DIVERGENCE AND ADAPTATION IN BERINGIAN BIRDS

by

Caitlyn C. Oliver Brown

B.S. San José State University, 2019

A Thesis submitted in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in

Biological Sciences

University of Alaska Fairbanks

August 2024

APPROVED:

Kevin Winker, Committee Chair

Naoki Takebayashi, Committee Member

Stefanie Ickert-Bond, Committee Member

Diane Wagner, Chair

Department of Biology and Wildlife

Karsten Hueffer, Dean

College of Natural Science and Mathematics

Richard Collins, Director

Graduate School

Abstract

Beringia is a high-latitude hotspot of avian divergence and speciation. The unique biogeography of Beringia impacted avian speciation in two ways: through the cyclic appearance of the land bridge between the Asian and North American continents and through glacial refugia. These cyclic processes repeatedly split and connected avian populations, alternately reducing and increasing opportunities for gene flow between populations. In this thesis, I examine how this dynamic system impacted Beringian avian taxa using population genomic analyses. First, I examine broad patterns of divergence and gene flow across 11 lineages of birds using ultraconserved elements (UCEs), which are a multi-locus subsampling of the nuclear genome. These bird lineages contain two or more sister taxa at the population, subspecies, or species level that were likely impacted by the Bering land bridge and/or by glacial refugia. I tested models that provided key demographic information, such as population size, gene flow, and divergence time estimates. Demographic modeling showed gene flow in all cases at a wide range of rates between pairwise comparisons, and all inferred models included a divergence event during the Quaternary. Next, I focus on one species, the Song Sparrow (*Melospiza melodia*), in the Beringian part of its range. Five subspecies of the Song Sparrow reside in southern Alaska, from the Aleutian Islands to southeast Alaska, and have a wide range of body sizes. Using whole-genomic sequencing and morphology, I examine the phenotypic and genomic differences in these subspecies. I quantified the morphological differences, showing that the western subspecies are significantly larger than the eastern subspecies. I then determined that two candidate genes are under positive selection in the most isolated subspecies, *M. m. maxima*. Finally, I reconstructed a phylogeny and found that *M. m. maxima* is sister to the other *M. melodia* subspecies. These results highlight how the unique biogeography of Beringia impacted the generation of avian diversity in the region.

Acknowledgments

I would like to thank my advisor, Kevin Winker, and committee members, Steffi Ickert-Bond and Naoki Takebayashi, for their comments on this thesis and help throughout my career at UAF. Many fellow graduate students supported me during my time at UAF. A few deserve special consideration. Keiler Collier and Fern Spaulding welcomed me to the Bird Lab, taught me so much of what I know now, and coauthored one chapter. I also thank the rest of the UAMN Bird Lab: Jack Withrow, Morgan Lynn, Min Jang, and Symcha Gillette for their help, support, and enjoyable days in the prep lab. I am also forever indebted to my parents and family for always believing in me.

I thank the University of Alaska Museum and the Burke Museum of Natural History and Culture for tissue specimen loans and the museum collectors who collected specimens used in this project. Funding for these projects was provided by the Kessel Fund for Northern Ornithology and the Friends of Ornithology.

I acknowledge the Alaska Native nations upon whose ancestral lands I reside. At the University of Alaska Fairbanks, the Troth Yeddha' Campus is located on the ancestral lands of the Dena people of the lower Tanana River. Specimens used in this thesis came from across Alaska and beyond, from the native lands of Inupiat, Yup’ik/Cup’ik, Unangax̂, Sugpiaq, Dena, Dena’ina Ełnena, Eyak, Tlingit, and Haida.

Table of Contents

Copyright iii

Abstract iv

Acknowledgments v

Table of Contents vi

List of Figures ix

List of Tables x

Chapter 1: General Introduction 1

1.1 References 2

Chapter 2: Gene flow accompanies divergence in Beringian birds 4

2.1 Abstract 4

2.2 Introduction 4

2.3 Methods 6

2.3.1 Study Design 6

2.3.2 Sampling and Laboratory 7

2.3.3 Bioinformatics Pipeline 7

2.3.4 Analyses 8

2.4 Results 10

2.4.1 Population Statistics 10

2.4.2 Demographic Analyses 11

2.5 Discussion 12

2.6 References 14

2.7 Figures 20

2.8 Tables 23

Chapter 3: Evidence of positive selection and a novel phylogeny among five subspecies of song sparrow (*Melospiza melodia*) in Alaska 30
3.1 Abstract 30
3.2 Introduction 30
3.3 Methods 32
3.3.1 Phenotypic Analyses 32
3.3.2 Sampling and Whole-genome Sequencing 33
3.3.3 Bioinformatics 34
3.3.4 Selection Analyses 34
3.3.5 Phylogeny Reconstruction 35
3.4 Results 35
3.4.1 Phenotypic Analyses 35
3.4.2 Summary Statistics of Whole-genome Sequencing 36
3.4.3 Selection Analyses 36
3.4.4 Phylogenetic Reconstruction 36
3.5 Discussion 36
3.5.1 Phenotypic Variation 37
3.5.2 Selection Analyses 38
3.5.3 Phylogenetic Reconstruction 39
3.5.4 Conclusion 40
3.6 References 41
3.7 Figures 48
3.8 Tables 52
Chapter 4: General Conclusions 54
4.1 References 55

Appendix A: Supplementary Material for Chapter 2 56

Appendix B: Supplemental Protocol 73

Appendix C: Supplementary Material for Chapter 3 75

List of Figures

Fig. 2.1: Map of Beringia during Last Glacial Maximum 20
Fig. 2.2: Models of divergence tested with δaδi on two populations. 21
Fig. 2.3: Number of pairwise comparisons per best-fit and runner-up demographic models, with avian orders shown in different colors. 22
Fig. 3.1: Distribution of five subspecies of song sparrow (*Melospiza melodia*) in Alaska. 48
Fig. 3.2: Museum specimens of five *M. melodia* subspecies of interest that occur in Alaska 49
Fig. 3.3: PCA plot of phenotypic data for the five *Melospiza melodia* subspecies. 50
Fig. 3.4: Maximum-likelihood phylogeny based on 4793 UCE loci. 51

List of Tables

Table 2.1: Summary of lineages compared in this study.……..….23
Table 2.2: Between-population F_{ST} values…….....24
Table 2.3: Two-population model comparisons for each of 14 pairwise comparisons…….25
Table 2.4: Results of two-population δaδi analyses with best-fit models……..27
Table 3.1: *p*-value results from McDonald-Kreitman tests……..52
Table 3.2: Mean (± standard deviation) for the phenotypic traits compared across the five song sparrow subspecies of interest……...53

Chapter 1: General Introduction

Beringia is a high-latitude hotspot of avian divergence and speciation due to, in part, glacial cycles in the Pleistocene. This region roughly extends from the Lena River in Siberia, across the Bering and Chukchi seas to the Mackenzie River in Canada, and south to the tip of the Kamchatka Peninsula (West, 1998). Two important processes impacted avian diversity in Beringia: the cyclic appearance of the Bering Land Bridge and isolation in glacial refugia (Winker et al., 2023). The Bering-Chukchi platform was intermittently exposed and submerged due to cycles of glacial and interglacial periods (Hopkins, 1967). During the Last Glacial Maximum (LGM), sea levels dropped and exposed a land bridge that was approximately 1600 km wide, from north to south. The Bering Land Bridge enabled many species to disperse from one continent to the other and created an opportunity for previously separated populations to be reunited. Additionally, Beringian habitats were unglaciated, and large continental glaciers separated Beringia from some North American terrestrial populations, causing Beringia to be a glacial refugium for some populations (Hewitt, 2004; Hopkins, 1967; Pielou, 1991). This biogeographical history has resulted in a unique, high-latitude environment to study divergence, gene flow, and local adaptation.

In this thesis, I examine how these glacial cycles and refugia affected the diversification of modern-day Beringia birds. In Chapter 2, I examined the broad patterns of divergence and speciation in Beringia in 11 lineages from five avian orders (Oliver Brown et al., 2024b). Allopatric speciation, or speciation with no gene flow, was historically considered the dominant mode of avian speciation (Mayr, 1963; Coyne & Orr, 2004). With the proliferation of evidence from genomic sequencing data, we find that speciation can occur with gene flow (Seehausen et al., 2014). To determine the history of divergence and gene flow between avian populations, subspecies, and species, I tested demographic models to determine the ways in which they diverged and the presence or absence of gene flow and its levels (if any), using ultraconserved elements (UCEs) from the nuclear genome. I found models of divergence with gene flow were best supported across all taxa, though the specific models varied among lineages. Overall, a divergence with symmetrical migration (gene flow) model was the best supported in a majority of pairwise comparisons across all taxa.

In Chapter 3, I narrowed my focus to one species of Alaska's birds, the Song Sparrow (*Melospiza melodia*), which, in Beringia, is distributed along the southern edge of Alaska,

throughout the Aleutian Islands and to southeast Alaska. Previous studies determined that multiple avian species inhabited one or more glacial refugia in the Aleutian Islands during the Pleistocene (Holder et al., 1999; Pruett et al., 2018; Pruett & Winker, 2005). Additionally, island systems are natural experiments used to understand the genetic underpinnings of local adaptation. The Aleutian Islands of the North Pacific Ocean provide a unique system to observe local adaptation in avian species due to the relatively young and phenotypically differentiated subspecies (Pruett & Winker, 2010). Song Sparrow subspecies show a range of body sizes, with the most isolated subspecies (*M. m. maxima*) being the largest. Because of these obvious phenotypic differences, I expected to see differences in the genome. In this study, I categorized phenotypic and genomic differences between Alaska Song Sparrow subspecies (Oliver Brown et al., 2024a). I quantified the staggering morphological differences in this system. I then used a candidate gene approach to determine which genes (if any) were under positive selection. Finally, I uncovered a novel phylogenetic relationship among subspecies in this system which suggests a different colonization history for the westernmost subspecies, *M. m. maxima,* than was previously hypothesized.

1.1 References

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates Inc.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Holder, K., Montgomerie, R., & Friesen, V. L. (1999). A test of the glacial refugium hypothesis using patterns of mitochondrial and nuclear DNA sequence variation in Rock Ptarmigan (*Lagopus mutus*). *Evolution*, *53*(6), 1936–1950. https://doi.org/10.1111/j.1558-5646.1999.tb04574.x

Hopkins, D. (Ed.). (1967). *The Bering Land Bridge*. Stanford University Press.

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press. https://doi.org/10.4159/harvard.9780674865327

Oliver Brown, C. C., Collier, K. A., Mills, K. K., Spaulding, F., Glenn, T. C., Pruett, C. L., & Winker, K. (2024a). Evidence of positive selection and a novel phylogeny among five subspecies of song sparrow (*Melospiza melodia*) in Alaska. *bioRxiv*, 2024.05.21.595201. https://doi.org/10.1101/2024.05.21.595201

Oliver Brown, C. C., Glenn, T. C., & Winker, K. (2024b). Gene flow accompanies divergence in Beringian birds. [Unpublished manuscript]. Department of Biology and Wildlife, University of Alaska Fairbanks.

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

Pruett, C. L., Li, T., & Winker, K. (2018). Population genetics of Alaska Common Raven show dispersal and isolation in the world's largest songbird. *The Auk*, *135*(4), 868–880. https://doi.org/10.1642/AUK-17-144.1

Pruett, C. L., & Winker, K. (2005). Biological impacts of climatic change on a Beringian endemic: Cryptic refugia in the establishment and differentiation of the Rock Sandpiper (*Calidris ptilocnemis*). *Climatic Change*, *68*(1), 219–240. https://doi.org/10.1007/s10584-005-1584-4

Pruett, C. L., & Winker, K. (2010). Chapter 13: Alaska Song Sparrows (*Melospiza melodia*) demonstrate that genetic marker and method of analysis matter in subspecies assessments. *Ornithological Monographs*, 67(1), 162–171. https://doi.org/10.1525/om.2010.67.1.162

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176–192. https://doi.org/10.1038/nrg3644

West, F. H. (Ed.). (1998). *American Beginnings: The Prehistory and Palaeoecology of Beringia*. University of Chicago Press.

Winker, K., Withrow, J. J., Gibson, D. D., & Pruett, C. L. (2023). Beringia as a high-latitude engine of avian speciation. *Biological Reviews*, *98*(4), 1081–1099. https://doi.org/10.1111/brv.12945

Chapter 2: Gene flow accompanies divergence in Beringian birds

2.1 Abstract

The generation and maintenance of biodiversity are driven by population divergence and speciation. We investigated divergence, gene flow, and speciation in Beringia, a region at the top of the North Pacific Ocean with a history of dramatic landscape alteration through glacial cycles. These cycles repeatedly split and connected the Asian and North American continents through the Pleistocene, separating and reconnecting avian populations. Glacial refugia within Beringia also isolated some populations for a time before potentially enabling them to reunite during interglacial periods. Prior work suggests gene flow plays an important role in the divergence of Beringian birds. To improve our understanding of the generation of avian diversity in Beringia, we tested models of demographic history in 11 lineages from five orders (Anseriformes, Gaviiformes, Charadriiformes, Piciformes, and Passeriformes) using population-, subspecies-, and species-level pairwise comparisons. We sequenced an average of 3,710 ultraconserved element (UCEs) loci from the nuclear genomes of these taxa to examine genetic differentiation and test models of divergence through diffusion analysis for demographic inference (δaδi). All of the inferred best-fit models of divergence included gene flow. Together with prior work, this corroborates that divergence with gene flow is the predominant mode of divergence and speciation in Beringian birds.

2.2 Introduction

Population divergence and speciation are key elements in generating and maintaining biodiversity. Historically, allopatry was thought to be the primary mode of speciation between populations, and this process requires long-term separation without gene flow (Coyne & Orr, 2004; Mayr, 1963; Price, 2008). However, with increasing genomic data, there is evidence that speciation often occurs with gene flow (Bush, 1994; Nosil, 2008; Seehausen et al., 2014). As two populations diverge, they acquire genetic differences, but if these populations are in contact and are not reproductively isolated, gene flow can cause reduced genetic differentiation, or reticulation (Feder et al., 2012; Seehausen et al., 2014). Understanding how divergence with gene flow occurs is important for furthering our knowledge of the generation of biodiversity.

Beringia is a high-latitude hotspot of avian speciation. This region occurs at the top of the Pacific Ocean and stretches from eastern Russia, across the Bering and Chukchi seas, to western

Canada (Fig. 2.1; West, 1998). During the Pleistocene (2.6 million years ago [Mya] to 11 thousand years ago [kya]), glacial and interglacial cycles repeatedly exposed and submerged a shallow shelf between the Asian and North American continents (Hopkins, 1967; Winker et al., 2023). As these cycles connected and split the two continents repeatedly, continental populations could reunite, initiating or renewing gene flow. The land bridge was exposed an estimated nine times, though some estimates reach upwards of 20 (Hopkins, 1967). Repeated exposure would allow mobile taxa, such as birds, to reconnect and experience gene flow between these continental populations. During these cycles, glacial refugia could also become established. Much of Beringia itself was largely ice-free during glacial maxima, and smaller refugia provided isolated habitats for some populations (Fig. 2.1; Winker et al., 2023). Winker et al. (2023) suggested that these refugia might have played a more important role in avian diversification than the opening and closing of the land bridge. For example, genetic evidence suggests several bird populations, such as the common raven (*Corvus corax*), might have persisted in refugia along the Commander – Aleutian Archipelago (Pruett et al., 2018; Winker et al., 2023). These refugia split populations for a time before gene flow was reestablished. Thus far, divergence with gene flow might be the predominant mode of diversification in Beringia birds (McLaughlin et al., 2020; Spaulding et al., 2022, 2023; Winker et al., 2018).

We further test this hypothesis among 11 lineages representing a cross-section of divergence depths encompassing the speciation continuum in five avian orders (Anseriformes, Gaviiformes, Charadriiformes, Piciformes, and Passeriformes). We used the same molecular markers and analytical methods as previous studies (McLaughlin et al., 2020; Spaulding et al., 2022, 2023; Winker et al., 2018) to compare our results with prior work. We sequenced ultraconserved elements (UCEs) from these taxa and determined models of divergence using diffusion analysis for demographic inference (δaδi; Gutenkunst et al., 2009). In addition to reconstructing the divergence histories between population (or lineage) pairs, these analyses also produce important demographic estimates, such as effective population size, rates of gene flow, and divergence time. We also consider how our results inform current taxonomy with respect to species limits. Overall, we aimed to determine how different types of divergence and levels of gene flow have been involved in the divergence and speciation of Beringian birds.

2.3 Methods

2.3.1 Study Design

We chose Beringian lineages from various stages of the divergence process, from populations to subspecies and species. We predict that these lineages were affected by the glacial-interglacial cycles of the Quaternary (Winker et al., 2023). For each lineage, we chose a pair of populations, subspecies, or species, using currently accepted taxonomic designations (Gibson & Withrow, 2015). We examined populations of red-throated loon (*Gavia stellata*), common murre (*Uria aalge*), thick-billed murre (*U. lomvia*); subspecies of common merganser (*Mergus merganser merganser / M. m. americanus*), dunlin (*Calidris alpina articola / C. a. pacifica*), American pipit (*Anthus rubescens japonicus / A. r. pacificus*), eastern yellow wagtail (*Motacilla tschutschensis tschutschensis/ M. t. simillima)*; and species of Eurasian and American three-toed woodpecker (*Picoides tridactylus* / *P. dorsalis*), and black and ruddy turnstones (*Arenaria melanocephala / A. interpres*).

Additionally, for two lineages, we included pairwise comparisons between three populations. For the common raven (*Corvus corax*), we compared populations from Asia, mainland Alaska, and the Near Islands in the westernmost Aleutians (shortened to Attu henceforth). These represent three subspecies: *C. c. pricipalis* (North America), *C. c. kamtschaticus* (Russian Far East, Aleutian Islands), and *C. c. corax* (central and western Russia, Mongolia*;* Boarman & Heinrich, 2020). Our *C. corax* samples from Asia came from two separate subspecies (nominate *corax* and *C. c. kamtschaticus*), but we treated these samples as a single population. Prior work showed that the populations of ravens on the Near Islands of the westernmost Aleutians are genetically distinct from other populations (Pruett et al., 2018). We also examined comparisons of arctic warbler subspecies (*Phylloscopus borealis borealis / P. b. kennicotti*) and between the arctic warbler and the Kamchatka leaf warbler (*P. b. borealis / P. examinandus*), which were recently split into two species (Alström et al., 2011; Lowther & Sharbaugh, 2020). In total, we examine 14 pairwise comparisons, across four populations, seven subspecies, and three species.

We used ultraconserved elements (UCEs), orthologous loci from the nuclear genome, as our molecular markers to reconstruct the history of divergence in each lineage. Using orthologous loci, such as UCEs, enables us to compare divergence among multiple lineages and to examine divergence at both shallow and deeper levels (Faircloth et al., 2012; Harvey et al., 2016). We aimed to include a minimum of five individuals in each taxon but, due to DNA quality, some

sample groups contained only four individuals. We used high coverage and called both alleles, which effectively doubles sample size. It has been demonstrated that the parameter estimates in lineages with divergence levels of subspecies and species are generally resilient to even lower sample sizes (McLaughlin & Winker, 2020).

To compare these results with prior studies, we used the same molecular markers and analytical methods. This includes using ultraconserved elements (UCEs) as a molecular marker and similar data analyses using diffusion analysis for demographic inference (δaδi v2.3.0; Gutenkunst et al., 2009). This enables the results from prior studies using the same methods (McLaughlin et al., 2020; Spaulding et al., 2022, 2023; Winker et al., 2018) to be directly comparable. Although we used similar data and analytical methods, we developed entirely new datasets for avian taxa that had not yet been studied in this way, including lineages from two previously unexamined avian orders (Gaviiformes and Piciformes).

2.3.2 Sampling and Laboratory

Samples were extracted from muscle tissue of vouchered museum specimens from the University of Alaska Museum and the University of Washington Burke Museum (Table A.1). We extracted DNA using the QIAGEN DNeasy Blood + Tissue Extraction Kit following manufacturer's protocol (QIAGEN, 2006).

Our library preparation and sequencing procedures followed the iTru methods of Glenn et al. (2019). In short, we prepared Illumina dual-indexed DNA libraries which were checked for size distribution on 1.5% agarose gels and quantified using a Qubit fluorimeter (Invitrogen, Inc., Carlsbad, CA, USA). Libraries that were of insufficient quality or quantity using iTru methods were repeated using iNextEra protocols (Appendix B). We then enriched pools of 8-12 samples with 5,060 UCE loci using the Tetrapods-UCE 5Kv1 kit from MYcroarray following manufacturer's protocol version 5.01 (https://arborbiosci.com/mybaits-manual). The resulting pools were then sequenced on an Illumina HiSeq 2500 to obtain paired-end 150 base (PE150) reads. FASTQ files are deposited on NCBI Sequence Read Archive under BioProject PRJNA1129545.

2.3.3 Bioinformatics Pipeline

Our pipeline followed Winker et al. (2018) with two modifications. In short, raw and untrimmed FASTQ sequences were cleaned using Illumiprocessor (v.2.0, Faircloth, 2013) which

incorporates Trimmomatic (Bolger et al., 2014). Next, we used PHYLUCE (v. 1.7.1; Faircloth, 2016) to identify ultraconserved elements in reference sequences. To make reference sequences, we chose two individuals from each lineage (e.g., one from each population, subspecies, or species) and concatenated the raw files together. References were assembled *de novo* using SPAdes (Bankevich et al., 2012) on the online server Galaxy (Afgan et al., 2016). PHYLUCE was used to identify UCE loci in the reference using the probe set provided here: https://raw.githubusercontent.com/faircloth-lab/uce-probe-sets/master/uce-5k-probe-set/uce-5k-probes.fasta. Next, for each individual, we combined singleton files with READ1 FASTQ files. We used BWA-MEM (Li, 2013; Li & Durbin, 2009) and SAMTools (Li et al., 2009) to index the reference files and then aligned the reads from each individual's FASTQ files to the reference. We used Genome Analysis Toolkit (GATK v. 4.2.5, McKenna et al., 2010) and HaplotypeCaller to call single nucleotide polymorphisms (SNPs) and restricted called SNPs to high-quality sites (Q30).

