades hybrid zone between inland *townsendi* and *occidentalis* (chapter 3). This RALY peak is also the strongest peak of genomic differentiation between the coastal *townsendi* and *occidentalis* (Figure 4.4), providing corroborating evidence that the RALY region is causally related to plumage differences between *townsendi* and *occidentalis*.

4.4.4 Caveats and future directions

We found correlated mitonuclear evolutionary response related to fatty acid metabolism and climate adaption. Future study should investigate whether these mitonuclear genes ‘coadapt’ to climate of coastal *townsendi* habitat, as opposed to being independently selected under the same climatic condition. Although the tight functional association (Figure 4.5B) among the correlatedly evolved mitonuclear genetic regions supports mitonuclear coadaptation, more tests are needed to confirm this possibility. If there is mitonuclear coadaptation, there should be (1) epistasis effect of mtDNA and nDNA on fatty acid metabolic phenotypes; (2) the high frequency fatty acid metabolic phenotypes within each site should be more fit for local climate than foreign climate.

4.5 Conclusion

Examination of genomic differentiation in this young species group has revealed signatures of climate-correlated coadaptation among mito-nuclear genes for fatty-acid metabolism. Consistent with the mtDNA pattern, the coastal *townsendi* demonstrate a nuclear genomic pattern consistent with ancient admixture between inland *townsendi* and a geographically structured ancient *occidentalis* population. Three genetic clusters of coastal *townsendi* are characterized by a mixed genetic ancestry between the parental populations (*occidentalis* and inland *townsendi*), providing natural replicates for examining the role of selection in shaping genomic differentiation. These three coastal *townsendi* clusters exhibit parallel differentiation from inland *townsendi* at two of the three most differentiated genomic regions (on chr5 and chrZ) between inland *townsendi* and *occidentalis*. These two genetic regions are both involved in mitochondrial fatty acid metabolism. The geographic distributions of the fatty acid metabolism-related mito-nuclear genetic combinations are associated with geographic variation in climate, suggesting mt-nDNA coevolution may have occurred in response to selection for climate adaptation. Such climate-related mito-nuclear selection could be an important force driving population differentiation in this species complex.

Chapter 5 Conclusion

5.1 Genetic architecture of complex speciation

My thesis revealed a clear pattern of ‘selection with gene flow’, in which a few targets of selection remain distinct between species, while the rest of the genome is similar due to gene flow between species (Wu 2001; Via 2009; Feder et al. 2012a). This process generates the same genomic pattern as ‘divergence with gene flow’ model, however the underlying process is different. The latter assumes gradual elevation of differentiation at sympatry or parapatry, while the initial differentiation in this system was due to isolation associated with Pleistocene glaciation (Rohwer and Wood 1998; Rohwer et al. 2001b; Weir and Schluter 2004; Krosby and Rohwer 2009). There is a phenomenal contrast of divergence between the differentiated and undifferentiated regions: the three targets of divergence on chr5, chr20, and chrZ are almost fixed while the genome-wide F_{ST} is only ~ 0.03 . I also found evidence of selection for these diverged regions (F_{ST} near 1): the peak at RALY locus on chr20 is involved in plumage divergence (explaining $> 50\%$ of variation in three plumage traits that differ between species), which demonstrates a stable narrow cline across the hybrid zone over decades (chapter 3). The other two regions are related to mitonuclear coadaptation for fatty acid metabolism (chapter 4), and covary with mtDNA ancestry in the coastal *townsendi* range. These three genomic regions of large effect under selection may, in the future, allow genome-wide differentiation to accumulate through elevating differentiation of their neutral flanking regions and/or their co-functioning regions.

Is speciation usually initiated by a few genetic regions of large effects? This pair of warblers are at the early stage of divergence with very limited divergence genome-wide providing a great opportunity to address this question. In this case, the genome-wide

incompatibility has not been reached, thus in the face of gene flow, the few genomic regions have been maintained by selection while the neutral regions have diffused across species boundaries. This genomic differentiation could expand via linkage (i.e. divergent hitchhiking) (Felsenstein 1974; Hartl 1977; Via 2012; Payseur and Rieseberg 2016) and/or linkage disequilibrium among the selected regions to form greater reproductive isolation and/or with other co-functioning genes (i.e. genomic hitchhiking) (Feder et al. 2013). The warbler system thus supports the scenario in which a few regions of large effect initiated speciation. Further comparative study of species pairs at the early stage along the speciation continuum would reveal how common such a scenario is.

5.2 Genetic basis of mitonuclear coadaptation and its role in speciation

The mitonuclear correlated evolution represents the most outstanding genomic divergence in the genomic background. Despite the lack of fixed differences in the nuclear genome, the *occidentalis* and inland *townsendi* demonstrate distinct mitochondrial haplotypes that are ~0.5 million years divergent, that occur in habitats of distinct climatic conditions. The mitochondrial co-functioning nuclear genes harbor almost fixed differences between species. Both mitonuclear ancestries exist in the admixed population that inhabit an intermediate climatic environment, in which the site-level mitonuclear ancestry is correlated with climatic variation among sites. It is thus likely that mitochondrial climatic adaptation led to the indirect selection on its co-functioning nuclear variants.

The selection for mitonuclear coadaptation would be more effective if the mitochondrial co-functioning nuclear genes are Z-linked. This is because 100% of the female (ZW) Z genes are from the father (Hill and Johnson 2013), such that if there is any incompatibility between the maternal mt and paternal Z genes the incompatibility will be fully revealed in the daughters at F1

generation (Hill and Johnson 2013). The chrZ mitochondrial co-functioning gene BBOX1 might be a target under strong selection against incompatibility. Such strong selection could counteract the prevalent gene flow in early stage of speciation.

At least in this warbler system in the early stage in speciation, the evidence is consistent with mitonuclear coadaptation dominating the genome-wide divergence. It is possible that mitonuclear incompatibilities are a substantial source of Bateson-Dobzhansky-Muller incompatibilities (Bateson 1909; Dobzhansky 1941; Orr 1996) as argued before (Burton and Barreto 2012; Burton et al. 2013). The mitochondrial genome evolves much faster than the nuclear genome in animals (Allio et al. 2017), while co-functioning extensively with the nuclear genome (Rand et al. 2004; Dowling et al. 2008). Thus mtDNA can be the frontier of genomic differentiation between diverging lineages. To maintain mitonuclear cofunctions, there has to be compensatory evolution in the nuclear co-functioning regions (Rand et al. 2004; Dowling et al. 2008), forming distinctive mitonuclear combinations among species (Burton and Barreto 2012; Hill 2017; Hill et al. 2019).

5.3 Genetic basis of plumage and social signaling and its role in divergence

The plumage gene divergence in this system is not likely simply due to color-centric species delineation in birds (see chapter 1), but instead a form of reproductive isolation maintained by selection. If there is neutral plumage genetic variants that mix at secondary contact, the resulting plumage cline should widen over time according to neutral diffusion. However, I found a significantly narrower plumage cline and even narrower RALY cline than expected from eutral diffusion. This observation suggests selection (divergent selection or selection against hybrids) maintaining narrow plumage and RALY clines. Despite genome-wide similarity between *occidentalis* and inland *townsendi*, divergence and selection at the RALY

locus might allow genome-wide differentiation to build up. The RALY pleiotropy (one genetic region that affects multiple phenotypes) allows three plumage traits to highly covary and stay distinct between species. If the pleiotropy influences other traits that is not detected by the admixture mapping, the divergence of the plumage traits driven by selection will lead to increased variance of these other traits modulated by the same gene in the hybrids (Barton 2001).

The mechanism of selection that is maintaining the plumage boundary between species might be sexual selection for mate preference and mate competition. Avian plumage signals are important targets of sexual selection (Bennett et al. 2002; Uy et al. 2009; Delhey et al. 2017). At secondary contact, if there is strong incompatibility and selection against hybrids, there will be selection against heterospecific mating via reinforcement (Dobzhansky 1937b), which leads to divergence of mating/competition signals (i.e. character displacement) (Brown and Wilson 1956).

5.4 The relation of plumage and mitochondrial loci

The genomic regions of divergence related to mitonuclear cofunction and plumage pigmentation could be related. They are likely synergistically selected to strengthen the reproductive barrier, i.e. genomic hitchhiking (Felsenstein 1981; Barton 1983). More specifically, the plumage gene difference might be selected to signal different mitonuclear genotypes to ensure mating among individuals with compatible mitonuclear genotypes (Hill 2018). Since any mismatch in co-functioning mtDNA and nDNA can be detrimental to the fitness of individuals, visual signals and mate preference are expected to evolve to avoid mating with the individuals with incompatible mtDNA and nDNA (Hill 2018), as a particular avenue of reinforcement (Dobzhansky 1937a).

If so, a few predictions (Hill 2018) can be evaluated in this system: (1) hybrids with dysfunctional mitonuclear genotypes should show poor condition-dependent signaling; (2) genes for species-specific plumage signal and mitonuclear genotype should be physically linked, and potentially Z-linked; (3) the transition zones of mitonuclear genotypes should match the transition zones of plumage signals. More functional experiments are needed to fully test these predictions. For individuals in the coastal *townsendi* populations with mismatched mitonuclear ancestries, they are genetically *townsendi*-like of RALY, although the same is true for individuals carrying matched mitonuclear ancestries. Prediction 2 is rejected in that RALY and the mitochondrial co-functioning nuclear genes are on different chromosomes. Prediction 3 is partly supported in this system: the boundary of mitonuclear genotypes between *occidentalis* and inland *townsendi* is consistent with the boundary of the RALY genotype. However, this prediction is also partly rejected because such alignment in mitonuclear and RALY genotypic boundaries falls apart in the coastal *townsendi* range. In Haida Gwaii, a coastal *townsendi* population that harbors almost complete *occidentalis* mitonuclear genotypes has *townsendi*-like RALY genotype. This is probably due to the lack of risk for mitonuclear mismatch in populations that harbor predominantly the concordant mtDNA and nDNA ancestries, thus reduced selection for mitonuclear-dependent plumage signaling. However, high risk of mitonuclear mismatch occurs for other coastal *townsendi* populations that contain ~50% of the alternative mtDNA and nDNA ancestry, where the *townsendi*-like RALY are fixed. If RALY difference is selected to signal mitonuclear concordance, it might not be applicable for coastal *townsendi*, which could be due to the greater variability in climate, thus variable selection pressures at different sites. Altogether, RALY genotypes might be a signal selected to reflect mitonuclear concordance, but such a signal might be region/habitat-specific.

