GENETIC POPULATION STRUCTURE OF THE LEAST FLYCATCHER (*EMPIDONAX MINIMUS*): IMPLICATIONS FOR EVALUATING MIGRATORY CONNECTIVITY.

SARA CASTRO GARCÍA

Bachelor of Science, University of Guanajuato, 2022

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

MASTER OF SCIENCE

In

BIOLOGICAL SCIENCES

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

Thesis Examination Committee
Date of Defence: August 7th, 2025

Dr. Theresa M. Burg Thesis Supervisor	Professor	Ph.D.
Kevin J. Kardynal Thesis Examination Committee Member	Research Scientist at Environment and Climate Change Canada	M.Sc.
Dr. Andrew Iwaniuk Thesis Examination Committee Member	Professor	Ph.D.
Dr. Shelley Hoover Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Dmytro Yevtushenko Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Dedication

To my six angels, whose presence has accompanied me throughout my life. Even if you are not here physically, I have felt your love every time I find myself lonely. In this work, and in everything I do, I think of you, and I love you.

Gustavo Castro Rojas

Martha Antonia Arenal Hernández

Yolanda Castro Arenal

Diana Armida Castro Arenal

María Guadalupe Zepeda Gutiérrez

Roberto Zepeda Gutiérrez

Abstract

Over the last six decades, avian aerial insectivores in North America have experienced an overall decline of ~60% in their population across their breeding ranges. Understanding their population genetic structure and spatial connections across the annual cycle is critical to determining potential factors driving these trends; however, this information is limited for many species. This study uses RADseq data to examine population structure and connectivity in least flycatchers (*Empidonax minimus*) across North America. Supported by field and museum sampling, blood, feather, and/or tissue samples were collected from 14 sites throughout the breeding range as well as during migration in Canada and the USA, and on the non-breeding grounds in Mexico. The results from principal components analysis (PCA), pairwise F_{ST}, STRUCTURE, and least-cost corridor analyses indicate high levels of gene flow among breeding populations, with weak genetic structure observed between two groups that exhibit an east-west split, which is enhanced using outlier loci. A genetic stock identification analysis was conducted to determine the breeding origin of the non-breeding samples, utilizing outlier loci. Twelve out of 29 non-breeding samples were successfully assigned to a breeding population. In two of the three non-breeding locations, individuals of mixed origin were observed, indicating weak migratory connectivity. The results indicate that the weak migratory connectivity and young age of the species might contribute to the low levels of population structure observed. This study allowed to increase the comprehension of the genetic structure and migratory connectivity of the least flycatcher.

Acknowledgements

Pursuing my master's degree in a field like this in a country such as Canada has filled me with joy and gratitude. This experience has truly changed my life and opened my eyes to the many opportunities, adventures, and friends that can be found in places that seem far away from home. Over the past two years, I have felt the love and support of many people, and I will forever cherish the impact that all of them have had on my life and work.

First and foremost, I want to thank my supervisor, Dr. Theresa Burg, for giving me the opportunity to work in her lab. I have gained knowledge and skills that extend beyond molecular ecology and will impact both my professional and personal life. Thank you for guiding me in telling the story that needs to be told. A special thanks goes to Keith Hobson, whose knowledge and ideas filled this research until it was time for him to go. I would also like to extend my gratitude to my committee members, Dr. Andrew Iwaniuk, Dr. Shelley Hoover and Dr. Kevin Kardynal, for helping me delineate the questions and analyses needed for this research.

I want to thank mitacs, the University of Lethbridge, Environment and Climate Change Canada, and Juventudes Guanajuato for their financial support that allowed me to complete this degree. Thanks to all the museums and banding stations that donated samples to this project: CMN, CUMV, ECCC, ECOSUR, IBUNAM, NMNH, and NYSM. It changed the project's trajectory for the better.

I also want to thank everyone in my lab for their support, advice, and for being an ear and a second set of eyes whenever I needed it. Thanks to Teegi for always answering my questions, guiding me through the many issues I faced, and sharing our fieldwork adventure (with Anisha too!). You truly made my thesis what it is. Thanks to Anisha, Ploy, and Milan for not only sharing the office but also the stress that comes with learning about bioinformatics and population genetics. Sharing a workspace with all of you and getting to know you has been wonderful, and I feel happy knowing I have places to visit across the world—and that you have a place to visit in Mexico. I would also like to thank Luca Montana for helping me understand the analyses and issues that can arise with RADseq sequences; without your help, I would probably still be stuck with some of my analyses. Thanks to everyone in the lab, including Melissa, Clint, and Genevieve Chelak, for sharing meals and joyful conversations whenever needed. And thanks to everyone in the water building for creating such a welcoming environment, for the questions and chats that made me feel accompanied.

I also want to thank all the friends and relationships that have helped me outside my workspace and have made Lethbridge feel like home. Sofi, Blake, Thamara, Razeen and Enona, thank you for being a little family and filling my days with laughter and delicious food. Jemma, thank you for being a friend who is always there, for all the adventures and moments we've shared. Amanda, thank you for being so welcoming and funny, and for gossiping with me in Spanish whenever we needed it. To the people from the McCune lab for the parties and conversations, and for being a wonderful company.

I want to thank everyone who has supported me from afar. In hard days, being able to talk to you gave me the strength I needed to continue. To my family, Juan Carlos, Raquel and Rebeca, you are my heart. To my friends, Ximena, Ricardo, Josu, Isa, Gemma, and Jessi, thank you for making the physical distance the only space between us. César, thank you for appearing right in the middle of my master's degree and becoming the music of my days, as well as an ongoing source of support who believed in me even when I didn't.

Finally, I would like to thank all the birds that, without their knowledge, donated their DNA for this project. I hope it can be helpful in protecting and caring for the wonderful species that is the least flycatcher. I hope that they can continue to migrate and enjoy the same countries that I do for many years to come.

Table of Contents

List of tables

Table of figures

List of Abbreviations

°C	Degrees Celsius
3D-GBS	3D-Genotyping By Sequencing
AUC	Area under the curve
Bp	Base pairs
BR	Breeding (sample)
DNA	Deoxyribonucleic acid
F_{ST}	Fixation index
GSI	Genetic stock identification
IBD	Isolation by distance
IBIS	Institute of Integrative Biology and Systems
LCC	Least-cost corridor
LGM	Last glacial maximum
M	Migratory (sample)
MCMC	Markov chain Monte Carlo
MEM	Maximal Exact Match
NBR	Non-breeding (sample)
NGS	Next-Generation Sequencing
PCA	Principal components analysis
PofZ	Probability of assignment
RADseq	Restriction site Associated DNA sequencing
SDM	Species distribution model
SNP	Single nucleotide polymorphism
VCF	Variant call format
z-score	Standard score

Sampling locations

CAB	Central Alberta
ME	Maine
MT	Montana
MX	Mexico
NAB	Northern Alberta
ND	North Dakota
NY	New York
ON	Ontario
SAB	Southern Alberta
SEAB	Southeast Alberta
SEBC	Southeast British Columbia
SK	Saskatchewan
WV	West Virginia
YT	Yukon Territories

Museums and banding stations

CMN	Canadian Museum of Nature
CUMV	Cornell University Museum of Vertebrates

ECCC	Environment and Climate Change Canada
ECOSUR	El Colegio de la Frontera Sur
IBUNAM	Instituto de Biología de la Universidad Nacional Autónoma de México
NMNH	National Museum of Natural History
NYSM	New York State Museum

Chapter 1: General Introduction

1.1 Population genetics

Ecological-evolutionary interactions at the population level determine the survival, abundance, density, and dispersal of individuals between populations (Lowe et al., 2017). To connect the evolutionary processes with the ecological interactions that shape them, population genetics are useful in explaining the effects of demography on phenotypes and allele frequencies, and provide information on population structure, dynamics, adaptation, and speciation (Casillas & Barbadilla, 2017; Hohenlohe et al., 2021; Lowe et al., 2017). By studying a population's genetic information, we can assess its size and connectivity, genetic variation, and mean fitness (Hohenlohe et al., 2021; Lowe et al., 2017).

One of the most important concepts in population genetics is gene flow: the movement of individuals from one population to another that alters genetic variation within a population (Hohenlohe et al., 2021; Lowe et al., 2017; Tigano & Friesen, 2016). The effects of gene flow can vary depending on the population itself. For example, it may disrupt local adaptation through outbreeding depression, which decreases the fitness of individuals (Sexton et al., 2014). In this scenario, gene flow is harmful as it hinders local adaptation by countering selection (Ellstrand, 2014). Conversely, gene flow can enhance genetic variation in small populations with high inbreeding rates, enabling new adaptations that can ultimately increase both population size and fitness (Frankham, 2015; Sexton et al., 2014). By enhancing genetic diversity, gene flow proves beneficial and can promote the spread of advantageous alleles (Ellstrand, 2014).

Gene flow and individual factors, like recombination, mutation, and evolutionary history, affect the distribution of genetic variation within and among populations (Hohenlohe et al., 2021; Renfrew et al., 2022; Schowalter, 2022). How genetic variation is distributed is defined as population structure (Barroso et al., 2020).

A population is considered genetically structured when genetic variation (e.g. allele frequency) differs among populations consisting of several individuals with genetic characteristics divergent from each other (Corander et al., 2008). When populations are isolated from each other, genetic variation reflects this separation, so genetic studies can provide insights into the history of the population (Corander et al., 2008) and infer demographic independence or reproductive isolation (Renfrew et al., 2022).

1.2 Migration

In migratory species, the movement of individuals across their annual range can invoke gene flow (Webster et al., 2002). Although migratory behavior is found in many taxa, it is especially prominent in birds. Despite the different migratory strategies observed across bird species, it is accepted that migration evolved in an early ancestor and is now present in many different orders (Liedvogel & Delmore, 2018). Changes in environmental resource availability drive seasonal migration, thus creating an endogenously controlled movement of birds between their breeding and non-breeding grounds (Gu et al., 2024)(Salewski & Bruderer, 2007). Through these regular, seasonal movements, birds are exposed to different environmental conditions in the breeding and non-breeding grounds that can influence their fitness (Culp et al., 2017). Studying a species' migratory behaviour is necessary to understand how different events throughout the annual cycle affect subsequent events at both the individual and population levels (Webster et al., 2002).

Knowing the routes, times, and sites that birds visit during their breeding, migratory, and non-breeding periods are a central question when studying seasonal migration. One way to gain knowledge about seasonal migration is by studying migratory connectivity, which describes the spatiotemporal linkages of individuals between breeding and non-breeding grounds and is determined by the co-occurrence of birds at non-breeding grounds originating from different breeding sites (Piironen et al., 2023; Webster et al., 2002). Strong migratory connectivity occurs when most individuals from a breeding population move to the same wintering region, ensuring that populations do not mixthroughout the annual cycle (Desaix et al., 2019; Webster et al., 2002). Weak migratory connectivity occurs when individuals from different breeding populations mix on the non-breeding grounds (Desaix et al., 2019). The degree of migratory connectivity can predict the response of migrants to environmental change (Finch et al., 2017). For example, populations with weak migratory connectivity are potentially adapted to the various wintering areas used by different individuals, thus being able to respond to changes in various breeding and wintering grounds. On the other hand, populations with strong connectivity may be locally adapted to specific conditions of breeding and wintering areas they inhabit (Webster et al., 2002).

1.3 Next Generation Sequencing

To study population structure and migratory connectivity, the field of population genetics has benefitted from molecular biology advances such as Next-Generation Sequencing (NGS) (Fig.1.1). NGS encompasses different sequencing protocols, all of which share the common feature of sequencing millions of DNA molecules simultaneously (Hu et al., 2021). One of the main features of NGS is that it can decrease the sequencing coverage to specific parts of the genome by using restriction enzymes, thus being cost-effective for large-scale studies. By sequencing parts of the genome, molecular markers such as single nucleotide polymorphisms (SNPs), a single-base sequence variant at a particular point in the genome, can be identified (Butler, 2012).

Using SNPs, Ruegg et al. (2014) identified genetically distinct groups of Wilson's warblers (*Wilsonia pusilla*) and found novel region-specific migratory routes and timetables of migration along the Pacific flyway. This research also noted most significant genetic differences were found between the eastern and western populations (Ruegg et al., 2014). Thus, genetic markers can be used to understand a species' population structure, connectivity, and migration patterns (Renfrew et al., 2022). Genetic analyses and methods can be used to measure and conserve biodiversity at the individual, population, and species levels, which is especially useful in species under decline (Kardos, 2021).

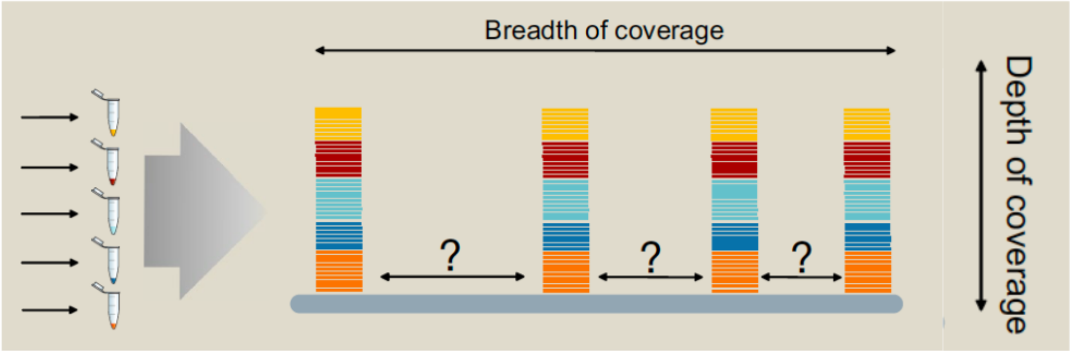


Figure 1.1. Next-generation sequencing (NGS). In 3D-GBS, extracted DNA samples are digested with restriction enzymes to reduce genome complexity, and repetitive genome regions can be avoided (Elshire, 2011). By selecting specific restriction enzymes, the same parts of the genome are selected and sequenced in all the samples (de Ronne et al., 2023).

1.4 Study species

Migrant aerial insectivores have declined by 59% since 1970 (North American Bird Conservation Initiative, 2019), which can be attributed to various factors, such as habitat loss, broad-scale changes in insect populations, agricultural intensification, and pesticide use. As migratory birds, they are more vulnerable to climate change than year-round residents because they are exposed to climate variation in different places throughout their annual cycle (Michel et al., 2021).

The least flycatcher (*Empidonax minimus*) is an aerial insectivore whose population has decreased by ~53% since 1970 (Spiller & Dettmers, 2019). It is a long-distance Nearctic-Neotropical migrant that breeds across Canada and the northern US from the Yukon to Wyoming in the west to Atlantic Canada and northeastern US (Fig. 1.2). During the breeding season, the least flycatcher typically nests in clusters found in the lower to mid-canopy of deciduous forests in various trees species, including birches, maples, poplars, and pines (Tarof & Briskie, 2020). The clumped breeding behaviour of the species suggests that it is a highly sociable species, and that nesting close to neighbors might be key to the survival of individuals (Tarof & Briskie, 2020). From the breeding grounds, it migrates through the central Plains and overwinters in southern and coastal Mexico, the Gulf Coast in the US, and Central America (Godfrey, 1979). The least flycatcher is susceptible to forest disturbance, environmental pollution, and local disturbances (Holmes & Sherry, 2001; Tarof & Briskie, 2020). It is especially sensitive to forest disturbances due to their aggregated nesting behaviour (Dellasala & Rabe, 1987; Tarof & Ratcliffe, 2000).

Like many other aerial insectivores, the decline in the least flycatcher population follows a geographical gradient, with the probability of decline being highest in the northeast area of its breeding range (Nebel et al., 2010). Research on the least flycatcher has focused mainly on its breeding ecology. For example, Darveau et al. (1993) studied the nesting success, nest sites, and parental care of the species in a declining sugar maple forest in Quebec. Similarly, Tarof and Ratcliffe (2000) studied the pair formation and population and copulation behaviour of clusters of the least flycatcher, finding that extra-pair partners were common. However, little is known about its population genetics and how gene flow is affected by the movement of individuals between breeding and non-breeding regions. By studying the population structure of the species and how it is affected by its migratory connectivity, we can understand the genetic effects of declines in different locations of the range of the least flycatcher—i.e., how these declines could be affecting the population structure of the species.

1.5 Thesis objectives

The primary purpose of my research is to understand the population structure and migratory connectivity of the least flycatcher. The objectives are:

1. To analyze DNA sequences of least flycatchers from blood and tissue samples and identify genetically distinct groups using single-nucleotide polymorphisms (SNPs).
2. To determine the migratory connectivity and population structure of each genetically distinct group by sampling both migratory and non-breeding individuals.
3. To identify the abiotic factors, such as geographical barriers and environmental variables, that influence the genetic population structure of the least flycatcher.

This study will increase our knowledge of how least flycatcher populations are connected. This will allow us to identify breeding populations' migratory pathways and wintering sites and estimate their connectivity. Even though the least flycatcher is listed as a species of least concern by the IUCN (IUCN, 2025), the rate at which the species has decreased indicates that it is necessary to understand the reasons behind this. By studying the population structure of the least flycatcher, we can assess the fitness of the populations, and by using genetic diversity, landscape, geographical barriers, and migration, we can pinpoint populations at a higher risk of declining.

The hypothesis for this thesis is that the least flycatcher will show a similar population structure to that found in other migratory bird species, such as the prothonotary warbler (*Protonotaria citrea*) (Desaix et al., 2019): weak population structure, caused by geographical barriers such as the Rocky Mountains. I expect to find two genetically distinct populations, divided by the Rocky Mountains. I also expect to find strong migratory connectivity between those two genetically distinct groups and the grounds they overwinter in.

.

1.6 References

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, *17*(2), 81-92. https://doi.org/10.1038/nrg.2015.28

Barroso, G. V., Moutinho, A. F., & Dutheil, J. Y. (2020). A population genomics lexicon. In *Statistical Population Genomics* (ed J. Y. Dutheil), 3-17.