With the resulting VCF file, we used VCFTools (v. 0.1.16; Danecek et al., 2011) to reduce our data set to a minimum genotype quality (GQ) of 10. The high-quality VCF file was made biallelic by filtering out loci that had more than two alleles and then thinned to one SNP per locus. For our demographic analyses, we removed loci associated with the Z chromosome because sex chromosomes have a different inheritance pattern than autosomes (Garrigan et al., 2007; Jorde et al., 2000). To do this, we used BLASTn through the command line to identify sex-linked (Z-linked) loci in chromosome-level reference sequences. The reference sequences used were different depending on the focal taxon (Table A.2). These putatively Z-linked loci were removed from our thinned, biallelic VCF file using the script find_chrom.py (https://github.com/jfmclaughlin92/beringia_scripts).

2.3.4 Analyses

To calculate population statistics, we used VCFTools to calculate coverage depths and both SNP- and locus-specific F_{ST}. For other population statistics, we converted the biallelic VCF file to an appropriate file type using PGDSpider (Lischer & Excoffier, 2012) before using the R package "adegenet" in RStudio (Jombart & Ahmed, 2011; v. 2022.12.0.353; Posit Team, 2022; R Core Team, 2022). "Adegenet" was used to test for Hardy-Weinberg equilibrium in each

population, calculate observed and expected heterozygosity, calculate Nei's pairwise F_{ST}, and determine assignment probabilities using discriminant analysis of principal components.

We used diffusion analysis for demographic inference (δaδi v2.3.0; Gutenkunst et al., 2009) to identify best-fit demographic models. For demographic analyses, we used the thinned VCF file with Z-linked loci removed. We tested eight models of divergence: A) "neutral", non-divergence model; B) "split with no migration" (divergence without gene flow); C) "split with migration" (divergence with gene flow that is bidirectionally symmetric); D) "split with bidirectional migration" (divergence with gene flow that is bidirectionally asymmetric; E) "secondary contact with migration" (after a period of no contact, migration is re-established); F) "secondary contact with bidirectional migration"; G) split with exponential population growth and no migration ("island"); H) split with exponential population growth and migration ("IM"; Fig. 2.2). Parameters were optimized by running each model repeatedly until the highest maximum log composite likelihood value was observed across multiple runs with the same parameter bounds. Following this series of optimization runs, the best five log-likelihood scores from each set of runs were averaged to summarize that model, and we used the Akaike Information Criterion (AIC) to determine the best-fit model. We then ran the best-fit model with 100 bootstrap replicate data sets to provide a 95% confidence interval around each parameter estimated. Scripts are available at https://github.com/coliverbrown/uce-dadi-script.

To interpret δaδi output in biological terms, we followed Winker et al. (2018). We obtained estimates of substitutions per site per generation using BLASTn to compare our reference FASTA sequences to a closely related genome on NCBI and using time since most recent common ancestor estimates from Claramunt & Cracraft (2015; Table A.3). Generation time was defined as $G = \alpha + (s/(1-s))$, where α is the age of first breeding and s is annual survival, following Sæther et al. (2005). α and s were determined from the literature (Table A.3). Calculations for substitution rates, generation time, and adjusted length of sequences surveyed were used with the best-fit model parameter estimates to provide biological estimates of ancestral population size (N_{ref}), size of populations (*nu1, nu2*), time since split (T), time since secondary contact (T_{sc}), migration (gene flow in individuals/generation, derived from m), migration from population 1 into population 2 (individuals/generation, derived from *m21*), migration from population 2 to population 1 (individuals/generation, derived from *m12*), and size of population after split (S for population 1; 1-S for population 2). Ancestral population size was

derived from the output of θ from δaδi, where $\theta = 4*(N_{ref})*$(substitution rate)*(adjusted length of sequences).

We also analyzed the collective results of similar studies of avian divergence in this region (this study; McLaughlin et al., 2020; Spaulding et al., 2022, 2023; and Winker et al., 2018). We tallied best-fit and runner-up models by avian order and plotted the totals in R (v. 4.4.0; R Core Team, 2021) and RStudio (v. 2022.12.0.353; Posit Team, 2022) using the package "tidyverse" (v.1.3.2; Wickham et al., 2019). To accommodate the lower sample sizes that would occur with overly narrower bins, we binned the models into model families ("split" (Fig. 2.2C; 2.2D), "secondary contact" (Fig. 2.2E; 2.2F), and "IM" (Fig. 2.2H). We also binned the avian orders into waterbirds (Anseriformes, Charadriiformes, Gaviiformes) and landbirds (Piciformes and Passeriformes). We ran two Chi-squared tests, one to determine whether one or more model families were represented more than expected and the other to determine whether model families differed between waterbirds and landbirds. We performed these Chi-squared tests in R (v. 4.4.0; R Core Team, 2021). If the Chi-squared test suggested a significant difference, we ran a pairwise comparison post-hoc test through the R package "RVAideMemoire" (v. 0.9-83-7; Herve, 2024).

2.4 Results

2.4.1 Population Statistics

We obtained >700 million reads, ranging from 2,613,961 to 18,027,183 (average = 6,100,217) reads per individual. Our reference assembly sequences from two individuals in each lineage produced 54,127 to 320,689 contigs, with a total of 22,798,932 – 56,200,041 bp (Table A.4). We identified between 2,147 – 4,131 UCE loci per reference, with an average contig length between 1,197 – 1,623 bp (Table A.5).

Coverage across all SNPs averaged 39.9×, ranging between 32.5× and 47.9× coverage (Table 2.1). Expected heterozygosity (H_E) ranged from 0.008 to 0.26, and observed heterozygosity (H_O) ranged between 0.009 to 0.26 (Table A.6). In eight pairwise comparisons there were significant differences between H_E and H_O (*Mergus merganser* subspp., *Arenaria interpres/A. melanocephala, Picoides tridactylus/P. dorsalis, Corvus corax* all populations*, P. borealis borealis / P. examinandus*, *Anthus rubescens* subspp., and *Motacilla tschutschensis* subspp.; Table A.6). For all but one of these, H_E was larger than H_O, which suggests hybridization (Table

A.6). The other comparison (*C. corax* Asia–Attu), had lower H_E than H_O, which means there is less genetic diversity than expected (Table A.6).

F_{ST} values ranged from 0.0008 to 0.65 and were higher in subspecies and species lineage pairs (Table 2.2). In nine pairwise comparisons F_{ST} values were significantly non-zero, suggesting differentiation (*M. merganser* subspp., *A. interpres / A. melanocephala*, *Calidris alpina* subspp., *G. stellata*, *P. tridactylus / P. dorsalis*, *C. corax* North America–Attu, *P. borealis borealis / P. examinandus*, *A. rubescens* subspp., and *M. tschutschensis* subspp.; Table 2.2). The remaining five lineage pairs had F_{ST} values that were insignificant, which suggests no differentiation between populations using this metric for these data (*Uria aalge, U. lomvia*, C. *corax* Asia–North America, *C. corax* Asia–Attu, and *P. borealis* subspp.; Table 2.2).

2.4.2 Demographic Analyses

Although the best-fit models varied among lineages, all models included gene flow (Table 2.3). Our no-gene-flow or neutral models were not supported in any comparison. In all but two comparisons, there was a single, unambiguously supported best-fit model (i.e., $\Delta AIC > 10$; Table 2.3). *Calidris alpina* subspp. had two statistically equivalent best-fit models: divergence with symmetrical migration (gene flow) and divergence with secondary contact and symmetrical migration (Fig. 2.2C and 2.2E, respectively; Table 2.3). The two *Phylloscopus* spp. had both secondary contact models as the best fit (Fig. 2.2E, 2.2F; Table 2.3).

The most common model of divergence among these taxa was divergence with symmetrical migration (split-migration; Fig. 2.2C; Table 2.3). This model was the best fit for seven pairwise comparisons: *G. stellata, C. alpina* subspp.*, U. aalge, U. lomvia, C. corax* Attu–North America, *C. corax* Asia-North America, and *P. borealis* subspp. In three lineages the best-fit model was the isolation with exponential growth and migration model (*P. tridactylus/ P. dorsalis, A. rubescens* subspp., and *A. interpres/ A. melanocephala*; Fig. 2.2H; Table 2.3). Divergence with secondary contact was the best-fit model for five comparisons, either with one migration parameter, indicating symmetrical gene flow (*P. examinandus / P. b. borealis, C. alpina* subspp.; Fig. 2.2E, Table 2.3), or with two migration parameters, indicating asymmetric gene flow (*P. examinandus / P. b. borealis*, *M. tschutschensis* subspp., *M. merganser* subspp.; Fig. 2.2F, Table 2.3). Finally, one comparison was supported by the divergence with asymmetrical migration model (*C. corax* Asia–Attu; Fig. 2.2D, Table 2.3).

Effective ancestral population size (N_{ref}) estimates ranged from 6,823 (*C. corax* Attu–North America) to 501,197 (*M. tschutschensis* subspp.; Table 2.4). Estimates of current effective population sizes ranged from 10,380 (*U. aalge*, Asia) to 3,565,675 (*P. b. borealis;* Table 2.4). For the three comparisons where the isolation with exponential growth and migration model (Fig. 2.2H) was the best fit, estimates of size after divergence ranged from 5,195 (*Picoides* spp.) to 40,670 (*A. rubescens* subspp.; Table 2.4). Time-since-divergence-estimates ranged from 50,296 yr (*C. corax* Attu–North America) to 3,320,557 yr (*C. corax* Asia–North America; Table 2.4). In the pairwise comparisons in which the secondary contact models were the best fit, time of secondary contact ranged from 9,028 yr (*M. merganser* subspp.) to 1,185,359 yr (*M. tschutschensis* subspp.; Table 2.4). Estimates of migration ranged from 0.006 inds. /gen. (*Arenaria* spp.) to 206.5 inds. /gen. (*P. borealis* subspp.; Table 2.4).

Including this and previous studies, a total of 30 pairwise comparisons of Beringian avian taxa were analyzed using δaδi. These resulted in a total of 46 best-fit and runner-up models (Fig. 2.3). The split with symmetrical migration model was the most supported model (18 out of 46; Fig. 2.3), and variation among model families was significant ($\chi^2 = 15.70$, $df = 2$, p-value = 0.0004). Supported models were independent of avian classification ($\chi^2 = 4.35$, $df = 2$, p-value = 0.11). The split and secondary contact model families were supported significantly more than the IM model family (pairwise post-hoc test; p-value = 0.00016, 0.00097, respectively).

2.5 Discussion

We surveyed 11 Beringian bird lineages and made 14 pairwise contrasts from populations to full species to determine the modes of divergence that predominate in this region. Although these lineages were from five avian orders and thus have different life histories, we found similar divergence patterns across all taxa. Each best-fit demographic model showed gene flow occurring after the split or during a period of secondary contact (Fig. 2.2C-F, 2.2H). While we expected gene flow to be present between populations, we found gene flow was present in all comparisons, including between full species (Table 2.4). Gene flow is thus occurring at all stages of divergence, and strict allopatry is not the primary mode of divergence and speciation among avian taxa in Beringia.

Contrasting our results with similar work on avian divergence in Beringia (see McLaughlin et al., 2020; Spaulding et al., 2022, 2023; Winker et al., 2018) shows that the split-with-

migration model of divergence (with one migration parameter; Fig. 2.2C) is the predominant mode of diversification in this region. In 46 pairwise comparisons of Beringian birds, 18 were best supported by the split-with-migration model (Fig. 2.3). Symmetrical migration is likely due to similar dispersal abilities between each of the two populations, whereas asymmetrical migration could be due to source/sink dynamics (Oswald et al., 2017). The IM (isolation with migration and exponential growth, Fig. 2.2H) is the least commonly supported model. Our analyses showed no difference in the number of split-with-migration and secondary-contact best-fit models but a significant difference between split-with-migration and IM models. The IM model includes an exponential increase in effective population size (N_e) after divergence (Fig. 2.2H), whereas the split-with-migration and secondary-contact models assume no changes in effective population sizes (Fig. 2.2B-F). N_e affects the rate of change in a population caused by genetic drift and can be influenced by population size, genetic structure, and inbreeding, among other things (Charlesworth, 2009; Wright, 1931). While we cannot determine why only three pairwise comparisons were supported by the IM model (*Arenaria* spp., *Picoides* spp., *A. rubescens* subspp.; Table 2.3), it is likely that in these taxa some event caused a dramatic increase to population sizes after the initial split.

Divergence time estimates placed the population splits during the Pleistocene (Table 2.4). This dynamic period would have connected and separated populations across Beringia, thus causing genetic differentiation in many avian taxa. However, we cannot determine which glacial cycles corresponded with each divergence event for several reasons. Our divergence time estimates differ greatly from mitochondrial DNA (mtDNA) estimates from prior research. For example, Zink et al. (1995) estimated *P. tridactylus* and *P. dorsalis* split around 2.75 Mya, whereas our estimates placed the divergence event around 160 Kya (Table 2.4). Gene flow has also likely affected these estimates, resulting in younger divergence time estimates (McLaughlin et al., 2020; Winker et al., 2023). Finally, avian divergence and speciation generally requires time spans that include multiple glacial-interglacial cycles (Price 2008). However, even though our divergence times are estimates, we can confidently state that Pleistocene glacial cycles were involved in the divergence of Beringian avian populations.

Current taxonomy is not completely in agreement with our genomic results. The two subspecies of *A. rubescens* should theoretically have the opportunity for gene flow in Beringia, but we found gene flow to be substantially less than one individual per generation (Table 2.4;

Fig. A.1). Past work using mtDNA showed a relatively deep divergence between the *japonicus* and *pacificus* subspecies and suggested splitting them into two species (Doniol-Valcroze et al., 2023; Zink et al., 1995). Considering our multilocus results together with these other studies, we support splitting these pipits into two species, *Anthus japonicus* and *A. rubescens*.

Our *Corvus corax* populations and subspecies showed unique clustering patterns. The principal components analysis initially showed a clustering of all individuals except that two Asian samples were outliers (Fig. A.2A). When the outliers were removed, the Asian and mainland North American individuals were clustered together, and the Attu individuals were separate (Fig. A.2B). Additionally, we estimate less than one individual per generation migrating between North America and Attu (Table 2.4). Because *C. corax* has little morphological variation between populations (Boarman & Heinrich, 2020), we would expect moderate to high levels of gene flow between populations and subspecies. Yet, our results align with prior research showing that *C. corax* on Attu is isolated from other North American populations (Pruett et al., 2018). This confirms that the Attu population of *C. corax* is an isolated population.

We demonstrated that gene flow had a profound impact on divergence among Beringia's birds. The geological history of Beringia, including the cyclic appearance of the land bridge and the presence of glacial refugia, caused populations to become isolated and then reconnect and possibly establish or reestablish gene flow. Divergence with symmetrical gene flow was the best demographic model of lineage history in half of our pairwise comparisons (seven of 14; Table 2.3), and in over a third (18 of 46; Fig. 2.3) of the comparisons made in all similar studies in this system. Strict allopatric divergence (i.e., isolation with no gene flow throughout divergence history) is not present in Beringian birds studied thus far.

2.6 References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., & Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, *44*(W1), W3–W10. https://doi.org/10.1093/nar/gkw343

Alström, P., Saitoh, T., Williams, D., Nishiumi, I., Shigeta, Y., Ueda, K., Irestedt, M., Björklund, M., & Olsson, U. (2011). The Arctic Warbler *Phylloscopus borealis*– three anciently separated cryptic species revealed. *Ibis*, *153*(2), 395–410. https://doi.org/10.1111/j.1474-919X.2011.01116.x

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Boarman, W. I., & Heinrich, B. (2020). Common Raven (*Corvus corax*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.comrav.01species_shared.bow.project_name

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bush, G. L. (1994). Sympatric speciation in animals: New wine in old bottles. *Trends in Ecology & Evolution, 9*(8), 285–288. https://doi.org/10.1016/0169-5347(94)90031-0

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195–205. https://doi.org/10.1038/nrg2526

Claramunt, S., & Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, *1*(11), e1501005. https://doi.org/10.1126/sciadv.1501005

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates Inc.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Doniol-Valcroze, P., Coiffard, P., Alström, P., Robb, M., Dufour, P., & Crochet, P.A. (2023). Molecular and acoustic evidence support the species status of *Anthus rubescens rubescens* and *Anthus [rubescens] japonicus* (Passeriformes: Motacillidae). *Zootaxa*, *5343*(2), 173-192. https://doi.org/10.11646/zootaxa.5343.2.4

Faircloth, B. C. (2013). illumiprocessor: A trimmomatic wrapper for parallel adapter and quality trimming. http://dx.doi.org/10.6079/J9ILL

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, *28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Garrigan, D., Kingan, S. B., Pilkington, M. M., Wilder, J. A., Cox, M. P., Soodyall, H., Strassmann, B., Destro-Bisol, G., de Knijff, P., Novelletto, A., Friedlaender, J., & Hammer, M. F. (2007). Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics*, 177(4), 2195–2207. https://doi.org/10.1534/genetics.107.077495

Gibson, D. D., & Withrow, J. J. (2015). Inventory of the species and subspecies of Alaska birds, second edition. *Western Birds*, *46*(2), 94–185.

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., Leon, F. J. G.-D., Portilla, M. A. del R., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Belanger, M., … Faircloth, B. C. (2019). Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ*, *7*, e7755. https://doi.org/10.7717/peerj.7755

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, *65*(5), 910–924. https://doi.org/10.1093/sysbio/syw036

Hendricks, P., & Verbeek, N. A. (2020). American Pipit (*Anthus rubescens*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.amepip.01species_shared.bow.project_name

Herve, M (2023). RVAideMemoire: Testing and plotting procedures for biostatistics. R package version 0.9-83-7, https://CRAN.R-project.org/package=RVAideMemoire

Hopkins, D. M. (Ed.). (1967). *The Bering Land Bridge*. Stanford University Press.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., & Batzer, M. A. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *The American Journal of Human Genetics,* 66(3), 979–988. https://doi.org/10.1086/302825

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. https://doi.org/10.48550/arXiv.1303.3997

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Lowther, P. E., & Sharbaugh, S. (2020). Arctic Warbler (*Phylloscopus borealis*), version 1.0. In S. M. Billerman (Ed.), *Birds of the World*. Cornell Lab of Ornithology. https://doi.org/10.2173/bow.arcwar1.01

Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequence data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

McLaughlin, J. F., Faircloth, B. C., Glenn, T. C., & Winker, K. (2020). Divergence, gene flow, and speciation in eight lineages of trans-Beringian birds. *Molecular Ecology*, 1–17. https://doi.org/10.1111/mec.15574

McLaughlin, J. F., & Winker, K. (2020). An empirical examination of sample size effects on population demographic estimates in birds using single nucleotide polymorphism (SNP) data. *PeerJ*, *8*, e9939. https://doi.org/10.7717/peerj.9939

Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology, 17*(9), 2103–2106. https://doi.org/10.1111/j.1365-294X.2008.03715.x

Oswald JA, Overcast I, Mauck III WM, Andersen MJ, Smith BT. 2017. Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. *Molecular Ecology* 26:1386–1400. https://doi.org/10.1111/mec.14013

Price, T. (2008). *Speciation in Birds*. Roberts and Company Publishers.

Pruett, C. L., Li, T., & Winker, K. (2018). Population genetics of Alaska Common Raven show dispersal and isolation in the world's largest songbird. *The Auk*, *135*(4), 868–880. https://doi.org/10.1642/AUK-17-144.1

R Core Team. (2022). R: A language and environment for statistical computing. (4.1.1). R Foundation for Statistical Computing. https://www.R-project.org/

Sæther, B.-E., Lande, R., Engen, S., Weimerskirch, H., Lillegård, M., Altwegg, R., Becker, P. H., Bregnballe, T., Brommer, J. E., McCleery, R. H., Merilä, J., Nyholm, E., Rendell, W., Robertson, R. R., Tryjanowski, P., & Visser, M. E. (2005). Generation time and temporal scaling of bird population dynamics. *Nature*, *436*(7047), 99–102. https://doi.org/10.1038/nature03666

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176–192. https://doi.org/10.1038/nrg3644

Spaulding, F., McLaughlin, J. F., Cheek, R. G., McCracken, K. G., Glenn, T. C., & Winker, K. (2023). Population genomics indicate three different modes of divergence and speciation with gene flow in the green-winged teal duck complex. *Molecular Phylogenetics and Evolution*, *182*, 107733. https://doi.org/10.1016/j.ympev.2023.107733

Spaulding, F., McLaughlin, J. F., Glenn, T. C., & Winker, K. (2022). Estimating movement rates between Eurasian and North American birds that are vectors of avian influenza. *Avian Diseases*, *66*(2), 155–164. https://doi.org/10.1637/aviandiseases-D-21-00088

West, F. H. (Ed.). (1998). *American Beginnings: The Prehistory and Palaeoecology of Beringia*. University of Chicago Press.

Winker, K., Glenn, T. C., & Faircloth, B. C. (2018). Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*, *6*, e5735. https://doi.org/10.7717/peerj.5735

Winker, K., Withrow, J. J., Gibson, D. D., & Pruett, C. L. (2023). Beringia as a high-latitude engine of avian speciation. *Biological Reviews*, *98*(4), 1081–1099. https://doi.org/10.1111/brv.12945

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.