5.5 Broad implication

This warbler system with ongoing and ancient admixture and differentiation at the early stages of speciation provides a unique angle looking into the genomic mechanism of speciation. It exhibited a clear pattern of 'selection with gene flow', in which a few potentially inter-related regions under strong selection pioneered differentiation across the genome, while the rest of the genome remains undifferentiated due to gene flow. This thesis also highlighted the role of mitonuclear adaptation and selection on pigmentation in providing strong genomic targets of selection for early divergence to initiate.

References

- Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle
BT - Selected Papers of Hirotugu Akaike. Pp. 267–281 in Second International Symposium
on Information Theory.
- Alcaide, M., E. S. C. Scordato, T. D. Price, and D. E. Irwin. 2014. Genomic divergence in a ring
species complex. *Nature* 511:83–85.
- Allio, R., S. Donega, N. Galtier, and B. Nabholz. 2017. Large variation in the ratio of
mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity
and the use of mitochondrial DNA as a molecular marker. *Mol. Biol. Evol.* 34:2762–2772.
- Anderson, C. N., and G. F. Grether. 2010. Interspecific aggression and character displacement of
competitor recognition in *Hetaerina* damselflies. *Proc. R. Soc. London, Ser. B* 277:549–55.
- Atkin, O. K., and D. Macherel. 2009. The crucial role of plant mitochondria in orchestrating
drought tolerance.
- Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn. 2007. GenABEL: An R library for
genome-wide association analysis. *Bioinformatics* 23:1294–1296.
- Ballard, J. W. O., and M. C. Whitlock. 2004. The incomplete natural history of mitochondria.
Mol. Ecol. 13:729–744.
- Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific
phylogenies. *Mol. Biol. Evol.* 16:37–48.
- Bar-Yaacov, D., Z. Hadjivasiliou, L. Levin, G. Barshad, R. Zarivach, A. Bouskila, and D.
Mishmar. 2015. Mitochondrial involvement in vertebrate speciation? The Case of mito-
nuclear genetic divergence in chameleons. *Genome Biol. Evol.* 7:3322–3336.
- Baris, T. Z., D. N. Wagner, D. I. Dayan, X. Du, P. U. Blier, N. Pichaud, M. F. Oleksiak, and D.
L. Crawford. 2017. Evolved genetic and phenotypic differences due to mitochondrial-
nuclear interactions. *PLoS Genet.* 13:e1006517.
- Barrera-Guzmán, A. O., A. Aleixo, M. D. Shawkey, and J. T. Weir. 2018. Hybrid speciation
leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proc. Natl.
Acad. Sci.* 115:E218–E225.
- Barton, N. H. 1979a. Gene flow past a cline. *Heredity (Edinb).* 43:333–339.
- Barton, N. H. 1983. Multilocus Clines. *Evolution (N. Y).* 37:454–471.
- Barton, N. H. 1979b. The dynamics of hybrid zones. *Heredity (Edinb).* 43:341–359.
- Barton, N. H. 2001. The role of hybridization in evolution. *Mol. Ecol.* 10:551–568.
- Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pp. 13–45 in R. G.
Harrison, ed. *Hybrid zones and the evolutionary process*. Oxford Univ. Press, Oxford, UK.
- Barton, N. H., and G. M. Hewitt. 1985. Analysis of Hybrid Zones. *Annu. Rev. Ecol. Syst.*
16:113–148.
- Barton, N. H., and G. M. Hewitt. 1989. Adaptation, speciation and hybrid zones. *Nature*
341:497–503.
- Bateson, W. 1909. *Heredity and variation in modern lights*. Pp. 85–101 in *Darwin and Modern
Science*. Cambridge University Press, Cambridge.
- Beaudet, A. L. 2017. Brain carnitine deficiency causes nonsyndromic autism with an extreme
male bias: A hypothesis. *BioEssays* 39:1–25.
- Bennett, A. T. D., I. C. Cuthill, J. C. Partridge, and K. Lunau. 2002. Ultraviolet plumage colors
predict mate preferences in starlings. *Proc. Natl. Acad. Sci.*, doi: 10.1073/pnas.94.16.8618.
- Boratyński, Z., T. Ketola, E. Koskela, and T. Mappes. 2016. The Sex Specific Genetic Variation

- of Energetics in Bank Voles, Consequences of Introgression? *Evol. Biol.* 43:37–47.
- Bradshaw, H. D., and D. W. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426:176–178.
- Brelsford, A., and D. E. Irwin. 2009. Incipient speciation despite little assortative mating: The yellow-rumped warbler hybrid zone. *Evolution (N. Y.)* 63:3050–3060.
- Brelstaff, J., T. Lashley, J. L. Holton, A. J. Lees, M. N. Rossor, R. Bandopadhyay, and T. Revesz. 2011. Transportin1: A marker of FTLN-FUS. *Acta Neuropathol.* 122:591–600.
- Brown, W. L., and E. O. Wilson. 1956. Character Displacement. *Syst. Zool.* 5:49–64.
- Buerkle, C. A., and C. Lexer. 2008. Admixture as the basis for genetic mapping.
- Buggs, R. J. A. 2007. Empirical study of hybrid zone movement. *Heredity (Edinb.)* 99:301–312.
- Burton, R. S., and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities?
- Burton, R. S., R. J. Pereira, and F. S. Barreto. 2013. Cytonuclear Genomic Interactions and Hybrid Breakdown. *Annu. Rev. Ecol. Evol. Syst.* 44:281–302.
- Bush, G. L. 1994. Sympatric speciation in animals: new wine in old bottles.
- Calvo, S. E., and V. K. Mootha. 2010. The Mitochondrial Proteome and Human Disease. *Annu. Rev. Genomics Hum. Genet.* 11:25–44.
- Chang, N., C. Sutherland, E. Hesse, R. Winkfein, W. B. Wiehler, M. Pho, C. Veillette, S. Li, D. P. Wilson, E. Kiss, and M. P. Walsh. 2007. Identification of a novel interaction between the Ca(2+)-binding protein S100A11 and the Ca(2+)- and phospholipid-binding protein annexin A6. *Am. J. Physiol. Cell Physiol.* 292:C1417-30.
- Chunco, A. J. 2014. Hybridization in a warmer world.
- Clemente, F. J., A. Cardona, C. E. Inchley, B. M. Peter, G. Jacobs, L. Pagani, D. J. Lawson, T. Antão, M. Vicente, M. Mitt, M. DeGiorgio, Z. Faltyskova, Y. Xue, Q. Ayub, M. Szpak, R. Mägi, A. Eriksson, A. Manica, M. Raghavan, M. Rasmussen, S. Rasmussen, E. Willerslev, A. Vidal-Puig, C. Tyler-Smith, R. Villems, R. Nielsen, M. Metspalu, B. Malyarchuk, M. Derenko, and T. Kivisild. 2014. A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. *Am. J. Hum. Genet.* 95:584–589.
- Crawford, J. E., and R. Nielsen. 2013. Detecting adaptive trait loci in nonmodel systems: Divergence or admixture mapping? *Mol. Ecol.* 22:6131–6148.
- Curat, M., and L. Excoffier. 2005. The effect of the Neolithic expansion on European molecular diversity. *Proc. R. Soc. B Biol. Sci.* 272:679–688.
- Curry, C. M. 2015. An Integrated Framework for Hybrid Zone Models. *Evol. Biol.* 42:359–365.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
- Darwin, C. 1859. *On the Origin of the Species*.
- De Queiroz, K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–886.
- Delhey, Peters, and Kempnaers. 2017. Cosmetic Coloration in Birds: Occurrence, Function, and Evolution. *Am. Nat.* 1:S145-58.
- Dijkstra, P. D., and S. E. Border. 2018. How does male-male competition generate negative frequency-dependent selection and disruptive selection during speciation? *Curr. Zool.* 64:89–99. Oxford University Press.
- Dobzhansky, T. 1941. *Genetics and the origin of species*.
- Dobzhansky, T. 1937a. *Genetics and the Origin of Species*. Columbia University Press, Columbia.

- Dobzhansky, T. 1937b. *Genetics and the Origin of Species*. Columbia University Press, New York.
- Dowling, D. K., U. Friberg, and J. Lindell. 2008. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol. Evol.* 23:546–554.
- Eaton, M. D., and S. M. Lanyon. 2003. The ubiquity of avian ultraviolet plumage reflectance. *Proc. R. Soc. B Biol. Sci.* 270:1721–1726.
- Elgvin, T. O., C. N. Trier, O. K. Tørresen, I. J. Hagen, S. Lien, A. J. Nederbragt, M. Ravinet, H. Jensen, and G. P. Sætre. 2017. The genomic mosaicism of hybrid speciation. *Sci. Adv.* 3:e1602996.
- Ellegren, H. 2010. Evolutionary stasis: the stable chromosomes of birds. *Trends Ecol. Evol.* 25:283–291. Elsevier Ltd.
- Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379.
- Endler, J. A. 1977a. Geographic variation, speciation, and clines. *Monogr. Popul. Biol.* 10:1–246.
- Endler, J. A. 1977b. Properties of Clines. Pp. 62–65 *in* *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.
- Feder, J. L., S. P. Egan, and P. Nosil. 2012a. The genomics of speciation-with-gene-flow.
- Feder, J. L., S. M. Flaxman, S. P. Egan, A. A. Comeault, and P. Nosil. 2013. Geographic Mode of Speciation and Genomic Divergence. *Annu. Rev. Ecol. Evol. Syst.* 44:73–97.
- Feder, J. L., R. Gejji, S. Yeaman, and P. Nosil. 2012b. Establishment of new mutations under divergence and genome hitchhiking. *Philos. Trans. R. Soc. B Biol. Sci.* 367:461–474.
- Felsenstein, J. 1981. Skepticism Towards Santa Rosalia, or Why are There so Few Kinds of Animals? *Evolution (N. Y.)*. 35:124–138.
- Felsenstein, J. 1974. The evolution advantage of recombination. *Genetics* 78:737–756.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, New York, NY.
- Fisher, R. A. 1919. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburgh* 52:399–433.
- Forister, M. L. 2005. Independent inheritance of preference and performance in hybrids between host races of *Mitoura* butterflies (Lepidoptera: Lycaenidae). *Evolution (N. Y.)*. 59:1149–1155.
- Fumagalli, M., I. Moltke, N. Grarup, F. Racimo, P. Bjerregaard, M. E. Jørgensen, T. S. Korneliussen, P. Gerbault, L. Skotte, A. Linneberg, C. Christensen, I. Brandslund, T. Jørgensen, E. Huerta-Sánchez, E. B. Schmidt, O. Pedersen, T. Hansen, A. Albrechtsen, and R. Nielsen. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science (80-.)*. 349:1343–1347.
- Gagnaire, P. A., E. Normandeau, and L. Bernatchez. 2012. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Mol. Biol. Evol.* 29:2909–2919.
- Gao, X., J. Starmer, and E. R. Martin. 2008. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* 32:361–369.
- Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, New Jersey.