Bertola, L. D., Brüniche-Olsen, A., Kershaw, F., Russo, I. R. M., MacDonald, A. J., Sunnucks, P., Bruford, M. W., Cadena, C. D., Ewart, K. M., de Bruyn, M., Eldridge, M. D. B., Frankham, R., Guayasamin, J. M., Grueber, C. E., Hoareau, T. B., Hoban, S., Hohenlohe, P. A., Hunter, M. E., Kotze, A., . . . Segelbacher, G. (2024). A pragmatic approach for integrating molecular tools into biodiversity conservation. *Conservation Science and Practice*, *6*(1), e13053 . https://doi.org/10.1111/csp2.13053

Butler, J. M. (2012). Single Nucleotide Polymorphisms and applications. In J. M. Butler (Ed.), *Advanced Topics in Forensic DNA Typing: Methodology* (pp. 347-369). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-374513-2.00012-9

Casillas, S., & Barbadilla, A. (2017). Molecular population genetics. *Genetics*, *205*(3), 1003-1035. https://doi.org/10.1534/genetics.116.196493

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences*, *114*(30), E6089-E6096. https://doi.org/10.1073/pnas.1704949114

Corander, J., Sirén, J., & Arjas, E. (2008). Bayesian spatial modeling of genetic population structure. *Computational Statistics*, *23*(1), 111-129. https://doi.org/10.1007/s00180-007-0072-x

Culp, L. A., Cohen, E. B., Scarpignato, A. L., Thogmartin, W. E., & Marra, P. P. (2017). Full annual cycle climate change vulnerability assessment for migratory birds. *Ecosphere*, 8(3), e01565. https://doi.org/10.1002/ecs2.1565

de Ronne, M., Légaré, G., Belzile, F., Boyle, B., & Torkamaneh, D. (2023). 3D-GBS: a universal genotyping-by-sequencing approach for genomic selection and other high-throughput low-cost applications in species with small to medium-sized genomes. *Plant Methods*, *19*(1), 13.

Darveau, M., Gauthier, G., DesGranges, J.-L., & Mauffette, Y. (1993). Nesting success, nest sites, and parental care of the least flycatcher in declining maple forests. Canadian Journal of Zoology, 71(8), 1592-1601. https://doi.org/10.1139/z93-225

Desaix, M. G., Bulluck, L. P., Eckert, A. J., Viverette, C. B., Boves, T. J., Reese, J. A., Tonra, C. M., & Dyer, R. J. (2019). Population assignment reveals low migratory connectivity in a weakly structured songbird. *Molecular Ecology*, *28*(9), 2122-2135. https://doi.org/10.1111/mec.15083

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, *17*(7), 422-433. https://doi.org/10.1038/nrg.2016.58

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, *101*(5), 737-753. https://doi.org/https://doi.org/10.3732/ajb.1400024

Fedorka, K. M., Winterhalter, W. E., Shaw, K. L., Brogan, W. R., & Mousseau, T. A. (2012). The role of gene flow asymmetry along an environmental gradient in constraining local adaptation and range expansion. *Journal of Evolutionary Biology*, 25(8), 1676-1685. https://doi.org/10.1111/j.1420-9101.2012.02552.x

Finch, T., Butler, S. J., Franco, A. M. A., & Cresswell, W. (2017). Low migratory connectivity is common in long-distance migrant birds. *Journal of Animal Ecology*, *86*(3), 662-673. https://doi.org/10.1111/1365-2656.12635

Frankham, R. (2015). Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*(11), 2610-2618. https://doi.org/10.1111/mec.13139

Godfrey, E. W. (1979). *Birds of Canada*.

Gu, Z. R., Dixon, A., & Zhan, X. J. (2024). Genetics and evolution of bird migration. *Annual Review of Animal Biosciences*, *12*, 21-43. https://doi.org/10.1146/annurev-animal-021122-092239

Haig, S. M. (1998). Molecular contributions to conservation. *Ecology*, *79*(2), 413-425. https://doi.org/10.2307/176942

Harbron, S., & Rapley, R. (2011). *Molecular analysis and genome discovery*. John Wiley & Sons.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62-82. https://doi.org/10.1111/mec.15720

Holmes, R. T., & Sherry, T. W. (2001). Thirty-year bird population trends in an unfragmented temperate deciduous forest: Importance of habitat change. *Auk*, *118*(3), 589-609. https://doi.org/10.1642/0004-8038(2001)118[0589:Tybpti]2.0.Co;2

Hu, T. S., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801-811. https://doi.org/10.1016/j.humimm.2021.02.012

IUCN. 2025. *The IUCN Red List of Threatened Species. Version 2025-1.* https://www.iucnredlist.org. Accessed on 17/08/2025.Kardos, M. (2021). Conservation genetics. *Current Biology*, *31*(19), R1185-R1190. https://doi.org/https://doi.org/10.1016/j.cub.2021.08.047

Kennington, W. J., Gockel, J., & Partridge, L. (2003). Testing for asymmetrical gene flow in a *Drosophila melanogaster* body-size cline. *Genetics*, 165(2), 667-673. https://doi.org/10.1093/genetics/165.2.667

Li, T., Zhang, M., Qu, Y., Ren, Z., Zhang, J., Guo, Y., Heong, K. L., Villareal, B., Zhong, Y., & Ma, E. (2011). Population genetic structure and phylogeographical pattern of rice grasshopper, *Oxya hyla intricata*, across Southeast Asia. *Genetica*, 139(4), 511-524. https://doi.org/10.1007/s10709-011-9573-1

Liedvogel, M., & Delmore, K. (2018). (Micro)evolutionary changes and the evolutionary potential of bird migration. In D. T. Tietze (Ed.), *Bird species: how they arise, modify and vanish* (pp. 109-127). Springer International Publishing Ag. https://doi.org/10.1007/978-3-319-91689-7_7

Lowe, W. H., Kovach, R. P., & Allendorf, F. W. (2017). Population genetics and demography unite ecology and evolution. *Trends in Ecology & Evolution*, *32*(2), 141-152. https://doi.org/10.1016/j.tree.2016.12.002

Michel, N. L., Hobson, K. A., Morrissey, C. A., & Clark, R. G. (2021). Climate variability has idiosyncratic impacts on North American aerial insectivorous bird population trajectories. *Biological Conservation*, *263*, 109329. https://doi.org/https://doi.org/10.1016/j.biocon.2021.109329

Mapelli, F. J., Mora, M. S., Mirol, P. M., & Kittlein, M. J. (2012). Population structure and landscape genetics in the endangered subterranean rodent *Ctenomys porteousi*. *Conservation Genetics*, 13(1), 165-181.

Milà, B., Girman, D. J., Kimura, M., & Smith, T. B. (2000). Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1447), 1033-1040. https://doi.org/10.1098/rspb.2000.1107

Morin, P. A., Luikart, G., Wayne, R. K., & The Snp Workshop, G. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, *19*(4), 208-216.

Nebel, S., A. Mills, J. D. McCracken, and P. D. Taylor. 2010. Declines of aerial insectivores in North America follow a geographic gradient. *Avian Conservation and Ecology - Écologie et conservation des oiseaux* 5(2): 1. http://www.ace-eco.org/vol5/iss2/art1/

Newton, I. (2023). The migration ecology of birds. *Elsevier*.

North American Bird Conservation Initiative. (2019). *The State of Canada's Birds* (Environment and Climate Issue).

Piironen, A., Fox, A. D., Kampe-Persson, H., Skyllberg, U., Therkildsen, O. R., & Laaksonen, T. (2023). When and where to count? Implications of migratory connectivity and nonbreeding distribution to population censuses in a migratory bird population. *Population Ecology*, 65(2), 121-132. https://doi.org/10.1002/1438-390x.12143

Pironon, S., Papuga, G., Villellas, J., Angert, A. L., García, M. B., & Thompson, J. D. (2017). Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. *Biological Reviews*, *92*(4), 1877-1909. https://doi.org/10.1111/brv.12313

Ralston, J., Fitzgerald, A. M., Burg, T. M., Starkloff, N. C., Warkentin, I. G., & Kirchman, J. J. (2021). Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. *Ornithology*, 138(3), 1-16. https://doi.org/10.1093/ornithology/ukab018

Renfrew, R. B., Perlut, N., Maxwell, L. M., Cadman, M., Kim, D. H., Clucas, G. V., & Kovach, A. I. (2022). Population structure of a grassland songbird (*Dolichonyx oryzivorus*) to inform conservation units. *Biodiversity and Conservation*, *31*(1), 77-96. https://doi.org/10.1007/s10531-021-02321-5

Rimbert, H., Darrier, B., Navarro, J., Kitt, J., Choulet, F., Leveugle, M., Duarte, J., Rivière, N., Eversole, K., Le Gouis, J., Davassi, A., Balfourier, F., Le Paslier, M.-C., Berard, A., Brunel, D., Feuillet, C., Poncet, C., Sourdille, P., & Paux, E. (2018). High throughput SNP discovery and genotyping in hexaploid wheat. *Plos One*, *13*(1), e0186329. https://doi.org/10.1371/journal.pone.0186329

Ringbauer, H., Kolesnikov, A., Field, D. L., & Barton, N. H. (2018). Estimating barriers to gene flow from distorted isolation-by-distance patterns. *Genetics*, *208*(3), 1231-1245. https://doi.org/10.1534/genetics.117.300638

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., B. Siegel, R., Desanre, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology 23,* 5726-5739. https://doi.org/10.1111/mec.12977

Salewski, V., & Bruderer, B. (2007). The evolution of bird migration—a synthesis. *Naturwissenschaften*, *94*(4), 268-279. https://doi.org/10.1007/s00114-006-0186-y

Salgotra, R. K., & Chauhan, B. S. (2023). Genetic diversity, conservation, and utilization of plant genetic resources. *Genes*, *14*(1), 174. https://doi.org/10.3390/genes14010174

Schowalter, T. D. (2022). Section II - Population ecology. In T. D. Schowalter (Ed.), *Insect Ecology (Fifth Edition)* (pp. 207-209). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-323-85673-7.00024-1

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution*, *68*(1), 1-15. https://doi.org/10.1111/evo.12258

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589-4621. https://doi.org/10.1111/j.1365-294x.2010.04828.x

Smith-Keune, C., & Van Oppen, M. (2006). Genetic structure of a reef-building coral from thermally distinct environments on the Great Barrier Reef. *Coral Reefs*, 25(3), 493-502. https://doi.org/10.1007/s00338-006-0129-2

Spiller, K. J., & Dettmers, R. (2019). Evidence for multiple drivers of aerial insectivore declines in North America. *Condor*, *121*(2), duz010. https://doi.org/10.1093/condor/duz010

Tarof, S. A., & Ratcliffe, L. M. (2000). Pair formation and copulation behavior in least flycatcher clusters. *The Condor*, 102(4), 832-837. https://doi.org/10.1093/condor/102.4.832

Tarof, S. A., Ratcliffe, L. M., Kasumovic, M. M., & Boag, P. T. (2005). Are least flycatcher (*Empidonax minimus*) clusters hidden leks? *Behavioral Ecology*, 16(1), 207-217. https://doi.org/10.1093/beheco/arh155

Tarof, S. A., & Briskie, J. V. (2020). *Least flycatcher (Empidonax minimus), version 1.0. In Birds of the World.* Cornell Lab of Ornithology, Ithaca, NY, USA.

Taylor, C. M., & Stutchbury, B. J. M. (2016). Effects of breeding versus winter habitat loss and fragmentation on the population dynamics of a migratory songbird. *Ecological Applications*, *26*(2), 424-437. https://doi.org/10.1890/14-1410

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*(10), 2144-2164. https://doi.org/10.1111/mec.13606

Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., & Holmes, R. T. (2002). Links between worlds: Unraveling migratory connectivity. *Trends in Ecology & Evolution*, *17*(2), 76-83. https://doi.org/10.1016/s0169-5347(01)02380-1

Chapter 2. Genetic population structure of the least flycatcher (*Empidonax minimus*): Implications for evaluating migratory connectivity

2.1 Introduction

A population has strong migratory connectivity when most individuals from a breeding population migrate to the same non-breeding site, which could result in them being adapted to specific breeding and non-breeding habitats (Webster et al., 2002). On the other hand, a population exhibits weak migratory connectivity when individuals from different breeding grounds mix on non-breeding grounds. In cases of weak migratory connectivity, individuals may be more flexible in their choice of breeding sites, migratory routes, and non-breeding sites, resulting in higher levels of gene flow (Desaix et al., 2019; Trierweiler et al., 2014). The level of migratory connectivity in a population can help to predict the response of migrants to environmental change, as it can be related to individuals being locally or broadly adapted to different habitats in their wintering range (Finch et al., 2017).

Traditional methods for tracking migration patterns are time-consuming, expensive, and logistically complicated when studying small, long-distance migrant birds, such as passerines (Berthold, 2001). Although miniaturized tracking devices have been developed, these technologies still present logistical challenges, such as tracker recovery, latitudinal errors, mortality, and high tag or harness failure rates (McKinnon & Love, 2018). In contrast, genetic markers have been effective for studying migration across various species (Desaix et al., 2019; Harrigan et al., 2023; Wilson et al., 2022) as they can be used on samples that are easily collected and can be used to pinpoint evolutionary or ecological relationships between organisms (Monsen-Collar & Dolcemascolo, 2010; Ruegg et al., 2014). Genetic markers have proven useful to study genetic differences between populations (Adachi et al., 2023; Clegg et al., 2003). In migratory species, populations are spread across highly variable sites, which likely affects gene flow and genetic variation between them (Desaix et al., 2019). Thus, genetic markers are a useful method for assessing connectivity between populations on breeding and non-breeding grounds, thereby contributing to the understanding of their evolutionary histories (Desaix et al., 2019).

Using genomic data for population assignment and migratory connectivity has been widely employed across different species (Desaix et al., 2019; Ruegg et al., 2014), and selecting the appropriate markers is an essential step to assign unknown individuals to their origin populations accurately. As this process compares allele frequencies, it works more effective when the markers used are highly differentiated, as they accentuate the differences between the origin sites. For example, when using 96 highly divergent SNPs, Ruegg et al. (2014) were able to correctly self-assign individuals back to their breeding locations with 80%- 100% accuracy with the leave-one-out method in Wilson's warblers (*Cardellina pusilla*). Research with the prothonotary warbler (*Protonotaria citrea*) tested the effect of the number of markers on assignment accuracy and found that using 600 highly differentiated SNPs, their assignment accuracy was 83.6% (Desaix et al., 2019). On the other hand, Colston-Nepali et al. (2020) used 6,614 SNPs for population assignment of Atlantic northern fulmars (*Fulmarus glacialis*), and found that the accuracy ranged from 37%-99%.

Studying the migratory connectivity and how it is connected to population structure is essential for declining migratory species, one of which is the least flycatcher (*Empidonax minimus*). This species is a migratory aerial insectivore whose population has declined by ~53% since 1970 (Spiller & Dettmers, 2019). It breeds in in Canada and the northern US, migrates through the Central Plains and eastern US and overwinters in southern and coastal Mexico and Central America (Godfrey, 1979). The least flycatcher is susceptible to forest disturbances and environmental pollution (Holmes & Sherry, 2001; Tarof & Briskie,

2020) and is undergoing population declines across its breeding range (Spiller & Dettmers, 2019), so understanding how its populations are connected throughout its range is essential for conservation management. However, nothing is known about the population genetics of the species, an imperative component to understanding how these declines are affecting the least flycatchers at a population and species level (Kardos, 2021)

This study analyzed single-nucleotide polymorphisms (SNPs) from individuals in breeding locations to identify population structure. Samples from non-breeding individuals were analyzed to assign them to breeding populations. Finally, environmental and geographic analyses—a species distribution model (SDM) and least-cost corridor (LCC)—were conducted to identify possible abiotic factors affecting the population structure and migratory connectivity of the birds.

2.2 Methods

2.2.1 Sample collection, DNA extraction and sequencing, and data preprocessing

Blood samples were collected in the field in Alberta, Saskatchewan and British Columbia in 2022 and 2024. Birds were attracted using call playback, caught in mist nets, and a small (<50 µL) blood sample was taken from the brachial vein using a capillary tube. The birds were banded and released at the capture sites. The samples were stored in a 2 ml tube with ethanol at ambient temperature until we returned to the lab, and were then kept at -20°C. Tissue and feather samples were collected through collaboration with Canadian Migration Monitoring Network (CMMN) banding stations, Environment and Climate Change Canada, and several museums (Royal Saskatchewan Museum, Burke Museum, Royal British Columbia Museum, Cornell University Museum of Vertebrates, Canadian Museum of Nature, New York State Museum, National Museum of Natural History, El Colegio de la Frontera Sur, Instituto de Biología de la Universidad Autónoma de México) to expand the sampling coverage across the least flycatcher's non-breeding range (Fig. 1.3).

Overall,92 samples were obtained from 14 locations across North America, with three intentions: (1) to cover most of the breeding range of the least flycatcher (2) to have sites on either side of known geographical barriers, and (3) to have sites in the migratory and wintering locations of the species. For individual sample information and sites, refer to Appendix 1. DNA was extracted using one of two methods: a modified salting-out procedure (Aljanabi & Martinez, 1997) used for samples collected before 2024, whereas the Qiagen DNeasy blood and tissue kit was used for samples collected in 2024. DNA was standardized to 20 ng/µL and sequenced by RADseq (restriction-site associated DNA sequencing) using three enzymes (*Pst*I, *Msp*I, and *Nsi*I) to obtain genomic data for genotype-by-sequencing (de Ronne et al., 2023). Library preparation was done by the Platforme d'analyse génomique at the Institute of Integrative Biology and Systems (IBIS) at Laval University. The samples were then sent to Genome Quebec for sequencing on an Illumina NovaSeq 6000 S4 PE100 with paired-end sequencing.

The sequenced reads were demultiplexed based on unique barcodes using Sabre/1.00 (Joshi, 2011). Data quality was evaluated using FastQC 0.11.5 and low-quality reads (Phred<30) were removed to ensure a genotyping accuracy of 99.9% (Andrews, 2010). Adapters and barcodes were removed with Cutadapt 1.16, and the sequences were trimmed to 80 bp using q 30,30 to eliminate any ends with a Phred score<30 and to make all sequences the same length (Martin, 2011). Trimmed sequences were aligned to the least flycatcher genome (B10K-DU-023-27_Empidonax_minimus) sequenced by the B10K project using the Maximal Exact Match (MEM) algorithm in BWA-MEM (Li & Durbin, 2009), and indexed using SAMtools (Danecek et al., 2021). To ensure the reliability of the sequences, the BAM files were validated using bamUtil v1.0.14 (Jun et al., 2015), and a file containing the sequence information for each sample was generated. Sequences with a mapping rate below 45% were removed.