Zink, R. M., Rohwer, S., Andreev, A. V., & Dittmann, D. L. (1995). Trans-Beringia comparisons of mitochondrial DNA differentiation in birds. *The Condor*, *97*(3), 639–649. https://doi.org/10.2307/1369173

2.7 Figures

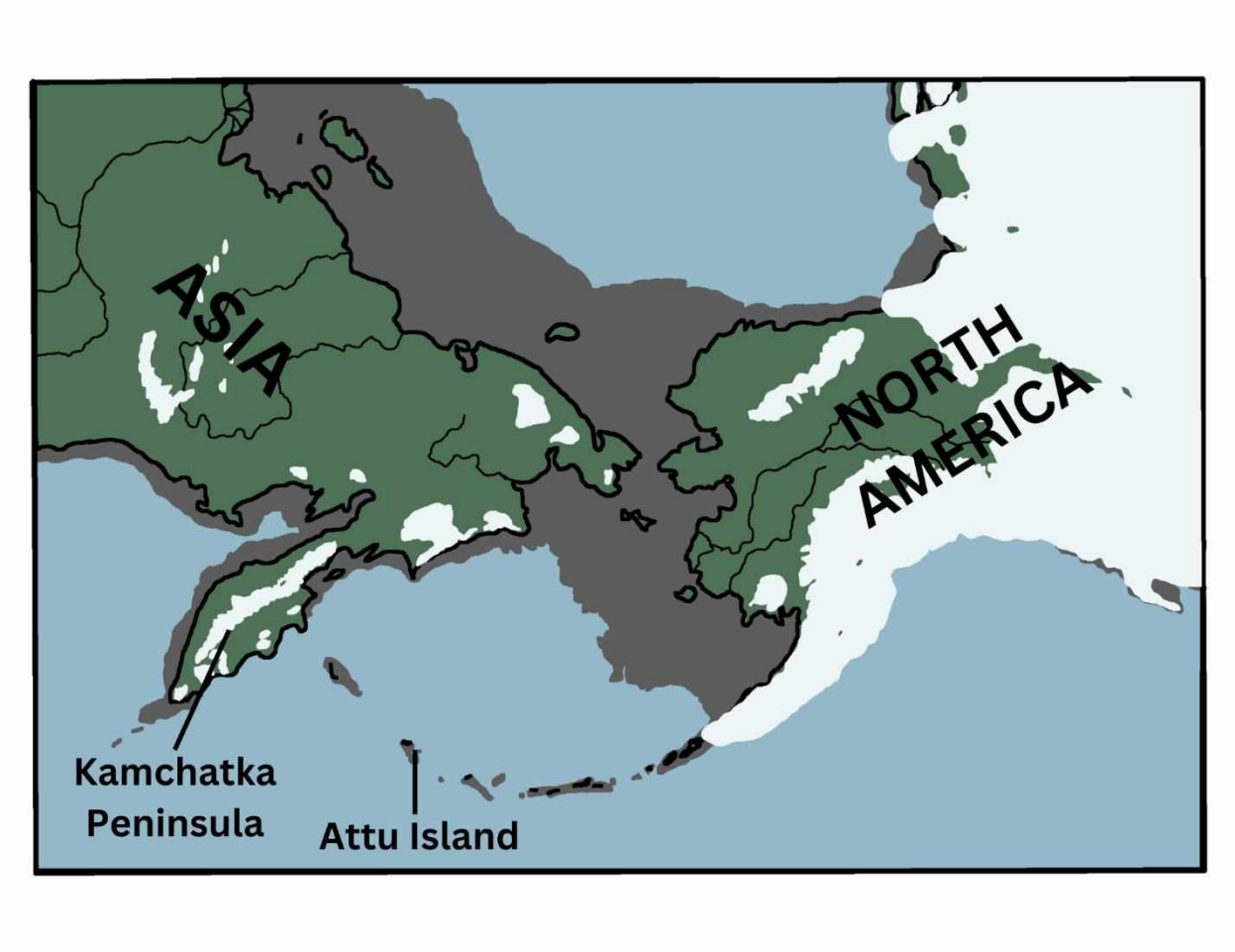

Fig. 2.1: Map of Beringia during Last Glacial Maximum (LGM; About Beringia [U.S. National Park Service], 2023). Green areas denote modern-day land, gray shading denotes shoreline during the LGM, and white areas denote glaciation. Locations of interest (Attu Island and Kamchatka Peninsula) are shown.

.

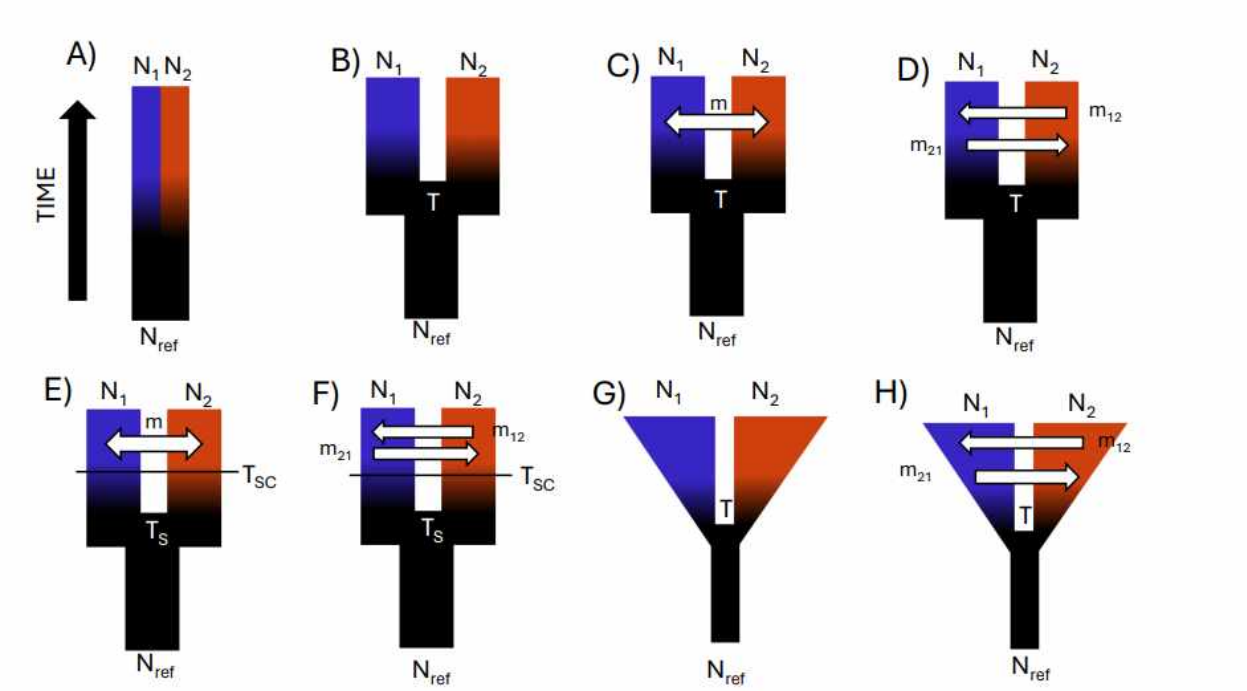

Fig. 2.2: Models of divergence tested with δaδi on two populations. A) neutral (populations never diverge); B) split with no migration (divergence without gene flow); C) split with migration (divergence with gene flow that is bidirectionally symmetric); D) split with bidirectional migration (divergence with gene flow that is bidirectionally asymmetric); E) secondary contact with migration; F) secondary contact with bidirectional migration; G) split with exponential population growth and no migration; H) split with exponential population growth and migration. Letters refer to parameters used in models: ancestral population size (N_{ref}), effective population size (N_1, N_2), time since split (T, T_S), time of secondary contact (T_{SC}), migration (m), migration from population 1 to population 2 (m_{21}), migration from population 2 to population 1 (m_{12}). Single arrow bars indicate models with gene flow that is bidirectionally symmetrical; double arrow bars are models with different levels of gene flow in each direction. Rectangles indicate unchanging population sizes and triangles indicate population growth. Colors suggest increasing population differentiation, except in the neutral model (A).

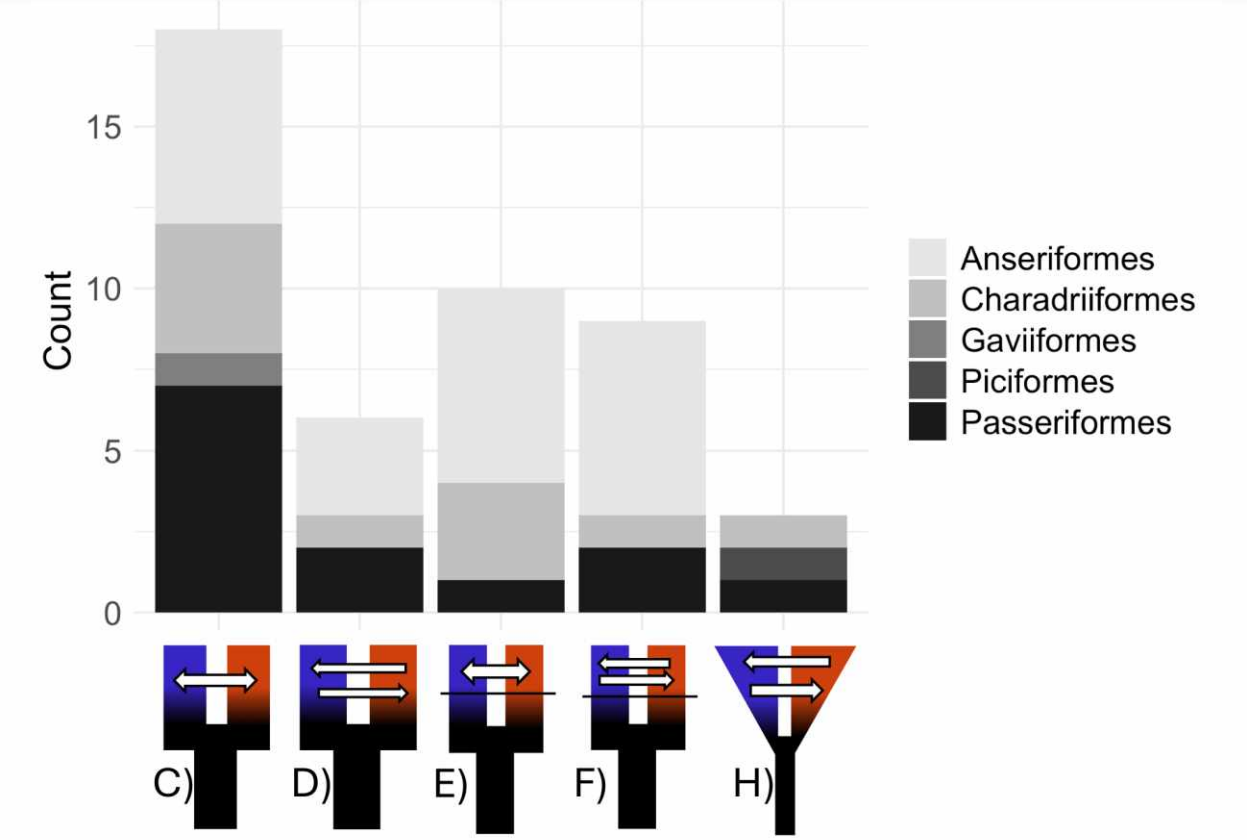

Fig. 2.3: Number of pairwise comparisons per best-fit and runner-up demographic models, with avian orders shown in different colors. Letters correspond to Fig. 2.2. These include results from prior research (McLaughlin et al., 2020; Spaulding et al., 2022, 2023; Winker et al., 2018).

2.8 Tables

Table 2.1: Summary of lineages compared in this study, including number of samples, total single nucleotide polymorphisms (SNPs), average coverage depth per SNP, and total variable loci included in analyses. The order of numbers in *N* refers to the order in the populations column.

	Populations (pop1/pop2)	*N* (pop1:pop2)	Total SNPs	Average coverage depth	Total variable loci
Anseriformes					
Mergus merganser subspp.	*merganser/americanus*	6:4	7,012	39.00	2,913
Charadriiformes					
Arenaria interpres / A. melanocephala	*interpres/melanocephala*	5:5	30,216	32.81	3,996
Calidris alpina subspp.	*arcticola/pacifica*	5:4	6,533	45.64	1,818
Uria aalge	Asia/North America	4:5	18,472	32.60	3,729
Uria lomvia	Asia/North America	5:5	21,865	43.72	3,655
Gaviiformes					
Gavia stellata	Asia/North America	5:5	11,385	43.68	3,346
Piciformes					
Picoides tridactylus / P. dorsalis	*tridactylus/dorsalis*	5:5	8,072	35.87	3,009
Passeriformes					
Corvus corax	Asia/North America/Attu	7:4:5	93,278	44.99	3,991
Phylloscopus borealis subspp. / *P. examinandus*	*borealis/kennicotti/ examinandus*	5:5:4	29,428	47.97	3,914
Anthus rubescens subspp.	*japonicus/pacificus*	5:5	21,400	40.02	3,808
Motacilla tschutschensis subspp.	*tschutschensis /simillima*	5:5	32,734	33.54	2,045

Table 2.2: Between-population F_{ST} values estimated using biallelic one-SNP-per-locus datasets. *P*-values are from *G*-tests run with 99 simulations. Asterisks (*) denote significance.

	Taxonomic Depth	F_{ST}	*p*-value
Anseriformes			
Mergus merganser subspp.	Subspecies	0.269	*0.01
Charadriiformes			
Arenaria interpres / A. melanocephala	Species	0.645	*0.01
Calidris alpina subspp.	Subspecies	0.014	*0.03
Uria aalge	Population	0.005	0.11
Uria lomvia	Population	0.001	0.39
Gaviiformes			
Gavia stellata	Population	0.017	*0.03
Piciformes			
Picoides dorsalis / P. tridactylus	Species	0.431	*0.01
Passeriformes			
Corvus corax (Asia – North America)	Subspecies	-0.037	0.42
Corvus corax (Asia – Attu)	Population	0.013	0.17
Corvus corax (North America – Attu)	Subspecies	0.176	*0.01
Phylloscopus borealis borealis / *P. examinandus*	Species	0.116	*0.04
Phylloscopus borealis subspp.	Subspecies	0.008	0.16
Anthus rubescens subspp.	Subspecies	0.386	*0.01
Motacilla tschutschensis subspp.	Subspecies	0.017	*0.05

Table 2.3: Two-population model comparisons for each of 14 pairwise comparisons. Two lineages (*Corvus corax* and *Phylloscopus* spp.) include three and two comparisons, respectively. Maximum log composite likelihood (MLCL) is averaged from the top five runs of the optimized model presented, followed by AIC values in parentheses. Bold numbers indicate best-fit models and italic numbers indicate runner-up models (ΔAIC < 10).

	Neutral (Fig. 2A)	Split, no migration (Fig. 2B)	Split, one migration parameter (Fig. 2C)	Split migration, two parameters (Fig. 2D)	Secondary contact, single migration parameter (Fig. 2E)	Secondary contact, two migration parameters (Fig. 2F)	Isolation, population growth, no migration (Fig. 2G)	Isolation, population growth, migration (Fig. 2H)
Mergus merganser subspp.	-2075.64 (4155.27)	-761.50 (1529.00)	-416.20 (840.40)	-384.65 (779.30)	-361.40 (732.80)	**-304.86 (621.72)**	-1023.32 (2054.64)	-348.65 (709.30)
Arenaria interpres / Arenaria melanocephala	-12043.77 (24091.55)	-423.33 (852.66)	-216.90 (441.80)	-214.66 (439.33)	-214.18 (438.36)	-215.90 (443.79)	-277.76 (563.52)	**-138.01 (288.02)**
Calidris alpina subspp.	-433.82 (871.63)	-731.54 (1469.07)	**-101.41 (210.83)**	-144.19 (298.37)	*-100.93 (211.86)*	-167.12 (346.24)	-1007.84 (2023.69)	-179.79 (371.57)
Uria aalge	-202.62 (409.25)	-2788.64 (5583.27)	**-147.42 (302.84)**	-382.00 (774.00)	-183.16 (376.33)	-457.00 (926.00)	-3591.12 (7190.25)	-426.24 (864.48)
Uria lomvia	-933.03 (1870.06)	-2010.32 (4026.65)	**-142.61 (293.22)**	-433.87 (877.74)	-163.99 (337.99)	-468.23 (948.47)	-2418.78 (4845.55)	-489.02 (990.04)
Gavia stellata	-453.76 (911.52)	-1562.83 (3131.67)	**-156.98 (321.96)**	-209.11 (428.22)	-175.94 (361.88)	-265.80 (543.60)	-2082.70 (4173.41)	-278.96 (569.92)
Picoides dorsalis/ Picoides tridactylus	-3662.79 (7329.58)	-306.50 (618.99)	-209.83 (427.67)	-209.66 (429.33)	-204.43 (418.86)	-204.98 (421.96)	-273.42 (554.84)	**-136.41 (284.82)**
Corvus corax (Asia vs N America)	-6706.57 (13417.14)	-3931.67 (7869.35)	**-3102.03 (6212.06)**	-3884.66 (7779.33)	-3119.49 (6248.98)	-3258.83 (6529.66)	-4735.57 (9479.15)	-4174.73 (8361.46)
Corvus corax (Asia vs Attu)	-5503.39 (11010.79)	-3491.45 (6988.91)	-2753.15 (5514.30)	**-2673.47 (5356.94)**	-2810.36 (5630.72)	-2728.08 (5468.15)	-4183.84 (8375.68)	-3334.69 (6681.38)
Corvus corax (Attu vs N America)	-209.10 (422.20)	-149.83 (305.66)	**-81.25 (170.50)**	-96.31 (202.62)	-93.01 (196.01)	-96.43 (204.87)	-233.35 (474.69)	-99.12 (210.24)
Phylloscopus borealis subspp.	-401.13 (806.26)	-851.73 (1709.46)	**-127.89 (263.78)**	-318.16 (646.32)	-145.83 (301.66)	-320.13 (652.25)	-1965.19 (3938.38)	-374.53 (761.06)

Table 2.3 continued

Phylloscopus borealis borealis / Phylloscopus examinandus	-1003.26 (2010.52)	-974.24 (1954.47)	-267.22 (542.44)	-274.44 (558.87)	**-244.37 (498.74)**	*-246.26 (504.52)*	-2821.26 (5650.51)	-298.38 (608.75)
Anthus rubescens subspp.	-4483.97 (8971.93)	-789.60 (1585.19)	-634.92 (1277.85)	-631.87 (1273.74)	-610.48 (1230.95)	-606.88 (1225.75)	-748.51 (1505.02)	**-447.29 (906.58)**
Motacilla tschutschensis subspp.	-2266.08 (4536.15)	-1531.53 (3069.06)	-1188.92 (2385.84)	-1129.43 (2268.87)	-1199.34 (2408.67)	**-1108.81 (2229.62)**	-2005.33 (4018.65)	-1190.76 (2393.52)

Table 2.4: Results of two-population δaδi analyses with best-fit models. Included are effective population sizes (N_e) of each population, effective size of the ancestral population (N_{ref}), time since divergence in years and, for applicable models, time since secondary contact, migration rates using the size of each population, and effective size of populations after initial split, where applicable. For models with one migration parameter, two migration rates were calculated with the effective population sizes of both populations. For lineages in which multiple models fit, both models are presented in separate rows (split-mig = split with symmetrical migration, split bidir mig = split with asymmetrical migration, SC = secondary contact model, IM = isolation with migration model; see Figure 2). All estimates include 95% CI calculated with 100 δaδi bootstraps.

	Population 1	Population 2	N_e (pop1)	N_e (pop2)	N_{ref}	Time since divergence (yr)	Time since secondary contact (yr)	Migration (from pop1)	Migration (from pop2)	Size after split (pop1)	Size after split (pop2)
Mergus merganser subspp.	*merganser*	*americanus*									
SC, 2m (Fig. 2F)			56,779 (±3,434)	38,525 (±3,058)	15,355 (±3,021)	215,623 (±11,431)	9,028 (±29,647)	0.234 (±0.18)	5.648 (±0.37)		
Arenaria spp.	*interpres*	*melanoce-hala*									
IM (Fig. 2H)			269,742 (±23,152)	37,095 (±3,395)	38,852 (±12,583)	1,185,456 (±53,887)		0.006 (±0.07)	0.044 (±0.28)	21,681 (±10,622)	17,171 (±6,764)
Calidris alpina subspp.	*arcticola*	*pacifica*									
Split-mig (Fig. 2C)			390,483 (±44,076)	84,873 (±49,412)	28,484 (±2,465)	316,605 (±40,859)		33.67 (±2.83)	7.32 (±0.61)		
SC, 1m (Fig. 2E)			289,009 (±26,710)	89,853 (±6,870)	28,935 (±2,651)	228,360 (±47,314)	88,296 (±48,328)	29.957 (±2.37)	9.314 (±8.28)		
Uria aalge	Asia	North America									
Split-mig (Fig. 2C)			10,380 (±9,583)	35,025 (±22,760)	10,008 (±953)	958,448 (±206,395)		5.19 (±0.45)	17.51 (±1.51)		

Table 2.4 continued

Uria lomvia	Asia	North America							
Split-mig (Fig. 2C)			81,137 (±32,530)	192,971 (±49,204)	16,164 (±1,767)	511,856 (±70,708)	26.763 (±2.51)	63.651 (±5.96)	
Gavia stellata	Asia	North America							
Split-mig (Fig. 2C)			41,465 (±26,080)	25,468 (±12,257)	13,255 (±732)	254,318 (±74,210)	18.250 (±0.81)	11.210 (±0.50)	
Picoides spp.	*tridactylus*	*dorsalis*							
IM (Fig. 2H)			87,948 (±5,707)	93,224 (±4,952)	16,712 (±1,191)	159,179 (±12,541)	0.086 (±0.08)	0.039 (±0.11)	11,517 (±7,538)	5,195 (±3,760)
Corvus corax	Asia	North America							
Split-mig (Fig. 2C)			456,502 (±2,727)	10,797 (±140,925)	96,550 (±15,297)	3,320,557 (±491,658)	1.219 (±0.27)	0.029 (±0.006)	
Corvus corax	Asia	Attu							
Split bidir mig (Fig. 2D)			284,384 (±29,917)	25,535 (±27,738)	196,231 (±13,443)	2,319,213 (±280,702)	0.021 (±0.008)	1.635 (±0.11)	
Corvus corax	North America	Attu							
Split-mig (Fig. 2C)			12,778 (±389)	374,941 (±8,322)	6,823 (±361)	50,296 (±4,733)	0.02 (±0.11)	0.73 (±3.33)	
Phylloscopus borealis subspp.	*borealis*	*kennicotti*							
Split-mig (Fig. 2C)			3,565,675 (±260,407)	85,835 (±276,145)	81,019 (±8,423)	339,406 (±67,997)	206.519 (±18.70)	4.971 (±0.45)	

Table 2.4 continued

Phylloscopus spp.	*borealis*	*examinand-us*								
SC, 1m (Fig. 2E)			235,707 (±16,727)	313,604 (±13,626)	86,664 (±4,251)	348,384 (±28,594)	27,456 (±66,324)	4.948 (±0.32)	6.583 (±0.44)	
SC, 2m (Fig. 2F)			297,400 (±25,673)	222,744 (±13,851)	80,579 (±9,412)	351,939 (±59,297)	62,216 (±104,086)	1.500 (±0.20)	3.975 (±0.21)	
Anthus rubescens subspp.	*japonicus*	*pacificus*								
IM (Fig. 2H)			293,853 (±22,773)	391,200 (±31,390)	70,046 (±7,269)	305,382 (±39,633)		0.178 (±0.16)	0.322 (±0.37)	29,376 (±10,867)	40,670 (±19,066)
Motacilla tschutschensis subspp.	*simillima*	*tschutsche-nsis*								
SC, 2m (Fig. 2F)			1,467,728 (±224,562)	326,926 (±60,243)	501,197 (±49,505)	2,447,329 (±353,138)	1,185,359 (±218,058)	0.060 (±0.10)	4.140 (±0.30)	

Chapter 3: Evidence of positive selection and a novel phylogeny among five subspecies of song sparrow (*Melospiza melodia*) in Alaska

3.1 Abstract

Under local adaptation, populations evolve traits in response to the local environment. Isolated island populations often experience different selection pressures than their mainland counterparts, which enables the study of how phenotypes and genotypes respond to different selection regimes. We used a group of five phenotypically differentiated subspecies of song sparrow (*Melospiza melodia*) in Alaska to examine the effects of local adaptation. Song sparrows occur across southern Alaska from Attu Island in the western Aleutian Islands, to southeast Alaska. Moving from western to eastern Alaska these populations demonstrate striking body size differences (larger-to-smaller) and a change from a sedentary to a migratory/partially migratory life-history strategy. We examined the phenotypic attributes of these populations and used whole-genomic data to determine relationships and test candidate loci for evidence of selection. Phenotypic measurements of museum specimens (n = 227) quantified the dramatic size differences among these populations, with westernmost *M. m. maxima* being ~1.6 times larger than easternmost *M. m. rufina*. Ultraconserved elements (UCEs) were extracted for phylogenetic reconstruction and candidate genes were extracted for selection testing. We analyzed 26 candidate genes for body size, migration and dispersal, color, and salt tolerance. Two of the candidate genes showed signs of positive selection: *BCO1* (associated with plumage color) and *KCTD21* (associated with dispersal). Phylogenetic analysis of UCEs showed *M. m. maxima* as sister to the other Alaska *M. melodia* subspecies. This suggests *M. m. maxima* colonized earliest, perhaps before the last glacial maximum, and that Alaska was later recolonized by ancestors of the remaining four subspecies.