- Gay, L., P. A. Crochet, D. A. Bell, and T. Lenormand. 2008. Comparing clines on molecular and phenotypic traits in hybrid zones: A window on tension zone models. *Evolution* (N. Y.) 62:2789–2806.
- Gray, S. M., and J. S. McKinnon. 2007. Linking color polymorphism maintenance and speciation. *Trends Ecol. Evol.* 22:71–79.
- Grether, G. F., K. S. Peiman, J. A. Tobias, and B. W. Robinson. 2017. Causes and Consequences of Behavioral Interference between Species. *Trends Ecol. Evol.* 32:760–772.
- Hartl, D. L. 1977. How does the Genome Congeal? Pp. 65–82 *in* *Measuring Selection in Natural Populations*.
- Henze, K., and W. Martin. 2003. Evolutionary biology: essence of mitochondria. *Nature* 426:127–128.
- Hill, G. E. 2006. Female mate choice for ornamental coloration. *Bird Color. Vol. II. Funct. Evol.*
- Hill, G. E. 2019. *Mitonuclear Ecology*. Oxford University Press, Oxford.
- Hill, G. E. 2018. Mitonuclear Mate Choice: A Missing Component of Sexual Selection Theory? *BioEssays* 40.
- Hill, G. E. 2015. Sexiness, Individual Condition, and Species Identity: The Information Signaled by Ornaments and Assessed by Choosing Females. *Evol. Biol.* 42:251–259.
- Hill, G. E. 2017. The mitonuclear compatibility species concept. *Auk* 134:393–409.
- Hill, G. E., J. C. Havird, D. B. Sloan, R. S. Burton, C. Greening, and D. K. Dowling. 2019. Assessing the fitness consequences of mitonuclear interactions in natural populations. *Biol. Rev.* 94:consequences of mitonuclear interactions in natu.
- Hill, G. E., and J. D. Johnson. 2013. The mitonuclear compatibility hypothesis of sexual selection. *Proc. R. Soc. B Biol. Sci.* 280:20131314.
- Hill, G. E., and K. J. McGraw. 2004. Correlated changes in male plumage coloration and female mate choice in cardueline finches. *Anim. Behav.* 67:27–35.
- Hothorn, T., K. Hornik, M. A. van De Wiel, and A. Zeileis. 2008. Implementing a Class of Permutation Tests COIN Package. *J. Stat. Softw.* 28:1–23.
- Hugall, A. F., and D. Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. *Nature* 485:631–634.
- Hunt, S. E., W. McLaren, L. Gil, A. Thormann, H. Schuilenburg, D. Sheppard, A. Parton, I. M. Armean, S. J. Trevanion, P. Flicek, and F. Cunningham. 2018. *Ensembl variation resources. Database* (Oxford). 2018.
- Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol. Ecol.* 25:4488–4507.
- Irwin, D. E., B. Milá, D. P. L. Toews, A. Brelsford, H. L. Kenyon, A. N. Porter, C. Grossen, K. E. Delmore, M. Alcaide, and J. H. Irwin. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. *Mol. Ecol.* 27:4839–4855.
- Jackson, W. M., C. S. Wood, and S. Rohwer. 1992. Age-Specific Plumage Characters and Annual Molt Schedules of Hermit Warblers and Townsend’s Warblers. *Condor* 94:490–501.
- Jewett, S. G. 1944. Hybridization of Hermit and Townsend warblers. *Condor* 46:23–24.
- Johnson, N. A., and A. H. Porter. 2000. Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J. Theor. Biol.* 205:527–542.
- Johnson, R. C., G. W. Nelson, J. L. Troyer, J. A. Lautenberger, B. D. Kessing, C. A. Winkler, and S. J. O’Brien. 2010. Accounting for multiple comparisons in a genome-wide association

- study (GWAS). *BMC Genomics* 11:724. BioMed Central Ltd.
- Key, K. 1968. The Concept of Stasipatric Speciation. *Syst. Zool.* 17:14–22.
- Kim, K. W., B. C. Jackson, H. Zhang, D. P. L. Toews, S. A. Taylor, E. I. Greig, I. J. Lovette, M. M. Liu, A. Davison, S. C. Griffith, K. Zeng, and T. Burke. 2019. Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nat. Commun.* 10:1852.
- Kingma, S. A., I. Szentirmai, T. Székely, V. Bókony, M. Bleeker, A. Liker, and J. Komdeur. 2008. Sexual selection and the function of a melanin-based plumage ornament in polygamous penduline tits *Remiz pendulinus*. *Behav. Ecol. Sociobiol.* 62:1277–1288.
- Knief, U., C. M. Bossu, N. Saino, B. Hansson, J. Poelstra, N. Vijay, M. Weissensteiner, and J. B. W. Wolf. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat. Ecol. Evol.* 3:570–576.
- Konishi, M., and K. Takata. 2004. Impact of asymmetrical hybridization followed by sterile F 1 hybrids on species replacement in *Pseudorasbora*. *Conserv. Genet.* 5:463–474.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.
- Kronforst, M. R., L. G. Young, D. D. Kapan, C. McNeely, R. J. O’Neill, and L. E. Gilbert. 2006. Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. *Proc. Natl. Acad. Sci.* 103:6575–6580.
- Krosby, M., and S. Rohwer. 2009. A 2000 km genetic wake yields evidence for northern glacial refugia and hybrid zone movement in a pair of songbirds. *Proc. R. Soc. B Biol. Sci.* 276:615–621.
- Krosby, M., and S. Rohwer. 2010. Ongoing movement of the hermit warbler X Townsend’s Warbler Hybrid Zone. *PLoS One* 5:e14164.
- Kühlbrandt, W. 2015. Structure and function of mitochondrial membrane protein complexes.
- Lane, N. 2011. Mitonuclear match: Optimizing fitness and fertility over generations drives ageing within generations. *BioEssays* 33:860–869.
- Lawniczak, M. K. N., S. J. Emrich, A. K. Holloway, A. P. Regier, M. Olson, B. White, S. Redmond, L. Fulton, E. Appelbaum, J. Godfrey, C. Farmer, A. Chinwalla, S. P. Yang, P. Minx, J. Nelson, K. Kyung, B. P. Walenz, E. Garcia-Hernandez, M. Aguiar, L. D. Viswanathan, Y. H. Rogers, R. L. Strausberg, C. A. Sasaki, D. Lawson, F. H. Collins, F. C. Kafatos, G. K. Christophides, S. W. Clifton, E. F. Kirkness, and N. J. Besansky. 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* (80-.). 330:512–514.
- Legendre, P. 1998. Numerical ecology, 2nd English Edition. Elsevier Science.
- Leigh, J. W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6:1110–1116.
- Li, H. 2010. Aligning new-sequencing reads by BWA BWA : Burrows-Wheeler Aligner. Slides, doi: 10.1002/pssa.200673542.
- Longo, N., M. Frigeni, and M. Pasquali. 2016. Carnitine transport and fatty acid oxidation. *Biochim. Biophys. Acta - Mol. Cell Res.* 1863:2422–2435.
- Lovett, M., Z. Y. Cheng, E. M. Lamela, T. Yokoi, and C. J. Epstein. 1987. Molecular markers for the agouti coat color locus of the mouse. *Genetics* 115:747–754.
- Matsubayashi, K. W., and H. Katakura. 2009. Contribution of multiple isolating barriers to reproductive isolation between a pair of phytophagous ladybird beetles. *Evolution* (N. Y). 63:2563–2580.

- Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.*, doi: 10.1017/S0016672300014634.
- Mayr, E. 1970. *Populations, Species, and Evolution*. Harvard University Press, Cambridge, MA, USA.
- Mayr, E. 1942. *Systematics and the origin of species*. Columbia University Press, New York.
- McEntee, J. P., J. G. Burleigh, and S. Singhal. 2018. Dispersal predicts hybrid zone widths across animal diversity: Implications for species borders under incomplete reproductive isolation. *bioRxiv*, doi: 10.1101/472506.
- McGraw, K. J. 2000. Carotenoid-based ornamentation and status signaling in the house finch. *Behav. Ecol.* 11:520–527.
- Mehta, S. 2013. Oxidation of Fatty Acids - via Beta-Oxidation | *Biochemistry Notes* | PharmaXChange.info.
- Mendel, G. 1869. Ueber einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. (On Hieracium hybrids obtained by artificial fertilisation). *Verh. Naturf. Ver. Brünn.* 8:26–31.
- Michaud, E. J., S. J. Bultman, M. L. Klebig, M. J. van Vugt, L. J. Stubbs, L. B. Russell, and R. P. Woychik. 1994. A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. *Proc. Natl. Acad. Sci.* 91:2562–2566.
- Mikami, O. K., M. Kohda, and M. Kawata. 2004. A new hypothesis for species coexistence: Male-male repulsion promotes coexistence of competing species.
- Milá, B., T. B. Smith, and R. K. Wayne. 2007. Speciation and rapid phenotypic differentiation in the yellow-rumped warbler *Dendroica coronata* complex. *Mol. Ecol.* 16:159–173.
- Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm. 2011. How many species are there on earth and in the ocean? *PLoS Biol.* 9:e1001127.
- Morales, H. E., A. Pavlova, N. Amos, R. Major, A. Kilian, C. Greening, and P. Sunnucks. 2018. Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat. Ecol. Evol.* 2:1258–1267.
- Morales, H. E., P. Sunnucks, L. Joseph, and A. Pavlova. 2017. Perpendicular axes of differentiation generated by mitochondrial introgression. *Mol. Ecol.* 26:3241–3255.
- Nadeau, N. J., F. Minvielle, S. Ito, M. Inoue-Murayama, D. Gourichon, S. A. Follett, T. Burke, and N. I. Mundy. 2008. Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian ASIP (agouti) gene. *Genetics* 178:777–786.
- Nosil, P. 2012. A Genetic Mechanism to Link Selection to Reproductive Isolation. Pp. 109–138 *in* *Ecological Speciation*. Oxford Univ. Press, Oxford.
- Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: Causes and consequences. *Philos. Trans. R. Soc. B Biol. Sci.* 367:332–342.
- Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009a. Divergent selection and heterogeneous genomic divergence.
- Nosil, P., L. J. Harmon, and O. Seehausen. 2009b. Ecological explanations for (incomplete) speciation.
- Orr, H. A. 1996. Dobzhansky, Bateson, and the genetics of speciation.
- Orr, H. A. 1995. The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics* 139:1805–1813.
- Orr, H. A., and M. Turelli. 2001. The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution (N. Y.)*, doi: 10.1111/j.0014-

3820.2001.tb00628.x.

- Owen-Ashley, N. T., and L. K. Butler. 2004. Androgens, interspecific competition and species replacement in hybridizing warblers. *Proc. R. Soc. B Biol. Sci.* 271:S498–S500.
- Pagliarini, D. J., S. E. Calvo, B. Chang, S. A. Sheth, S. B. Vafai, S. E. Ong, G. A. Walford, C. Sugiana, A. Boneh, W. K. Chen, D. E. Hill, M. Vidal, J. G. Evans, D. R. Thorburn, S. A. Carr, and V. K. Mootha. 2008. A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell* 134:112–123.
- Paul, H. S., G. Sekas, and S. A. Adibi. 1992. Carnitine biosynthesis in hepatic peroxisomes: Demonstration of γ -butyrobetaine hydroxylase activity. *Eur. J. Biochem.* 203:599–605.
- Payseur, B. A., and L. H. Rieseberg. 2016. A genomic perspective on hybridization and speciation.
- Pearson, S. F. 2000. Behavioral asymmetries in a moving hybrid zone. *Behav. Ecol.* 11:84–92.
- Pearson, S. F., and S. Rohwer. 2000. Asymmetries in male aggression across an avian hybrid zone. *Behav. Ecol.* 11:93–101.
- Pearson, S. F., S. Rohwer, B. Museum, and B. Museum. 2000. Asymmetries in male aggression across an avian hybrid zone. *Behav. Ecol.* 11:93–101.
- Pebesma, E. J., and R. S. Bivand. 2005. Classes and methods for spatial data in R. *R News* 5:9–13.
- Pfennig, K. S., and D. W. Pfennig. 2009. Character displacement: ecological and reproductive responses to a common evolutionary problem. *Q. Rev. Biol.* 84:253–276.
- Poelstra, J. 2013. The Genetics of Speciation and Colouration in Carrion and Hooded Crows.
- Poelstra, J. W., H. Ellegren, and J. B. W. Wolf. 2013. An extensive candidate gene approach to speciation: Diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity (Edinb)*. 111:467–473.
- Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, and J. B. W. Wolf. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science (80-.)*. 344:1410–1414.
- Price, T. 2007. *Speciation In Birds*. 1st ed. Roberts and Company, Greenwood Village, CO.
- Price, T., H. L. Gibbs, L. De Sousa, and A. D. Richman. 1998. Different timing of the adaptive radiations of North American and Asian warblers. *Proc. R. Soc. B Biol. Sci.*, doi: 10.1098/rspb.1998.0527.
- Price, T., I. J. Lovette, E. Bermingham, H. L. Gibbs, and A. D. Richman. 2000. The imprint of history on North American and Asian warblers. *Am. Nat.* 156:354–367.
- Pryke, S. R., and S. Andersson. 2003. Carotenoid-based epaulettes reveal male competitive ability: Experiments with resident and floater red-shouldered widowbirds. *Anim. Behav.* 66:217–224.
- R Core team. 2015. R Core Team.
- Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Datasets. *Genetics* 197:573–589.
- Rand, D. M., R. A. Haney, and A. J. Fry. 2004. Cytonuclear coevolution: The genomics of cooperation. *Trends Ecol. Evol.* 19:645–653.
- Rice, W. R., and E. E. Hostert. 1993. Laboratory Experiments on Speciation: What Have We Learned in 40 Years? *Evolution (N. Y)*. 47:1637.
- Rieseberg, L. H. 1997. Hybrid origins of plant species.
- Rohwer, S., E. Bermingham, and C. Wood. 2001a. Plumage and mitochondrial DNA haplotype variation across a moving hybrid zone. *Evolution (N. Y)*. 55:405–422.

- Rohwer, S., E. Bermingham, and C. Wood. 2001b. Plumage and Mitochondrial Dna Haplotype Variation Across a Moving Hybrid Zone. *Evolution* (N. Y). 55:405.
- Rohwer, S., and C. Wood. 1998. Three Hybrid Zones Between Hermit and Townsend ' S Warblers in Washington and Oregon. *Auk* 115:284–310.
- Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence. *PLoS Biol.*, doi: 10.1371/journal.pbio.2000234.
- Ryan, S. F., J. M. Deines, J. M. Scriber, M. E. Pfrender, S. E. Jones, S. J. Emrich, and J. J. Hellmann. 2018. Climate-mediated hybrid zone movement revealed with genomics, museum collection, and simulation modeling. *Proc. Natl. Acad. Sci.* 201714950.
- Sætre, G. P., T. Moum, S. Bureš, M. Král, M. Adamjan, and J. Moreno. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* 387:589–592.
- Sallam, T., M. C. Jones, T. Gilliland, L. Zhang, X. Wu, A. Eskin, J. Sandhu, D. Casero, T. Q. D. A. Vallim, C. Hong, M. Katz, R. Lee, J. Whitelegge, and P. Tontonoz. 2016. Feedback modulation of cholesterol metabolism by the lipid-responsive non-coding RNA LeXis. *Nature* 534:124–128.
- Sambatti, J. B. M., D. Ortiz-Barrientos, E. J. Baack, and L. H. Rieseberg. 2008. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol. Lett.* 11:1082–1091.
- Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. *Proc. Natl. Acad. Sci.* 106:9955–9962.
- Schumer, M., R. Cui, D. L. Powell, G. G. Rosenthal, and P. Andolfatto. 2016. Ancient hybridization and genomic stabilization in a swordtail fish. *Mol. Ecol.* 25:2661–2679.
- Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G. P. Sætre, C. Bank, Å. Brännström, A. Brelsford, C. S. Clarkson, F. Eroukhmanoff, J. L. Feder, M. C. Fischer, A. D. Foote, A. D. Foote, P. Franchini, C. D. Jiggins, F. C. Jones, A. K. Lindholm, K. Lucek, M. E. Maan, D. A. Marques, D. A. Marques, S. H. Martin, B. Matthews, J. I. Meier, J. I. Meier, M. Möst, M. Möst, M. W. Nachman, E. Nonaka, D. J. Rennison, J. Schwarzer, J. Schwarzer, J. Schwarzer, E. T. Watson, A. M. Westram, and A. Widmer. 2014. Genomics and the origin of species. *Nat. Rev. Genet.* 15:176–192.
- Seehausen, O., and D. Schluter. 2004. Male-male competition and nuptial-colour displacement as a diversifying force in Lake Victoria cichlid fishes. *Proc. R. Soc. B Biol. Sci.* 271:1345–1353.
- Seutin, G., B. N. White, and P. T. Boag. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* 69:82–90.
- Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman. 2010. Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America.
- Sharpe, R. B. 1909. A Hand-list of the Genera and Species of Birds. British Museum of Natural History, London, UK.
- Shriner, D. 2017. Overview of admixture mapping. *Curr. Protoc. Hum. Genet.* 2017:1.23.1-1.23.8.
- Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* 75:733–756.
- Szymura, J. M. ., and N. H. Barton. 1986. Genetic Analysis of a Hybrid Zone Between the Fire-