After validation, genotypes were called, and SNPs were identified using the Stacks 2.3e pipeline (Catchen et al., 2013). Loci were called if they were present in 50% of the sampling sites and had a minimum allele count of three; a random SNP was called for each locus. The resulting variant call format (VCF) file was processed and filtered using VCFtools/0.1.16 (Danecek et al., 2011). The filters allowed for 30% missing data for SNPs and up to 95% missing data for individuals.

2.2.2 Population genetic structure

To analyze the population structure of the least flycatcher, three analyses were done: 1) a principal component analysis (PCA) as an exploratory analysis, to simplify the data and observe possible clusters between samples; 2) a pairwise fixation index (F_{ST}) analysis, that measures genetic differentiation between two populations to assess how similar they are to one another (Freeland, 2020); and 3) a STRUCTURE analysis, which infers the genetic structure of a population by grouping individuals into clusters based on their genetic similarities (Porras-Hurtado et al., 2013). The PCA was performed using the Adegenet v.2.1.11 package in R (Jombart, 2008) and visualized with ggplot2 3.5.1 (Wickham, 2016). A second PCA was performed using the same method, conducted only in individuals sampled in the breeding sites, to identify possible genetic structure among breeding individuals only.

Genetic differentiation among breeding populations was assessed by calculating pairwise F_{ST} values using the R package dartR 1.0.8 (Mijangos et al., 2022). F_{ST} values range from 0 to 1, where a value closer to zero would mean there is low genetic differentiation, and a value closer to one means that there is high genetic differentiation between the populations compared (Freeland, 2020). Associated 95% confidence intervals were computed from 9999 bootstraps, and p-values were corrected using the Benjamini-Hochberg method to account for multiple comparisons (Benjamini Hochberg, 1995). Only sample sites with more than four individuals were retained to reflect the allele frequencies of the sampled populations accurately, and to prevent statistical errors resulting from lower sample sizes (Loh et al., 2013; Weir & Cockerham, 1984).

Finally, STRUCTURE v2.3.4 (Pritchard et al., 2000) was run to infer the genetic groups in the samples using the admixture model, independent allele frequencies, and loc priors (Hubisz et al., 2009). It was run for five iterations, with 100,000 burn-in and 300,000 post-burn-in steps, for each genetically distinct group (K) from 1 to 6. The most probable number of clusters was determined using the Evanno delta K method (Evanno et al., 2005). A second STRUCTURE run was performed using the same process, but using only samples from the breeding sites, to identify possible genetic structure among breeding individuals.

2.2.3 Outlier analysis

Outlier loci were selected to investigate genetic population structure among the breeding populations (Narum & Hess, 2011). To identify the outlier SNPs, the breeding populations were divided into two groups: eastern samples (ME, ON, NY, and WV) and western samples (SEBC, CAB, SAB, SEAB, SK, MT, and ND) based on their location east or west of the Mississippi River and the Great Lakes (Appendix 1). Then, a Weir and Cockerham F_{ST} analysis (Weir & Cockerham, 1984) was conducted using VCFtools/0.1.16 between these two groups. Subsequently, a Manhattan plot, used to visualize SNPs, was generated to visualize the resulting F_{ST} values of 11,837 SNPs using the ggplot package in R (R Core team, 2024). The SNPs that showed a F_{ST} value > 0.1 were selected, based on their location in the Manhattan plot—points with a value over 0.1 were outliers in the graph. A total of 111 markers were retained, and PCA and STRUCTURE analyses were executed to visualize the genetic structure using outlier loci, using the same settings as before. Outlier SNPs may be candidates for divergent selection and can be more effective in identifying genetic structure and assigning individuals to breeding regions (Woodings et al., 2018).

2.2.4 Population assignment to assess migratory connectivity

Individuals from migratory and wintering populations were assigned to their breeding population of origin by a population assignment analysis using the R package Rubias v.03.4 (Anderson et al., 2008) with the genetic stock identification (GSI) method. This method is typically used to identify the source population in mixed-stock fisheries, but has also been used for studying migratory connectivity in birds (Anderson et al., 2008; Colston-Nepali et al., 2020). Rubias assigns samples of unknown origin to a source population based on the probability of assignment (PofZ), which is calculated by comparing the allele frequencies of the source and mixture populations (Moran & Anderson, 2018).

To accurately assign individuals to populations, the samples were divided into two categories: mixture, comprising non-breeding individuals, and source, comprising breeding samples. The source samples were split into reporting units (east and west) and collections—the locations from which the samples were collected. The collection information was removed from the mixture individuals to allow for correct assignment to their origin. The analysis was conducted between the reporting units—i.e., the *a priori* groups east and west of the Mississippi River and Great Lakes—following the patterns indicated in the PCA and F_{ST} analyses using only the outlier SNPs with the highest F_{ST} values ($F_{ST} > 0.1$). The accuracy of the assignments was tested using a leave-one-out method that simulates a random population mixture and calculates the likelihood that each source population belongs to the reporting unit to which it was assigned. The simulations were conducted with 100 replicates and 200 simulated mixture individuals. The PofZ value was used to verify the accuracy of the assignment for each population, and the standard score (z-score) distribution was recorded.

Next, the origin of migratory and non-breeding individuals was estimated using the "infer_mixture" function from Rubias, performed with 200,000 Markov chain Monte

Carlo iterations and 40,000 burn-ins. The z-score was calculated from the genotype log-likelihoods. A z-score outside of the source distribution's range indicates that the mixture individual may originate from a breeding site that was not sampled in the study. The z-score distributions of the source and mixture samples were compared, and a threshold of -3 to 2, based on the z-score distributions of the breeding individuals in the leave-one-out method, was used to filter the mixture individuals that did not fit the reference distribution. Samples with a PofZ lower than 0.8 and z-score outside of the cutoff range (-3 to 2) were not assigned to any reporting units.

2.2.5 Landscape genetics

To understand and predict how the least flycatcher is distributed across landscapes in their breeding grounds, and how these distributions might change under different conditions, a species distribution model (SDM) was constructed. The SDM quantifies the correlation between environmental factors and the occurrence of a species (Miller, 2010). Input data for the model were prepared in ArcGIS Pro using the SDMtoolbox v2.2b (Brown et al., 2017). Least flycatcher occurrence data were obtained from the Global Biodiversity Information Facility (GBIF; http://data.gbif.org/). The occurrences were filtered to select those in the breeding grounds (Canada and the northern United States) during the breeding months (June and July), and any information prior to 1970 was excluded to avoid non-contemporary data. This resulted in 300,330 occurrences being selected. The density of the occurrence points was reduced (rarefied) with an input distance of 15 km to avoid any inflated performance values (Brown et al., 2017). Environmental data were accessed through the WORLDCLIM dataset (Fick & Hijmans, 2017), which comprises 19 different climate variables. A Pearson's correlation calculation was performed to remove highly correlated variables (R > 0.7). After the correlation calculation, the 10 layers included: mean diurnal range, temperature annual range, mean temperature of the wettest quarter, mean temperature of the warmest quarter, mean temperature of the coldest quarter, precipitation seasonality, precipitation of the wettest quarter, precipitation of the driest quarter, precipitation of the warmest quarter, and precipitation of the coldest quarter. All the layers were clipped to the same extent (North and Central America) and projected in the World Geodetic System 1984. The environmental layers and the rarefied occurrence data were exported into Maxent v.3.4.4 (Hijmans et al., 2005) to create the model. The settings were: regularization multiplier = 1, 500 maximum iterations, and a maximum number of background points = 10,000.

To identify the dispersion routes among the breeding samples, a least-cost corridor (LCC) analysis was conducted using SDMtoolbox v2.2b (Brown et al., 2017) in ArcGIS Pro. In an LCC, each landscape unit or grid cell is assigned a friction value based on its impact (positive or negative) on the dispersal process (Adriaensen et al., 2003). As LCC utilizes occurrence and environmental data, it aims to represent the landscape's effects on movement and connectivity (Sawyer et al., 2011). The SDM was inverted to create a friction layer, and the geographic coordinates for each breeding sample were input as decimal values.

2.3 Results

2.3.1 Sequencing and preprocessing of data

A total of 129,763 SNPs were identified using Stacks 2.3e. After checking the mapping rate and filtering with VCFtools, 92 individuals from 14 populations and 11,837 SNPs were retained. The mean coverage of the samples that were kept was 27.3x per sample, ranging from 1.56x to 126.7x.

2.3.2 Population genetic structure

Two PCA analyses were performed to visualize the genetic structure. For the first one, all 92 samples and 11,837 SNPs were analyzed. In the first two components (Fig. 2.1.A), all populations appeared to cluster together along PC1 (3.36%), except for two samples from Mexico. However, along PC2 (2.23%), some structure is evident, with populations from YK, SEBC, NAB, CAB, SAB, SEAB, SK, MT, ND and two ON samples clustering together; and a second, looser cluster formed by samples from ME, ON, NY, and WV. Except for the two Mexico (MX) samples that were large outliers, the rest of the MX samples clustered between the two groups, slightly closer to the eastern cluster. Overall, this first PCA analysis shows slight structure between western (YT, SEBC, NAB, CAB, SAB, SEAB, SK, MT, and ND) and eastern (ME, ON, NY, WV) samples. The other two components, PC3 and PC4, account for lower levels of variance, at 1.81% and 1.76%, respectively (Fig. 2.1.B). The western samples clustered together with samples from NY, ON, and most samples from MX. No other cluster was apparent, and samples from WV, ON, ME, NY and MX were spread throughout both axes (PC3 and PC4).

The second PCA (Fig. 2.2) included only the breeding samples (n = 63, 11 populations). The first plot (Fig. 2.2.A) (PC1: 3.17% vs. PC2: 2.56%) showed a loose cluster formed by SEBC, CAB, SAB, SEAB, SK, MT, and ND. ME, ON, NY and WV samples pull out from the first cluster. The slight separation between eastern and western locations in the first PCA with all samples is also present in this second analysis. In PC3 (2.50%) and PC4 (2.41%) (Fig. 2.2.B), samples from SEBC, CAB, SAB, SK, SEAB, MT, ND, ON, and NY are clustered around the origin. Samples from SEAB, NY, ON, ME, and WV pull out from this cluster, similar to the pattern seen in PC1 and PC2 (Fig. 2.2.A), with the western samples showing a tight clustering, while the eastern samples form a loose cluster.

For the pairwise F_{ST} analysis (Fig. 2.3), only populations with more than four samples were used (SAB, SEAB, SK, ON, NY and WV). The F_{ST} values ranged from F_{ST} = 0.015, between SEAB and WV and SAB and NY, to F_{ST} = 0, between SEAB and SAB. The highest F_{ST} values were between eastern and western populations: WV and SAB (F_{ST} = 0.015), WV and SEAB (F_{ST} = 0.014), WV and SK (F_{ST} = 0.01), NY and SAB (F_{ST} = 0.015), and NY and SK (F_{ST} = 0.009), supporting the PCA results of higher differentiation between eastern and western samples. Only three out of fifteen comparisons were nonsignificant and between geographically close sites: SAB and SEAB, SEAB and SK, and ON and NY.

After conducting the STRUCTURE analyses, no clustering or structure was observed in any of the K values across all analyses (including all samples and breeding samples); all samples exhibited varying levels of ancestry with no geographic correlation (Fig. 2.4).

2.3.3 Outlier analysis

To select outlier SNPs, a cutoff value of FST > 0.1 was used, resulting in 111 SNPs for the 63 breeding individuals. A PCA was performed to identify genetic structure. An east-west split was evident in the first four components plotted (Fig. 2.5). The first two principal components had a higher percentage of variation in the data (PC1 = 24.36% and PC2 = 7.42%) compared to the PCA analysis using all SNPs (PC1 = 3.17% and PC2 = 2.56%). In the first plot (Fig. 2.5.A), samples from the western regions cluster together on PC1, as well as two ON samples. Meanwhile, eastern samples from WV, ON, ME, and NY are spread along both principal components, with the ME sample being the closest to the western cluster. In PC3 and PC4 (Fig. 2.5.B), which explain 5.77% and 4.98% of the variance, respectively, a loose cluster is formed by samples from SK, WV, MT, SEBC, ND, SAB, and SEAB. Samples from ME, ON, WV, and NY are spread throughout the plot.

The STRUCTURE analysis using only the outlier SNPs (Fig. 2.6) showed the same genetic differentiation between eastern and western samples as found in both PCA and the F_{ST} analysis using all SNPs. The most probable number of clusters was determined as K=2 (Fig. 2.6A), where there is some differentiation between the west group (SEBC, CAB, SAB, SEAB, SK, MT, and ND) and samples from the east group (ON, NY, and WV). Three samples from the eastern group, two ON samples and the ME sample, have a lower ancestry to the second cluster—the one that groups the eastern samples—and seem to have a more similar ancestry to the western group. At K =3 (Fig. 2.6B) the same east-west split is evident, with eastern samples showing higher levels of admixture. However, it does not show any clustering that is different from what is observed at K = 2 (Fig. 2.6A).

2.3.4 Population assignment and migratory connectivity

The leave-one-out method, which tests assignment accuracy for the reporting units using the known breeding locations, and outlier SNPs (Anderson et al., 2008), yielded an average accuracy of 92.3% for the eastern reporting unit and 99.9% for the western reporting unit. All the breeding samples from the western group (n = 40) were correctly assigned to the west reporting unit; in the eastern group, two ON samples were assigned to the west reporting unit, while the rest of the eastern samples (n = 23) were correctly assigned to the east reporting unit (Table 2.3). If the PofZ value was outside of the range (-3 to 2) the assignment was considered incorrect.

The assignment test done for the mixture (non-breeding) individuals, and the z-scores of each individual were compared to the z-score of the source (breeding) individuals (Fig. 2.7). The distributions of these two groups were not similar, indicating that some non-breeding individuals might breed in areas that were not sampled and might be part of a genetic cluster that was not observed in our analyses, or that the resolution is not enough to assign them to a breeding population.

After completing the "infer_mixture" function and selecting individuals with a PofZ > 0.8 and a z-score between -3 and 2, only 12 of the 29 migratory and non-breeding birds were confidently assigned to a reporting unit (Table 2.2). One YT sample was assigned to the western reporting unit; two of the New York migrating birds were assigned to a different reporting unit, eastern and western. The nine retained samples from Mexico were sampled in two different places, Quintana Roo and Yucatan. The three Quitana Roo samples were assigned to the eastern reporting unit, and the six Yucatan samples were evenly assigned between the eastern and western reporting units. The mix of eastern and western birds may indicate that birds from different breeding locations are mixing on the wintering grounds and using the Atlantic flyway during migration.

2.3.5 Landscape genetics

The SDM was modelled with a rarefied dataset of 3,655 occurrence records, using 10 of the 19 environmental layers that remained after the correlation analysis. The environmental layers that contributed most to the model were: mean temperature of the warmest quarter (43.1%) and mean temperature of the coldest quarter (38.6%). The performance of the modelled SDM had an AUC of 0.809—the closer the value is to one, the more accurate it is. The SDM was designed to accommodate the breeding range of the least flycatcher (Fig. 2.8), where the model closely matches the bird's breeding range (Fig. 1.3).

An LCC (Fig. 2.9) was modelled to find the most likely dispersion routes between the breeding sampling sites. Overall, the analysis revealed varying levels of connectivity between the sampled locations, with the lowest resistance observed between samples from SAB, SEAB, SK, and ND. The eastern samples (ON, NY, ME, and WV) exhibited higher resistance between sampling locations than the western samples. Some of the breeding locations with a lower sample size (CAB, SEBC, MT, and ME) appear more isolated in contrast to the previous analyses, where those samples appear to be genetically similar to their geographic clusters, likely due to the small sample size or peripheral location. The LCC analysis did not reveal any fragmentation due to distance; samples from both sides of the Mississippi River and the Great Lakes exhibited moderate resistance between them. For example, there are moderate levels of connectivity between the ND and ON sites. However, closer to each of the mountain ranges surrounding the sampling sites (the Rocky Mountains and the Appalachian Mountains), there are some signs of low connectivity.

2.4 Discussion

This is the first study to examine the genetic structure and migratory connectivity of the least flycatcher. As genetic factors influence demographic factors such as population growth, it is essential to study the genetic structure of a species to pinpoint areas of interest (Kardos, 2021). Moreover, this study provides evidence that using different types of markers (outlier loci versus all SNPs) is effective in identifying genetic structure that is otherwise undetected by STRUCTURE (Figure 2.4) and successfully assigning individuals of unknown origin to their source location. My hypothesis was partially supported, as two genetic groups were found, but the split did not correspond to the Rocky Mountains and, the level of migratory connectivity observed was lower than expected.

Genetic structure among samples on the eastern and western sides of the Mississippi River and the Great Lakes was observed by the PCA and F_{ST} analyses (Figure 2.2 and Figure 2.3); this was not supported by the STRUCTURE analysis done with all SNPs (Figure 2.4), suggesting that the genetic differentiation observed is too low to be detected using this approach. The weak genetic structure was enhanced when analyzed using 111 outlier SNPs, which revealed the same east-west pattern in both the PCA and STRUCTURE analyses (Figure 2.5 and Figure 2.6). The LCC supported the genetic structure observed between the samples on the breeding grounds.

By selecting 111 outlier SNPs, 12 out of 29 birds from non-breeding locations were confidently assigned to one of the two identified genetic groups. The leave-one-out method allowed me to accurately assign individuals of known origin to their correct locations, which in turn provided a level of confidence with the assignments of 12 non-breeding birds to their breeding populations. However, the inability to assign 17 of the 29 non-breeding birds to a breeding population could indicate that there is a third genetic group that was not sampled in this study, possibly in the Northwest part of their range.

The population assignment revealed low levels of migratory connectivity among the populations, suggesting that individuals from different breeding locations may migrate to the same wintering location. The weak migratory connectivity observed may be due to the least flycatcher having a relatively small non-breeding range compared to its breeding range, which could promote interpopulation mixing (Finch et al., 2017). It is not surprising that both analyses conclude with weak genetic structure and weak migratory connectivity, as seasonal migration can influence dispersal behaviours and genetic diversity within a species due to site fidelity and seasonal timing (Pegan et al., 2025; Wilson et al., 2022).