3.2 Introduction

Our current understanding of evolution suggests that adaptation to local conditions enables populations to persist and that geographic variation among populations often reflects these processes. Local adaptation is defined as the process by which a local population evolves traits best suited for its environment (Turesson, 1922; Mayr, 1963; Williams, 1966; Kawecki & Ebert, 2004; Savolainen, Lascoux & Merilä, 2013). Local adaptation to a specific environment can

potentially lead to speciation (Savolainen, Lascoux & Merilä, 2013; Tiffin & Ross-Ibarra, 2014), but this process of adaptation can be hindered by the homogenizing effects of dispersal and gene flow (Kawecki & Ebert, 2004; Savolainen, Lascoux & Merilä, 2013). While it is still challenging to identify the genetic basis of local adaptation, the advent of high-throughput genomic data can be used to identify potential genes under selection (Tiffin & Ross-Ibarra, 2014; Bomblies & Peichel, 2022). Phenotypically differentiated island populations can be used to examine the genetic basis of local adaptation. This has been demonstrated in many terrestrial organisms, such as lizards (Losos, Warheitt & Schoener, 1997), birds (Grant, 1981), spiders (Gillespie, 2002), and plants (Choi et al., 2021).

Island systems have been influential in our understanding of the processes of evolution (MacArthur & Wilson, 1967; Losos & Ricklefs, 2009; Warren et al., 2015). Islands, especially archipelagos, are natural laboratories to study evolution, given their relatively young geological age and geographic isolation from the mainland (Losos & Ricklefs, 2009; Warren et al., 2015). Because of these conditions, island populations provide a way to determine how local adaptation accrues. The majority of research on local adaptation has occurred on tropical or mid-latitude islands, however there has been a recent focus on high-latitude island systems (Whittaker & Fernadez-Palacios, 2007). For example, the Aleutian Islands in the North Pacific Ocean constitute a volcanic archipelago that extends ~1,800 km from western Alaska towards eastern Russia (Murie, 1959). During the Pleistocene, Alaska experienced multiple cycles of glaciation and glaciers covered much of the continent. However, ice-free zones, or refugia, were present which could have allowed populations to persist, and the Aleutian Islands are hypothesized to contain some of these refugia (Pruett & Winker, 2005a; Winker et al., 2023). Isolated avian populations on these islands often exhibit both lower genetic diversity than mainland counterparts and phenotypic differences, such as larger body size (Pruett, Li & Winker, 2018). One such species is the song sparrow (*Melospiza melodia*, Passeriformes: Passerellidae), a widely distributed songbird found only in North America with ~25 recognized subspecies (Aldrich, 1984; Patten & Pruett, 2009). These subspecies exhibit a wide range of phenotypes and life histories and reside in a variety of environments across North America (Arcese et al., 2020; Carbeck et al., 2023). Because of this, song sparrows are an excellent candidate for the study of local adaptation.

Five song sparrow subspecies occur across southern Alaska, from the western Aleutian Islands to southeast Alaska. These subspecies, from west to east, are *M. m. maxima, M. m. sanaka, M. m. insignis, M. m. caurina*, and *M. m. rufina* (Fig. 3.1; Gibson & Withrow, 2015). Prior genetic research suggested that ancestral song sparrows colonized the Aleutian Islands sequentially, from east to west (Pruett & Winker, 2005b). Across this complex, these subspecies exhibit different life history traits, presumably due to differences in their environments. The subspecies that occur furthest west (*maxima* and *sanaka*) are non-migratory residents on the Aleutian Islands, a relatively harsh maritime environment. These two subspecies are found in grassy slopes along marine beaches, a habitat considered marginal for continental subspecies (Murie, 1959; Aldrich, 1984). The subspecies that occur furthest east in Alaska (*caurina* and *rufina*) are seasonal migrants, residing largely on continental North America and migrate south for the winter (Aldrich, 1984; Patten & Pruett, 2009). One subspecies (*insignis*) is primarily a year-round resident of the Kodiak Archipelago (Patten & Pruett, 2009). There are obvious phenotypic and natural history differences among these subspecies. The most notable phenotypic characteristics are differences in plumage coloration and body size, wherein the most western subspecies in the Aleutian Islands are much larger than the eastern subspecies in southeast Alaska (Aldrich, 1984; Pruett & Winker, 2005b; Patten & Pruett, 2009). These phenotypic differences suggest these five subspecies have adapted to their local environments (Fig. 3.2; Winker, 2010).

We studied phenotypic differentiation relevant to aspects of life history characteristics among the subspecies of the Alaska song sparrow complex to examine adaptive evolution of high-latitude island populations. We used a combination of genomic and morphological analyses using museum specimens. We asked three questions. First, which morphological traits are significantly different among the five *M. melodia* subspecies in Alaska? Second, are candidate genes, associated with key phenotypic attributes, under positive selection in the most extreme variant, *M. m. maxima*? Third, what are the phylogenetic relationships among these subspecies?

3.3 Methods

3.3.1 Phenotypic Analyses

To examine the phenotypic variation among the five *Melospiza melodia* subspecies, we obtained data from 227 vouchered specimens from the University of Alaska Museum (Table

C.1). We restricted our analyses to adult males with complete phenotypic and locality data. Phenotypic data included: mass, wing chord length, tail length, tarsus length, bill length, bill height, bill width, and skull length (Fig. C.1; Winker, 2000). Measurements were taken to the nearest 0.1 millimeter (mm) or to 0.1 gram (g) for mass prior to specimen preparation (Winker, 2000). Data were reviewed and analyzed using R (v. 4.1.1; R Core Team, 2021) and RStudio (v. 2022.12.0.353; Posit Team, 2022) and accompanying packages, "devtools" (v.2.4.5; Wickham et al., 2022) and "tidyverse" (v.1.3.2; Wickham et al., 2019). We removed outliers that were three standard deviations away from the mean to account for human measurement error. Phenotypic data were then visualized with boxplots and a principal components analysis (PCA) using the packages "ggplot2" (v.3.4.1; Wickham et al., 2019), "ggpubr" (v. 0.6.0; Kassambara, 2023), and "ggord" (v. 1.1.7; Beck, 2022). ANOVA tests were used for each measurement and pairwise Tukey post-hoc tests were used to determine which subspecies pairs were different. R scripts and raw phenotypic data are available at https://github.com/coliverbrown/melospiza-melodia-phenotype.

3.3.2 Sampling and Whole-genome Sequencing

For genomic analysis, we sampled high-quality vouchered tissue samples from wild individuals archived at the University of Alaska Museum (Table C.2). Our sample size included one individual from each of the five *M. melodia* subspecies that occur in Alaska. Two taxa were included as outgroups: the swamp sparrow (*M. georgiana*) and the Lincoln's sparrow (*M. lincolnii*). DNA extractions followed standard protocol for animal tissues using the QIAGEN DNeasy Blood + Tissue Extraction Kit (Valencia, California).

Libraries were prepared using the iTru library protocols described in Glenn et al. (2019). Briefly, we sheared the genomic DNA using a Bioruptor (Diagenode, Denville, NJ, USA) targeting a 500bp average fragment size. The sheared DNA was end-repaired, adenylated, and ligated to iTru adapters followed by limited-cycle PCR of iTru primers to add indexes and complete the library molecules using Kapa Library Preparation Kit reagents (Kapa Biosystems [Roche, Basel, Switzerland]). We sequenced the library on an Illumina HiSeq 2500 (Illumina, San Diego, CA, USA) at the HudsonAlpha Institute for Biotechnology (Huntsville, AL, USA) to obtain paired-end (PE) 100 bp reads. Raw sequence data have been deposited in NCBI Sequence Read Archive under BioProject PRJNA1114297.

3.3.3 Bioinformatics

After sequencing, we constructed whole genome assemblies for each of the sequenced individuals. Our bioinformatics pipeline centered on the package PHYLUCE (Faircloth, 2016). Raw and untrimmed FASTQ data that contained low-quality bases were removed using Illumiprocessor (Faircloth, 2013), which incorporates Trimmomatic (Bolger, Lohse & Usadel, 2014). Raw paired end reads (read1, read2, and singleton files) were then mapped to a dark-eyed junco (*Junco hyemalis;* Passeriformes: Passerellidae) reference sequence (MLZ69236; Friis et al., 2018). Resulting sequences were indexed using BWA (Li & Durbin, 2009) and SAMtools (Li et al., 2009). Next, we used PICARD (Broad Institute, 2019) to clean the alignments, add read group header information, and remove PCR and sequencing duplicates. Single nucleotide polymorphisms (SNPs) were called for each individual against the reference, and Genome Analysis Toolkit (GATK, v.4.2.1; McKenna et al., 2010) was used to call and realign around indels, call and annotate SNPs, and filter SNPs around indels. We restricted the data to high-quality SNPs by adding a quality filter (Q30) before converting the resulting VCF file to a FASTA file using GATK.

3.3.4 Selection Analyses

To determine whether genes associated with phenotype were under positive selection, we identified candidate genes and tested for signs of selection in *M. m. maxima*. Candidate genes were chosen through a review of literature that identified genes that showed correlations with phenotypes of interest. We chose 34 candidate genes associated with body size (Liu et al., 2015), bill size and shape (Lamichhaney et al., 2015; Chaves et al., 2016; Huang et al., 2022), migration (Vallone et al., 2007; Ruegg et al., 2014; Delmore et al., 2015), dispersal (San-Jose et al., 2023), plumage coloration (Walsh et al., 2011), and salt tolerance (Walsh et al., 2019). Sequences for each gene were obtained from NCBI GenBank (Table C.3). Once reference sequences were obtained, we extracted corresponding sequences in our outgroup taxa *M. georgiana* and *M. m. maxima* using a custom BLAST search in Geneious Prime 2023.0.4 (www.geneious.com). Additionally, we extracted gene sequences from a *M. m. maxima* reference sequence (UAM31500; Louha et al., 2020). Sequences with no hits or had an alignment score of less than 60% were discarded. If one transcript yielded multiple hits, only the highest-scoring hit was retained for further analysis. The sequences were aligned by gene using MUSCLE (Edgar, 2004).

We tested 26 candidate genes associated with body size (n = 5), bill size (n = 5), migration (n = 8), dispersal (n = 3), plumage color (n = 3), and salt tolerance (n = 2; Table 3.1) using the McDonald-Kreitman (MK) test, which compares the polymorphisms in one species or population to the divergence between multiple species (McDonald & Kreitman, 1991; Egea, Casillas & Barbadilla, 2008). The MK test was performed using an online tool (http://mkt.uab.es/mkt/mkt.asp; Egea, Casillas & Barbadilla, 2008).

3.3.5 Phylogeny Reconstruction

To reconstruct the phylogeny of the *M. melodia* complex and outgroups, we extracted ultraconserved elements (UCEs) from the whole-genome FASTA sequence files using PHYLUCE (Faircloth, 2016). We used UCEs because they are used to capture phylogenomic information from both shallow and deep time depths (Faircloth et al., 2012). We included the *Melospiza* individuals we sampled and the *J. hyemalis* reference sequence. Individual UCE loci were aligned using MUSCLE (Edgar, 2004) and were edge trimmed. We filtered this dataset to include loci that had all the *Melospiza* individuals and the *J. hyemalis* reference represented in a 95% matrix. The sequences were aligned by loci using MUSCLE and then concatenated into one dataset using PHYLUCE (Edgar, 2004; Faircloth, 2016). We obtained branch supports with the ultrafast bootstrap approximation (Hoang et al., 2018) implemented in IQ-TREE software (v.2.1.4; Minh et al., 2020). The phylogenetic tree was visualized and edited with FigTree (v.1.4.4; http://tree.bio.ed.ac.uk/software/Figtree/)

3.4 Results

3.4.1 Phenotypic Analyses

Phenotypic variation among the subspecies was pronounced. Overall, subspecies that occur in the western range (*maxima, sanaka*, and *insignis*) were larger than eastern Alaska subspecies (*rufina* and *caurina*). *Melospiza melodia maxima* was approximately 1.6 times larger in body mass than *rufina* (Table 3.2; Fig. C.2). For all other phenotypic measurements, differences were ~1.2 times larger (Table 3.2). In all measurements, there was no significant difference between *rufina* and *caurina* (Table 3.2). Additionally, *maxima* and *sanaka* showed differences only between bill height and bill width (Table 3.2).

PCA analysis revealed that 55% of the variation among the five subspecies of *M. melodia* was explained in PC1 by five phenotypic traits: mass, wing chord, tail length, tarsus length, and skull length. Around 13% of the variation was explained in PC2 by the remaining three traits: bill length, bill width, and bill height (Fig. 3.3). In the PCA graph, individuals tend to clump into two groups, mostly divided along PC1. The three western subspecies (*maxima, sanaka,* and *insignis*) group together, and the two remaining eastern subspecies (*caurina* and *rufina*) also group together (Fig. 3.3).

3.4.2 Summary Statistics of Whole-genome Sequencing

We obtained >400 million reads, ranging from 12,566,699 to 84,504,717 per individual (average = 60,119,612; Table C.4). Coverage averaged 6.7x, ranging from 2.1x to 9.7x per individual (Table C.4).

3.4.3 Selection Analyses

Of the 26 candidate genes tested for signals of positive selection, two showed significance: *BCO1* ($p = 0.037$) and *KCTD21* ($p = 0.032$; Table 3.2). These genes are associated with plumage color and dispersal, respectively (Ruegg et al., 2014; Toews, Hofmeister & Taylor, 2017; San-Jose et al., 2023).

3.4.4 Phylogenetic Reconstruction

We constructed a phylogenetic tree of this song sparrow complex and outgroups. After filtering UCE loci present in the *Melospiza* and *J. hyemalis* genomes to obtain a 95% matrix, the final alignment contained 4793 loci. All nodes within the *M. melodia* clade received 100% support from the ultrafast bootstrap (Fig. 3.4). The phylogenetic tree shows that two of the western subspecies (*sanaka* and *insignis*) and the two eastern subspecies (*caurina* and *rufina*) form their own clades. However, *maxima* is a sister group to all other *M. melodia* subspecies (Fig. 3.4).

3.5 Discussion

Our results show that the westernmost subspecies of song sparrow (*M. m. maxima*) is approximately 1.6 times larger than eastern *rufina* (Table 3.2). We found evidence for positive

selection in two genes (*BCO1* and *KCTD21*) in *M. m. maxima* (Table 3.2). We reconstructed a novel phylogeny for the Alaska song sparrow complex, which places *M. m. maxima* outside the rest of the *M. melodia* clade (Fig. 3.4). This suggests that song sparrows colonized the Aleutian Islands early and underwent divergence prior to the eastern portion of the species' current Alaska range was recolonized.

3.5.1 Phenotypic Variation

The strong increase in body size among resident populations in the western Aleutians (Table 3.2) is concordant with previous research (Aldrich, 1984) and follows Bergmann's Rule, by which endothermic animals tend to be larger in cooler climates (Bergmann, 1847, Meiri & Dayan, 2003; Carbeck et al. 2023). When examining the PCA results (Fig. 3.3), the three western subspecies (*maxima, sanaka*, and *insignis*) grouped together and the remaining two eastern subspecies (*caurina* and *rufina*) grouped together. These groups are consistent with the different life-history strategies, with the larger, western subspecies being sedentary and the smaller, eastern subspecies being migratory. Sedentary birds tend to comply with Bergmann's rule more than migratory species because migratory species can seasonally escape more severe climatic conditions (Meiri & Dayan, 2003; but see Ashton, 2002). Likewise, island living could also have an impact on body size. Birds and mammals experience the "island rule", whereby large organisms on islands evolve towards smaller sizes and small organisms evolve toward larger sizes (Losos & Ricklefs, 2009).

Melospiza melodia maxima and *sanaka* were similar in all measurements except for bill height and bill width (Table 3.2). *M. m. sanaka* had significantly narrower and shallower bills than *maxima,* which is consistent with previous research (Patten & Pruett, 2009; Gibson & Withrow, 2015). Slender bills are associated with invertebrate consumption, while stouter bills are useful for eating seeds (Aldrich, 1984). However, current knowledge does not include differences in diet between these two subspecies that might explain this difference in morphology (Aldrich, 1984). It is possible that the thicker bills of *maxima* were an adaptation for living in a glacial refugium during the Last Glacial Maximum, but the exact reason is unknown.

3.5.2 Selection Analyses

We found two genes of 26 that showed signs of positive selection: *BCO1* and *KCTD21* (Table 3.2). *BCO1* is associated with coloration and carotenoid breakdown (Toews, Hofmeister & Taylor, 2017). Carotenoids, which are obtained from an animal's diet, must be metabolized to be incorporated into tissues. Beta-carotene (β-carotene) oxygenases (i.e., *BCO1* and *BCO2*) are some of the best-known genes involved in carotenoid breakdown. *BCO1* facilitates the separation of vitamin A from β-carotene, whereas *BCO2* is more directly involved in coloration (Eriksson et al., 2008; Toews, Hofmeister & Taylor, 2017). After carotenoid breakdown, the pigments can be deposited into feathers and skin. Many studies examine the expression of these candidate genes in colorful birds. For example, Walsh et al. (2011) found that *BCO1* was expressed in developing feathers and skin of queleas, another bird species. However, there are many different strategies for carotenoid deposition and numerous genes associated with carotenoid coloration, some that we did not include in our study design (Eriksson et al., 2008; Walsh et al. 2011).

Potassium channel tetramerization domain, *KCTD21*, is associated with dispersal, migration, and the circadian clock (San-Jose et al., 2023). Dispersal and migration are two separate, but possibly intertwined, processes. Migration is the seasonal go-and-return movement between breeding and non-breeding grounds, whereas dispersal is the movement from the place of birth to breed somewhere else (Dingle & Drake, 2007). In a study of the genetic underpinnings of dispersal in reptiles, San-Jose et al. (2023) found that multiple genes associated with the circadian clock and migration were differently expressed in dispersers. This suggests that some of the same processes are involved in migration and dispersal. The McDonald-Kreitman test only specifies if divergence is higher than the polymorphisms and thus assumes positive selection (McDonald & Kreitman, 1991; Egea, Casillas & Barbadilla, 2008). It is impossible to know, using this method, if this gene is being expressed or suppressed. We have two hypotheses as to why *KCTD21* is under positive selection. First, this gene could remain under positive selection due to ancestral *maxima* individuals dispersing to the western Aleutian Islands. Second, *KCTD21* could contain an allele that suppresses its activity, thus remaining under positive selection to prevent *maxima* individuals from dispersing. Given their remote island habitat, the latter seems more likely.

Recent studies have examined the song sparrow complex to identify genes under selection. Carbeck et al. (2023) found evidence of natural selection acting on genes associated with body size. The authors found nine candidate genes associated with body size showed divergence between *maxima* and *rufina* song sparrows and then validated these findings by predicting the genotypes of five small song sparrows in California (Carbeck et al., 2023). However, our study did not test for selection of any candidate genes that Carbeck et al. (2023) found.

We expected to see more candidate genes under positive selection in *maxima*, such as genes related to body size and salt tolerance. However, we only tested the protein-coding regions of the genome. The regulatory regions surrounding the protein-coding regions can be influenced by linked selection, but these regions were not tested. Additionally, the standard McDonald-Kreitman test has limitations. Slightly deleterious mutations, where selection acts weakly, can make it difficult to identify adaptive evolution (Andolfatto, 2008). Andolfatto (2008) suggested using simulations to reduce type-I errors when running the MK test. Additionally, small sample sizes can overestimate adaptive substitutions and increasing the sample size can reduce this bias (Andolfatto, 2008; Subramanian, 2016). However, the MK test is a standard metric to identify potential adaptive selection in genomic regions. Our results are an excellent first look into how *M. m. maxima* might have locally adapted to the Aleutian Islands. Future work should increase sample size to include more than two sequences tested, use coalescent simulations to determine significance levels (Andolfatto, 2008; Subramanian, 2016), and test more candidate genes that other studies have identified (Carbeck et al., 2023).

3.5.3 Phylogenetic Reconstruction

Our phylogenetic tree shows that the westernmost subspecies *maxima* is a sister taxon to all other Alaska *M. melodia* subspecies (Fig. 3.4). Further, the eastern Alaska subspecies *caurina* and *rufina* are grouped together and two of the western subspecies (*sanaka* and *insignis*) form another clade (Fig. 3.4). This suggests that *maxima* colonized the western Aleutian Islands first, underwent differentiation from other forms, and then at some later time, the eastern part of the Alaskan range was colonized by ancestors of the remaining subspecies. One possibility for this colonization history is that following an initial colonization of Alaska, *maxima* survived in an Aleutian glacial refugium (Pruett & Winker, 2005a; Winker et al., 2023) while other Alaska populations were extirpated, and then Alaska was re-colonized post-glacially. This hypothesis is

inconsistent with the colonization history postulated by Pruett and Winker (2005b), which suggested a probable postglacial stepwise process from east to west.