- Bellied Toads, *Bombina bombina* and *B. variegata*, Near Cracow in Southern Poland. *Evolution* (N. Y). 40:1141–1159.
- Tarof, S. A., P. O. Dunn, and L. A. Whittingham. 2005. Dual functions of a melanin-based ornament in the common yellowthroat. *Proc. R. Soc. B Biol. Sci.* 272:1121–1127.
- Tars, K., J. Leitans, A. Kazaks, D. Zelencova, E. Liepinsh, J. Kuka, M. Makrečka, D. Lola, V. Andrianovs, D. Gustina, S. Grinberga, E. Liepinsh, I. Kalvinsh, M. Dambrova, E. Loza, and O. Pugovics. 2014. Targeting carnitine biosynthesis: Discovery of new inhibitors against γ -butyrobetaine hydroxylase. *J. Med. Chem.* 57:2213–2236.
- Taylor, S. A., E. L. Larson, and R. G. Harrison. 2015. Hybrid zones: Windows on climate change.
- Taylor, S. A., T. A. White, W. M. Hochachka, V. Ferretti, R. L. Curry, and I. Lovette. 2014a. Climate-mediated movement of an avian hybrid zone. *Curr. Biol.* 24.
- Taylor, S. A., T. A. White, W. M. Hochachka, V. Ferretti, R. L. Curry, and I. Lovette. 2014b. Report Climate-Mediated Movement of an Avian Hybrid Zone. *Curr. Biol.* 24:1–6. Elsevier Ltd.
- Team, R. C. 2014. R core team (2014). R A Lang. *Environ. Stat. Comput. R Found. Stat. Comput.* Vienna, Austria. URL <http://www.R-project.org> ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals.
- Toews, D. P. L., A. Brelsford, C. Grossen, B. Milá, and D. E. Irwin. 2016a. Genomic variation across the Yellow-rumped Warbler species complex. *Auk* 133:698–717.
- Toews, D. P. L., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J. Lovette. 2016b. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Curr. Biol.* 26:2313–2318.
- Turner, T. L., M. W. Hahn, and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* 3:1572–1578.
- Uy, J. A. C., R. G. Moyle, C. E. Filardi, and Z. A. Cheviron. 2009. Difference in Plumage Color Used in Species Recognition between Incipient Species Is Linked to a Single Amino Acid Substitution in the Melanocortin-1 Receptor. *Am. Nat.* 174:244–254.
- van Doorn, G. S., U. Dieckmann, and F. J. Weissing. 2004. Sympatric Speciation by Sexual Selection: A Critical Reevaluation. *Am. Nat.* 163:709–725.
- van Riemsdijk, I., R. K. Butlin, B. Wielstra, and J. W. Arntzen. 2019. Testing an hypothesis of hybrid zone movement for toads in France. *Mol. Ecol.* 28:1070–1083.
- Vaz, F. M., S. Van Gool, R. Ofman, L. Ijlst, and R. J. A. Wanders. 1998. Carnitine biosynthesis: Identification of the cDNA encoding human γ -butyrobetaine hydroxylase. *Biochem. Biophys. Res. Commun.* 250:506–510.
- Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow.
- Via, S. 2009. Natural selection in action during speciation. *Proc. Natl. Acad. Sci.* 106:9939–9946.
- Wang, L., K. Luzynski, J. E. Pool, V. Janoušek, P. Dufková, M. M. Vyskočilová, K. C. Teeter, M. W. Nachman, P. Munclinger, M. MacHolún, J. Piúlek, and P. K. Tucker. 2011. Measures of linkage disequilibrium among neighbouring SNPs indicate asymmetries across the house mouse hybrid zone. *Mol. Ecol.* 20:2985–3000.
- Wang, T., A. Hamann, D. L. Spittlehouse, and T. Q. Murdock. 2012. ClimateWNA-high-

- resolution spatial climate data for western North America. *J. Appl. Meteorol. Climatol.* 51:16–29.
- Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, S. Searle, S. White, A. J. Vilella, S. Fairley, A. Heger, L. Kong, C. P. Ponting, E. D. Jarvis, C. V. Mello, P. Minx, P. Lovell, T. A. F. Velho, M. Ferris, C. N. Balakrishnan, S. Sinha, C. Blatti, S. E. London, Y. Li, Y. C. Lin, J. George, J. Sweedler, B. Southey, P. Gunaratne, M. Watson, K. Nam, N. Backström, L. Smeds, B. Nabholz, Y. Itoh, O. Whitney, A. R. Pfenning, J. Howard, M. Völker, B. M. Skinner, D. K. Griffin, L. Ye, W. M. McLaren, P. Flicek, V. Quesada, G. Velasco, C. Lopez-Otin, X. S. Puente, T. Olender, D. Lancet, A. F. A. Smit, R. Hubley, M. K. Konkel, J. A. Walker, M. A. Batzer, W. Gu, D. D. Pollock, L. Chen, Z. Cheng, E. E. Eichler, J. Stapley, J. Slate, R. Ekblom, T. Birkhead, T. Burke, D. Burt, C. Scharff, I. Adam, H. Richard, M. Sultan, A. Soldatov, H. Lehrach, S. V. Edwards, S. P. Yang, X. Li, T. Graves, L. Fulton, J. Nelson, A. Chinwalla, S. Hou, E. R. Mardis, and R. K. Wilson. 2010. The genome of a songbird. *Nature* 464:757–762.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* (N. Y.) 38:1358.
- Weir, J. T., and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. *Proc. R. Soc. B Biol. Sci.* 271:1881–1887.
- Weise, M., and J. R. Meyer. 1979. Juvenile dispersal and development of site-fidelity in the Black-capped Chickadee. *Auk* 96:40–55.
- West-Eberhard, M.-J. 1984. Sexual selection, competitive communication, and species-specific signals in insects. Pp. 283–324 *in* T. Lewis, ed. *Insect Communication*. Academic Press.
- Whibley, A. C., N. B. Langlade, C. Andalo, A. I. Hanna, A. Bangham, C. Thébaud, and E. Coen. 2006. Evolutionary paths underlying flower color variation in *Antirrhinum*. *Science* (80-.). 313:963–966.
- Wielstra, B., T. Burke, R. K. Butlin, A. Avci, N. Uz, E. Bozkurt, K. Olgun, and A. J. W. 2017a. A genomic footprint of hybrid zone movement in crested newts. 1:93–101.
- Wielstra, B., T. Burke, R. K. Butlin, A. Avci, N. Üzümlü, E. Bozkurt, K. Olgun, and J. W. Arntzen. 2017b. A genomic footprint of hybrid zone movement in crested newts. *Evol. Lett.* 1:93–101.
- Williams, L. M., and M. F. Oleksiak. 2011. Ecologically and evolutionarily important SNPs identified in natural populations. *Mol. Biol. Evol.* 28:1817–1826.
- Wolf Horrell, E. M., M. C. Boulanger, and J. A. D’Orazio. 2016. Melanocortin 1 receptor: Structure, function, and regulation.
- Wolf, J. B. W., and H. Ellegren. 2017. Making sense of genomic islands of differentiation in light of speciation.
- Wright, S. 1931. Evolution in Mendelian Populations. *Genetics* 16:97–159.
- Wu, C. I. 2001. The genic view of the process of speciation.
- Wu, X., and M. Watson. 2009. CORNA: Testing gene lists for regulation by microRNAs. *Bioinformatics* 25:832–833.
- Zerbino, D. R., P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A.

- Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, and P. Flicek. 2018. Ensembl 2018. *Nucleic Acids Res.* 46:D754–D761.
- Zhang, G., C. Li, Q. Li, B. Li, D. M. Larkin, C. Lee, J. F. Storz, A. Antunes, M. J. Greenwold, R. W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R. R. Da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J. T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M. V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M. N. Romanov, R. Borges, J. P. Machado, I. Khan, M. S. Springer, J. Gatesy, F. G. Hoffmann, J. C. Opazo, O. Håstad, R. H. Sawyer, H. Kim, K. W. Kim, H. J. Kim, S. Cho, N. Li, Y. Huang, M. W. Bruford, X. Zhan, A. Dixon, M. F. Bertelsen, E. Derryberry, W. Warren, R. K. Wilson, S. Li, D. A. Ray, R. E. Green, S. J. O'Brien, D. Griffin, W. E. Johnson, D. Haussler, O. A. Ryder, E. Willerslev, G. R. Graves, P. Alström, J. Fjeldså, D. P. Mindell, S. V. Edwards, E. L. Braun, C. Rahbek, D. W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang, E. D. Jarvis, M. T. P. Gilbert, J. Wang, C. Ye, S. Liang, Z. Yan, M. L. Zepeda, P. F. Campos, A. M. V. Velazquez, J. A. Samaniego, M. Avila-Arcos, M. D. Martin, R. Barnett, A. M. Ribeiro, C. V. Mello, P. V. Lovell, D. Almeida, E. Maldonado, J. Pereira, K. Sunagar, S. Philip, M. G. Dominguez-Bello, M. Bunce, D. Lambert, R. T. Brumfield, F. H. Sheldon, E. C. Holmes, P. P. Gardner, T. E. Steeves, P. F. Stadler, S. W. Burge, E. Lyons, J. Smith, F. McCarthy, F. Pitel, D. Rhoads, and D. P. Froman. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* (80-.). 346:1311–1320.
- Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.
- Zoladz, J. A., A. Koziel, I. Broniarek, A. M. Woyda-Ploszczyca, K. Ogrodna, J. Majerczak, J. Celichowski, Z. Szkutnik, and W. Jarmuszkiewicz. 2017. Effect of temperature on fatty acid metabolism in skeletal muscle mitochondria of untrained and endurance-trained rats. *PLoS One* 12:e0189456.

Appendices

Appendix A Supplementary material for Chapter 2

Table S2.1 Transformation functions for each plumage variable for historical samples that we did not have recent access to. In each function, the y was the recent plumage rescored value and x was historical plumage score).

Trait	Regression function	Parameter estimates			R ²	Pruning
		a	b	c		
Lower Flank	$y = a + b \ln(x + c)$	-24.5340	11.3090	8.6960	0.8189	If $y > 0.87$, $y = 1$; If $y < 0$, $y = 0$
Mid Flank		0.6563	5.1796	1.0681	0.7642	If $y > 0.68$, $y = 1$
Bib Corner		0.4660	10.3498	1.2917	0.5070	If $y > 0.7$, $y = 1$
Face		2.1208	0.0587	1.0148	0.9624	NA
Yellow (intensity)	$y = a \ln(b x + c)$	0.6311	3.8198	1.0675	0.8913	If $y > 1$, $y = 1$
Back		0.2322	83.5590	0.7740	0.5914	If $y > 0.78$, $y = 1$; If $y < 0$, $y = 0$
Crown		1.0133	-0.0511	NA	0.8540	If $y < 0$, $y = 0$
Yellow (extent)	$y = a x + b$	1.0424	0.0235	NA	0.9565	If $y > 1$, $y = 1$

Supplementary Method (Chapter 2)

To estimate genomic hybrid index, we estimated admixture proportion with Faststructure (Raj et al. 2014) with the input being the top 5% F_{ST} SNPs (SNP-specific F_{ST} between allopatric *townsendi* and *occidentalis* was computed with the VCFtools (Danecek et al. 2011), and top 5% F_{ST} SNPs were selected). We ran Faststructure with 10^7 iterations, with prior being logistic and $k = 2$. Then we compared individual admixture proportion with the genomic PC1 that naturally captures the variation among the hybrids in the divergence between *occidentalis* and *townsendi* (Figure 2, S1).