2.4.1 Genetic population structure on the breeding grounds

Broad geographical genetic clustering was observed within the sampled locations on breeding grounds, separating the eastern and western samples (Fig. 2.2, Fig. 2.4, and Fig. 2.5). Other studies with migratory birds also showed weak genetic structure among populations. Desaix et al. (2019) found weak regional genetic structure between the Mississippi River Valley and the Atlantic Seaboard populations of prothonotary warblers (*Protonotaria citrea*), with an overall $F_{ST} = 0.0055$, concluding that there was high genetic connectivity among those two regions. Similarly, in a study comparing the genetic structure of 35 migratory and resident boreal bird species, Pegan et al. (2025) found clearer spatial structure among nonmigratory birds. They also found higher spatial genetic differentiation in long-distance migrants compared to short-distance migratory species, suggesting that the latter group has the most extensive gene flow out of the three compared. Further, in their study they considered the least flycatcher as a short-distance migrant. In the present study, the pairwise F_{ST} values (Fig. 2.3) support the weak genetic structure found between the eastern and western regions, with the highest values found between eastern and western sites. These findings, along with other studies, suggest levels of gene flow that vary across the least flycatcher's range, with higher levels observed in the western samples.

Within the western groups, populations separated by the Rocky Mountains, a known physical barrier, showed no significant differences. The samples from SEBC (west of the Rocky Mountains) clustered with the western samples in the PCA and STRUCTURE analyses (Fig. 2.2, Fig. 2.4, and Fig. 2.5), suggesting that the SEBC samples are genetically similar to those from the east side of the mountain range. However, as only one breeding site was sampled for this study, it is challenging to accurately assess the effect of this geographical barrier on genetic structure. In contrast to those separated by the Rocky Mountains, the eastern individuals sampled around the Appalachian Mountains and exhibited higher genetic differentiation. The breeding birds from the eastern sites showed higher genetic differentiation between them than those from the western group, with almost no clustering of sampling locations in any of the PCAs (Fig. 2.1, 2.2 and 2.4) and relatively high pairwise F_{ST} values (Fig. 2.3), suggesting that the mountain range might be acting as a dispersal barrier between these locations. Another reason for the higher levels of genetic differentiation found between the eastern populations might be due to the higher decline of least flycatcher individuals in the eastern area of their range. With these declines, populations can become isolated due to low levels of gene flow between them.

Migratory behaviour can result in high levels of gene flow between breeding sites, especially if philopatry—the tendency for individuals to return to the same breeding location—is low (Pegan et al., 2025). When examining a species' genetic structure, it is essential to consider the vast distance that migratory birds travel each year, as well as philopatry. Sedgwick's (2004) research with willow flycatchers (*Empidonax traillii*) supported the idea that passerine birds have low natal philopatry and high breeding site fidelity, but noted that this fidelity might depend on breeding success from previous seasons. The results from my thesis indicate that the least flycatcher has a lower philopatry than other passerines, which could cause the weak migratory connectivity observed; however, the sample size of non-breeding individuals was too small to test this hypothesis.

2.4.2 Population assignment and migratory connectivity

Despite weak genetic structure, by using 111 outlier SNPs, the assignment accuracy for least flycatcher individuals was 92.3% and 99.9% for eastern and western populations, respectively. However, there was insufficient genetic variation between the breeding locations to assign individuals to their exact origin only to west or east, this also affected the level of precision for identifying the origin of non-breeding individuals.

The GSI analysis showed that 12 out of 29 non-breeding individuals could be successfully assigned to a specific breeding reporting unit. The results indicated that seven out of the twelve assigned individuals originated from the eastern reporting unit, while the other five came from the western breeding area. The assignment patterns observed in the New York and Yucatan samples, where individuals have different origins, suggest a pattern of weak or diffuse migratory connectivity between breeding and non-breeding sites. None of the western migratory samples from Northern Alberta, as well as four of the five Yukon birds, could be assigned to a reporting unit, likely because their breeding populations were not sampled, which could indicate a third genetically distinct northern population that was not sampled.

The different patterns often reflect the movement of juveniles, as many adult birds rely on prior knowledge to return to known locations, exhibiting stronger site fidelity and adult philopatry (Finch et al., 2017). In contrast, juveniles tend to migrate later than adults (Finch et al., 2017). Research on the least flycatcher's migratory behaviour has shown that adults tend to migrate almost a month ahead of juvenile birds (Van Brempt et al., 2025). Juveniles must rely on innate mechanisms when they migrate south for the first time, and they can be easily displaced from these predetermined trajectories by environmental conditions. If they are successful in surviving these displacements, they will likely do so by deviating from the pre-established route, altering their migration routes (Finch et al., 2017). Migratory timing and natal philopatry among juvenile least flycatchers (Van Brempt et al., 2025; Weatherhead & Forbes, 1994) support low migratory connectivity between the breeding and wintering regions.

2.4.3 Landscape genetics and geographical barriers

The least-cost corridor analysis incorporates geographical information and behavioural aspects to measure connectivity between patches (Adriaensen et al., 2003), in this case, the sampled breeding locations. The high resistance values shown among samples closer to the mountain ranges suggest that the landscape may have an impact on dispersal.

As the modelled LCC did not use any genetic information to create the connectivity corridors among the breeding samples, it provided an independent analysis to compare with the genetic results. The high connectivity among the western region samples matched the genetic analyses, as did the higher resistance and slightly higher genetic differentiation observed among the eastern samples (Fig. 2.8). The LCC showed that the Rocky Mountains and Appalachian Mountains could act as barriers between different breeding grounds, predicting low dispersal across the mountain ranges, which in part matches the genetic results—especially between the eastern sampling sites. However, as those mountains have a North-South orientation, and due to the migration and low philopatry, mixing can occur outside of the breeding season. Increasing the sample size for locations sampled closer to the mountain ranges for further studies may help in a more thorough understanding of the role they play in the dispersal and movement of the least flycatcher.

Widespread climatic processes, such as the last glacial maximum (LGM), around 20,000 years ago, still have an influence on the distribution and genetic structure of high-latitude species (Kimmitt et al., 2023). During this period, glacial cycles caused species to retreat into geographically isolated refugia, and significantly reduced the genetic diversity of species in northern latitudes (Stralberg et al., 2017; (Milà et al., 2000). Many eastern

North American boreal birds were displaced to a single southern refugium during glaciations, and rapidly colonized their current boreal range following these events, about 18,000 years ago (Ralston et al., 2021). This rapid expansion could have occurred too recently for mutations or genetic diversity to appear, causing low levels of genetic structure, such as those observed in the least flycatcher populations (Milà et al., 2000).

2.5 Conclusions

This study provided insights into the genetic structure and migratory connectivity of the least flycatcher, a migratory bird. Furthermore, the findings indicate that the eastern group exhibits higher genetic variation, which could be attributed to the higher population declines observed on the eastern side of the breeding range of the least flycatcher. Similarly, the levels of migratory connectivity found indicate the need to focus on the non-breeding range of the species, as it could affect the population structure of this bird by affecting the philopatry of the individuals. By measuring the population structure of the least flycatcher, I was able to start studying the genetic composition of the species, which can be used to identify areas that require further research. In this case, the eastern region of the breeding range, as well as the non-breeding range of the species.

2.6 References

Adachi, A. M. C. L., Roque, P. C. G., Hazin, F. H. V., Vianna, M., Rotundo, M. M., Oliveira, C., Foresti, F., & Cruz, V. P. (2023). Genetic population structure and diversity of the whitetail dogfish *Squalus albicaudus* (Chondrichthyes, Squaliformes) along the Brazilian Coast as identified by SNP markers. *Fishes*, *8*(7), 373. https://doi.org/10.3390/fishes8070373

Adams, R. V., & Burg, T. M. (2015). Gene flow of a forest-dependent bird across a fragmented landscape. *Plos One*, *10*(11), e0140938. https://doi.org/10.1371/journal.pone.0140938

Adriaensen, F., Chardon, J. P., De Blust, G., Swinnen, E., Villalba, S., Gulinck, H., & Matthysen, E. (2003). The application of 'least-cost' modelling as a functional landscape model. *Landscape and Urban Planning*, *64*(4), 233-247. https://doi.org/https://doi.org/10.1016/S0169-2046(02)00242-6

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655-1664. https://doi.org/10.1101/gr.094052.109

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research 25*(22), 4692-4963.

Anderson, E. C., Waples, R. S., & Kalinowski, S. T. (2008). An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(7), 1475-1486. https://doi.org/10.1139/f08-049

Andrews, S. (2010). *FastQC: a high quality control tool for high throughput sequence data (version 0.12.2)*. In https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Araujo, H. A., Candy, J. R., Beacham, T. D., White, B., & Wallace, C. (2014). Advantages and challenges of genetic stock identification in fish stocks with low genetic resolution. *Transactions of the American Fisheries Society*, 143(2), 479-488. https://doi.org/10.1080/00028487.2013.855258

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. http://www.jstor.org/stable/2346101

Berthold, P. (2001). *Bird migration: A general survey*. Oxford University Press.

Brown, J. L., Bennett, J. R., & French, C. M. (2017). SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *PeerJ*, 5, e4095. https://doi.org/10.7717/peerj.4095

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124-3140. https://doi.org/10.1111/mec.12354

Clegg, S. M., Kelly, J. F., Kimura, M., & Smith, T. B. (2003). Combining genetic markers and stable isotopes to reveal population connectivity and migration patterns in a Neotropical migrant, Wilson's warbler (*Wilsonia pusilla*). *Molecular Ecology*, *12*, 819-830.

Colston-Nepali, L., Provencher, J. F., Mallory, M. L., Franckowiak, R. P., Sun, Z., Robertson, G. J., & Friesen, V. L. (2020). Using genomic tools to inform management of the Atlantic northern fulmar. *Conservation Genetics*, *21*(6), 1037-1050. https://doi.org/10.1007/s10592-020-01309-y

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. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1-4. https://doi.org/10.1093/gigascience/giab008

Dellasala, D. A., & Rabe, D. L. (1987). Response of least flycatchers *Empidonax minimus* to forest disturbances. *Biological Conservation*, *41*(4), 291-299. https://doi.org/10.1016/0006-3207(87)90092-9

De Ronne, M., Légaré, G., Belzile, F., Boyle, B., & Torkamaneh, D. (2023). 3D-GBS: a universal genotyping-by-sequencing approach for genomic selection and other high-throughput low-cost applications in species with small to medium-sized genomes. *Plant Methods*, *19*(1), 13. https://doi.org/10.1186/s13007-023-00990-7

Desaix, M. G., Bulluck, L. P., Eckert, A. J., Viverette, C. B., Boves, T. J., Reese, J. A., Tonra, C. M., & Dyer, R. J. (2019). Population assignment reveals low migratory connectivity in a weakly structured songbird. *Molecular Ecology*, *28*(9), 2122-2135. https://doi.org/10.1111/mec.15083

Duali, J., DeLuca, W. V., Mackenzie, S. A., Tremblay, J. A., Drolet, B., Haché, S., Roberto-Charron, A., Holguín-Ruiz, M., Boardman, R., Cooke, H. A., Rimmer, C. C., McFarland, K. P., Marra, P. P., Taylor, P. D., & Norris, D. R. (2024). Range-wide post- and pre-breeding migratory networks of a declining neotropical–nearctic migratory bird, the blackpoll warbler. *Scientific Reports*, *14*(1), 30229. https://doi.org/10.1038/s41598-024-80838-9

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302-4315.

Finch, T., Butler, S. J., Franco, A. M. A., & Cresswell, W. (2017). Low migratory connectivity is common in long-distance migrant birds. *Journal of Animal Ecology*, *86*(3), 662-673. https://doi.org/10.1111/1365-2656.12635

Freeland, J. R. (2020). *Molecular ecology*. John Wiley & Sons.

Godfrey, E. W. (1979). *Birds of Canada*. National Museum of Natural Sciences, Natural Museums of Canada.

Haché, S., Bayne, E. M., Villard, M. A., Proctor, H., Davis, C. S., Stralberg, D., Janes, J. K., Hallworth, M. T., Foster, K. R., Chidambara-Vasi, E., Grossi, A. A., Gorrell, J. C., & Krikun, R. (2017). Phylogeography of a migratory songbird across its Canadian breeding range: Implications for conservation units. *Ecology and Evolution*, *7*(16), 6078-6088. https://doi.org/10.1002/ece3.3170

Harrigan, R. J., Rajbhandary, J., Bossu, C., Sanzenbacher, P., Dietsch, T., Gruppi, C., Katzner, T. E., Smith, T. B., & Ruegg, K. (2023). Utilizing high-resolution genetic markers to track population-level exposure of migratory birds to renewable energy development. *PLOS Sustainability and Transformation*, *2*(12), e0000088. https://doi.org/10.1371/journal.pstr.0000088

Holmes, R. T., & Sherry, T. W. (2001). Thirty-year bird population trends in an unfragmented temperate deciduous forest: importance of habitat change. *The Auk 118*, 589-609.

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332. https://doi.org/https://doi.org/10.1111/j.1755-0998.2009.02591.x

Jensen, A. M., O'Neil, N. P., Iwaniuk, A. N., & Burg, T. M. (2019). Landscape effects on the contemporary genetic structure of ruffed grouse (*Bonasa umbellus*) populations. *Ecology and Evolution*, *9*(10), 5572-5592. https://doi.org/10.1002/ece3.5112

Jombart, T. (2008). *Adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Joshi, N. (2011). Sabre: a barcode demultiplexing and trimming tool for FastQ files. *GitHub: San Francisco, CA, USA*.

Jun, G., Wing, M. K., Abecasis, G. R., & Kang, H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale

DNA sequence data. *Genome Research*, *25*(6), 918-925. https://doi.org/10.1101/gr.176552.114

Kardos, M. (2021). Conservation genetics. *Current Biology*, *31*(19), R1185-R1190. https://doi.org/https://doi.org/10.1016/j.cub.2021.08.047

Kimmitt, A. A., Pegan, T. M., Jones, A. W., Wacker, K. S., Brennan, C. L., Hudon, J., Kirchman, J. J., Ruegg, K., Benz, B. W., Herman, R., & Winger, B. M. (2023). Genetic evidence for widespread population size expansion in North American boreal birds prior to the Last Glacial Maximum. *Proceedings of the Royal Society B: Biological Sciences*, 290(1991), 1-12. https://doi.org/10.1098/rspb.2022.1334

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

Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, *193*(4), 1233-1254. https://escholarship.org/content/qt11f4706p/qt11f4706p.pdf

Marjakangas, E.-L., Bosco, L., Versluijs, M., Xu, Y., Santangeli, A., Holopainen, S., Mäkeläinen, S., Herrando, S., Keller, V., Voříšek, P., Brotons, L., Johnston, A., Princé, K., Stephen, Aghababyan, K., Ajder, V., Dawn, Bino, T., Kerem, . . . Lehikoinen, A. (2023). Ecological barriers mediate spatiotemporal shifts of bird communities at a continental scale. *Proceedings of the National Academy of Sciences*, *120*(23), 1-10. https://doi.org/10.1073/pnas.2213330120

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

McKinnon, E. A., & Love, O. P. (2018). Ten years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging. *The Auk*, *135*(4), 834-856. https://doi.org/10.1642/auk-17-202.1

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). *dartR* v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods in Ecology and Evolution*, *13*(10), 2150-2158. https://doi.org/10.1111/2041-210x.13918

Milà, B., Girman, D. J., Kimura, M., & Smith, T. B. (2000). Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. *Proceedings of the Royal Society of London*. Series B: Biological Sciences, 267(1447), 1033-1040. https://doi.org/10.1098/rspb.2000.1107

Miller, J. (2010). Species Distribution Modeling. *Geography Compass*, 4(6), 490-509. https://doi.org/10.1111/j.1749-8198.2010.00351.x

Monsen-Collar, K. J., & Dolcemascolo, P. (2010). Using molecular techniques to answer ecological questions. *Nature Education Knowledge, 3*(10):1. https://www.nature.com/scitable/knowledge/library/using-molecular-techniques-to-answer-ecological-questions-15643181/

Moran, B., & Anderson, E. (2018). Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*, 551-560. https://doi.org/10.1139/cjfas-2018-0016

Narum, S. R., & Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, *11*(s1), 184-194. https://doi.org/10.1111/j.1755-0998.2011.02987.x

Pegan, T.M., Kimmitt, A.A., Benz, B.W. et al. (2025). Long-distance seasonal migration to the tropics promotes genetic diversity but not gene flow in boreal birds. *Nat Ecol Evol* 9, 957–969. https://doi.org/10.1038/s41559-025-02699-3

Piironen, A., Fox, A. D., Kampe-Persson, H., Skyllberg, U., Therkildsen, O. R., & Laaksonen, T. (2023). When and where to count? Implications of migratory connectivity and nonbreeding distribution to population censuses in a migratory bird population. *Population Ecology*, *65*(2), 121-132. https://doi.org/10.1002/1438-390x.12143

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, *4,* 1-13.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959. https://doi.org/10.1093/genetics/155.2.945

Ralston, J., Fitzgerald, A. M., Burg, T. M., Starkloff, N. C., Warkentin, I. G., & Kirchman, J. J. (2021). Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. *Ornithology*, 138(3), 1-16. https://doi.org/10.1093/ornithology/ukab018

Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., Stanton, J. C., Panjabi, A., Helft, L., Parr, M., & Marra, P. P. (2019). Decline of the North American avifauna. *Science*, *366*(6461), 120-124. https://doi.org/10.1126/science.aaw1313

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., B. Siegel, R., Desanre, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology 23,* 5726-5739. https://doi.org/10.1111/mec.12977

Sawyer, S. C., Epps, C. W., & Brashares, J. S. (2011). Placing linkages among fragmented habitats: do least-cost models reflect how animals use landscapes? *Journal of Applied Ecology*, *48*(3), 668-678. https://doi.org/10.1111/j.1365-2664.2011.01970.x

Sedgwick, J. A. (2004). Site fidelity, territory fidelity, and natal philopatry in willow flycatchers (*Empidonax traillii*). *Auk*, *121*(4), 1103-1121. https://doi.org/10.1642/0004-8038(2004)121[1103:Sftfan]2.0.Co;2

Spiller, K. J., & Dettmers, R. (2019). Evidence for multiple drivers of aerial insectivore declines in North America. *Condor*, *121*(2), 13, duz010. https://doi.org/10.1093/condor/duz010

Stralberg, D., Matsuoka, S. M., Handel, C. M., Schmiegelow, F. K. A., Hamann, A., & Bayne, E. M. (2017). Biogeography of boreal passerine range dynamics in western North America: past, present, and future. *Ecography*, 40(9), 1050-1066. https://doi.org/10.1111/ecog.02393

Tarof, S. A., & Briskie, J. V. (2020). Least flycatcher *(Empidonax minimus),* version 1.0. *In Birds of the World.* Cornell Lab of Ornithology, Ithaca, NY, USA.