Interestingly, *maxima* and *sanaka* are similar in size and have similar life-history patterns, even though *M. m. sanaka* is more closely related to all of the non-*maxima* subspecies. This phenotypic similarity could be due to convergent evolution in similar environments, or, more likely, due to adaptive alleles for Aleutian life introgressing through gene flow from *maxima* to *sanaka* (Pruett & Winker, 2005b; Reding et al., 2008; Graham et al., 2021).

It is possible that gene flow could be obscuring our phylogenetic results, however we do not expect this to be the case. Models show that moderate to high levels of gene flow can degrade phylogenetic signal and thus produce inaccurate phylogenetic trees (Eckert & Carstens, 2008; Leaché et al., 2014). Previous research by Pruett and Winker (2005b) found gene flow between *maxima* and *sanaka* to be at approximately 1.20 – 2.90 individuals per generation (Fig. C.3). The exact amount of gene flow needed to obscure phylogeny is not noted but the current estimates for gene flow in the Aleutian Islands are not very high. Future research should determine if this obscuration is happening within this song sparrow complex.

3.5.4 Conclusion

Among song sparrows in Alaska, phenotypic differences between subspecies suggest that selective pressures have acted heterogeneously on different populations. We found two candidate genes under positive selection in *M. m. maxima*. Genes under selection associated with carotenoid breakdown and dispersal (*BCO1* and *KCTD21*, respectively), suggest that *maxima* has undergone selective pressure for these traits relative to other members of the species in this region. The phylogenetic relationship that we recovered among Alaska song sparrow subspecies suggests a new hypothesis for the species' biogeographic history. We suggest that ancestral populations of *M. m. maxima* colonized Alaska and the Aleutian Islands, and then persisted in a glacial refugium, likely while other populations were extirpated. Then after a substantial period of time the rest of the state was recolonized by ancestors of the remaining subspecies. This complex might be an ideal candidate for future studies of evolution of island populations and local adaptation.

3.6 References

Aldrich JW. 1984. Ecogeographical variation in size and proportions of song sparrows (*Melospiza melodia*). *Ornithological Monographs*:iii–134. DOI: 10.2307/40166779.

Andolfatto P. 2008. Controlling type-I error of the McDonald–Kreitman test in genomewide scans for selection on noncoding DNA. *Genetics* 180:1767–1771. DOI: 10.1534/genetics.108.091850.

Arcese P, Sogge MK, Marr AB, Patten MA (2020). Song Sparrow (*Melospiza melodia*), version 1.0. In *Birds of the World* (A. F. Poole and F. B. Gill, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. DOI: 10.2173/bow.sonspa.01

Ashton KG. 2002. Patterns of within-species body size variation of birds: strong evidence for Bergmann's rule. *Global Ecology and Biogeography* 11:505–523. DOI: 10.1046/j.1466-822X.2002.00313.x.

Beck MW. 2022. ggord: Ordination Plots with ggplot2. R package version 1.1.7. DOI: https://zenodo.org/badge/latestdoi/35334615

Bergmann C. 1847. Über die Verlhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Göttinger Studien* 3:595–708.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Bomblies K, Peichel CL. 2022. Genetics of adaptation. *Proceedings of the National Academy of Sciences* 119:e2122152119. DOI: 10.1073/pnas.2122152119.

Broad Institute. 2019. Picard Toolkit.

Carbeck K, Arcese P, Lovette I, Pruett C, Winker K, Walsh J. 2023. Candidate genes under selection in song sparrows co-vary with climate and body mass in support of Bergmann's Rule. *Nature Communications* 14:6974. DOI: 10.1038/s41467-023-42786-2.

Chaves JA, Cooper EA, Hendry AP, Podos J, De León LF, Raeymaekers JAM, MacMillan WO, Uy JAC. 2016. Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology* 25:5282–5295. DOI: 10.1111/mec.13743.

Choi JY, Dai X, Alam O, Peng JZ, Rughani P, Hickey S, Harrington E, Juul S, Ayroles JF, Purugganan MD, Stacy EA. 2021. Ancestral polymorphisms shape the adaptive radiation of *Metrosideros* across the Hawaiian Islands. *Proceedings of the National Academy of Sciences* 118:e2023801118. DOI: 10.1073/pnas.2023801118.

Delmore KE, Hübner S, Kane NC, Schuster R, Andrew RL, Câmara F, Guigó R, Irwin DE. 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology* 24:1873–1888. DOI: 10.1111/mec.13150.

Dingle H, Drake VA. 2007. What is migration? *BioScience* 57:113–121. DOI: 10.1641/B570206.

Eckert AJ, Carstens BC. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Molecular Phylogenetics and Evolution* 49:832–842. DOI: 10.1016/j.ympev.2008.09.008.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792–1797. DOI: 10.1093/nar/gkh340.

Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research* 36:W157–W162. DOI: 10.1093/nar/gkn337.

Eriksson J, Larson G, Gunnarsson U, Bed'hom B, Tixier-Boichard M, Strömstedt L, Wright D, Jungerius A, Vereijken A, Randi E, Jensen P, Andersson L. 2008. Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLOS Genetics* 4:e1000010. DOI: 10.1371/journal.pgen.1000010.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717–726. DOI: 10.1093/sysbio/sys004.

Faircloth BC. 2013. illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. DOI: 10.6079/J9ILL

Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32:786–788. DOI: 10.1093/bioinformatics/btv646.

Friis G, Fandos G, Zellmer AJ, McCormack JE, Faircloth BC, Milá B. 2018. Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology* 27:5137–5153. DOI: 10.1111/mec.14946.

Gibson DD, Withrow JJ. 2015. Inventory of the species and subspecies of Alaska birds, second edition. *Western Birds* 46:94–185.

Gillespie RG. 2002. Biogeography of spiders on remote oceanic islands of the Pacific: archipelagoes as stepping stones? *Journal of Biogeography* 29:655–662. DOI: 10.1046/j.1365-2699.2002.00714.x.

Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vásquez NJ, Finger JW, Pierson TW, Bentley KE, Hoffberg SL, Louha S, Garcia-De Leon FJ, del Rio Portilla MA, Reed KD, Anderson JL, Meece JK, Aggrey SE, Rekaya R, Alabady M, Belanger M, Winker K, Faircloth BC. 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* 7:e:7755. DOI: 10.7717/peerj.7755.

Graham AM, Peters JL, Wilson RE, Muñoz-Fuentes V, Green AJ, Dorfsman DA, Valqui TH, Winker K, McCracken KG. 2021. Adaptive introgression of the beta-globin cluster in two Andean waterfowl. *Heredity* 127:107–123. DOI: 10.1038/s41437-021-00437-6.

Grant PR. 1981. Speciation and the adaptive radiation of Darwin's Finches: The complex diversity of Darwin's finches may provide a key to the mystery of how intraspecific variation is transformed into interspecific variation. *American Scientist* 69:653–663.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* 35:518–522. DOI: 10.1093/molbev/msx281.

Huang J, Wang C, Ouyang J, Tang H, Zheng S, Xiong Y, Gao Y, Wu Y, Wang L, Yan X, Chen H. 2022. Identification of key candidate genes for beak length phenotype by whole-genome resequencing in geese. *Frontiers in Veterinary Science* 9:847481. DOI: 10.3389/fvets.2022.847481.

Kassambara A. 2023. *ggpubr: "ggplot2" Based Publication Ready Plots*.

Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecology Letters* 7:1225–1241. DOI: 10.1111/j.1461-0248.2004.00684.x.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, Grant BR, Grant PR, Webster MT, Andersson L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375. DOI: 10.1038/nature14181.

Leaché AD, Harris RB, Rannala B, Yang Z. 2014. The influence of gene flow on species tree estimation: a simulation study. *Systematic Biology* 63:17–30. DOI: 10.1093/sysbio/syt049.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. DOI: 10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. DOI: 10.1093/bioinformatics/btp352.

Liu R, Sun Y, Zhao G, Wang H, Zheng M, Li P, Liu L, Wen J. 2015. Identification of loci and genes for growth related traits from a genome-wide association study in a slow- × fast-growing broiler chicken cross. *Genes & Genomics* 37:829–836. DOI: 10.1007/s13258-015-0314-1.

Losos JB, Ricklefs RE. 2009. Adaptation and diversification on islands. *Nature* 457:830–836. DOI: 10.1038/nature07893.

Losos JB, Warheitt KI, Schoener TW. 1997. Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature* 387:70–73. DOI: 10.1038/387070a0.

Louha S, Ray DA, Winker K, Glenn TC. 2020. A high-quality genome assembly of the North American song sparrow, *Melospiza melodia*. *G3 Genes|Genomes|Genetics* 10:1159–1166. DOI: 10.1534/g3.119.400929.

MacArthur RH, Wilson EO. 1967. *The Theory of Island Biogeography*. Princeton University Press.

Mayr E. 1963. *Animal Species and Evolution*. Cambridge: Harvard University Press.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654. DOI: 10.1038/351652a0.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequence data. *Genome Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

Meiri S, Dayan T. 2003. On the validity of Bergmann's rule. *Journal of Biogeography* 30:331–351. DOI: 10.1046/j.1365-2699.2003.00837.x.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol., 37:1530-1534. DOI: 10.1093/molbev/msaa015

Murie OJ. 1959. *Fauna of the Aleutian Islands and Alaska Peninsula*. Washington, D.C.: US Fish and Wildlife Service.

Patten MA, Pruett CL. 2009. The song sparrow, *Melospiza melodia*, as a ring species: Patterns of geographic variation, a revision of subspecies, and implications for speciation. *Systematics and Biodiversity* 7:33–62. DOI: 10.1017/S1477200008002867.

Posit team. 2022. RStudio: Integrated Development Environment for R.

Pruett CL, Li T, Winker K. 2018. Population genetics of Alaska Common Raven show dispersal and isolation in the world's largest songbird. *The Auk* 135:868–880. DOI: 10.1642/AUK-17-144.1.

Pruett CL, Winker K. 2005a. Biological impacts of climatic change on a Beringian endemic: cryptic refugia in the establishment and differentiation of the Rock Sandpiper (*Calidris ptilocnemis*). *Climatic Change* 68:219–240. DOI: 10.1007/s10584-005-1584-4.

Pruett CL, Winker K. 2005b. Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology* 14:1421–1434. DOI: 10.1111/j.1365-294X.2005.02493.x.

R Core Team. R: A language and environment for statistical computing.

Reding DM, Foster JT, James HF, Pratt HD, Fleischer RC. 2008. Convergent evolution of 'creepers' in the Hawaiian honeycreeper radiation. *Biology Letters* 5:221–224. DOI: 10.1098/rsbl.2008.0589.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* 23:4757–4769. DOI: 10.1111/mec.12842.

San-Jose LM, Bestion E, Pellerin F, Richard M, Di Gesu L, Salmona J, Winandy L, Legrand D, Bonneaud C, Guillaume O, Calvez O, Elmer KR, Yurchenko AA, Recknagel H, Clobert J, Cote J. 2023. Investigating the genetic basis of vertebrate dispersal combining RNA-seq, RAD-seq and quantitative genetics. *Molecular Ecology* 32. DOI: 10.1111/mec.16916.

Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation. *Nature Reviews Genetics* 14:807–820. DOI: 10.1038/nrg3522.

Subramanian S. 2016. The effects of sample size on population genomic analyses – implications for the tests of neutrality. *BMC Genomics* 17:123. DOI: 10.1186/s12864-016-2441-8.

Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* 29:673–680. DOI: 10.1016/j.tree.2014.10.004.

Toews DPL, Hofmeister NR, Taylor SA. 2017. The evolution and genetics of carotenoid processing in animals. *Trends in Genetics* 33:171–182. DOI: 10.1016/j.tig.2017.01.002.

Turesson G. 1922. The species and the variety as ecological units. *Hereditas* 3:100–113. DOI: 10.1111/j.1601-5223.1922.tb02727.x.

Vallone D, Frigato E, Vernesi C, Foà A, Foulkes NS, Bertolucci C. 2007. Hypothermia modulates circadian clock gene expression in lizard peripheral tissues. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 292:R160–R166. DOI: 10.1152/ajpregu.00370.2006.

Walsh J, Benham PM, Deane-Coe PE, Arcese P, Butcher BG, Chan YL, Cheviron ZA, Elphick CS, Kovach AI, Olsen BJ, Shriver WG, Winder VL, Lovette IJ. 2019. Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters* 3:324–338. DOI: 10.1002/evl3.126.

Walsh N, Dale J, McGraw KJ, Pointer MA, Mundy NI. 2011. Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. *Proceedings of the Royal Society B: Biological Sciences* 279:58–66. DOI: 10.1098/rspb.2011.0765.

Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, Morlon H, Mouquet N, Rosindell J, Casquet J, Conti E, Cornuault J, Fernández-Palacios JM, Hengl T, Norder SJ, Rijsdijk KF, Sanmartín I, Strasberg D, Triantis KA, Valente LM, Whittaker RJ, Gillespie RG, Emerson BC, Thébaud C. 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters* 18:200–217. DOI: 10.1111/ele.12398.

Whittaker RJ, Fernadez-Palacios JM. 2007. *Island Biogeography: Ecology, Evolution, and Conservation*. Oxford: Oxford University Press.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4:1686. DOI: 10.21105/joss.01686.

Wickham H, Hester J, Chang W, Bryan J. 2022. devtools: Tools to Make Developing R Packages Easier. https://devtools.r-lib.org/

Williams GC. 1966. *Adaptation and Natural Selection*. Princeton University Press.

Winker K. 2000. Obtaining, preserving, and preparing bird specimens. *Journal of Field Ornithology* 71:250–297. DOI: 10.1648/0273-8570-71.2.250.

Winker K. 2010. Chapter 1: Subspecies represent geographically partitioned variation, a gold mine of evolutionary biology, and a challenge for conservation. *Ornithological Monographs* 67:6–23. DOI: 10.1525/om.2010.67.1.6.

Winker K, Withrow JJ, Gibson DD, Pruett CL. 2023. Beringia as a high-latitude engine of avian speciation. *Biological Reviews* 98:1081–1099. DOI: 10.1111/brv.12945.

3.7 Figures

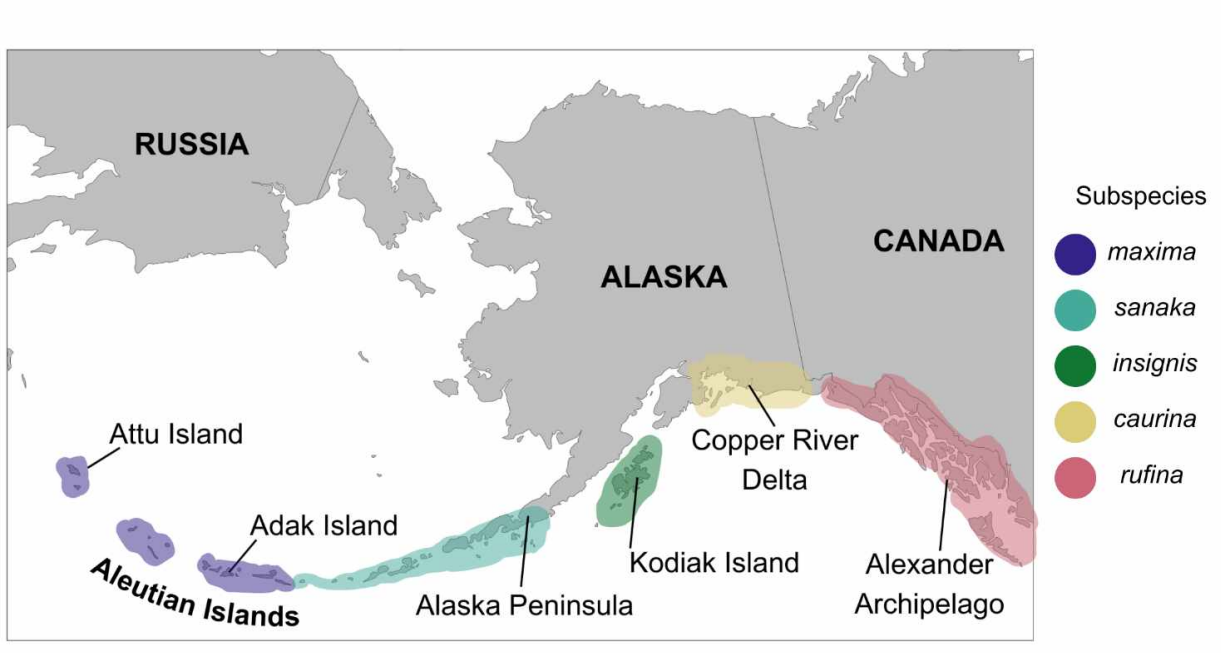

Fig. 3.1: Distribution of five subspecies of song sparrow (*Melospiza melodia*) in Alaska. From west to east: *M. m. maxima* (dark blue), *M. m. sanaka* (teal), *M. m. insignis* (green), *M. m. caurina* (yellow), and *M. m. rufina* (pink).

Fig. 3.2: Museum specimens of five *M. melodia* subspecies of interest that occur in Alaska. Colored circles correspond to the colors in Fig. 3.1. Photo credit: Caitlyn C. Oliver Brown

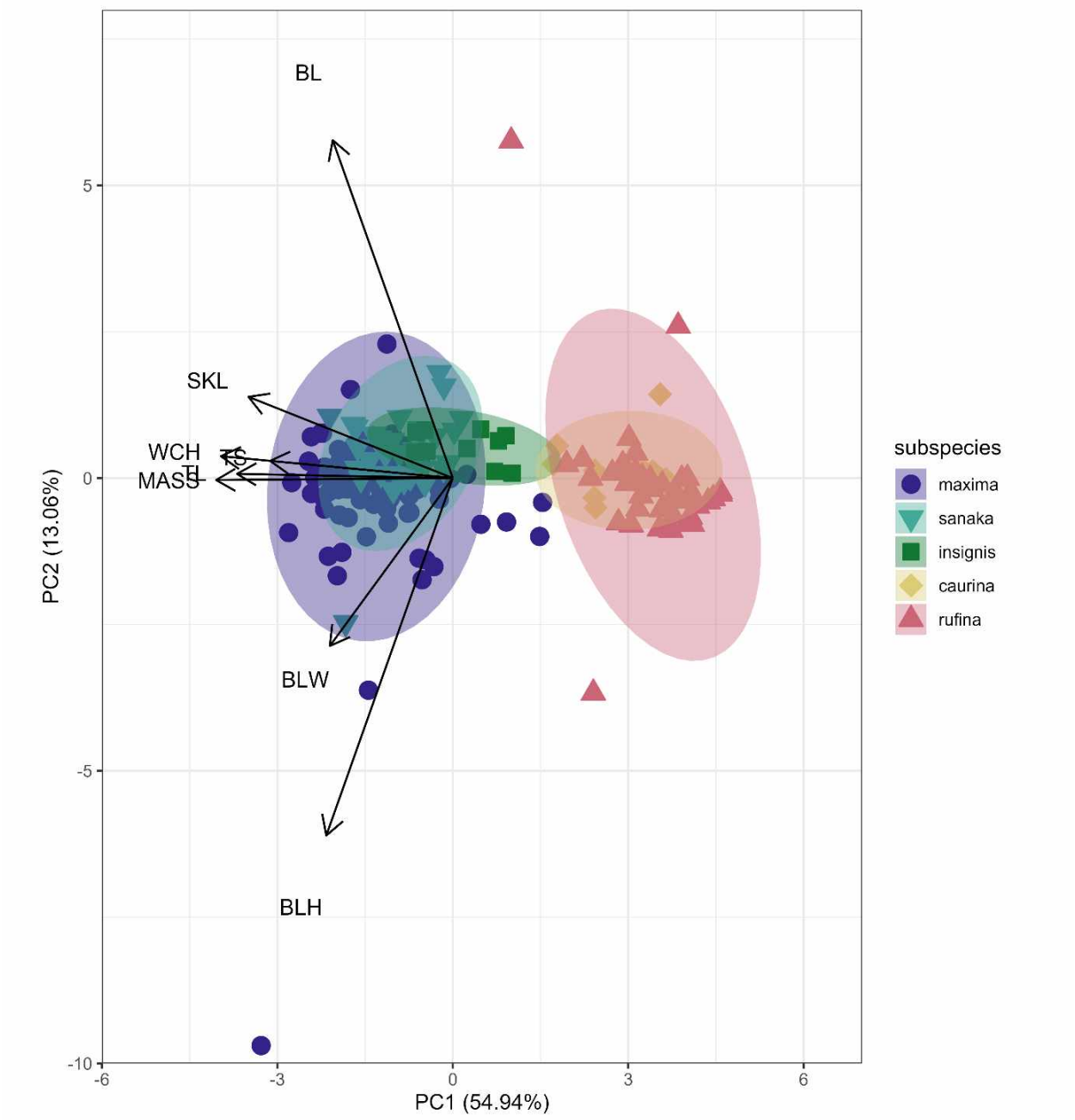

Fig. 3.3: PCA plot of phenotypic data for the five *Melospiza melodia* subspecies. PC1 explains approximately 55% of the variation with five phenotypic traits: mass (MASS), wing chord (WCH), tail length (TL), tarsus length (TS), and skull length (SKL). PC2 explains 13% of the variation with the three remaining traits: bill length (BL), bill width (BLW), and bill height (BLH). Three subspecies (*maxima, sanaka*, and *insignis*) group together and are larger from the second group of two subspecies (*caurina* and *rufina*). Colors correspond to Fig. 3.1.