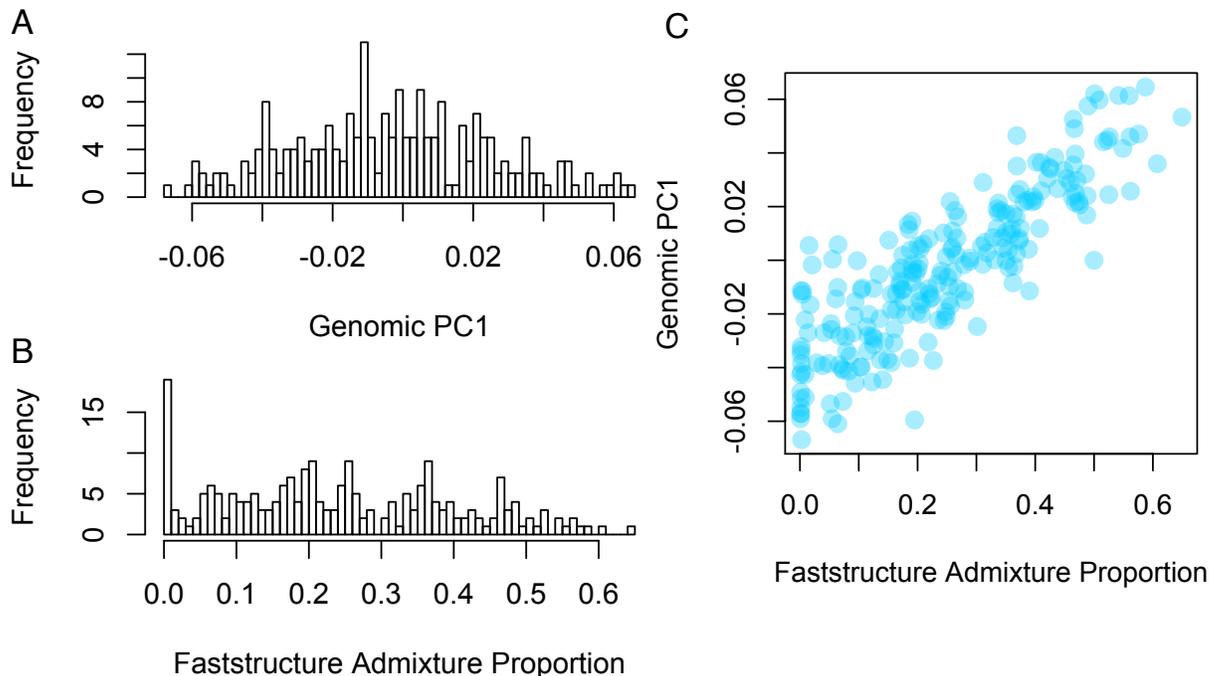


Figure S2.1 Comparison of estimates of hybrid index of individuals from the hybrid zone based on genomic PC1 (**A**) and Faststructure admixture proportion (**B**). We choose to employ genomic PC1 as the proxy of hybrid index because the Faststructure admixture proportion appears more arbitrarily binary than the genomic PC1, though these two metrics are highly correlated (**C**).

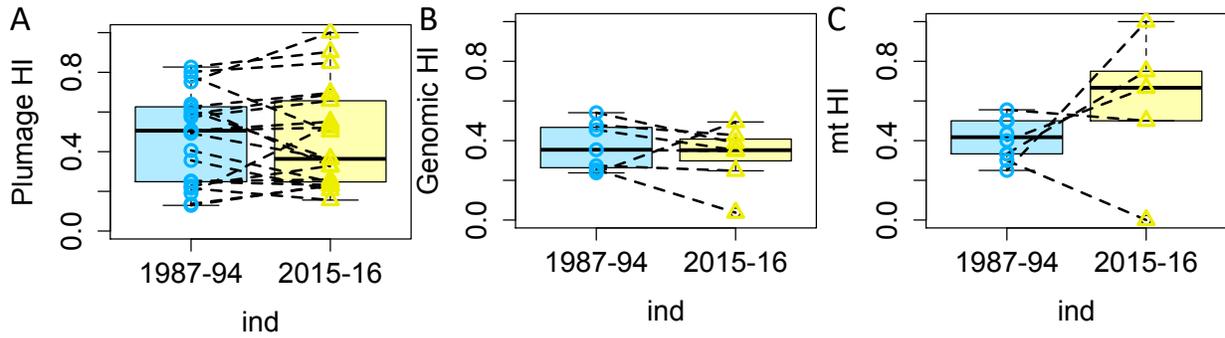


Figure S2.2 Boxplots showing site mean hybrid index (HI) across time periods in the center of the hybrid zone.

Appendix B Supplementary material for Chapter 3



Figure S3.1 Field photos of a hybrid male showing three different angles: 1) frontal with head tilted up showing throat badge and breast measurements), 2) profile showing the cheek, and 3) from above showing the crown.

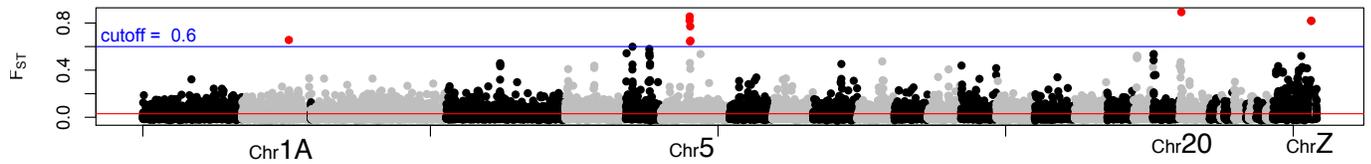


Figure S3.2 SNPs with $F_{ST} > 0.6$ are colored in red. High F_{ST} SNPs are distributed on chr1A, 5, 20, and Z.

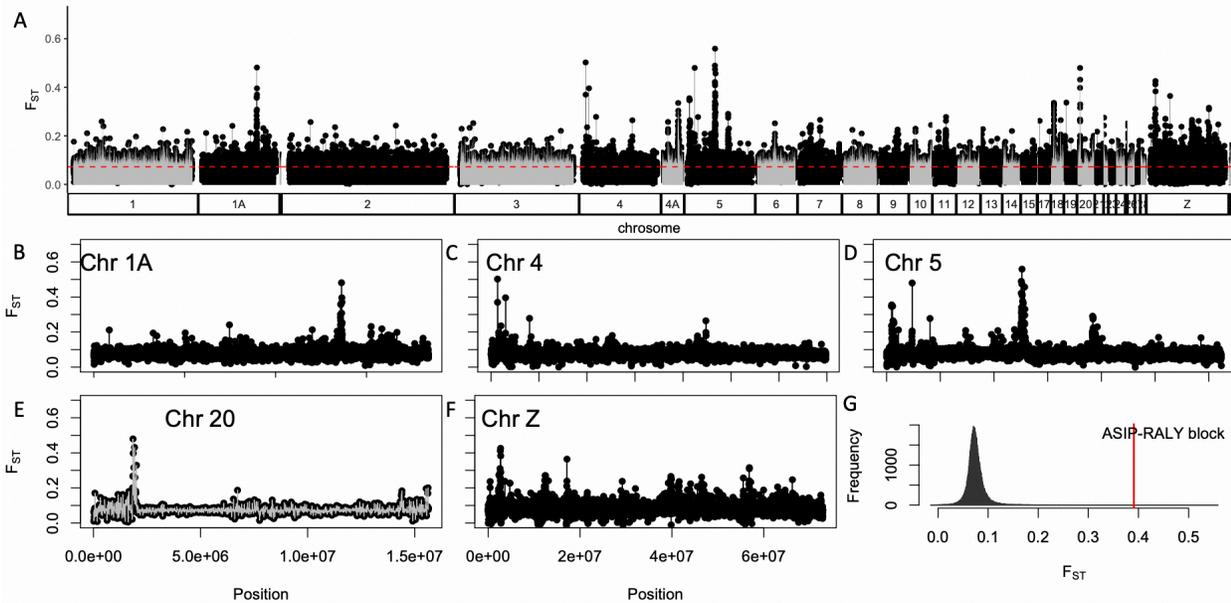


Figure S3.3 Weir Cockerham F_{ST} scan of with the WGS data, in which each dot represents a 10kb non-overlapping window (A), where peaks were found on chromosome 1A (B), 4 (C), 5 (D), 20 (E), and Z (F). G, the ASIP-RALY gene block demonstrates extremely high F_{ST} relative to the rest of the genome.