Trierweiler, C., Klaassen, R. H. G., Drent, R. H., Exo, K.-M., Komdeur, J., Bairlein, F., & Koks, B. J. (2014). Migratory connectivity and population-specific migration routes in a long-distance migratory bird. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778), 20132897. https://doi.org/10.1098/rspb.2013.2897

Van Brempt, M., Holroyd, G. L., & Hvenegaard, G. T. (2025). Phenological changes in sex- and age-differential autumn migration of the least flycatcher (*Empidonax minimus*). *The Wilson Journal of Ornithology*, *137*(1), 98-111. https://doi.org/10.1080/15594491.2024.2444030

Weatherhead, P. J., & Forbes, M. R. L. (1994). Natal philopatry in passerine birds - genetic or ecological influences. *Behavioral Ecology*, *5*(4), 426-433. https://doi.org/10.1093/beheco/5.4.426

Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S., & Holmes, R. T. (2002). Links between worlds: unraveling migratory connectivity. *Trends in Ecology & Evolution*, *17*(2), 76-83. https://doi.org/10.1016/s0169-5347(01)02380-1

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution 38(6)*, 1358-1370.

Wickham, H. (2016). Data analysis. In *ggplot2: Elegant graphics for data analysis* (pp. 189-201). Springer.

Wilson, R. E., Sonsthagen, S. A., DaCosta, J. M., Sorenson, M. D., Fox, A. D., Weaver, M., Skalos, D., Kondratyev, A. V., Scribner, K. T., Walsh, A., Ely, C. R., & Talbot, S. L. (2022). As the goose flies: migration routes and timing influence patterns of genetic diversity in a circumpolar migratory herbivore. *Diversity-Basel*, *14*(12), 1067. https://doi.org/10.3390/d14121067

Woodings, L. N., Murphy, N. P., Doyle, S. R., Hall, N. E., Robinson, A. J., Liggins, G. W., Green, B. S., Cooke, I. R., Bell, J. J., & Strugnell, J. M. (2018). Outlier SNPs detect weak regional structure against a background of genetic homogeneity in the Eastern Rock Lobster, *Sagmariasus verreauxi*. *Marine Biology*, *165*(12), 185. https://doi.org/10.1007/s00227-018-3443-7

Table 2.1 Number of birds for each location The source indicates how the samples were collected: fieldwork (Burg lab), Albert Creek Bird Observatory (ACBO), New York State Museum (NYSM), Canadian Museum of Nature (CMN), Cornell University Museum of Vertebrates (CUMV), El Colegio de la Frontera Sur (ECOSUR), Instituto de Biología de la Universidad Nacional Autónoma de México (IBUNAM), and National Museum of Natural History (NMNH).The season column indicates whether the individuals was collected during the breeding (BR) season or during the non-breeding (NBR) season.

Location	Number of birds	Population	Source	Season
Albert Creek, YT	5	YT	ACBO	NBR
Revelstoke, BC	4	SEBC	Burg lab	BR
Fort McMurray, AB	2	NAB	ECCC	NBR
Jasper, AB	2	CAB	Burg lab	BR
Waterton Lakes, AB	6	SAB	Burg lab	BR
Beauvais Lake, AB	1	SAB	Burg lab	BR
Lethbridge, AB	1	SAB	Burg lab	BR
Cypress Hills, AB	12	SEAB	Burg lab	BR
Dundurn, SK	6	SK	ECCC	BR
Prince Albert National Park, SK	1	SK	ECCC	BR
Saskatoon, SK	3	SK	ECCC	BR
Regina, SK	1	SK	ECCC	BR
Missoula, MT	1	MT	Burg lab	BR
Butte, MT	1	MT	Burg lab	BR
Bottineau, ND	1	ND	Burg lab	BR
Towner, ND	1	ND	Burg lab	BR
Somerset County, ME	1	ME	NYSM	BR
Ottawa, ON	9	ON	CMN	BR
Nassau County, NY	1	NY	CUMV	NBR
Tompkins County, NY	1	NY	CUMV	NBR
Franklin County, NY	1	NY	NYSM	BR
Greene County, NY	3	NY	NYSM	BR
Saratoga County, NY	1	NY	NYSM	BR
Highland, West Virginia	3	WV	NMNH	BR
Grayson, West Virginia	3	WV	NMNH	BR
Highland, West Virginia	2	WV	NMNH	BR
Othon P. Blanco, Quintana Roo, Mexico	3	MX	ECOSUR	NBR
Bacalar, Quintana Roo, Mexico	1	MX	ECOSUR	NBR

Puerto Felipe Carrillo, Quintana Roo, Mexico	8	MX	ECOSUR	NBR
El Cuyo, Yucatan, Mexico	1	MX	IBUNAM	NBR
Dzilam, Yucatan, Mexico	7	MX	IBUNAM	NBR

Table 2.2 Breeding samples were arranged by population and group (west or east), determined by geographic location relative to the Mississippi River and the Great Lakes.

Sample ID	Population	Group
SEBC 019	SEBC	West
SEBC 020	SEBC	West
SEBC 021	SEBC	West
SEBC 022	SEBC	West
CAB 001	CAB	West
CAB 002	CAB	West
SAB 001	SAB	West
SAB 002	SAB	West
SAB 003	SAB	West
SAB 004	SAB	West
SAB 005	SAB	West
SAB 006	SAB	West
SAB 009	SAB	West
SEAB L001	SEAB	West
SEAB 001	SEAB	West
SEAB 003	SEAB	West
SEAB 008	SEAB	West
SEAB 009	SEAB	West
SEAB 010	SEAB	West
SEAB 011	SEAB	West
SEAB 012	SEAB	West
SEAB 013	SEAB	West
SEAB 014	SEAB	West
SEAB 015	SEAB	West
SEAB 016	SEAB	West
SEAB 017	SEAB	West
SK 001	SK	West
SK 002	SK	West
SK 003	SK	West
SK 004	SK	West
SK 005	SK	West
SK 007	SK	West
SK 009	SK	West
SK 010	SK	West
SK 011	SK	West
SK 012	SK	West
MT M01	MT	West

MT W01	MT	West
ND S01	ND	West
ND T02	ND	West
ME 001	ME	East
ON 6172	ON	East
ON 6185	ON	East
ON 6187	ON	East
ON 6188	ON	East
ON 6189	ON	East
ON 6195	ON	East
ON 6196	ON	East
ON 6720	ON	East
ON 9743	ON	East
NY 007	NY	East
NY 009	NY	East
NY 010	NY	East
NY 011	NY	East
NY 012	NY	East
WV 001	WV	East
WV 002	WV	East
WV 003	WV	East
WV 004	WV	East
WV 005	WV	East
WV 006	WV	East
WV 007	WV	East
WV 008	WV	East

Table 2.3. List of breeding individuals with their original reporting unit, the reporting unit inferred by the leave-one-out method, and Z-score.

Sample ID	Population	Original reporting unit	Inferred reporting unit	Z-score
SEBC 019	SEBC	West	West	1.39
SEBC 020	SEBC	West	West	0.61
SEBC 021	SEBC	West	West	0.60
SEBC 022	SEBC	West	West	1.27
CAB 001	CAB	West	West	1.35
CAB 002	CAB	West	West	0.19
SAB 001	SAB	West	West	1.09
SAB 002	SAB	West	West	1.01
SAB 003	SAB	West	West	0.67
SAB 004	SAB	West	West	1.19
SAB 005	SAB	West	West	1.12
SAB 006	SAB	West	West	0.73
SAB 009	SAB	West	West	-0.09
SEAB L001	SEAB	West	West	0.60
SEAB 001	SEAB	West	West	0.73
SEAB 003	SEAB	West	West	0.94
SEAB 008	SEAB	West	West	-3.08
SEAB 009	SEAB	West	West	0.51
SEAB 010	SEAB	West	West	0.01
SEAB 011	SEAB	West	West	1.57
SEAB 012	SEAB	West	West	-1.23
SEAB 013	SEAB	West	West	1.06
SEAB 014	SEAB	West	West	-1.00
SEAB 015	SEAB	West	West	0.14
SEAB 016	SEAB	West	West	0.89
SEAB 017	SEAB	West	West	0.22
SK 001	SK	West	West	1.23
SK 002	SK	West	West	0.53
SK 003	SK	West	West	1.78
SK 004	SK	West	West	1.09
SK 005	SK	West	West	1.26
SK 007	SK	West	West	-1.34
SK 009	SK	West	West	0.92
SK 010	SK	West	West	0.72
SK 011	SK	West	West	1.34
SK 012	SK	West	West	-1.14

MT M01	MT	West	West	1.24
MT W01	MT	West	West	1.11
ND S01	ND	West	West	0.22
ND T02	ND	West	West	-1.30
ON 6172	ON	East	West	-3.16
ON 6185	ON	East	East	2.36
ON 6187	ON	East	West	0.01
ON 6188	ON	East	East	-0.49
ON 6189	ON	East	East	0.06
ON 6195	ON	East	East	0.31
ON 6196	ON	East	East	0.69
ON 6720	ON	East	East	-0.51
ON 9743	ON	East	East	0.45
NY 007	NY	East	East	0.13
NY 009	NY	East	East	2.19
NY 010	NY	East	East	-1.34
NY 011	NY	East	East	-1.78
NY 012	NY	East	East	0.62
WV 001	WV	East	East	0.59
WV 002	WV	East	East	-0.76
WV 003	WV	East	East	0.45
WV 004	WV	East	East	-0.73
WV 005	WV	East	East	-0.84
WV 006	WV	East	East	0.61
WV 007	WV	East	East	1.06
WV 008	WV	East	East	1.32

Table 2.4 List of non-breeding individuals with their assigned reporting unit, PofZ, and Z-score. Individuals in bold are those who were confidently assigned to a reporting unit if they had a PofZ > 0.8 and a Z-score value between -3 to 2.

Mixture collection	Individual	Rep. Unit	PofZ	Z-score
YT	YT 003	West	0.66	1.51
YT	YT 004	West	0.68	1.06
YT	**YT 005**	**West**	**0.80**	**-0.25**
YT	YT 006	West	0.68	-4.03
YT	YT 009	West	0.68	0.88
NAB	NAB 001	West	0.41	-0.09
NAB	NAB 003	West	0.57	1.05
NY	**NY 005**	**West**	**0.88**	**-1.00**
NY	**NY 006**	**East**	**0.96**	**0.33**
MX	MX 015	West	0.51	-2.22
MX	MX 016	West	0.71	-0.10
MX	MX 017	East	0.61	0.98
MX	MX 019	West	0.67	-0.78
MX	**MX 023**	**East**	**0.96**	**0.58**
MX	MX 024	West	0.59	-7.15
MX	MX 025	West	0.52	-29.45
MX	**MX 026**	**East**	**0.99**	**0.18**
MX	MX 027	West	0.99	-7.41
MX	**MX 028**	**East**	**0.98**	**0.59**
MX	MX 029	West	0.95	-17.37
MX	MX 030	East	0.54	-0.17
MX	MX 031	East	0.68	-0.65
MX	**MX 032**	**West**	**0.85**	**-1.85**
MX	**MX 033**	**West**	**0.99**	**-0.61**
MX	**MX 034**	**East**	**0.94**	**0.45**
MX	**MX 035**	**East**	**0.97**	**0.43**
MX	**MX 036**	**East**	**1.00**	**-0.53**
MX	MX0 37	West	0.52	-2.89
MX	**MX 038**	**West**	**0.98**	**0.69**

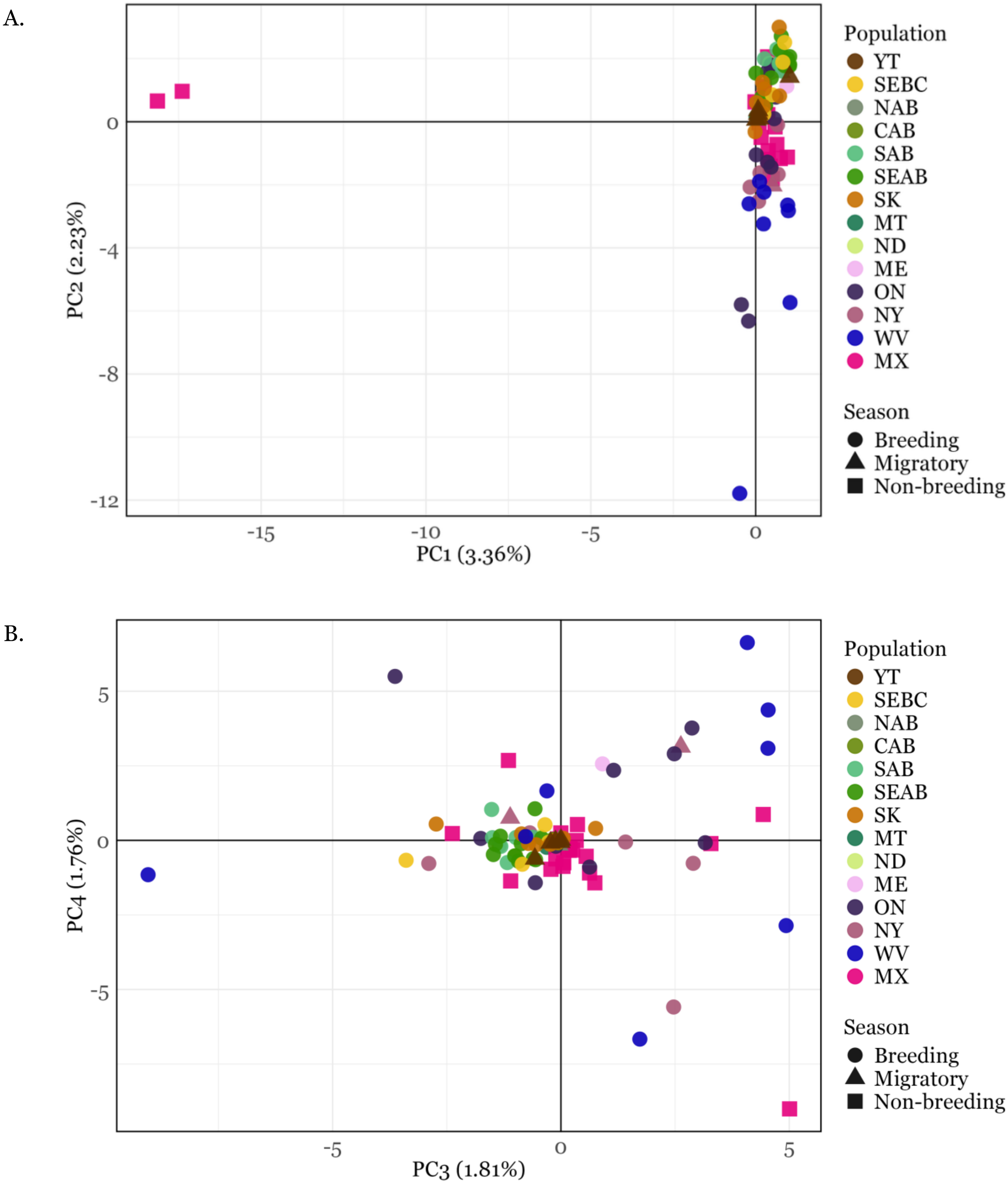


Figure 2.1 PCA of all least flycatcher sampled individuals with all SNPs (11,837) and all populations (14 populations, 92 individuals). Individuals were identified as: breeding (circle), and non-breeding: migratory (triangles) and non-breeding (squares).

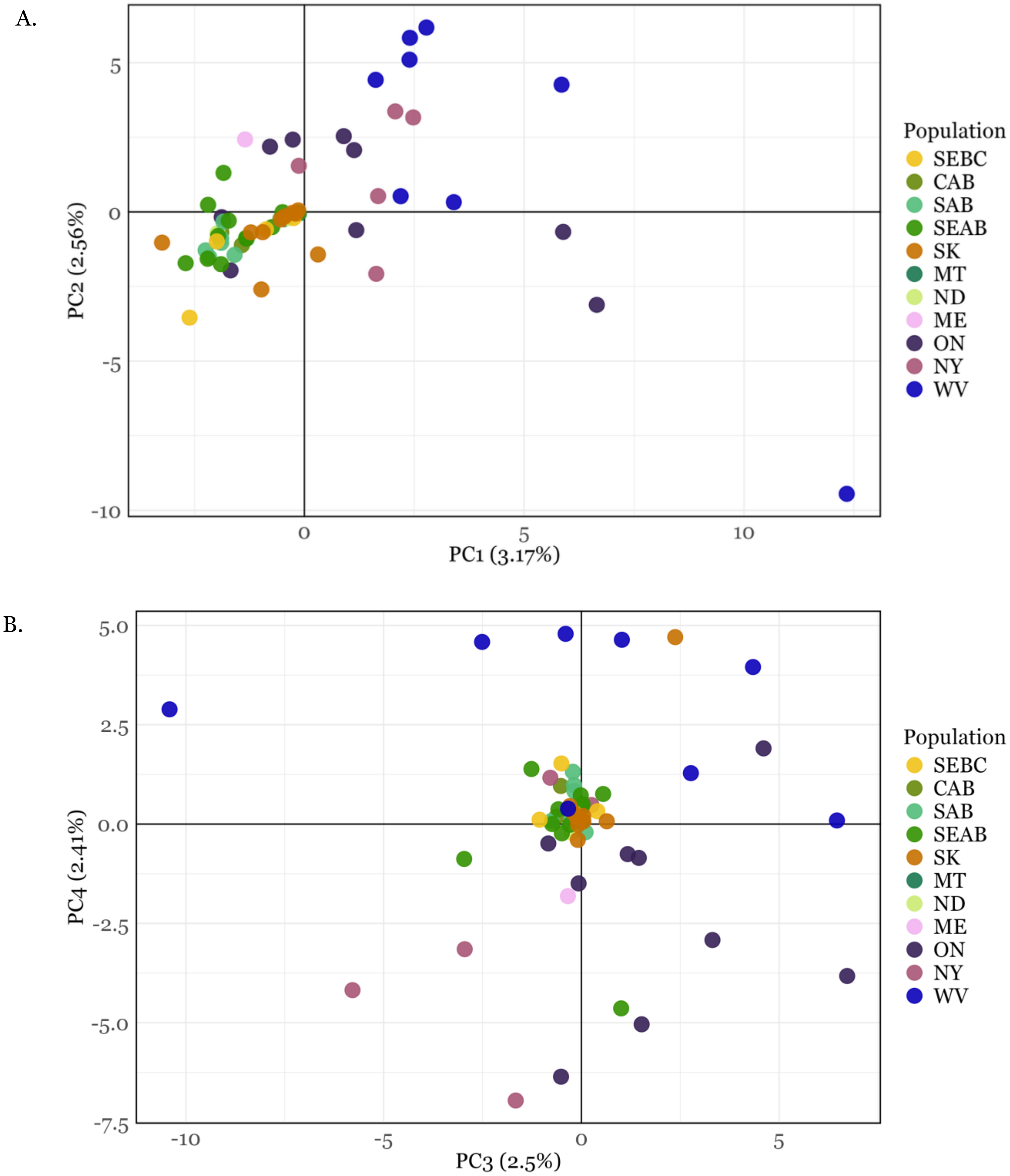


Figure 2.2 PCA with all SNPs (11,837) and only samples from breeding populations (11 populations, 63 individuals). Western samples are coloured in shades of yellow and green, while eastern samples are coloured in shades of pink and blue.