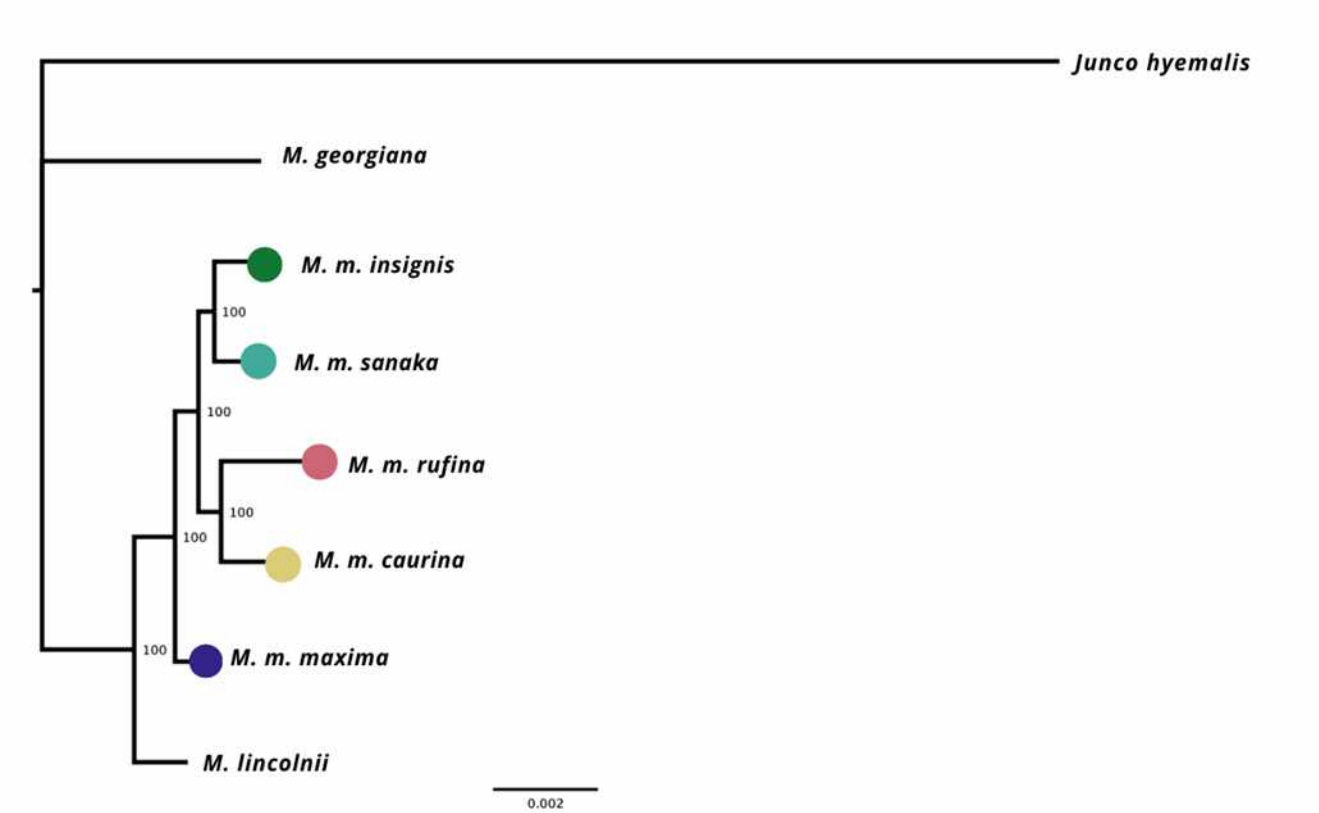

Fig. 3.4: Maximum-likelihood phylogeny based on 4793 UCE loci. Values on the nodes represent bootstrap values from 1000 iterations. *M. m. maxima* is shown to be a sister group to the other *M. melodia* subspecies.

3.8 Tables

Table 3.1: *p*-value results from McDonald-Kreitman tests. Gene names and *p*-values in bold denote significant results (< 0.05). Significance indicates a gene is under positive selection.

Phenotypic Trait	Candidate Genes	MK Test *p*-values	Citation
Bill size	*ALX1*	NULL	Lamichhaney et al., 2015
	CCDC149	0.781	Huang et al., 2022
	DLK1	0.611	Chaves et al., 2016
	HMGA2	0.145	Chaves et al., 2016
	LGI2	0.280	Huang et al., 2022
Body size	*AFAP1*	0.376	Liu et al., 2015
	GGA1	0.808	Liu et al., 2015
	LCORL	NULL	Liu et al., 2015
	QDPR	0.363	Liu et al., 2015
	SLIT2	0.792	Liu et al., 2015
Plumage color	*APOD1*	0.316	Walsh et al., 2011
	BCO1	**0.037**	Walsh et al., 2011
	CD36	0.924	Walsh et al., 2011
Dispersal	***KCTD21***	**0.032**	San-Jose et al., 2023
	SLC2A1	NULL	San-Jose et al., 2023
	TGFB2	0.264	San-Jose et al., 2023
Migration	*CLOCK*	0.849	Vallone et al., 2007
	CREB1	0.820	Ruegg et al., 2014
	CRY1	NULL	Vallone et al., 2007
	CRY2	0.675	Vallone et al., 2007
	NPAS3	0.300	Ruegg et al., 2014
	PER2	1.000	Vallone et al., 2007
	PER3	0.314	Ruegg et al., 2014
	YPEL1	0.279	Delmore et al., 2015
Salt tolerance	*MMP17*	0.156	Walsh et al., 2019
	SLC9A3	0.868	Walsh et al., 2019

Table 3.2: Mean (± standard deviation) for the phenotypic traits compared across the five song sparrow subspecies of interest. Values with different letters are significantly different based on a Tukey's post-hoc test.

	M. m. maxima	*M. m. sanaka*	*M. m. insignis*	*M. m. caurina*	*M. m. rufina*
Mass (g)	46.2 ± 2.71 (A)	45.4 ± 2.06 (A)	41.0 ± 2.61 (B)	28.4 ± 1.77 (C)	27.9 ± 1.75 (C)
Wing chord length (mm)	82.7 ± 3.04 (A)	83.3 ± 2.45 (A)	80.9 ± 2.14 (A)	72.6 ± 1.43 (B)	69.7 ± 3.03 (B)
Tail length (mm)	78.9 ± 4.27 (A)	77.8 ± 4.28 (A)	75.4 ± 4.52 (A)	67.1 ±4.28 (B)	65.8 ± 3.01 (B)
Tarsus length (mm)	27.6 ± 2.04 (A)	27.0 ± 0.99 (A)	26.3 ± 0.53 (AB)	24.2 ±1.32 (BC)	24.2 ± 1.41 (C)
Bill length (mm)	12.4 ± 1.34 (A)	12.4 ± 0.51 (A)	11.9 ± 0.57 (AB)	10.7 ± 0.60 (B)	10.8 ± 2.75 (B)
Bill height (mm)	7.5 ± 1.08 (A)	6.89 ± 1.04 (B)	6.5 ± 0.22 (AB)	6.1 ± 0.45 (AB)	6.37 ± 0.87 (B)
Bill width (mm)	5.9 ± 0.77 (A)	5.6 ± 0.56 (B)	5.3 ± 0.32 (BC)	5.1 ± 0.34 (BC)	5.1 ± 0.45 (C)
Skull length (mm)	38.8 ± 1.54 (A)	38.8 ± 0.80 (A)	38.0 ± 0.95 (A)	35.7 ± 1.93 (B)	34.8 ± 1.41 (B)

Chapter 4: General Conclusions

Beringia impacted avian speciation in two primary ways: through cyclic separation and connection of the Asian and North American continents and through glacial refugia providing isolation (Winker et al., 2023). In this thesis, I examined how this dynamic system affected the generation of avian diversity in Beringia. Firstly, I examined Asian and North American lineages to determine how the land bridge and/or refugia impacted divergence. Secondly, I narrowed my focus to one Alaskan species that may have inhabited a glacial refugium during the Pleistocene. My results provide key details about how Pleistocene biogeography impacted high-latitude divergence and speciation events.

In Chapter 2, I examined how the cyclic separation and connection of populations has impacted Beringian birds using ultraconserved elements (UCEs), a multilocus subsampling of the nuclear genome. I found that divergence with gene flow is the predominant process of divergence and speciation in Beringia. In seven pairwise comparisons of Beringia avian taxa, a divergence with symmetrical migration (gene flow) was the best-fit model (Table 2.3). Other best-fit models included secondary contact with both symmetrical and asymmetrical migration and isolation with exponential growth and migration (Table 2.3). Gene flow (as number of migrants per generation) ranged from less than one to over 200 individuals, and time estimates placed all divergences during the Pleistocene (Table 2.4). I also summarized the results from previous studies of avian divergence in this region and determined that a split with symmetrical migration model was the predominant mode of divergence across all taxa (Fig. 2.3). The IM model, which suggests divergence with exponential population growth, was the least commonly supported, being the best-fit model for only three pairwise comparisons (Table 2.3). Due to the IM model's change in effective population size parameter, these lineages likely experienced an event that dramatically increased population size. Overall, gene flow was common in all Beringian avifauna, and a history of strict allopatry (no gene flow) accompanying divergence and speciation was completely absent.

In Chapter 3, I examined morphology and whole-genome sequences of five subspecies of Song Sparrow (*Melospiza melodia*) that reside across southern Alaska (Fig. 3.1). It has been suggested that Song Sparrows colonized southern Alaska through a stepping-stone method from east to west (Pruett & Winker, 2005). The western-most subspecies, *M. m. maxima*, inhabits the

end of the Aleutian Island archipelago and is much larger than mainland subspecies (Fig. 3.2). I tested candidate genes associated with key phenotypic traits to determine if these genes were under positive selection in *M. m. maxima*. While I expected genes linked to body size to be under positive selection, I did not find evidence of this in the genes I tested. However, I found two genes, *BCO1* (associated with plumage color) and *KCTD21* (associated with dispersal), were under positive selection in *M. m. maxima* (Table 3.2). Using UCEs, I reconstructed a phylogeny of Alaska Song Sparrows and found that *M. m. maxima* is sister to all other *M. melodia* subspecies (Fig. 3.4). With these results, I determined that one subspecies (*M. m. maxima*) likely persisted through the last glacial maximum in a glacial refugium in the western Aleutian Islands. I hypothesize that Song Sparrows occupied their entire modern Alaska range before being mostly extirpated from it during the LGM. It seems likely that the *maxima* population was the only remnant until the rest of the range was recolonized by North American continental populations.

4.1 References

Pruett, C. L., & Winker, K. (2005). Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology*, *14*(5), 1421–1434. https://doi.org/10.1111/j.1365-294X.2005.02493.x

Winker, K., Withrow, J. J., Gibson, D. D., & Pruett, C. L. (2023). Beringia as a high-latitude engine of avian speciation. *Biological Reviews*, *98*(4), 1081–1099. https://doi.org/10.1111/brv.12945

Appendix A: Supplementary Material for Chapter 2

Table A.1: Specimen information used in this study. UAM = University of Alaska Museum; UWBM = University of Washington Burke Museum. All SRA reads are available under BioProject PRJNA1129545.

Taxon	Locality	Tissue Number	Catalogue Number	Sequence Read Archive
Mergus merganser americanus	USA: Washington, Douglas County	AEI105	UWBM 84934	SAMN42159334
Mergus merganser americanus	USA: Alaska, Kodiak Island	JJW2298	UAM34366	SAMN42159335
Mergus merganser americanus	UAS: Alaska, Interior, Delta	UAMX1058	UAM11037	SAMN42159336
Mergus merganser americanus	USA: Alaska, Kodiak Island	UAMX4957	UAM24305	SAMN42159337
Mergus merganser merganser	Russia: Magadan, Magadanskaya Oblast'	CSW4346	UWBM 43819	SAMN42159338
Mergus merganser merganser	USA: Alaska, Aleutian Islands, Shemya Island	JJW063	UAM25375	SAMN42159339
Mergus merganser merganser	USA: Alaska, Aleutian Islands, Kiska Island	JJW334	UAM25991	SAMN42159340
Mergus merganser merganser	USA: Alaska, Aleutian Islands, Kiska Island	JJW335	UAM25990	SAMN42159341
Mergus merganser merganser	USA: Alaska, Aleutian Islands, Kiska Island	JJW344	UAM28042	SAMN42159342
Mergus merganser merganser	USA: Alaska, Aleutian Islands, Attu Island	JJW762	UAM27012	SAMN42159343
Gavia stellata	Russia: Chukotskiy Avtonomnyy Okrug, Anadyr'	CSW4461	UWBM 43925	SAMN42159308
Gavia stellata	Russia: Chukotskiy Avtonomnyy Okrug, Anadyr'	DAB063	UWBM 44122	SAMN42159309
Gavia stellata	USA: Alaska, Seward Peninsula	DAR275	UAM13601	SAMN42159310
Gavia stellata	USA: Alaska, North Slope, Barrow	JJW3574	UAM40094	SAMN42159311
Gavia stellata	Russia: Avtonomnaya Respublika Yakutiya, Cherskiy	JMB1098	UWBM 44259	SAMN42159312
Gavia stellata	Russia: Avtonomnaya Respublika Yakutiya, Cherskiy	JMB1099	UWBM 44260	SAMN42159313
Gavia stellata	USA: Alaska, Interior, Alaska Range	KGM628	UAM20711	SAMN42159314
Gavia stellata	Russia: Magadanskaya Oblast', Talon	SAR6293	UWBM 44653	SAMN42159315
Gavia stellata	USA: Alaska, North Slope, Barrow	UAMX3952	UAM38314	SAMN42159317
Gavia stellata	USA: Alaska, North Slope, Barrow	UAMX6608	UAM40093	SAMN42159316

Table A.1 (continued)

Arenaria interpres	USA: Alaska, Seward Peninsula, Wooley Lagoon Road	DAR268	UAM14954	SAMN42159344
Arenaria interpres	USA: Alaska, Aleutian Islands, Chirikof Island	JJW676	UAM26977	SAMN42159345
Arenaria interpres	USA: Alaska, Kodiak archipelago, Ugak Island	JJW1637	UAM30797	SAMN42159346
Arenaria interpres	USA: Alaska, Alaska Peninsula, Cold Bay	JMM422	UAM20220	SAMN42159347
Arenaria interpres	USA: Alaska, Aleutian Islands, Attu Island	UAMX1744	UAM13527	SAMN42159348
Arenaria melanocephala	USA: Alaska, Kodiak archipelago, Ugak Island	JJW1153	UAM28748	SAMN42159349
Arenaria melanocephala	USA: Alaska, Seward Peninsula, Safety Sound	JJW3588	UAM40266	SAMN42159350
Arenaria melanocephala	USA: Alaska, Bering Coast, Jacksmith Bay	JMM326	UAM20142	SAMN42159351
Arenaria melanocephala	USA: Alaska, Bering Coast, Jacksmith Bay	JMM327	UAM20098	SAMN42159352
Arenaria melanocephala	USA: Alaska, Gulf of Alaska, Middleton Island, southend	JJW2157	UAM34087	SAMN42159353
Calidris alpina arcticola	USA: Alaska, North Slope, Barrow	JJW4409	UAM42687	SAMN42159354
Calidris alpina arcticola	USA: Alaska, North Slope, Barrow	JJW4410	UAM42688	SAMN42159355
Calidris alpina arcticola	USA: Alaska, North Slope, Barrow	JJW4411	UAM42689	SAMN42159356
Calidris alpina arcticola	USA: Alaska, North Slope, Barrow	UAMX3267	UAM20132	SAMN42159357
Calidris alpina arcticola	USA: Alaska, North Slope, Barrow	TB1751	UAM19228	SAMN42159358
Calidris alpina pacifica	USA: Alaska, Kodiak Island	JJW015	UAM26914	SAMN42159359
Calidris alpina pacifica	USA: Alaska, Kodiak Island	JJW016	UAM26913	SAMN42159360
Calidris alpina pacifica	USA: Alaska, Alaska Peninsula, Chirikof Island	JJW704	UAM27576	SAMN42159361
Calidris alpina pacifica	USA: Alaska, Kodiak Island	JJW1490	UAM30313	SAMN42159362

Table A.1 (continued)

Uria aalge inornata	Russia: Magadanskaya Oblast', Magadan	JMB994	UWBM 44407	SAMN42159363
Uria aalge inornata	Russia: Magadanskaya Oblast', Magadan	JMB996	UWBM 44409	SAMN42159364
Uria aalge inornata	Russia: Magadanskaya Oblast', Magadan	JMB997	UWBM 44410	SAMN42159365
Uria aalge inornata	Russia: Magadanskaya Oblast', Magadan	JMB998	UWBM 44411	SAMN42159366
Uria aalge inornata	USA: Alaska, Bering Sea, St. Matthew Island	KSW1629	UAM7729	SAMN42159367
Uria aalge inornata	USA: Alaska, Bering Sea, St. Matthew Island	KSW1631	UAM10687	SAMN42159368
Uria aalge inornata	USA: Alaska, Bering Sea, St. Matthew Island	KSW1632	UAM10688	SAMN42159369
Uria aalge inornata	USA: Alaska, Bering Sea, St. Matthew Island	KSW1633	UAM10689	SAMN42159370
Uria aalge inornata	USA: Alaska, Bering Sea, St. Matthew Island	KSW1634	UAM10690	SAMN42159371
Uria lomvia arra	North Pacific Ocean	BKS351	UWBM 55300	SAMN42159372
Uria lomvia arra	USA: Alaska, Kodiak archipelago, Taliudek Island	JJW1509	UAM30510	SAMN42159373
Uria lomvia arra	USA: Alaska, Kodiak archipelago, Taliudek Island	JJW1510	UAM30511	SAMN42159374
Uria lomvia arra	Japan: Honshu, off NE coast	JMB857	UWBM 55725	SAMN42159375
Uria lomvia arra	Russia: Magadanskaya Oblast', Magadan	JMB1001	UWBM 44177	SAMN42159376
Uria lomvia arra	Russia: Magadanskaya Oblast', Magadan	JMB1002	UWBM 44178	SAMN42159377
Uria lomvia arra	North Pacific Ocean: Bering Sea, Chirikov Basin	NCSM17754	UWBM 72702	SAMN42159378
Uria lomvia arra	USA: Alaska, North Slope, Barrow	UAMX387	UAM9608	SAMN42159379
Uria lomvia arra	USA: Alaska, Aleutian Islands, Aiktak Island	UAMX1553	UAM13445	SAMN42159380
Uria lomvia arra	USA: Alaska, Aleutian Islands, Bogoslof Island	UAMX3290	UAM20236	SAMN42159381

Table A.1 (continued)

Picoides dorsalis fasciatus	USA: Alaska, Kodiak Island	JJW2059	UAM31832	SAMN42159298
Picoides dorsalis fasciatus	USA: Alaska, SE Alaska, Hyder	JJW2842	UAM36600	SAMN42159299
Picoides dorsalis fasciatus	USA: Alaska, Interior, Chatanika R.	JJW3139	UAM38328	SAMN42159300
Picoides dorsalis fasciatus	USA: Alaska, SW Alaska, Dillingham	UAMX503	UAM8580	SAMN42159301
Picoides dorsalis fasciatus	USA: Alaska, Kenai Peninsula, Sterling	UAMX3273	UAM20114	SAMN42159302
Picoides tridactylus	Russia: Magadanskaya Oblast', Magadan	SAR6046	UWBM 44422	SAMN42159303
Picoides tridactylus	Russia: Magadanskaya Oblast', Khasynskiy Rayon	SVD144	UWBM 52660	SAMN42159304
Picoides tridactylus	Russia: Chukotskiy Avtonomnyy Okrug, Anadyrskiy Rayon	SVD3025	UWBM 82309	SAMN42159305
Picoides tridactylus	Mongolia: Selenge	TB1448	UAM17739	SAMN42159306
Picoides tridactylus	Mongolia: Hovd area	n.a.	UAM10581	SAMN42159307
Anthus rubescens japonicus	USA: Alaska, Aleutian Islands, Shemya Island	DDG1877	UAM13353	SAMN42159288
Anthus rubescens japonicus	USA: Alaska, Aleutian Islands, Shemya Island	JJW864	UAM27351	SAMN42159290
Anthus rubescens japonicus	USA: Alaska, Aleutian Islands, Shemya Island	JJW905	UAM28057	SAMN42159291
Anthus rubescens japonicus	USA: Alaska, Gulf of Alaska, Middleton Island	JJW3455	UAM39615	SAMN42159289
Anthus rubescens japonicus	USA: Alaska, Aleutian Islands, Attu Island	KSW3980	UAM15068	SAMN42159292
Anthus rubescens pacificus	USA: Alaska, Aleutian Islands, Frosty Peak	CLP200	UAM11452	SAMN42159293
Anthus rubescens pacificus	USA: Alaska, Kodiak Island	JJW2045	UAM31927	SAMN42159294
Anthus rubescens pacificus	USA: Alaska, Kodiak Island	JJW2046	UAM31928	SAMN42159295
Anthus rubescens pacificus	USA: Alaska, North Slope, Dalton Highway	MJM037	UAM19419	SAMN42159296
Anthus rubescens pacificus	USA: Alaska, Interior, Eagle Summit	RWD25949	UAM13540	SAMN42159297

Table A.1 (continued)

Corvus corax corax	Mongolia: Selenge	TB1445	UAM17737	SAMN42159331
Corvus corax corax	Mongolia: Arhangay Aymag, Hangayn Nuruu	BKS4041	UWBM 57899	SAMN42159320
Corvus corax corax	Mongolia: Arhangay Aymag, Hangayn Nuruu	BKS4052	UWBM 57909	SAMN42159321
Corvus corax corax	Russia: Tyumenskaya Oblast', Noyabr'sk	BKS3241	UWBM 56544	SAMN42159319
Corvus corax corax	Russia: Yamalo-Nenetskiy Avtonomnyy Okrug, Labytnangi	SVD1398	UWBM 59566	SAMN42159330
Corvus corax kamtschaticus	Russia: Magadanskaya Oblast', Yana River	BKS908	UWBM 46861	SAMN42159318
Corvus corax kamtschaticus	Russia: Magadanskaya Oblast', Tauy River	SVD231a	UWBM 71244	SAMN42159329
Corvus corax kamtschaticus	USA: Alaska, Aleutian Islands, Attu Island	DDG1749	UAM10021	SAMN42159322
Corvus corax kamtschaticus	USA: Alaska, Aleutian Islands, Shemya Island	JMM133	UAM17761	SAMN42159325
Corvus corax kamtschaticus	USA: Alaska, Aleutian Islands, Attu Island	RWD26009	UAM15040	SAMN42159328
Corvus corax kamtschaticus	USA: Alaska, Aleutian Islands, Attu Island	UAMX844	UAM9313	SAMN42159332
Corvus corax kamtschaticus	USA: Alaska, Aleutian Islands, Attu Island	UAMX1061	UAM11113	SAMN42159333
Corvus corax principalis	USA: Alaska, Kodiak archipelago, Sitkinak Island	JJW2991	UAM38000	SAMN42159323
Corvus corax principalis	USA: Alaska, Bering Sea, St. Matthew Islands	JJW4373	UAM41600	SAMN42159324
Corvus corax principalis	USA: Alaska, Aleutian Islands, Izembek Lagoon	KSW2715	UAM8803	SAMN42159326
Corvus corax principalis	USA: Alaska, Seward Peninsula, Unalakleet	KSW4763	UAM22252	SAMN42159327

Table A.1 (continued)

Motacilla tschutschensis simillima	USA: Alaska, Aleutian Islands, Attu Island	ABJ133	UAM13402	SAMN42159382
Motacilla tschutschensis simillima	USA: Alaska, Aleutian Islands, Attu Island	DDG1935	UAM15178	SAMN42159383
Motacilla tschutschensis simillima	USA: Alaska, Aleutian Islands, Attu Island	JJW216	UAM27366	SAMN42159384
Motacilla tschutschensis simillima	USA: Alaska, Aleutian Islands, Shemya Island	JJW520	UAM27364	SAMN42159385
Motacilla tschutschensis simillima	USA: Alaska, Aleutian Islands, Shemya Island	JJW629	UAM27367	SAMN42159386
Motacilla tschutschensis tschutschensis	USA: Alaska, Dalton Highway, Sagavanirktok R.	ABJ179	UAM14353	SAMN42159387
Motacilla tschutschensis tschutschensis	USA: Alaska, Seward Peninsula	JJW1179	UAM28715	SAMN42159388
Motacilla tschutschensis tschutschensis	USA: Alaska, Seward Peninsula	JJW1185	UAM28717	SAMN42159389
Motacilla tschutschensis tschutschensis	USA: Alaska, Seward Peninsula	JJW1317	UAM29362	SAMN42159390
Motacilla tschutschensis tschutschensis	USA: Alaska, North Slope, Dalton Highway	MJM122	UAM19448	SAMN42159391

Table A.1 (continued)

Phylloscopus borealis borealis	Russia: Magadanskaya Oblast', Magadan	CSW4345	UWBM 43818	SAMN42159392
Phylloscopus borealis borealis	Russia: Magadanskaya Oblast', Magadan	SAR6041	UWBM 44417	SAMN42159393
Phylloscopus borealis borealis	Russia: Magadanskaya Oblast', Magadan	SAR6042	UWBM 44418	SAMN42159394
Phylloscopus borealis borealis	Russia: Chukotskiy Avtonomnyy Okrug, Anadyrskiy Rayon	SVD2977	UWBM 82284	SAMN42159395
Phylloscopus borealis borealis	Russia: Chukotskiy Avtonomnyy Okrug, Anadyrskiy Rayon	SVD2978	UWBM 82285	SAMN42159396
Phylloscopus borealis kennicotti	USA: Alaska, Seward Peninsula, Snake River	JJW2241	UAM34615	SAMN42159397
Phylloscopus borealis kennicotti	USA: Alaska, SW Alaska, Kilbuck, Goodnews River	JMM337	UAM20149	SAMN42159398
Phylloscopus borealis kennicotti	USA: Alaska, SW Alaska, Kilbuck, Goodnews River	JMM338	UAM20150	SAMN42159399
Phylloscopus borealis kennicotti	USA: Alaska, Alaska Range, Mile 60 Denali Highway	KSW1456	UAM7372	SAMN42159400
Phylloscopus examinandus	USA: Alaska, Aleutian Islands, Shemya Island	DDG1991	UAM17665	SAMN42159401
Phylloscopus examinandus	Russia: Kamchatka, Petropavlosk	KSW5250	UAM29472	SAMN42159402
Phylloscopus examinandus	Russia: Kamchatka, Esso	KSW5336	UAM30464	SAMN42159403
Phylloscopus examinandus	Russia: Kamchatka, Klyuchivskaya Lake	KSW5354	UAM30462	SAMN42159404
Phylloscopus examinandus	Russia: Kamchatka, Izba	KSW5373	UAM30461	SAMN42159405

Table A.2: GenBank accession numbers for reference sequences used when identifying Z-linked loci.