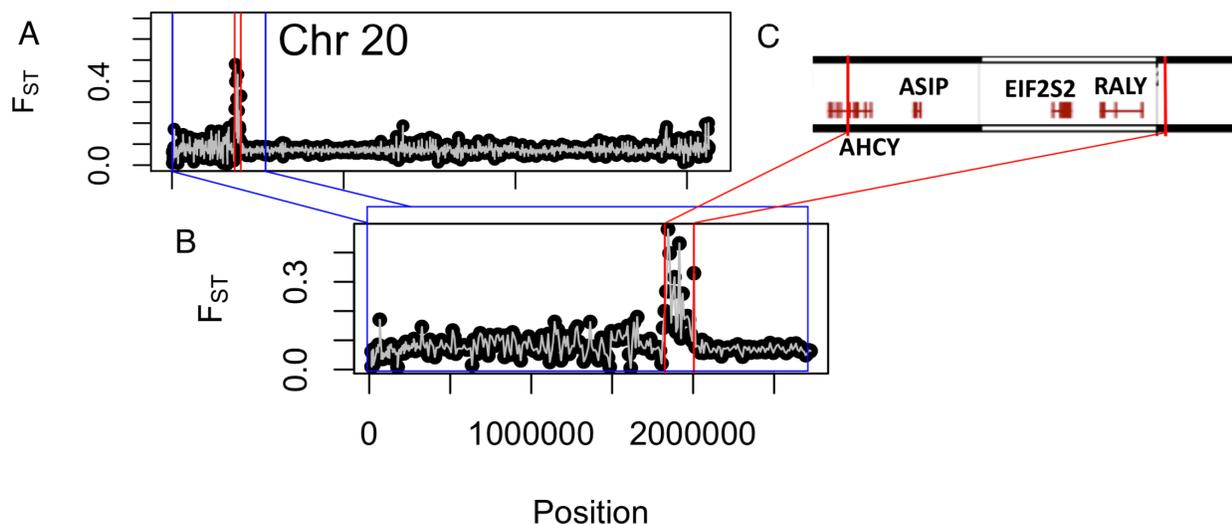


Figure S3.4 The F_{ST} peak on chromosome 20 (A) resides in the ASIP-RALY gene block (B, C). C, the position of the protein coding genes relative to the peak (bounded by the red vertical lines). C, within the genes, the vertical strikes are the coding regions flanked by the non-coding regions (horizontal lines).

Table S3.1 Allele frequencies of SNPs between RALY and ASIP in hybrid zone and parental zones.

	POS	allele1:freq	allele2:freq
Hybrid zone			
	1955244	T:0.946	A:0.054
	1972476	A:0.72	C:0.28
	1972481	T:0.917	C:0.083
	1981369	C:0.375	G:0.625
<i>townsendi</i> zone			
	1955244	T:1	A:0
	1972476	A:0.5	C:0.5
	1972481	T:0.656	C:0.344
	1981369	C:1	G:0
<i>occidentalis</i> zone			
	1955244	T:0.95	A:0.05
	1972476	A:0.944	C:0.056
	1972481	T:0.972	C:0.028
	1981369	C:0.048	G:0.952

Table S3.2 SNPs with $F_{ST} > 0.6$ and their position, association to genes and molecular functions

Chromosome	Position	FST	Relation to genes	Gene(s)	Function
1A	54442413	0.656	intergenic	GRM8, ENSTGUG00000004218	G protein-coupled receptor activity, glutamate receptor activity, telomere maintenance, DNA binding
5	25064223	0.822	intron	UBR1	ubiquitin-protein transferase activity
5	25174918	0.855	coding	ENSTGUG00000011205	DNA binding
5	25746680	0.644	intergenic	DPF3, RGS6	histone acetyltransferase activity; G protein-coupled receptor signaling pathway, intracellular signal transduction
5	25783847	0.772			
5	25875302	0.649	intron	RGS6	G protein-coupled receptor signaling pathway, intracellular signal transduction
20	1981369	0.892	intron	RALY	DNA and RNA binding, cholesterol biosynthesis
Z	66226657	0.818	intron	BBOX1, TNPO1	carnitine biosynthesis, oxidation-reduction; protein import into nucleus, intracellular protein transport

Appendix C Supplementary material for Chapter 4

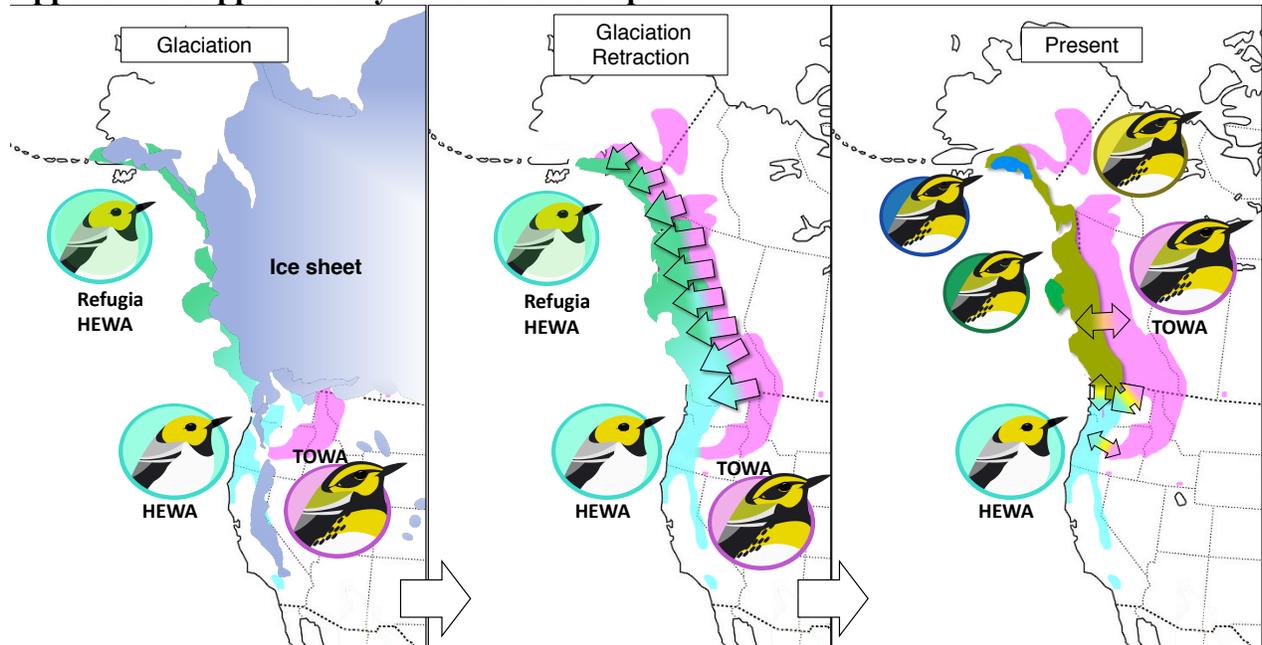


Figure S4.1 Illustration of the population differentiation and hybridization between *townsendi* and *occidentalis* during glacial expansion and retraction. Left: *occidentalis* (HEWA) distribute along the coastal North America. There might have been some north-to-west differentiation within *occidentalis* by distance. During the last glacial maxima, the *occidentalis* and *townsendi* (TOWA) populations resided in isolated glacial refugia. Center: after glacial retraction, the refugia *occidentalis* and inland *townsendi* expanded and hybridized along a broad inland-to-coastal front parallel to the coast. Right: the historical hybridization resulted in coastal *townsendi* populations with admixed mtDNA ancestry, while plumage and some nuclear genes resemble that of inland *townsendi*. There might be some population substructure within coastal *townsendi* due to refugia isolation.

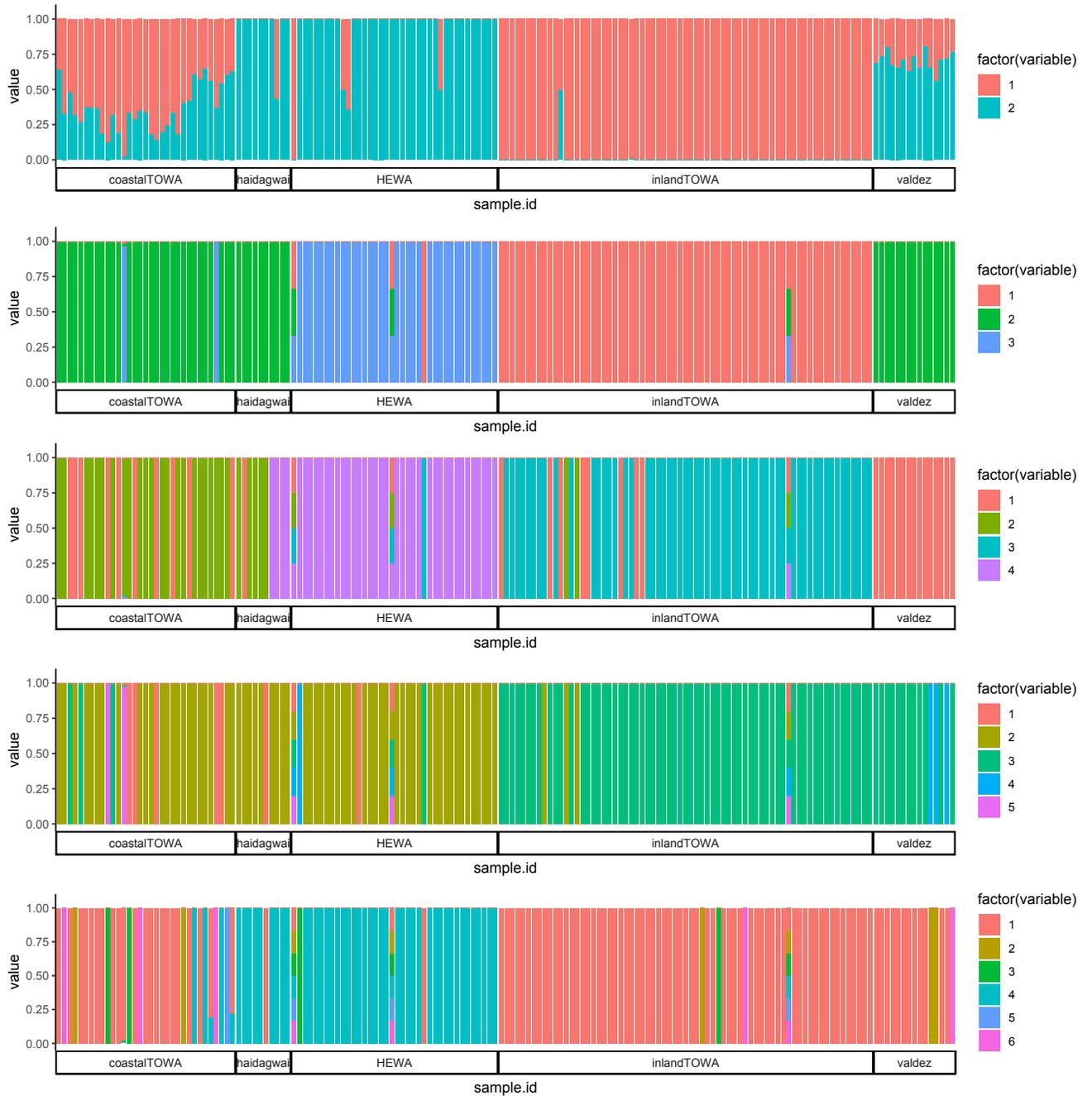


Figure S4.2 Faststructure analysis with $k = 1$ to 6 of the 165 individuals collected from *occidentalis* (HEWA), inland and coastal *townsendi* (TOWA), and Haida Gwaii as well as Valdez (coastal *townsendi*).