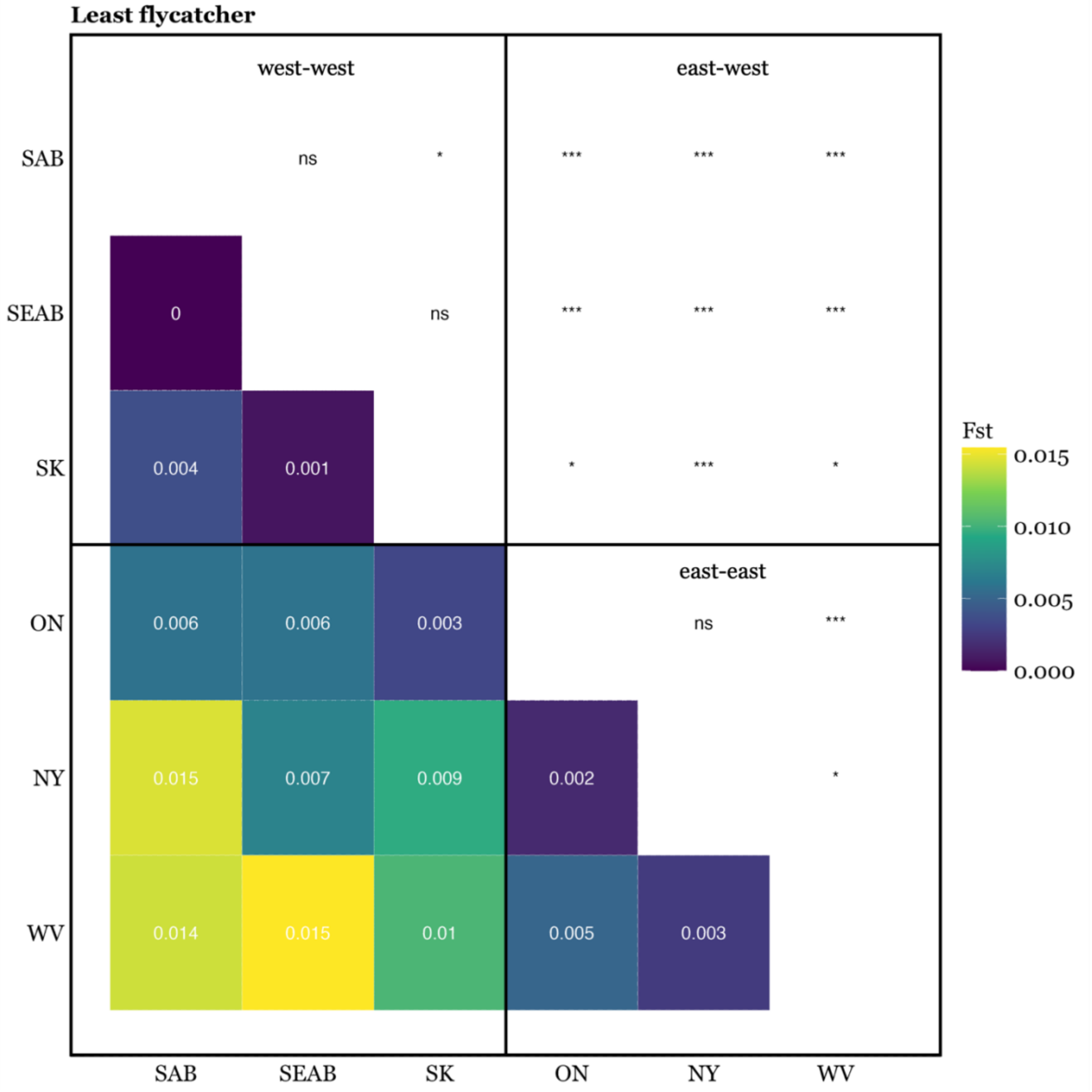


Figure 2.3 Heatmap of Fst values estimated between sampled breeding sites. Sites with fewer than five individuals were excluded from Fst estimates. Significance of Fst values is indicated in the upper triangle of the heatmap, where: *** P-values ≤ 0.001, * P-values ≤ 0.05, and ns are P-values > 0.05.

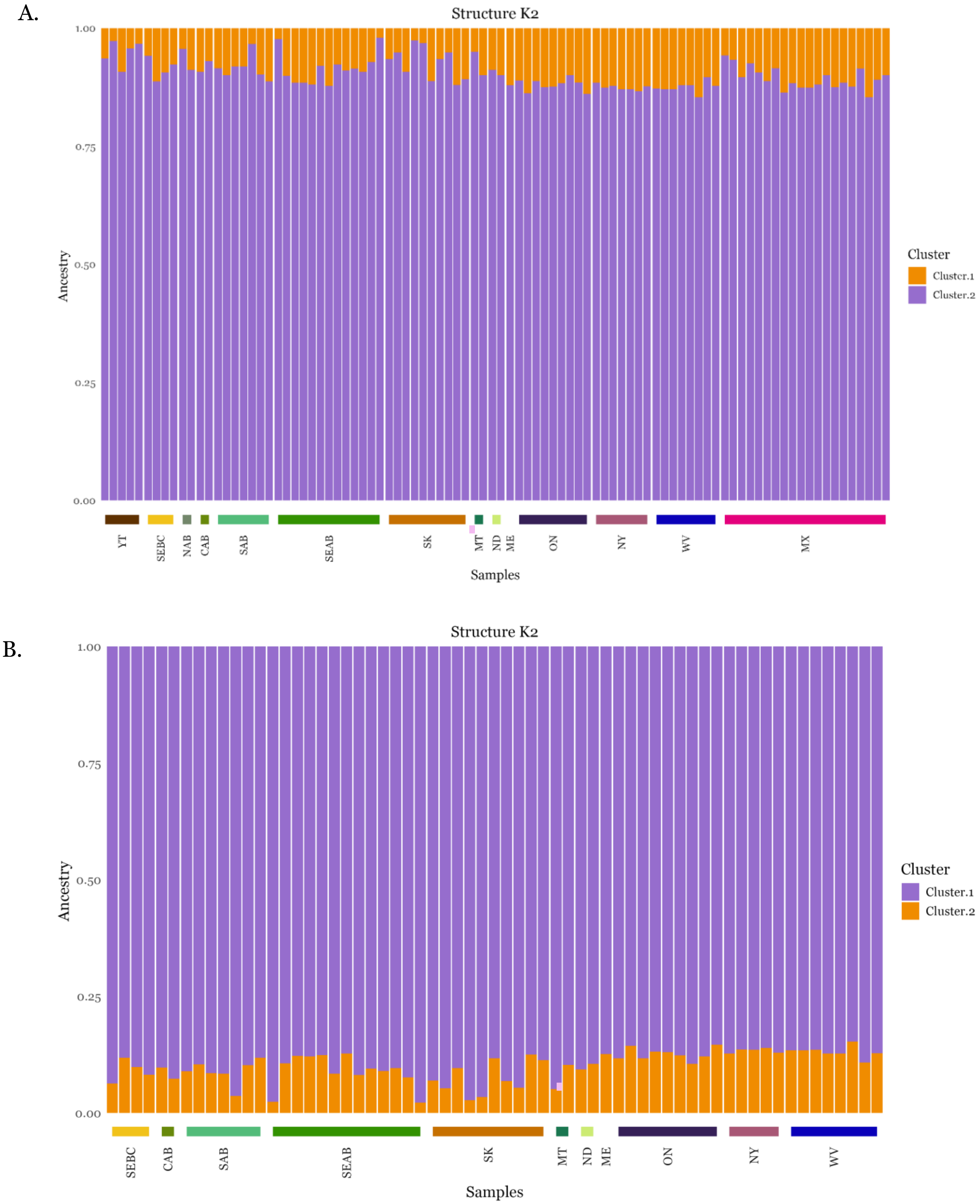


Figure 2.4 STRUCTURE plots at K =2 using all SNPS (11,837) and all 92 samples (A), and only breeding samples n=63 (B).

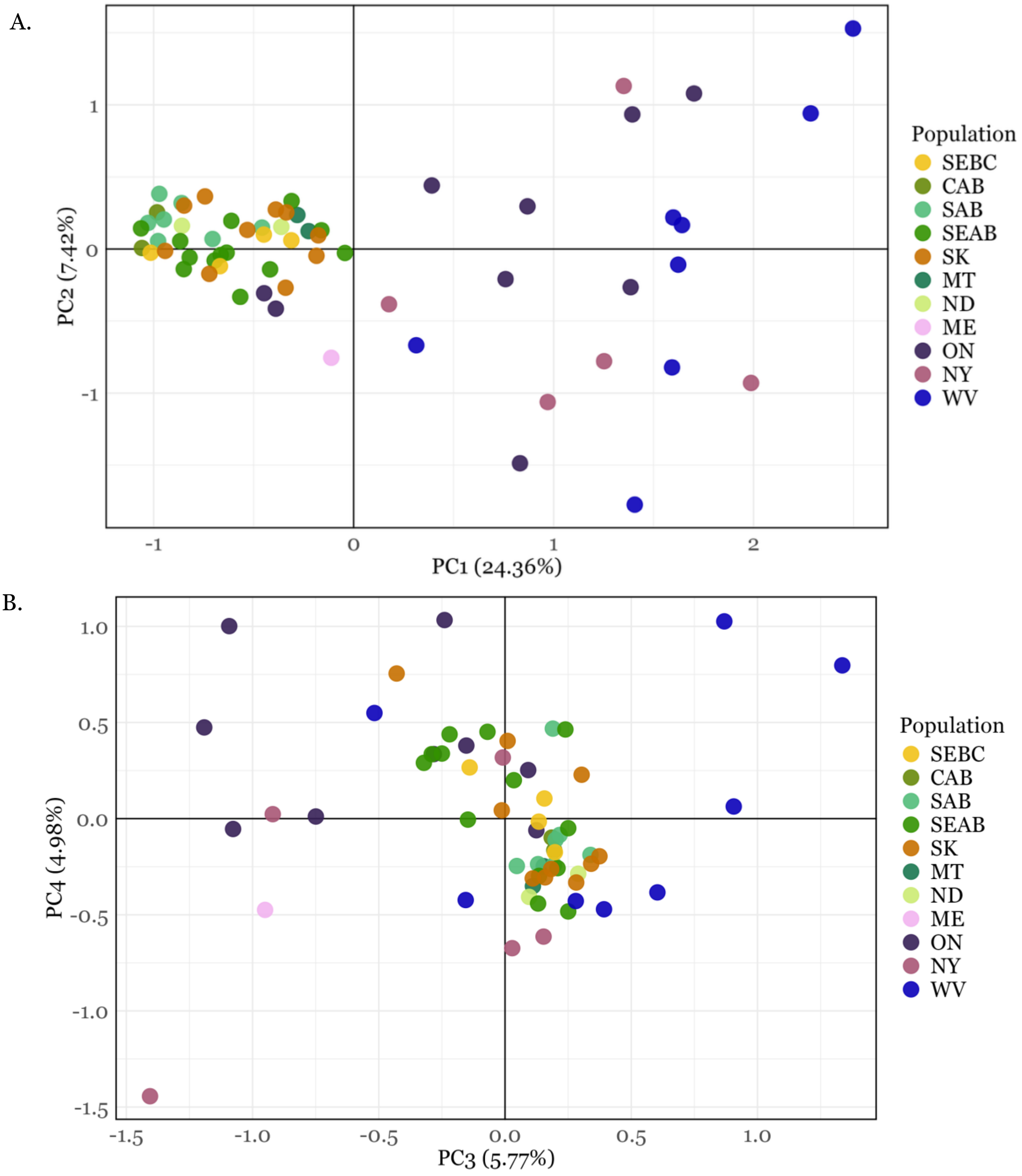


Figure 2.5 PCA with 111 outlier SNPs using only breeding populations (11 populations, 63 individuals).

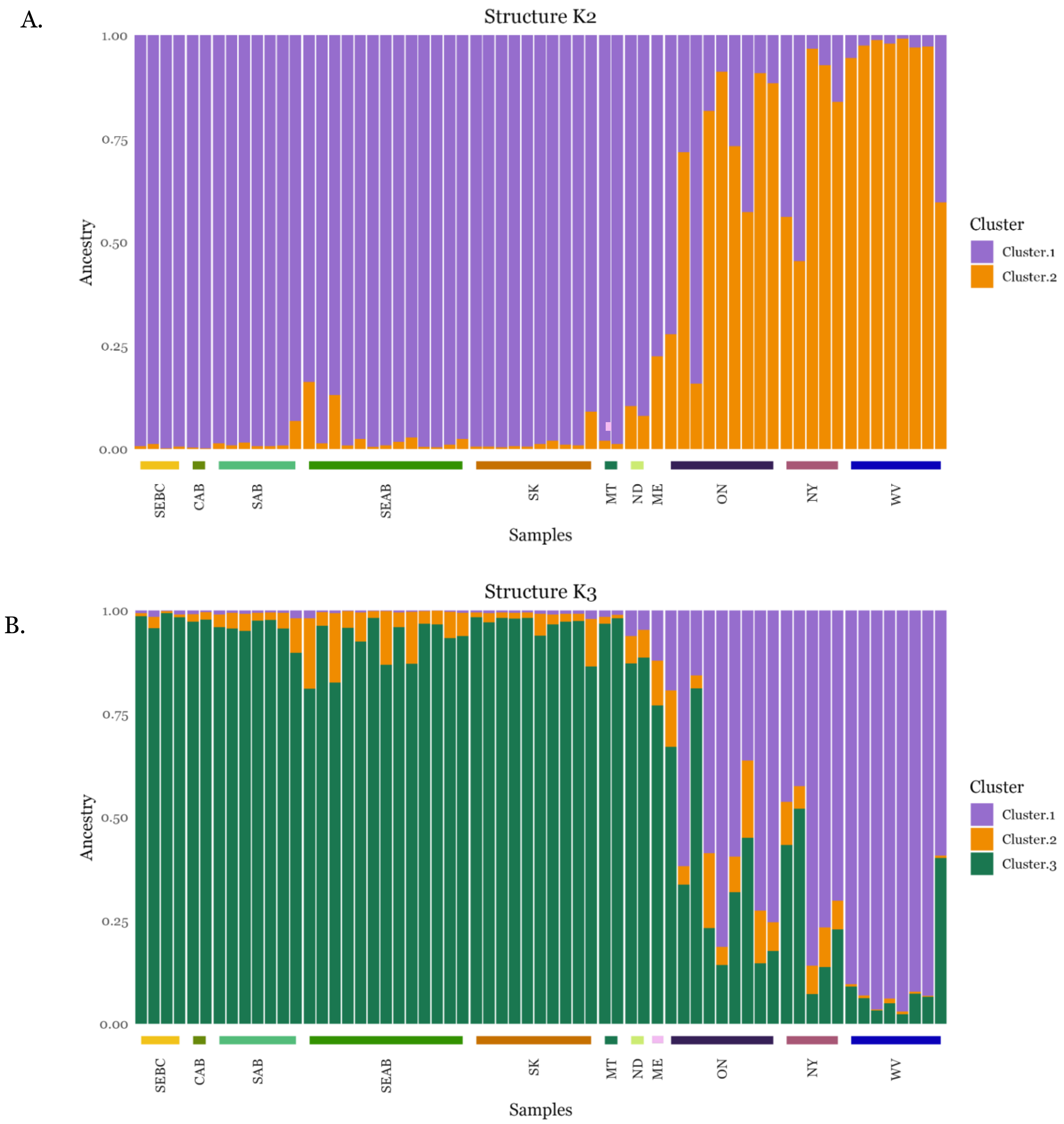


Figure 2.6 STRUCTURE plots for A. K = 2 and B. K = 3 for breeding individuals (63) using outlier SNPs (111).