Order of Taxa of Interest	Reference Species	NCBI GenBank Accession Number
Anseriformes	*Anas platyrhynchos*	GCF_015476345.1
Gaviiformes	*Gavia stellata*	GCF_030936135.1
Charadriiformes	*Sterna hirundo*	GCA_009819605.1
Piciformes	*Dryobates pubescens*	GCF_014839835.1
Passeriformes	*Taeniopygia guttata*	GCF_003957565.2

Table A.3: Genomes used for calculation of substitution rates and time since most recent common ancestor (TMRCA, from Claramunt & Cracraft, 2015), substitution rate, and generation time for each lineage.

	Genome	GenBank accession number	TMRCA (my)	Substitution rate (sub/site/gen)	Generation time (yrs)	Sources
Mergus merganser subspp	*Anser cygnoides*	GCA_002166845.1	28.132	1.86E-10	3.17	(Pearce et al., 2020)
Arenaria interpres / A. melanocephala	*Charadrius vociferus*	GCA_000708025.2	53.5153	7.74E-10	4.58	(Nettleship, 2020)
Calidris alpina subspp	*Charadrius vociferus*	GCA_000708025.2	53.5153	1.01E-10	4.7	(Warnock & Gill, 2023)
Uria aalge	*Larus argentatus*	GCA_028571125.1	26.5075	2.31E-11	23	(Ainley et al., 2021)
Uria lomvia	*Larus argentatus*	GCA_028571125.1	26.5075	3.30E-11	13.79	(Gaston & Hipfner, 2020)
Gavia stellata	*Aptendodytes forsteri*	GCA_000699145.1	63.0676	2.18E-11	16.5	(Rizzolo et al., 2020)
Picoides dorsalis/ P. tridactylus	*Indicator indicator*	GCA_027791375.2	35.6636	1.03E-09	2.95	(Pasinelli, 2006; Tremblay et al., 2020)
Corvus corax	*Taeniopygia guttata*	GCA_003957565.4	41.5132	1.35E-10	5.15	(Boarman & Heinrich 2020)
Phylloscopus borealis subspp	*Parus major*	GCA_001522545.3	37.6154	3.37E-10	1.89	(Pratt & Peach, 1991; Lowther & Sharbaugh, 2020)
Anthus rubescens subspp	*Taeniopygia guttata*	GCA_003957565.4	28.8236	4.24E-10	1.85	(Hendricks & Verbeek, 2020; Spaepen, 1988)
Motacilla tschutschensis subspp	*Taeniopygia guttata*	GCA_003957565.4	28.8236	4.51E-10	1.54	(Badyaev et al., 2020)

Table A.4: Overall results of assembly of two reference individuals in each lineage, including number of contigs, total bp, average length, minimum and maximum length, and number of contigs over 1 Kb.

Taxa	Contigs	Total bp	Average length	Min length	Max length	Contigs > 1Kb
Mergus merganser	54127	22798932	421.2118167	78	6512	4392
Gavia stellata	83745	26314437	314.2209923	78	8804	4036
Arenaria interpres / A. melanocephala	87900	37738433	429.3337088	78	7478	5101
Calidris alpina	284753	49863146	175.1101692	78	3264	3256
Uria aalge	71452	29021880	406.1730952	78	14398	4484
Uria lomvia	177818	36765270	206.7578648	77	10690	3234
Picoides tridactylus / P. dorsalis	79922	32447718	405.9923175	78	11067	4790
Anthus rubescens	66657	27793958	416.9698306	77	7522	4715
Corvus corax	73655	32351946	439.23625	78	17005	5257
Motacilla tschutschensis	320689	56200041	175.2477977	78	4530	3374
Phylloscopus borealis	79490	31327648	394.1080387	78	5036	4611

Table A.5: UCE reference assemblies used in each lineage, including number of UCEs, total bp, average length, minimum and maximum length, and number of contigs over 1 Kb.

Taxa	UCE loci identified	Total bp	Average length	Min length	Max length	Contigs > 1Kb
Mergus merganser	4056	6170039	1521.212771	235	5741	3695
Gavia stellata	4029	5762141	1430.166543	108	4304	3524
Arenaria interpres / A. melanocephala	4131	6705160	1623.132413	242	5287	3979
Calidris alpina	2168	2423574	1117.884686	134	3102	1459
Uria aalge	4071	6066822	1490.2535	229	3275	3647
Uria lomvia	3917	4593528	1172.715854	158	2966	2721
Picoides tridactylus / P. dorsalis	3986	5921854	1485.663322	229	3878	3535
Anthus rubescens	4083	6438673	1576.946608	173	3917	3723
Corvus corax	4102	6691820	1631.355436	210	5231	3883
Motacilla tschutschensis	2147	2571434	1197.687005	142	3462	1466
Phylloscopus borealis	4118	6309350	1532.139388	136	3780	3707

Table A.6: Expected (H_E) and observed heterozygosities (H_O) for each population and a statistical test of whether they are different through adegenet pairwise t-test (significance indicated by an asterisk, *).

	Comparisons	H_E	H_O	**t**	**df**	***p*** (H_E vs H_O)
Mergus merganser subspp	*merganser/americanus*	0.183 / 0.128	0.193 / 0.147	13.258	2912	< 2.2e-16 *
Arenaria interpres / A. melanocephala	*interpres / melanocephala*	0.173 / 0.070	0.189 / 0.076	37.417	3995	< 2.2e-16 *
Calidris alpina subspp	*arcticola / pacifica*	0.163 / 0.169	0.178 / 0.176	1.309	1817	0.0954
Uria aalge	Asia / North America	0.228 / 0.237	0.264 / 0.262	-9.309	3728	1
Uria lomvia	Asia / North America	0.164 / 0.165	0.190 / 0.179	-9.792	3654	1
Gavia stellata	Asia / North America	0.197 / 0.192	0.212 / 0.216	-4.886	3345	1
Picoides dorsalis / P. tridactylus	*tridactylus / dorsalis*	0.149 / 0.129	0.167 / 0.139	19.819	3008	< 2.2e-16 *
Corvus corax	Asia / North America	0.260 / 0.089	0.028 / 0.009	122.77	3990	< 2.2e-16 *
Corvus corax	Asia / Attu	0.260 / 0.011	0.028 / 0.012	123.64	3990	< 2.2e-16 *
Corvus corax	North America / Attu	0.009 / 0.011	0.009 / 0.012	1.849	3990	0.0322 *
Phylloscopus borealis borealis / P. examinandus	*borealis / examinandus*	0.144 / 0.163	0.158 / 0.178	3.7193	3913	1.012e-4 *
Phylloscopus borealis subspp.	*borealis / kennicotti*	0.144 / 0.132	0.158 / 0.149	-5.1811	3913	1
Anthus rubescens subspp	*japonicus / pacificus*	0.115 / 0.134	0.121 / 0.146	19.466	3807	< 2.2e-16 *
Motacilla tschutschensis subspp	Asia / North America	0.079 / 0.245	0.090 / 0.071	45.091	2044	< 2.2e-16 *

Table A.7: Raw parameter output from δaδi used to calculate biological estimates. The average of the top three runs output from the δaδi best-fit and runner-up models for each pairwise comparison is paired with the ±95% confidence interval around that parameter (in parentheses). Instances of a hyphen indicated that the parameter is not present in that best-fit model. These values were translated into biological estimates, which are listed in Table 2.4.

Pairwise comparison (nu1/nu2)	Best fit model via δaδi	nu1	nu2	m12	m21	m	T	T_{sc}	S	Theta
Mergus merganser americamus / M. m. merganser	SC 2m	2.51 (±0.19)	3.70 (±0.22)	0.19 (±0.15)	3.05 (±0.20)	-	2.21 (±0.12)	0.093 (±0.30)	-	223.53 (±43.98)
Arenaria interpres / A. melanocephala	IM	6.94 (±0.59)	0.95 (±0.09)	0.013 (±0.08)	0.014 (±0.16)	-	3.33 (±0.15)	-	0.56 (±0.05)	345.03 (±111.75)
Calidris alpina arcticola / C. a. pacifica	Split mig	13.71 (±1.55)	2.98 (±1.73)	-	-	4.91 (±0.41)	1.18 (±0.15)	-	-	150.31 (± 13.00)
	SC 1m	9.99 (±0.92)	3.11 (±0.24)	-	-	5.998 (±0.47)	0.84 (±0.17)	0.32 (±0.18)	-	152.69 (±13.99)
Uria aalge (Asia / North America)	Split mig	1.04 (±0.96)	3.50 (±2.27)	-	-	10.00 (±0.87)	2.08 (±0.45)	-	-	562.43 (±53.58)
Uria lomvia (Asia / North America)	Split mig	5.02 (±2.01)	11.94 (±3.04)	-	-	10.66 (±1.00)	1.15 (±0.16)	-	-	295.61 (±32.31)
Gavia stellata (North America / Asia)	Split mig	1.92 (±0.92)	3.13 (±1.97)	-	-	11.67 (±0.52)	0.58 (±0.17)	-	-	456.07 (±25.21)
Picoides dorsalis / P. tridactylus	IM	5.58 (±0.30)	5.26 (±0.34)	0.03 (±0.03)	0.01 (±0.04)	-	1.61 (±0.13)	-	0.31 (±0.03)	232.47 (±16.57)
Corvus corax Asia – North America	Split mig	4.73 (±0.04)	0.11 (±1.31)	-	-	0.52 (±0.11)	3.34 (±0.53)	-	-	385.62 (±54.26)
Corvus corax Asia – Attu	Split bi dir mig	1.45 (±0.15)	0.13 (±0.11)	2.26 (±0.16)	0.32 (±0.13)	-	1.15 (±0.14)	-	-	783.75 (±53.70)

Table A.7 continued

Corvus corax North America – Attu	Split mig	1.87 (±0.06)	54.96 (±1.22)	-	-	0.03 (±0.12)	0.72 (±0.07)	-	-	27.25 (±1.45)
Phylloscopus borealis borealis / P. b. kennicotti	Split mig	44.01 (±3.21)	1.06 (±3.41)	-	-	9.39 (±0.85)	1.11 (±0.22)	-	-	314.43 (±32.69)
Phylloscopus borealis borealis / P. examinandus	SC 1m	2.72 (±0.19)	3.62 (±0.14)	-	-	3.64 (±0.24)	1.06 (±0.09)	0.08 (±0.20)	-	336.34 (±16.50)
	SC 2m	3.69 (±0.31)	2.76 (±0.17)	0.81 (±0.11)	2.88 (±0.15)	-	1.16 (±0.19)	0.20 (±0.34)	-	312.73 (±36.53)
Anthus rubescens pacificus / A. r. japonicus	IM	5.58 (±0.45)	4.20 (±0.33)	0.12 (±0.13)	0.085 (±0.08)	-	1.18 (±0.15)	-	0.58 (±0.05)	442.13 (±45.89)
Motacilla tschutschensis simillima / M. t. tschutschensis	SC 2m	2.93 (±0.45)	0.65 (±0.12)	2.83 (±0.21)	0.18 (±0.29)	-	1.59 (±0.23)	0.77 (±0.14)	-	331.86 (±32.78)

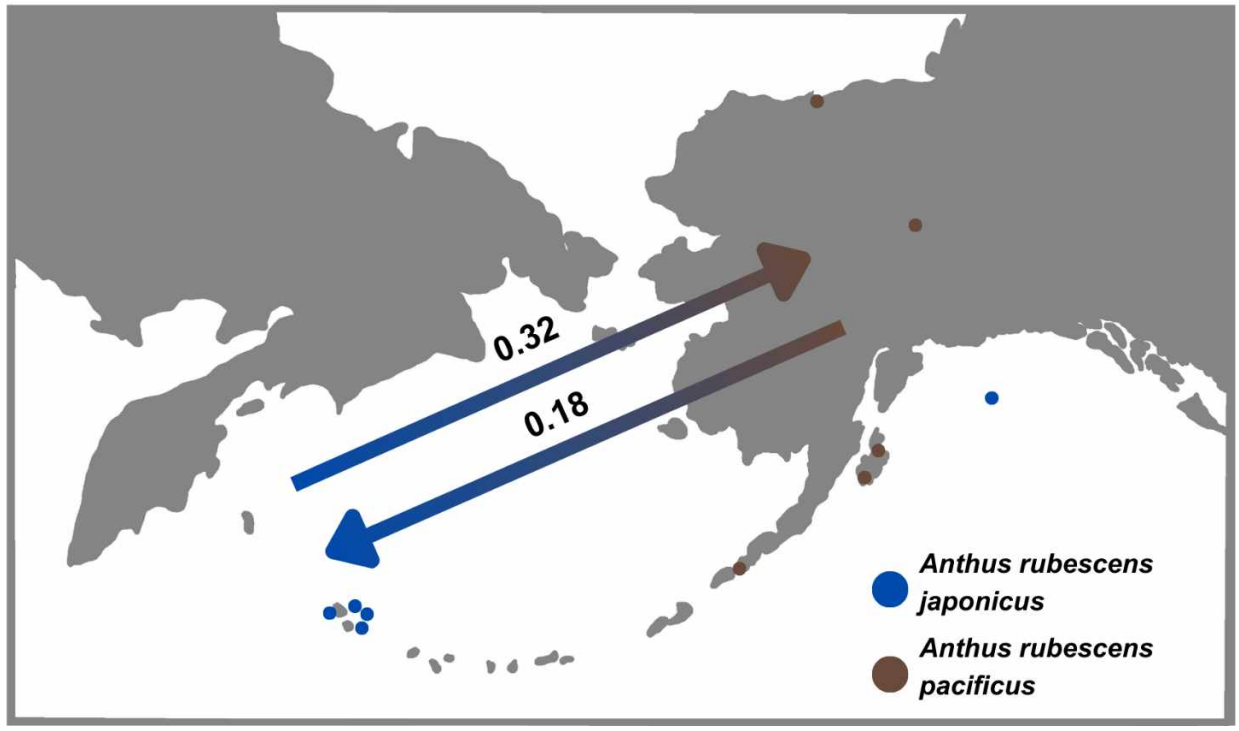

Fig. A.1: Maps of gene flow between *Anthus rubescens* subspecies. Dots represent sampling locations, and colors represent populations. Arrows represent directional gene flow between the two taxa with numbers denoting the number of individuals migrating per generation.

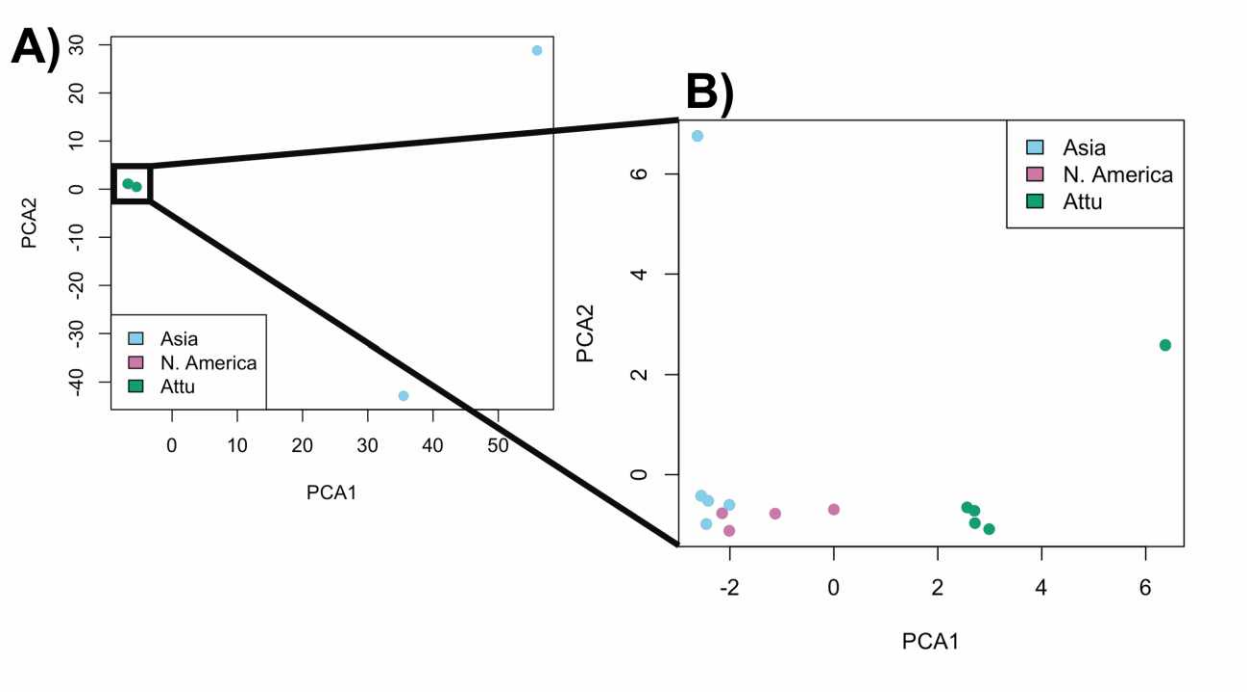

Fig. A.2: Principal components analysis (PCA) of *C. corax,* A) with all individuals and B) without two Asian outliers. The latter shows that Asia and North America populations are very similar, while the Attu population is different from the mainland counterparts.

Appendix A References

Ainley, D. G., Nettleship, D. N., & Storey, A. E. (2021). Common Murre (*Uria aalge*), version 2.0. *Birds of the World*. https://doi.org/10.2173/bow.commur.02species_shared.bow.project_name

Badyaev, A. V., Kessel, B., Gibson, D. D., del Hoyo, J., & Collar, N. (2020). Eastern Yellow Wagtail (*Motacilla tschutschensis*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.eaywag.01species_shared.bow.project_name

Boarman, W. I., & Heinrich, B. (2020). Common Raven (*Corvus corax*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.comrav.01species_shared.bow.project_name

Claramunt, S., & Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, 1(11), e1501005. https://doi.org/10.1126/sciadv.1501005

Gaston, A. J., & Hipfner, J. M. (2020). Thick-billed Murre (*Uria lomvia*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.thbmur.01species_shared.bow.project_name

Hendricks, P., & Verbeek, N. A. (2020). American Pipit (*Anthus rubescens*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.amepip.01species_shared.bow.project_name

Lowther, P. E., & Sharbaugh, S. (2020). Arctic Warbler (*Phylloscopus borealis*), version 1.0. In S. M. Billerman (Ed.), *Birds of the World*. Cornell Lab of Ornithology. https://doi.org/10.2173/bow.arcwar1.01

Nettleship, D. N. (2020). Ruddy Turnstone (*Arenaria interpres*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.rudtur.01species_shared.bow.project_name

Pasinelli, G. (2006). Population biology of European woodpecker species: A review. *Annales Zoologici Fennici*, *43*(2), 96–111.