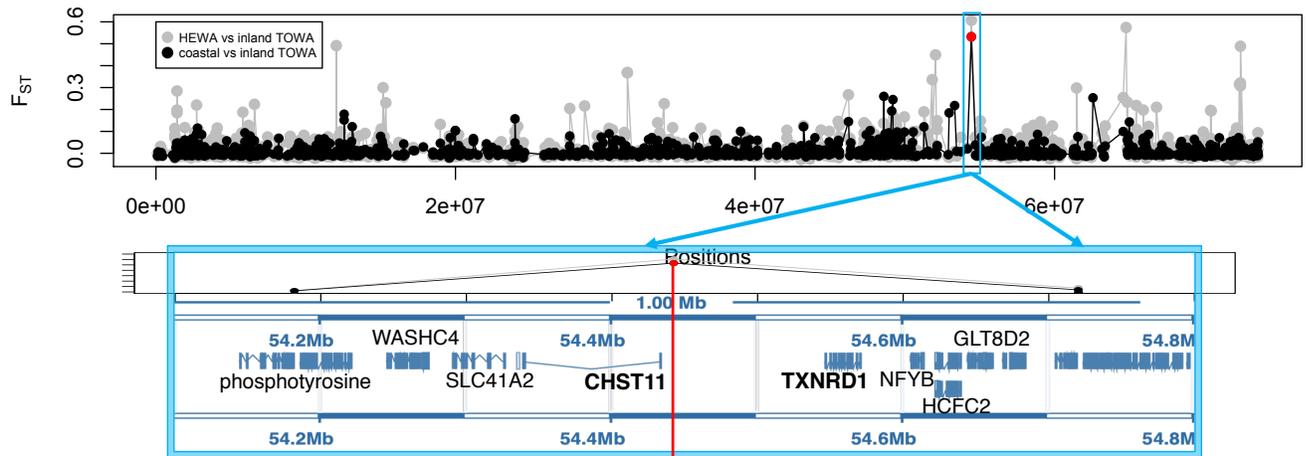


Figure S4.3 F_{ST} scan between *occidentalis* and inland *townsendi* (grey) and between inland and coastal *townsendi* (black) across chromosome 1A. The strongest F_{ST} peak (red dot) is concordant between the two comparisons (grey versus black). Zooming in around this peak (blue box), this peak is in the intergenic region between gene CHST11 and TXNRD1.

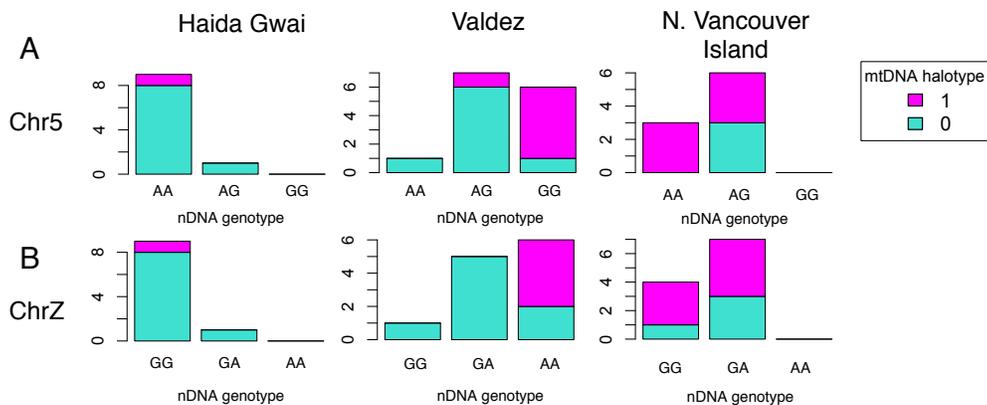


Figure S4.4 Bar plots showing the mtDNA-N ancestry association within Haida Gwaii and Valdez *townsendi* populations (A, chr5 marker; B, chr Z marker).

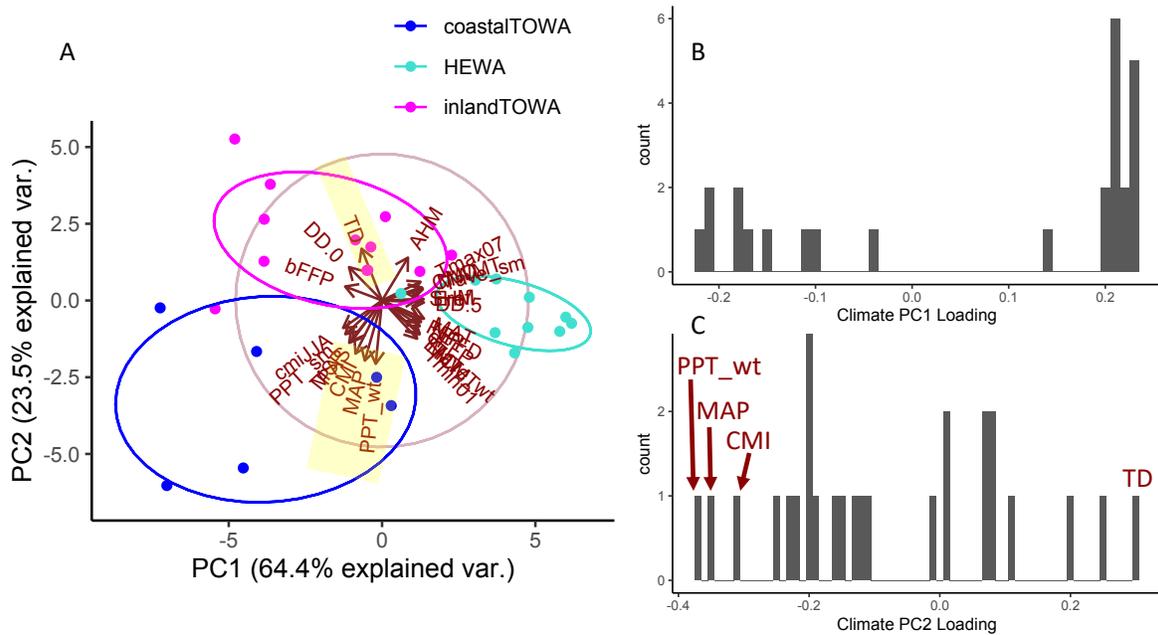


Figure S4.5 Dissecting climate PCA: **A**, biplot of PCA demonstrating loadings of the 26 climate variables in the PC space. **B**, **C**, histogram of variable loading for PC1 (**B**) and PC2 (**C**). Most of the variables demonstrates strong and even loading along PC1 (**B**), while there are 4 outstanding variables (highlighted in yellow, **A**) explaining PC2 (**C**): Temperature Difference (TD), Climate Moisture Index (CMI), Mean Annual Precipitation (MAP), Winter Precipitation (PPT_wt).

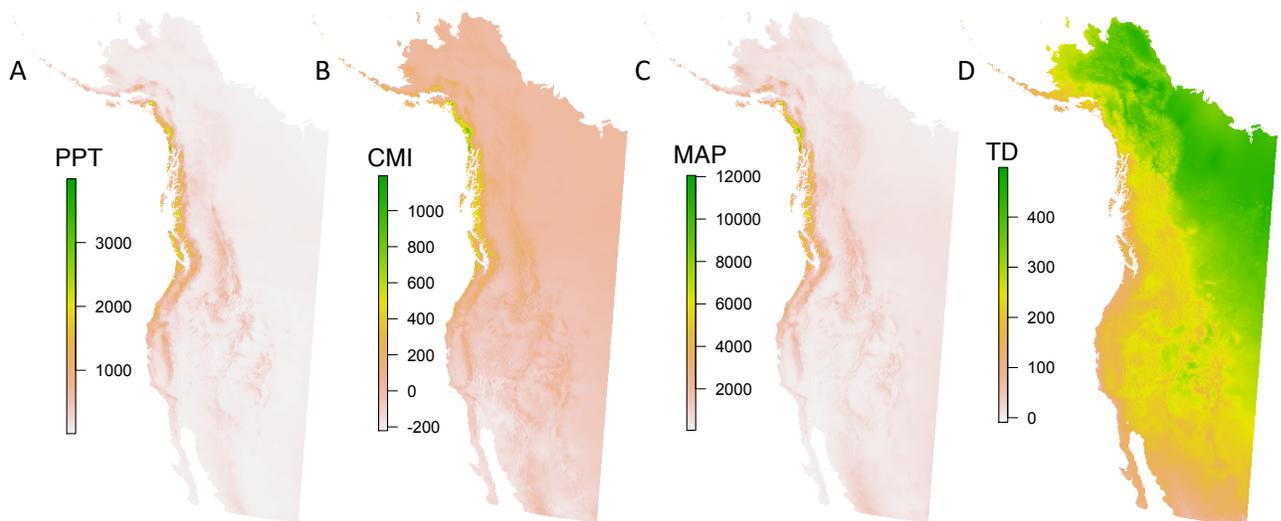


Figure S4.6 Map demonstrating spatial variation of the 4 key climate variables explaining climate PC2 (see Figure 7, S3): **A**, Winter Precipitation (PPT_wt), **B**, Climate Moisture Index (CMI), **C**, Mean Annual Precipitation (MAP), **D**, Temperature Difference (TD).