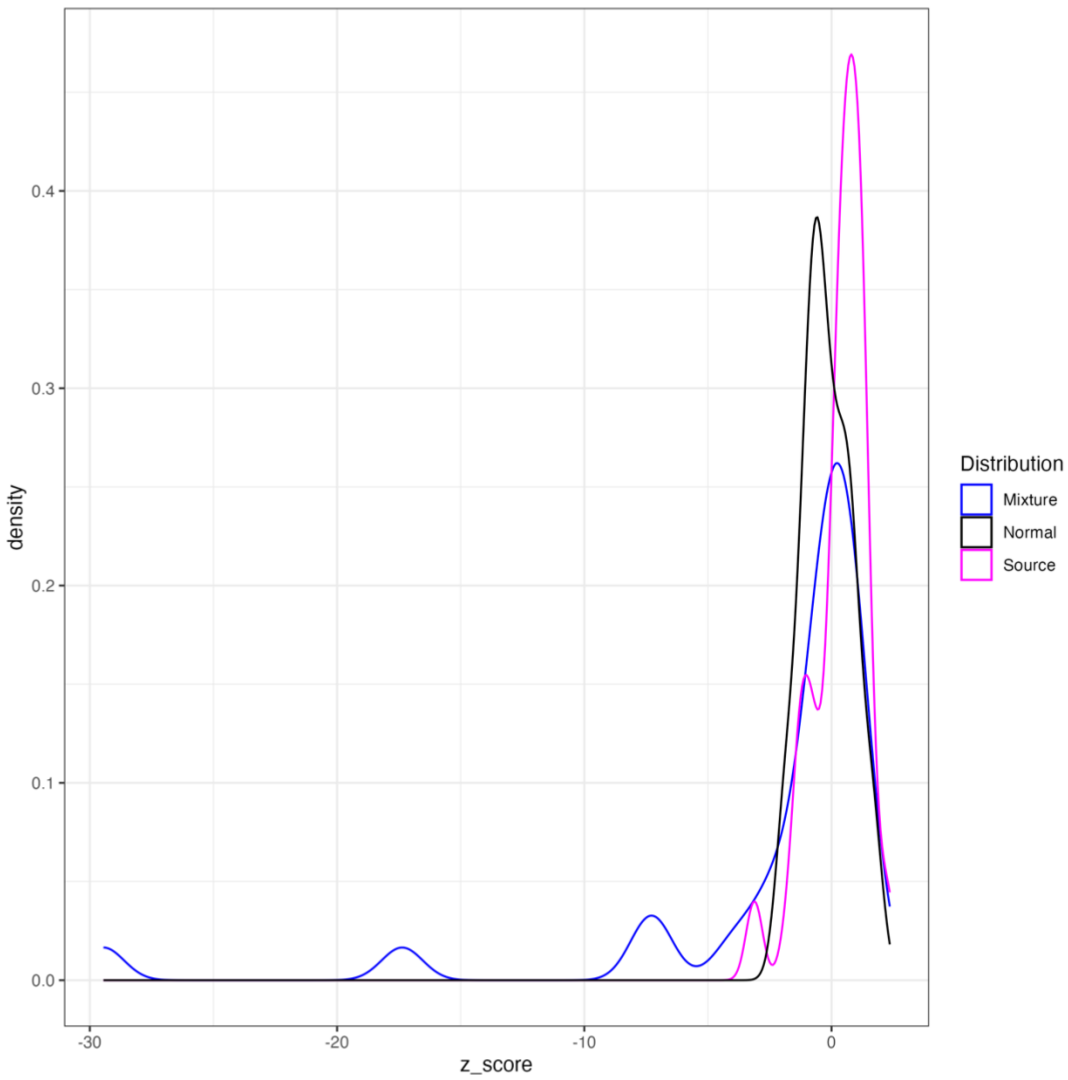


Figure 2.7 Z-score distribution comparison between self-assigned source individuals (pink line) mixture individuals (blue line), and a normal distribution for n=70 (black).

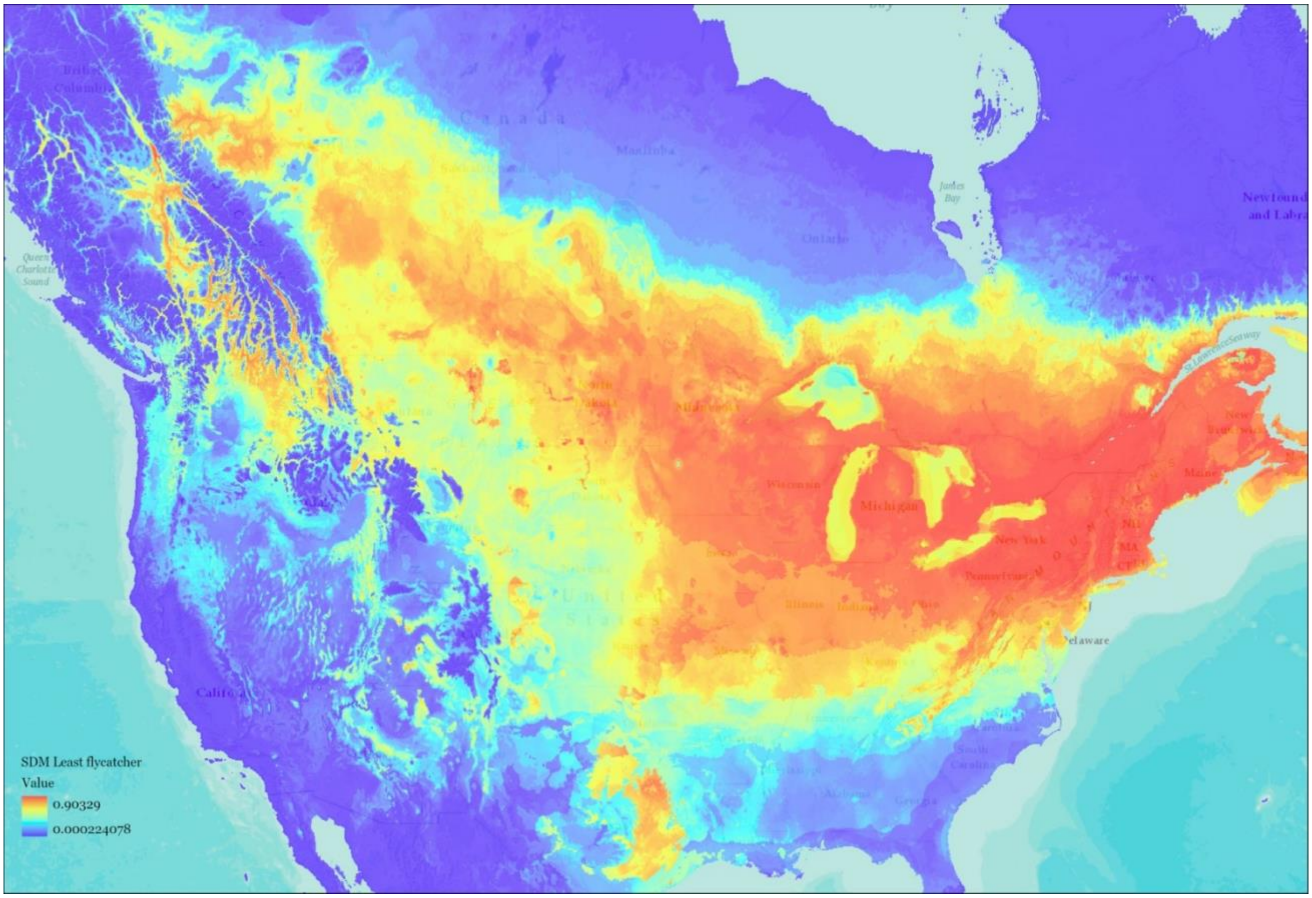


Figure 2.8. Species distribution model of the least flycatcher across its breeding range.3,655 rarefied points from the occurrence data and 10 environmental layers with low correlation (<0.7). The model shows the areas of suitable habitat for the least flycatcher (red-high, blue-low).

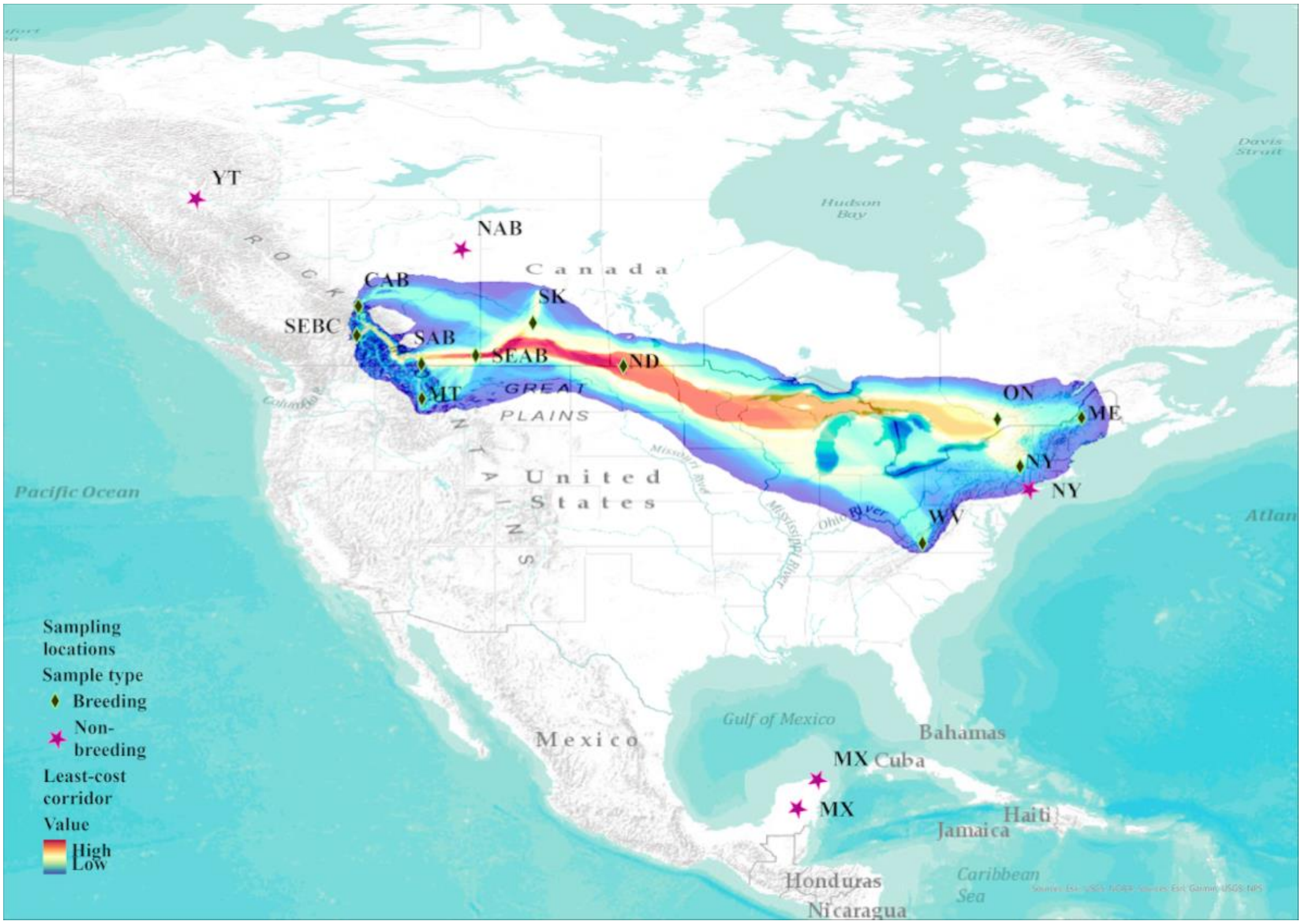


Figure 2.9 Least cost corridors between the sampling breeding sites of the least flycatcher showing the dispersal routes with the lowest resistance (red) based on the species' preferred environmental conditions as modelled by the SDM.

Chapter 3: General discussion

3.1 Population structure of the least flycatcher

The results of the genetic structure of the least flycatcher are similar to those observed in other species who are present in the east of North America, but not the west. Although the least flycatcher has been successful in colonizing most of Canada, its breeding range does not extend to the West Coast, and it only migrates through the east. For example, a study by Ferrari et al. (2018) in the Canada warbler (*Cardellina canadensis*), a species whose breeding range does not extend beyond the Rocky Mountains, found subtle genetic structure present between their sampled populations (North Carolina, New Hampshire, and Northwest Territories), with samples from New Hampshire and the Northwest Territories being more similar than the North Carolina samples. The similarities in the population structure of these species can be explained by a widespread climatic process that shaped the current ranges of North American species: the LGM (~20,000 years ago). Populations had to move into isolated refugia to survive, and after the LGM was over, they rapidly colonized their current ranges (Shafer et al., 2010). The weak genetic structure observed between these two species might be a result of them using the same southern refugium during the LGM, and when recolonizing their current ranges, the Rocky Mountains might have acted as a barrier to prevent further colonization (Ralston et al., 2021). Conversely, species with stronger population structure patterns, such as the bank swallows (*Riparia riparia),* probably had to use different refugia to survive, and it is likely that the populations remained separate during recolonization (Maddegamgoda Lekamlage, 2023). However, it is important to consider that in this study there were not enough sites from the western side of the Rocky Mountains to confidently assess their effect on the population structure of the least flycatcher.

Based on my results and the literature, it is probable that the population structure observed in the least flycatcher occurred more recently than the LGM. Similarly, the genetic variation observed within the eastern group could be an effect of another widespread climatic process: climate change. As it has been observed across the aerial insectivores, steeper declines are occurring in the east and north of North America (Nebel et al., 2010). Nebel et al. (2010) mention that some of the reasons for these declines might be: (1) acid precipitation, which affects aquatic food webs and insects, (2) regional differences in climate responses to atmospheric change, (3) contaminants, such as chemicals used in agriculture and heavy metals and sulfur from industrial activities, and (4) habitat loss due to urbanization (Spiller & Dettmers, 2019). And so, the lower levels of gene flow in the eastern group (when compared to the western genetic group) might be due to birds becoming isolated due to population reductions.

Besides landscape and phylogeography, migration distance also seems to affect population structure. In a comparative study with resident, short-distance, and long-distance migrants, Pegan et al. (2025) found that the migration distance affected the genetic structure of different species. The short-distance migrants, the chestnut-sided warbler (*Setophaga pensylvanica*), and the yellow-bellied flycatcher (*Empidonax flaviventris*), were the group that showed the lowest levels of genetic structure among the three groups compared (Pegan et al., 2025). Both of these species migrate similar distances and locations to the least flycatcher, thereby corroborating the results of my study. Besides demonstrating that migration and habitat play a huge role in shaping the population structure of a species, our study adds to the growing knowledge about aerial insectivores and can be used as a tool to create hypotheses regarding the population structure of species that have not been studied yet.

3.2 Migratory connectivity of the least flycatcher

Our results indicate that the least flycatcher has weak migratory connectivity among its breeding and wintering sites, which could be affecting—reducing—the population structure between the western and eastern genetic groups. However, this depends entirely on where birds fly to during pre-breeding migration, and whether they present adult site fidelity or different levels of natal philopatry, but there are very few studies that study the migratory biology of the least flycatcher (Van Brempt et al., 2025). Studies regarding the natal philopatry of least flycatcher have shown that juvenile birds have a very low return rate (low philopatry), but that could be due to individuals dying during their first migration, or that territorial older birds cause them to disperse to other areas (Weatherhead & Forbes, 1994). Observations from the Beaverhill Bird Observatory regarding nest reuse (Dykstra, 2024) found that at least 5 nests of least flycatcher were reused during two years, which could indicate site fidelity, but only the nests (and not the birds) were monitored.

Studies on the migratory biology of the willow flycatcher (*Empidonax trailii*) found that individuals presented low natal philopatry but high-breeding site fidelity, which is common in passerines Sedgwick (2004). However, they also found that breeding performance and residency explained site fidelity in females (Sedgwick, 2004). If a similar phenomenon occurs with least flycatchers, then the changes in site fidelity due to previous breeding performance could be affecting the migratory connectivity of the species, and in turn, its population structure. However, further studies that focus on site fidelity in least flycatchers are needed to test this.

3.3 Future directions

Although this study marks an important step in understanding the ecological and genetic patterns of the least flycatcher, further research is necessary to expand upon these findings. Disentangling the effects of migratory connectivity and the age of the population requires additional samples, specifically non-breeding birds, and higher resolution markers such as whole genome sequencing. Increasing the number of sampling locations on the western side of the Rocky Mountains would help to evaluate their role as geographical barriers, as the sample size here was too limited to draw definitive conclusions. However, the species range does not extend too far to the west beyond the Rocky Mountains, limiting options for sampling sites. Similarly, increasing the number of samples further north of the breeding range of the least flycatcher would permit the investigation of a third genetically defined group in that region.

Additional research into the age and timing of arrival of migratory birds would improve understanding of migratory connectivity, particularly since juvenile birds are presumed to exert the greatest influence. Last, while the current sampling sites were sufficient for initial investigation of the genetic structure and migratory connectivity, increasing the number of samples and locations in the migratory and wintering regions is crucial to confirm that these patterns are representative at the species level. The samples from migratory birds will provide us with a better understanding of how they utilize flyways to move between breeding and wintering areas, while samples from the wintering grounds will inform us whether birds from different breeding populations are intermixing.

Although the least flycatcher is not a species at risk, our results indicate that declines may have affected population genetics on the eastern side of their distribution. It is necessary to consider how to prevent further declines to avoid losing the potential genetic diversity that the species possesses. Similarly, it appears essential to consider migration routes and wintering grounds in management planning, as migratory connectivity plays a significant role in the gene flow of the species.

3.4 References

Dykstra, E. (2024). Least Flycatcher nest reuse in the Beaverhill Natural Area. Beaverhill Bird Observatory.

Ferrari, B. A., Shamblin, B. M., Chandler, R. B., Tumas, H. R., Hache, S., Reitsma, L., & Nairn, C. J. (2018). Canada warbler (*Cardellina canadensis*): novel molecular markers and a preliminary analysis of genetic diversity and structure. *Avian Conservation and Ecology*, *13*(1).1-8. https://doi.org/10.5751/ace-01176-130108

Maddegamgoda Lekamlage, T. T. (2023). *Landscape Genetics of Two Declining North American Aerial Insectivores* (Publication Number 30690326) [M.Sc., University of Lethbridge (Canada)]. ProQuest Dissertations & Theses Global. Canada -- Alberta, CA. http://cyber.usask.ca/login?url=https://www.proquest.com/dissertations-theses/landscape-genetics-two-declining-north-american/docview/2886452189/se-2?accountid=14739

Nebel, S., A. Mills, J. D. McCracken, and P. D. Taylor. 2010. Declines of aerial insectivores in North America follow a geographic gradient. *Avian Conservation and Ecology - Écologie et conservation des oiseaux* 5(2): 1. http://www.ace-eco.org/vol5/iss2/art1/

Pegan, T. M., Kimmitt, A. A., Benz, B. W., Weeks, B. C., Aubry, Y., Burg, T. M., Hudon, J., Jones, A. W., Kirchman, J. J., Ruegg, K. C., & Winger, B. M. (2025). Long-distance seasonal migration to the tropics promotes genetic diversity but not gene flow in boreal birds. *Nature Ecology & Evolution*, *9*(6), 957-969. https://doi.org/10.1038/s41559-025-02699-3

Ralston, J., Fitzgerald, A. M., Burg, T. M., Starkloff, N. C., Warkentin, I. G., & Kirchman, J. J. (2021). Comparative phylogeographic analysis suggests a shared history among eastern North American boreal forest birds. *Ornithology*, *138*(3), 1-16. https://doi.org/10.1093/ornithology/ukab018

Sedgwick, J. A. (2004). Site fidelity, territory fidelity, and natal philopatry in willow flycatchers (Empidonax traillii) [Article]. *Auk*, *121*(4), 1103-1121. https://doi.org/10.1642/0004-8038(2004)121[1103:Sftfan]2.0.Co;2

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*(21), 4589-4621. https://doi.org/10.1111/j.1365-294x.2010.04828.x

Spiller, K. J., & Dettmers, R. (2019). Evidence for multiple drivers of aerial insectivore declines in North America. *Condor*, *121*(2), 13, duz010. https://doi.org/10.1093/condor/duz010

Van Brempt, M., Holroyd, G. L., & Hvenegaard, G. T. (2025). Phenological changes in sex- and age-differential autumn migration of the least flycatcher (*Empidonax minimus*). *The Wilson Journal of Ornithology*, *137*(1), 98-111. https://doi.org/10.1080/15594491.2024.2444030

Weatherhead, P. J., & Forbes, M. R. L. (1994). Natal philopatry in passerine birds - genetic or ecological influences. *Behavioral Ecology*, *5*(4), 426-433. https://doi.org/10.1093/beheco/5.4.426

Appendices

Appendix 1. Individual sample information for the 63 breeding and 29 non-breeding individuals analyzed. The season column indicates whether the individuals was collected during the breeding (BR) season or during the non-breeding (NBR) season.