Pearce, J., Mallory, M. L., & Metz, K. (2020). Common Merganser (*Mergus merganser*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.commer.01species_shared.bow.project_name

Pratt, A., & Peach, W. (1991). Site tenacity and annual survival of a Willow Warbler *Phylloscopus trochilus* population in Southern England. *Ringing & Migration*, *12*(3), 128–134. https://doi.org/10.1080/03078698.1991.9674005

Rizzolo, D. J., Gray, C. E., Schmutz, J. A., Barr, J. F., Eberl, C., & McIntyre, J. W. (2020). Red-throated Loon (*Gavia stellata*), version 2.0. *Birds of the World*. https://doi.org/10.2173/bow.retloo.02species_shared.bow.project_name

Spaepen, J. F. (1988). Estimation of the survival rates of meadow pipits—A comparison of two different methods. *Ringing & Migration*, *9*(2), 117–130. https://doi.org/10.1080/03078698.1988.9673934

Tremblay, J. A., Leonard Jr., D. L., & Imbeau, L. (2020). American Three-toed Woodpecker (*Picoides dorsalis*), version 1.0. *Birds of the World*. https://doi.org/10.2173/bow.attwoo1.01species_shared.bow.project_name

Warnock, N. D., & Gill, R. E. (2023). Dunlin (*Calidris alpina*), version 2.0. *Birds of the World*. https://doi.org/10.2173/bow.dunlin.02species_shared.bow.project_name

Appendix B: Supplemental Protocol

Kelly Petersen, Piotr Tuczapski, and Travis Glenn
Created: 13 Feb 2023 Last updated: 20 May 2024

This protocol is adapted from: https://www.protocols.io/view/cost-conscious-generation-of-multiplexed-short-rea-14egnx27zl5d/v2 ; published as:

Jones, A., Stanley, D., Ferguson, S., Schwessinger, B., Borevitz, J., & Warthmann, N. (2023). Cost-conscious generation of multiplexed short-read DNA libraries for whole-genome sequencing. *PLoS ONE, 18*(1), e0280004. These should be cited if this protocol is used.

Reagents Needed:

Illumina DNA Prep Kit (Illumina 20060060)
NEB Q5 DNA polymerase (New England Biolabs M0491L)
1M Tris-HCl (pH 8.0) (e.g., Invitrogen 15568025)
1M MgCl2 (ThermoFisher AM9530G)
PEG-8000 (powder or 25% solution)
Molecular grade H_2O (e.g., Fisher BP2819-1)

Transposase on Beads Library Preparation (iNextEra Library Prep 2023)

1. Normalize all DNA samples to ~10 ng/µL (quantified on Qubit or plate reader; **see Note 1**). Place ≥ 20 µL in plates or strip tubes (less can be used, but this amount is recommended for handling ease). Ensure these (and all other tubes/plates below) are labeled clearly.
2. Prepare 2x TMP buffer in a 1.5 mL tube:
    a. 20 µL 1M Tris-HCl (pH 8.0)
    b. 10 µL 1M $MgCl_2$
    c. 640 µL 25% PEG-8000
    d. 330 µL molecular grade H_2O
3. Set thermal cycler to incubate at 53°C with the lid set to 80°C. Allow it to preheat.
4. Set up tagmentation. Prepare a master mix of TMP buffer with the following per-sample recipe:
    a. 3.0 µL 2x TMP custom buffer
    b. 0.5 µL Transposome (BLT)
    c. 2.5 µL genomic DNA (25 ng total)
5. Incubate samples at 53°C for 30 min. in a thermal cycler.
6. Set up PCR reactions. You will prepare a master mix with water, buffer, dNTPs and polymerase; add that to tagmented DNA, then add primers last.
    a. 40.25 µL molecular grade H_2O
    b. 9.5 µL Q5 reaction buffer
    c. 1.25 µL 10mM dNTPs
    d. 0.5 µL Q5 HF DNA Polymerase

51.5 µL Master Mix total per sample

Add 51.5 µl of the mastermix from above to 6.0 µL tagmented DNA from step 5

For each tube/well, add unique iNext Primers:

2.5 µL iNext5 index primer (5 µM)

2.5 µL iNext7 index primer (5 µM)

62.5 µL – grand total of each PCR

7. Cycle in 72°C for 3 min.; then 98°C for 3 min.; then 10 cycles of [98°C for 45 sec., 62°C for 30 sec., 72°C for 2 min.]; then 72°C for 1 min. Hold at 10°C.
8. Run 7.5 µL on agarose gel to ensure each sample amplified, yielding a smear in the desired size-range (300-700 bp).
9. Pool with normalization.
10. Purify the pools with SpeedBeads. Add 1.25x SpeedBeads to pooled DNA; skloosh; purify as normal and resuspend in TE buffer. Incubate 15 min. at room temperature, then place on magnet and pull off all liquid, leaving the beads behind.
11. Quantify the pool(s) with Qubit.
12. Sequence.

Notes:

1) The protocol works well for bird DNA & achieves similar results with DNA concentrations of 3-20 ng/µL. Because similar results are achieved with varying input in the range of 3-20 ng/µL, the precision of normalization is not super important. It is also not important which number you normalize to within the known range, but if possible, picking a number in range of 5-10 ng/µL seems best. We chose ~10 ng/µL because then you can off by ~3x (too high or too low) and still be within the range known to work well.
2) Lower concentrations of DNA (<1 ng/µL) can be used, but more cycles are required in step 7. You will want to batch samples with similar amounts of template DNA, so you can amplify them in groups with the same number of cycles (e.g., groups with 10, 12, or 14 cycles). We generally do not put samples back in for more cycles or use more than 14 cycles.

Appendix C: Supplementary Material for Chapter 3

Table C.1: Accession numbers and subspecies used in phenotypic analysis. UAM = University of Alaska Museum

UAM Number	Subspecies														
27914	maxima	10179	maxima	36618	maxima	15343	maxima	27911	maxima	30142	sanaka	11231	caurina	13911	rufina
27919	maxima	11229	maxima	36622	maxima	15354	maxima	27913	maxima	35424	sanaka	11104	caurina	13937	rufina
10188	maxima	15353	maxima	27852	maxima	14165	maxima	27910	maxima	35423	sanaka	11141	caurina	11824	rufina
10186	maxima	14611	maxima	38590	maxima	9305	maxima	27907	maxima	35421	sanaka	11361	caurina	11517	rufina
11175	maxima	15321	maxima	15116	maxima	15330	maxima	11363	maxima	27876	sanaka	11382	caurina	11516	rufina
11827	maxima	15298	maxima	36620	maxima	15342	maxima	11825	maxima	27877	sanaka	11384	caurina	13910	rufina
11176	maxima	15305	maxima	34715	maxima	8462	maxima	11171	maxima	27878	sanaka	11383	caurina	11552	rufina
11562	maxima	15367	maxima	31442	maxima	9302	maxima	22257	maxima	31879	sanaka	30365	caurina	13919	rufina
11561	maxima	10165	maxima	15117	maxima	15378	maxima	11381	sanaka	31880	sanaka	30364	caurina	11550	rufina
11558	maxima	15339	maxima	15115	maxima	15356	maxima	11362	sanaka	35415	sanaka	14941	caurina	11179	rufina
11267	maxima	10041	maxima	11173	maxima	15331	maxima	37189	sanaka	35416	sanaka	10652	caurina	11549	rufina
11559	maxima	10946	maxima	11174	maxima	15302	maxima	11236	sanaka	29243	sanaka	11046	caurina	11546	rufina
11177	maxima	13186	maxima	41447	maxima	15313	maxima	11237	sanaka	29246	sanaka	13944	rufina	11547	rufina
36278	maxima	15315	maxima	11270	maxima	15352	maxima	15374	sanaka	29239	sanaka	13438	rufina	13909	rufina
11178	maxima	15344	maxima	31441	maxima	8463	maxima	15334	sanaka	28578	insignis	13886	rufina	13908	rufina
11280	maxima	10170	maxima	34547	maxima	20672	maxima	18514	sanaka	35663	insignis	13941	rufina	11548	rufina
11268	maxima	11291	maxima	21908	maxima	10073	maxima	15384	sanaka	38749	insignis	13916	rufina	13080	rufina
36552	maxima	11295	maxima	34544	maxima	15363	maxima	15335	sanaka	30366	insignis	13915	rufina	13079	rufina
8461	maxima	11297	maxima	34545	maxima	20674	maxima	11238	sanaka	28552	insignis	13943	rufina	35795	rufina
15319	maxima	11294	maxima	27880	maxima	20518	maxima	13903	sanaka	30814	insignis	13945	rufina	14944	rufina
10942	maxima	11296	maxima	27858	maxima	15320	maxima	28555	sanaka	28549	insignis	13913	rufina	14942	rufina
11048	maxima	11289	maxima	27882	maxima	19048	maxima	10091	sanaka	35664	insignis	15193	rufina	14943	rufina
13059	maxima	13160	maxima	27884	maxima	10944	maxima	10171	sanaka	35666	insignis	13463	rufina	13920	rufina
10040	maxima	15323	maxima	27883	maxima	13128	maxima	11379	sanaka	31498	insignis	13938	rufina		
15326	maxima	8779	maxima	11227	maxima	11243	maxima	10187	sanaka	11182	caurina	13953	rufina		
11511	maxima	15318	maxima	8612	maxima	9418	maxima	12142	sanaka	11180	caurina	15086	rufina		
15340	maxima	27848	maxima	15299	maxima	11244	maxima	30144	sanaka	11223	caurina	14658	rufina		
15312	maxima	36621	maxima	15358	maxima	8781	maxima	30139	sanaka	11103	caurina	13952	rufina		
15307	maxima	36623	maxima	15304	maxima	9602	maxima	30141	sanaka	11101	caurina	14714	rufina		
												13912	rufina		

Table C.2: Sample information used in genomic analyses. UAM = University of Alaska Museum. All sequences generated in this study are archived under SRA BioProject PRJNA1114297. † denotes individual used as a reference sequence.

Species	Locality	Catalogue Number	Tissue Number	Sequence Archive
Junco hyemalis†	Virginia: Giles County, Mount Lake Biological Station	MLZ:Bird:69236	n.a.	PRJNA493001
Melospiza georgiana	Wisconsin: Burnett County, Saint Croix National Riverway	UAM11868	CLP550	SAMN41480621
Melospiza lincolnii	Alaska: Southeast, Gravina Island	UAM21920	ABJ048	SAMN41480622
Melodia melodia maxima	Alaska: Aleutian Islands, Attu Island	UAM31500	UAMX4047	PRJNA511035
Melospiza melodia maxima	Alaska: Aleutian Islands, Adak Island	UAM10946	CLP108	SAMN41480625
Melospiza melodia sanaka	Alaska: Shumagin Islands, Popof Island	UAM11585	CLP251	SAMN41480627
Melospiza melodia insignis	Alaska: Kodiak Island, Cape Chiniak	UAM14002	DDG1900	SAMN41480624
Melospiza melodia caurina	Alaska: Copper River Delta	UAM11384	CLP020	SAMN41480623
Melospiza melodia rufina	Alaska: Southeast, Hyder	UAM7343	KSW1374	SAMN41480626

Table C.3: Accession numbers, gene names, and associated phenotypic traits for sequences that were tested in gene selection. Asterisk (*) denotes sequences dropped from analysis due to poor BLAST results.

Associated Phenotypic Trait	Accession Number	Gene
Bill size	XM_002196007.3	*ALX1*
Bill size	XM_030270858.3	*CCDC149*
Bill size	XM_030274996.3	*DLK1*
Bill size	XM_004176091.5	*HMGA2*
Bill size	XM_041715714.1	*LGI2*
Body size	XM_030270679.3	*AFAP1*
Body Size	XM_002198985.6	*GGA1*
Body size	XM_030270815.3	*LCORL*
Body size	XM_002193889.6	*QDPR*
Body size	XM_030270827.3	*SLIT2*
Body size	XM_030276710.3	*WAPL**
Body size	XM_002193767.6	*LAP3**
Plumage color	XM_015637109.2	*APOD1*
Plumage color	XM_032750540.2	*BCO1*
Plumage color	XM_030291411.3	*CD36*
Plumage color	XR_003962894.3	*SCARB1**
Plumage color	XM_030273867.3	*HPS5**
Plumage color	XM_030256157.3	*STARD3**
Plumage color	XM_030290911.3	*BCO2**
Dispersal	XM_012569888.4	*KCTD21*
Dispersal	XM_004175099.5	*TGFB2*
Dispersal	XM_032752076.2	*SLC2A1*
Migration	XM_030271098.3	*CLOCK*
Migration	XM_030276951.3	*CREB1*
Migration	XM_030263133.3	*CRY1*
Migration	XM_012574176.4	*CRY2*
Migration	XM_032748493.2	*NPAS3*
Migration	XM_030280764.3	*PER2*
Migration	XM_030289516.3	*PER3*
Migration	XM_030285787.3	*YPEL1*
Salt tolerance	XM_041719360.1	*MMP17*
Salt tolerance	XM_030265641.3	*SLC9A3*
Salt tolerance	XM_030276092.3	*MYOF**
Salt tolerance	XM_030283214.3	*WNK2**

Table C.4: Summary statistics of sequenced datasets including total number of reads, percentage of properly paired reads, and mean depth of coverage.

Taxa	Tissue Number	Number of reads	% properly paired reads	Mean depth of coverage
Melospiza georgiana	CLP550	81,339,509	93.0%	7.04
Melospiza lincolnii	ABJ048	12,566,699	92.9%	2.10
Melospiza melodia maxima	CLP108	26,989,233	91.4%	3.33
Melospiza melodia sanaka	CLP251	84,504,717	94.0%	9.71
Melospiza melodia insignis	DDG1900	73,078,606	92.0%	8.62
Melospiza melodia caurina	CLP020	69,765,460	92.0%	8.12
Melospiza melodia rufina	KSW1374	72,593,062	90.6%	8.08

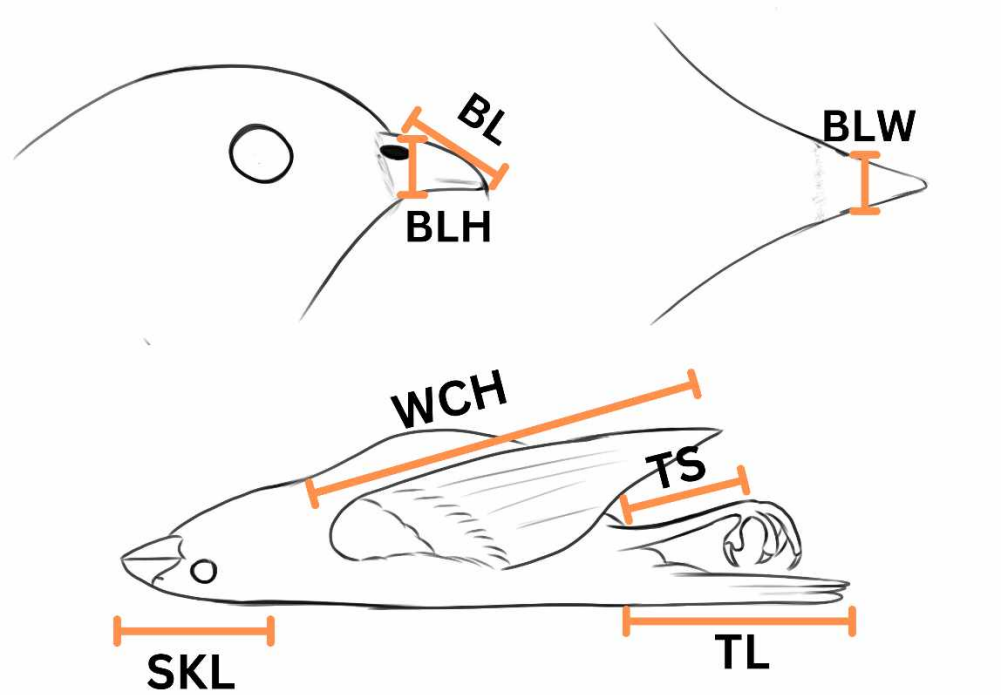

Fig. C.1: Diagram of phenotypic measurements taken from song sparrow specimens. These measurements are bill length (BL), bill height (BLH), bill width (BLW), wing chord length (WCH), tarsus length (TS), skull length (SKL), and tail length (TL). Mass, not shown, was also recorded.

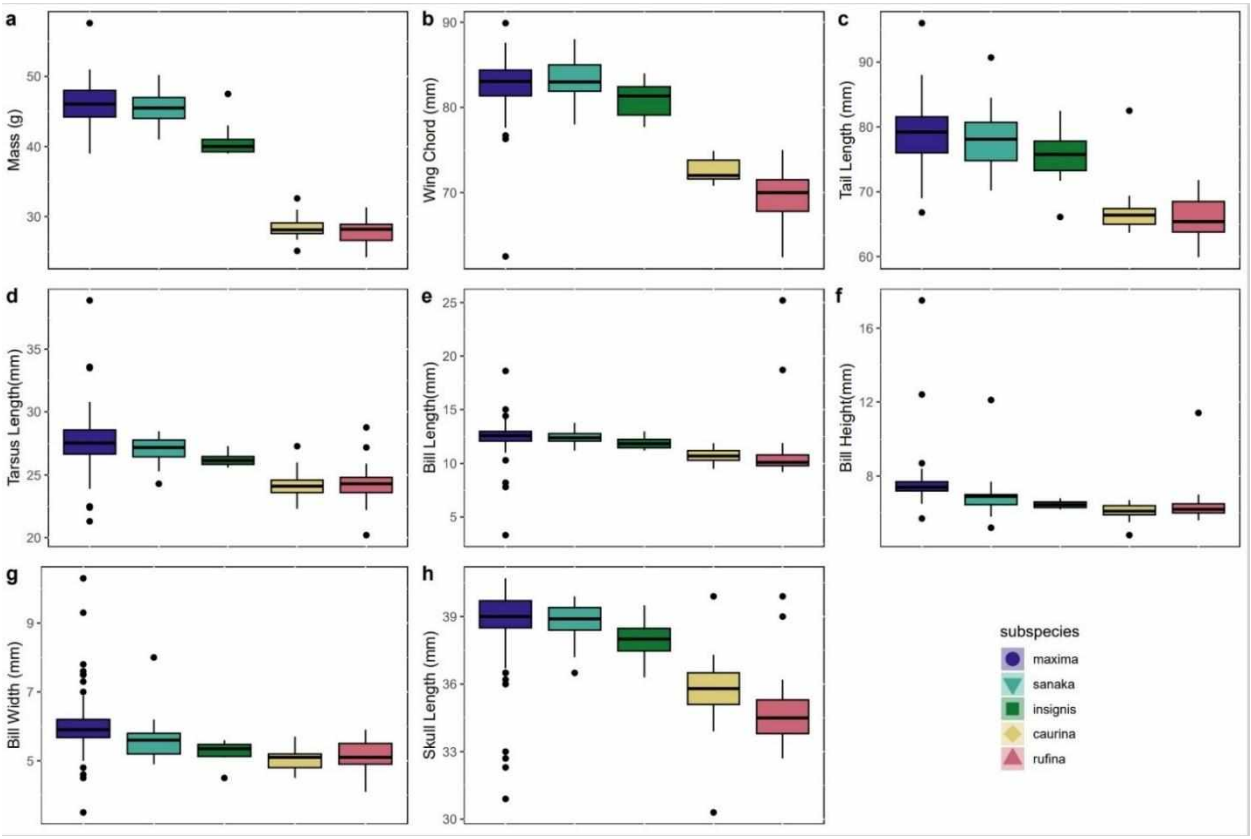

Fig. C.2: Boxplots of phenotypic data for each *Melospiza melodia* subspecies: a) mass (g), b) wing chord (mm), c) tail length (mm), d) tarsus length (mm), e) bill length (mm), f) bill height (mm), g) bill width (mm), and h) skull length (mm). Colors correspond to the colors in Fig. 3.1.

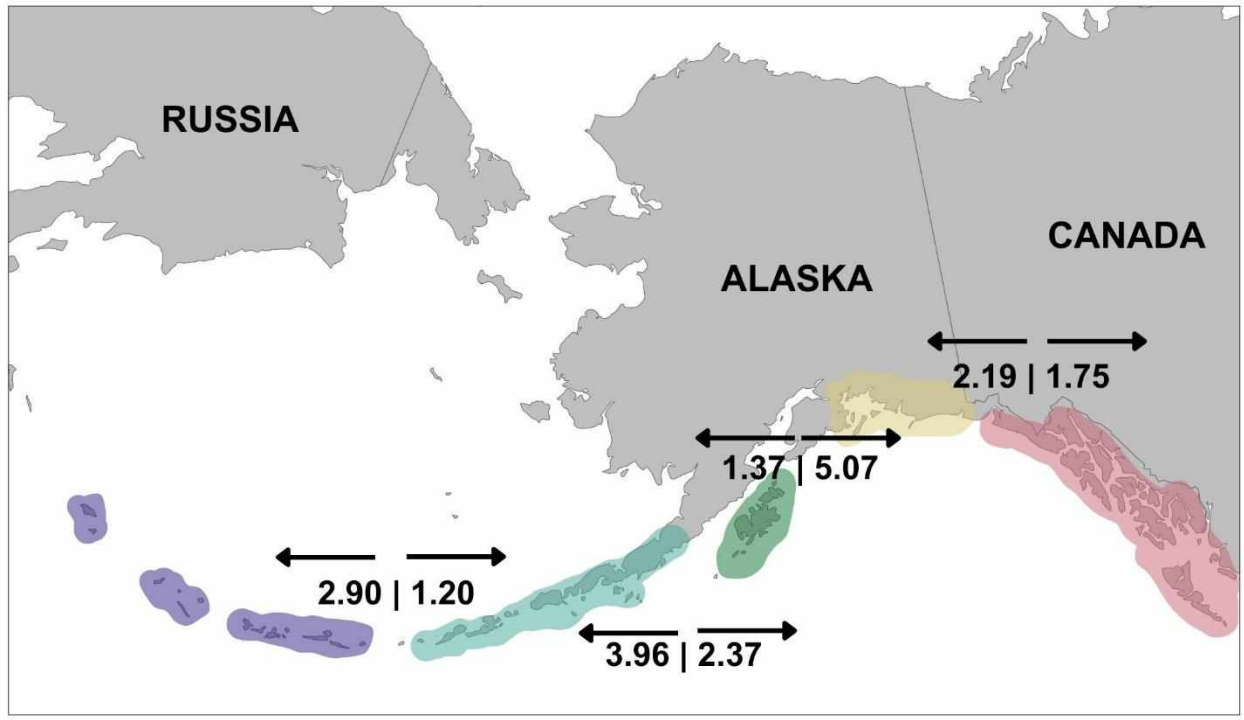

Fig. C.3: Map of Song Sparrow range in Alaska with pairwise estimates of directional gene flow (genetic migration, N_em) from Pruett & Winker (2005b). Values denote the number of migrant individuals per generation relative to effective population size, and the arrows denote direction of migration. For example, *M. m. maxima* is receiving 2.90 migrants per generation from *M. m. sanaka*. Colors correspond to Fig. 3.1.