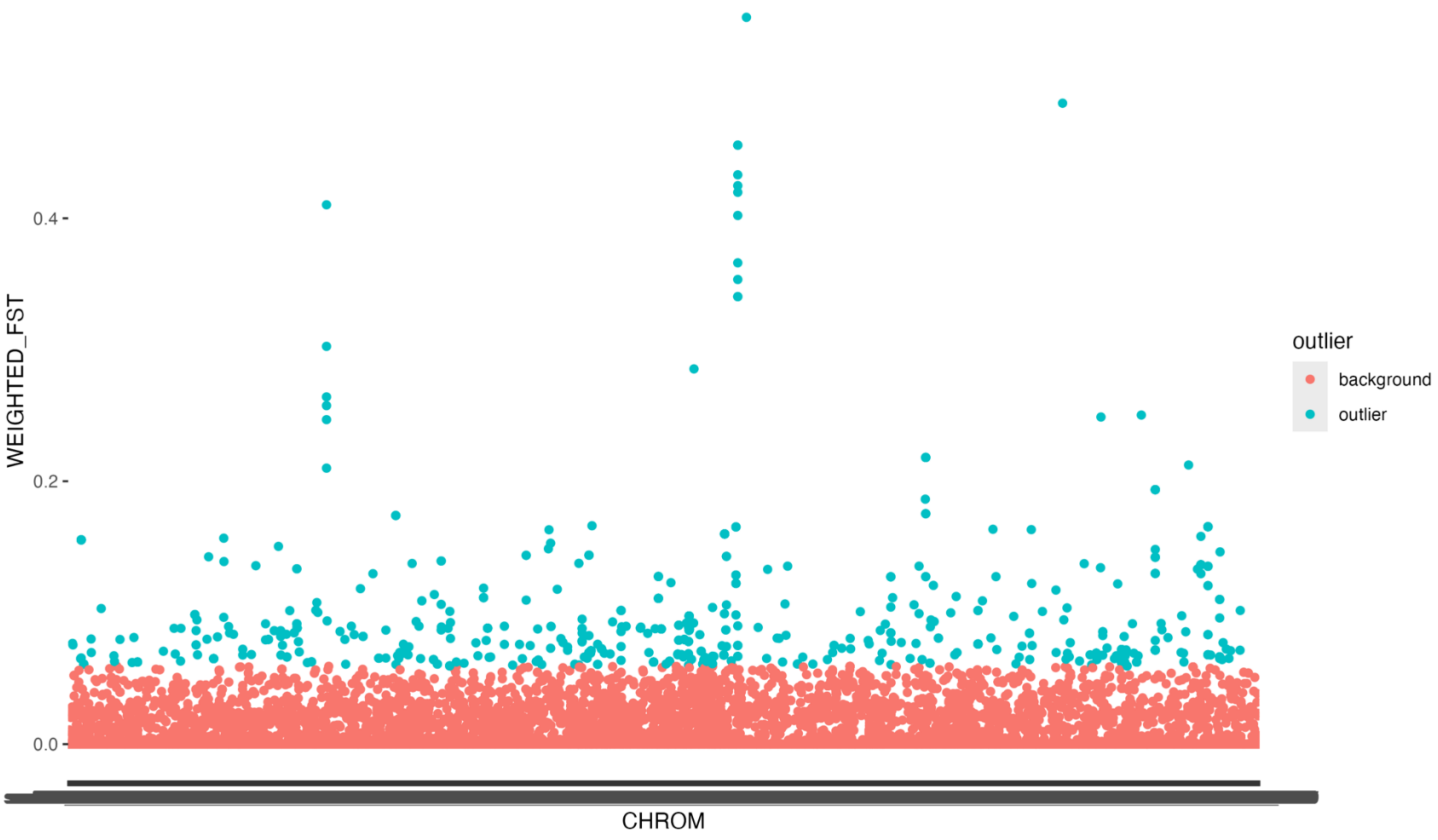
Location	Sample ID	Band/Museum ID	Population	Season	Latitude	Longitude
Albert Creek, YT	YT 003	2490-46111	YT	NBR	60.06181	-128.9165
Albert Creek, YT	YT 004	2490-46103	YT	NBR	60.06181	-128.9165
Albert Creek, YT	YT 005	2490-46100	YT	NBR	60.06181	-128.9165
Albert Creek, YT	YT 006	2490-46206	YT	NBR	60.06181	-128.9165
Albert Creek, YT	YT 009	2490-462900	YT	NBR	60.06181	-128.9165
Revelstoke, BC	SEBC 19	2710-78701	SEBC	BR	50.98145	-118.1995
Revelstoke, BC	SEBC 20	2710-78703	SEBC	BR	51.00502	-118.2278
Revelstoke, BC	SEBC 21	2710-78704	SEBC	BR	51.00502	-118.2278
Revelstoke, BC	SEBC 22	2710-78705	SEBC	BR		
Fort McMurray, AB	NAB 001	2630-89278	NAB	NBR	56.72144	-111.2845
Fort McMurray, AB	NAB 003	2630-89297	NAB	NBR	56.72144	-111.2845
Jasper, AB	CAB 001	LEFL 2	CAB	BR	52.91361	-118.0971
Jasper, AB	CAB 002	LEFL 3	CAB	BR	52.91325	-118.0951
Waterton Lakes, AB	SAB 001	LEFL 1	SAB	BR	49.07961	-113.85931
Waterton Lakes, AB	SAB 002	LEFL 2	SAB	BR	49.07961	-113.85931
Waterton Lakes, AB	SAB 003	LEFL 3	SAB	BR	49.07961	-113.85931
Waterton Lakes, AB	SAB 004	LEFL 4	SAB	BR	49.07961	-113.95931
Waterton Lakes, AB	SAB 005	LEFL 5	SAB	BR	52.91361	-118.09706
Waterton Lakes, AB	SAB 006	LEFL 6	SAB	BR	52.91361	-118.09506
Beauvais Lake, AB	SAB 009	2710-78671	SAB	BR	49.41525	-114.1033
Lethbridge, AB	SEAB L001	2710-78685	SEAB	BR	49.10203	-112.8619
Cypress Hills, AB	SEAB 001	LEFL002	SEAB	BR	49.6446	-110.0319
Cypress Hills, AB	SEAB 003	LEFL003	SEAB	BR	49.6446	-110.0319
Cypress Hills, AB	SEAB 008	LEFL004	SEAB	BR	49.65731	-110.1021

Cypress Hills, AB	SEAB 009	LEFL 1	SEAB	BR		
Cypress Hills, AB	SEAB 010	2710-78648	SEAB	BR	49.64561	-110.0302
Cypress Hills, AB	SEAB 011	2710-78649	SEAB	BR	49.64561	-110.0302
Cypress Hills, AB	SEAB 012	2710-78653	SEAB	BR	49.6631	-110.3039
Cypress Hills, AB	SEAB 013	2710-78687	SEAB	BR	49.66184	-110.3129
Cypress Hills, AB	SEAB 014	2710-78688	SEAB	BR	49.66184	-110.3129
Cypress Hills, AB	SEAB 015	2710-78689	SEAB	BR	49.66184	-110.3129
Cypress Hills, AB	SEAB 016	2710-78690	SEAB	BR	49.66184	-110.3129
Cypress Hills, AB	SEAB 017	2710-78691	SEAB	BR	49.66184	-110.3129
Dundurn, SK	SK 001	2080-29593	SK	BR		
Dundurn, SK	SK 002	2080-39589	SK	BR		
Dundurn, SK	SK 003	2080-29585	SK	BR		
Dundurn, SK	SK 004	2080-29584	SK	BR		
Dundurn, SK	SK 005	2060-41515	SK	BR		
Prince Albert National Park, SK	SK 007	2060-41566	SK	BR	53.851278	-106.085553
Saskatoon, SK	SK 009	24B-149	SK	BR	53.890932	-106.117999
Saskatoon, SK	SK 010	24B-150	SK	BR	52.031828	-106.703016
Saskatoon, SK	SK 011	24B-151	SK	BR	52.031828	-106.703016
Regina, SK	SK 012	24B-152	SK	BR	50.658482	-104.293988
Missoula MT	MT M01	LEFL 1	MT	BR	46.7923	-113.8837
Butte, MT	MT W01	LEFL 3	MT	BR	46.25953	-112.4781
Bottineau, ND	ND S01	LEFL 1	ND	BR	48.94575	-100.5078
Towner, ND	ND T02	ND17	ND	BR	48.47197	-100.3367
Somerset County, ME	ME 001	z0-15398	ME	BR	45.51451	-70.11713
Ottawa, ON	ON 6172	NBCC1016172	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6185	NBCC1016185	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6187	NBCC1016187	ON	BR	45.41278	-75.68867

Ottawa, ON	ON 6188	NBCC1016188	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6189	NBCC1016189	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6195	NBCC1016195	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6196	NBCC1016196	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 6720	NBCC1016720	ON	BR	45.41278	-75.68867
Ottawa, ON	ON 9743	NBCC1019743	ON	BR	45.41278	-75.68867
Nassau County, NY	NY 005 M	CU53016	NY	NBR		
Tompkins County, NY	NY 006 M	CU44354	NY	NBR		
Franklin County, NY	NY 007	zo-11147	NY	BR		
Greene County, NY	NY 009	zo-19312	NY	BR		
Greene County, NY	NY 010	zo-19313	NY	BR		
Greene County, NY	NY 011	zo-19626	NY	BR		
Saratoga County, NY	NY 012	zo-19673	NY	BR		
Highland, West Virginia	WV 001	USNM 587424	WV	BR		
Highland, West Virginia	WV 002	USNM 587425	WV	BR		
Highland, West Virginia	WV 003	USNM 587426	WV	BR		
Grayson, West Virginia	WV 004	USNM 601599	WV	BR		
Grayson, West Virginia	WV 005	USNM 601600	WV	BR		
Grayson, West Virginia	WV 006	USNM 601601	WV	BR		
Highland, West Virginia	WV 007	USNM 634223	WV	BR		
Highland, West Virginia	WV 008	USNM 634224	WV	BR		
Othon P. Blanco, Quintana Roo, Mexico	MX 015	ECO-CH-A-0343	MX	NBR	19.62039	-89.00489
Othon P. Blanco, Quintana Roo, Mexico	MX 016	ECO-CH-A-0347	MX	NBR	19.62039	-89.00489
Othon P. Blanco, Quintana Roo, Mexico	MX 017	ECO-CH-A-0504	MX	NBR	19.62039	-89.00489
Bacalar, Quintana Roo, Mexico	MX 019	ECO-CH-A-0753	MX	NBR	18.6845	-88.38525

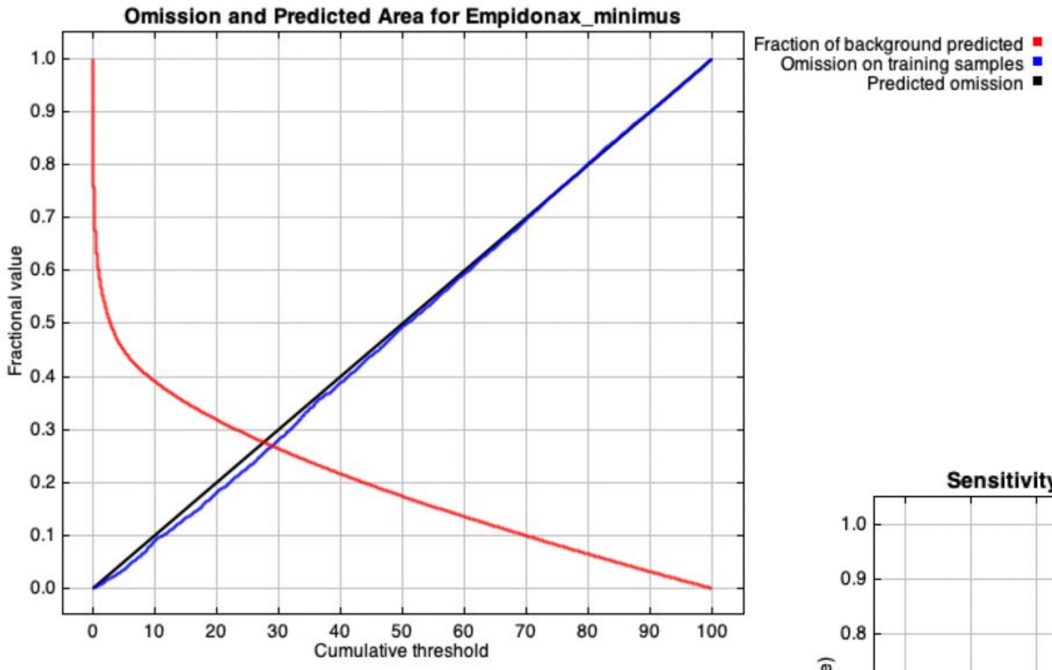
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 023	ECO-CH-A-0990	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 024	ECO-CH-A-1116	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 025	ECO-CH-A-1139	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 026	ECO-CH-A-1152	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 027	ECO-CH-A-1404	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 028	ECO-CH-A-1417	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 029	ECO-CH-A-1492	MX	NBR	19.57569	-88.05946
Puerto Felipe Carrillo, Quintana Roo, Mexico	MX 030	ECO-CH-A-1791	MX	NBR	19.57569	-88.05946
El Cuyo, Yucatan, Mexico	MX 031	12	MX	NBR	21.51785	-87.6766
Dzilam, Yucatan, Mexico	MX 032	181	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 033	192	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 034	194	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 035	198	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 036	199	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 037	216	MX	NBR	21.39327	-88.89426
Dzilam, Yucatan, Mexico	MX 038	217	MX	NBR	21.39327	-88.89426

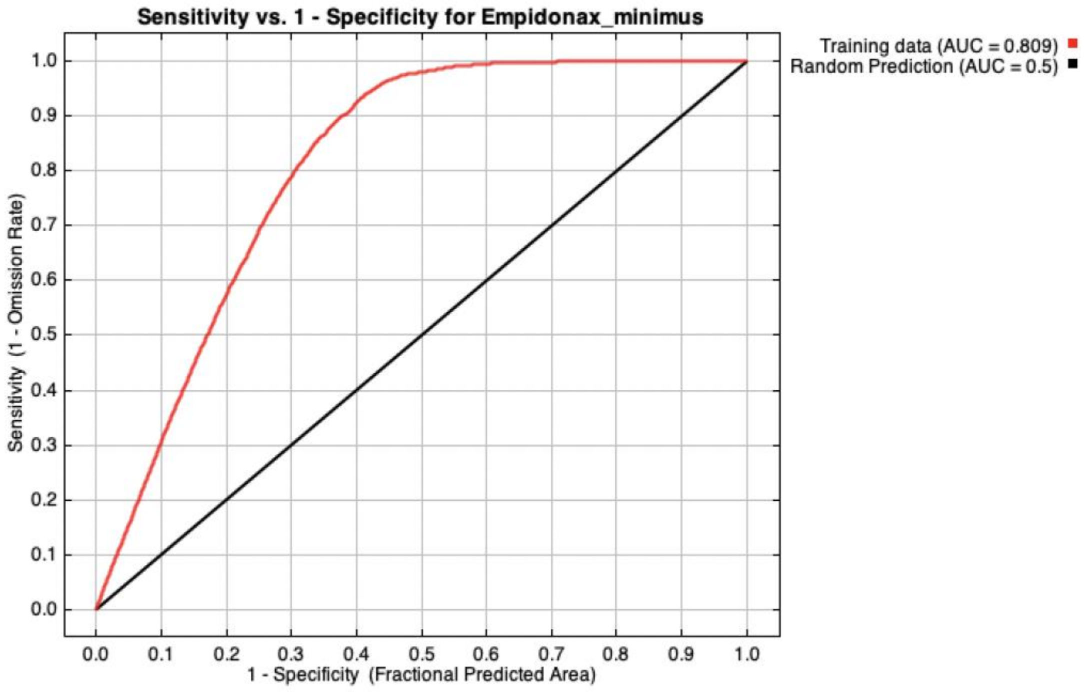
Appendix 2. Manhattan plot used to identify the outlier SNPs in the data. A cutoff of 0.1 was used, and SNPs that were selected are shown in blue.



Appendix 3. Estimates of relative contributions of the environmental variables to the Maxent model. Average/prediction curves and AUC curves are also included.

Variable	Percent contribution	Permutation importance
Mean temperature of warmest quarter	43.1	41.5
Mean temperature of coldest quarter	38.6	35.4
Temperature annual range	5.2	5
Mean diurnal range	3.9	5.9
Precipitation of driest quarter	3.2	0.2
Precipitation of wettest quarter	2.7	4
Precipitation of warmest quarter	1.6	2
Precipitation seasonality	0.9	2.3
Mean temperature of wettest quarter	0.8	1.8
Precipitation of coldest quarter	0.1	1.8


Omission and Predicted Area for Empidonax_minimus
Fraction of background predicted 🟥
Omission on training samples 🟦
Predicted omission ■
Fractional value
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
70
80
90
100
Cumulative threshold


Sensitivity vs. 1 - Specificity for Empidonax_minimus
Training data (AUC = 0.809) 🟥
Random Prediction (AUC = 0.5) ■
Sensitivity (1 - Omission Rate)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1 - Specificity (Fractional Predicted Area)