Genomic mechanisms underlying migration timing and population differentiation across North America in a long-distance migratory songbird

by

Evelien de Greef

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

Abstract

Understanding adaptive potential in migratory animals is important for anticipating population responses to rapid environmental changes. Because shifts in spring phenology can disrupt synchrony between individuals' movements and seasonal resources, it is crucial to examine mechanisms underlying migration timing and population genomics in vulnerable migratory species. In this thesis, I used the purple martin (*Progne subis*), a long-distance migratory songbird, as a model species to address the following objectives: 1) identify genomic regions underlying migration timing and corresponding to spring phenology, and 2) examine population structure across four North American breeding colonies and relate patterns of differentiation to natal dispersal. Using next-generation sequencing, I assembled a new draft reference genome and whole-genome data for 93 individuals. I then integrated genomic variation, direct migration timing data, first bloom dates, and continent-wide band encounter records. I recorded a high heritability in migration timing, demonstrating a strong genetic basis in this trait. I also identified loci associated with spring and fall migration near multiple genes, such as *MEF2D* and *CRY1*, which could play an important role in photoreception and circadian rhythms. Additionally, I found that timing of first bloom was associated with a high proportion of genomic variance and loci of polygenic effect. Although overall population differentiation was low, I documented isolation-by-distance and greatest distinction between the most northern (Alberta) and more southern colonies. Patterns of differentiation were consistent with observations of greater longitudinal than latitudinal movement in natal dispersal. Overall, my results advance our understanding of migration genomics by demonstrating a novel connection between genomic variation and direct migration timing phenotypes in a long-distance migratory songbird, while the new reference genome provides a foundation for purple martin genomic

research. The population structure and latitudinal patterns of differentiation I identified can inform conservation management of this species through delineation of distinct groups between northern and more southern breeding populations. Together, the strong genomic influences on migration timing and spatial limitations on latitudinal gene flow in the purple martin can be used as a framework to describe potential constraints on adaptability in changing climates.

Acknowledgements

I would like to thank my advisors Kevin Fraser and Kira Delmore for their mentorship throughout this project, and my committee members Colin Garroway and Debbie Kelly for their guidance and support. I am tremendously grateful for my partner Matthew Thorstensen for his unwavering intellectual guidance and emotional support at every step of the way.

I have been fortunate to be a part of two lab communities, receiving feedback and encouragement from Alicia Korpach, Saeedeh Bani Assadi, Leanne Neufeld, Hannah Justen, Wesley Brashear, Abby Kimmitt, and numerous undergraduate students. I also want to thank the staff, faculty, and graduate students in the Biological Sciences Department for their help with navigating the program and keeping up morale. Thank you to Kenneth Jeffries for generously sharing his laboratory space and equipment for sample extractions, as well as providing support throughout the program. I would also like to thank Chloe Schmidt for my initial training in extraction methods, Joey Wan for advice on samples, and Sharon Mullen for dependable assistance with finding shipment materials.

A big thank you to all the collaborators across North America who provided the samples used in this study, including: Kelly Applegate, John Barrow, Leslie Burgess, Louise Chambers, Ted Cheskey, Paul Clifton, Bruce Cousens, Bill Dietrich, Jason Fischer, Glen Hvenegaard, Pat Kramer, Paul Mammenga, Nanette Mickle, Michael North, Myrna Pearman, Jim Ray, Alisha Ritchie, Anne Savage, Tim Shaheen, Joe Siegrist, Cassandra Silverio, Bridget Stutchbury, John Tautin, and the UC Davis Museum of Wildlife and Fish Biology. I also thank the North American Bird Banding Program and the Bird Banding Office for providing encounter banding records from many contributors across the continent.

Thank you to the scientists and organizations involved in sequencing and bioinformatics resources: Alexander Suh, National Genomics Infrastructure in Sweden, the Science for Life Laboratory, Texas A&M AgriLife Genomics and Bioinformatics Service, Texas A&M High Performance Research Computing, and Compute Canada.

Thank you to all the agencies that funded this project: The University of Manitoba, Texas A&M University, Natural Sciences and Engineering Research Council, Research Manitoba, the Society of Canadian Ornithologists, and Sigma Xi.

I would also like thank my family and friends for their support and encouragement throughout the program, which helped make this thesis possible.

Table of Contents

List of Tables

List of Figures

Thesis format

This thesis is written in manuscript format. The first chapter introduces background for the thesis and outlines research questions. Chapters two (migration timing genomics) and three (population genomics) are presented as draft manuscripts, each containing their own abstract, introduction, methods, results, conclusion, references, tables, and figures. The fourth and final chapter concludes the thesis with a summary of findings, their significance, and directions for future research. The label for each table and figure includes its respective chapter number or appendix character.

Chapter 1: Thesis introduction

Climate change is a global threat affecting wildlife survival and diversity, with accelerating irreversible impacts worldwide. Warming global temperatures can disrupt evolved synchronies of birds' annual cycles with seasonal resources and critical habitats (Carey 2009; Galbraith et al. 2002). Though some birds are shifting the timing of seasonal activities (e.g. Helm et al. 2019; Jonzén et al. 2006), ranges (e.g. Hitch & Leberg 2007), or diets (e.g. Montevecchi & Myers 1997), these changes could be insufficient to keep up with the rapid rate of changing climates (Visser 2019). Over time, phenological mismatches can decrease long-term population viability, creating far-reaching consequences through interconnected species and ecosystems (Visser & Gienapp 2019). Many birds are challenged by the unprecedented pace of adjustments required to survive changing climates; these adjustments can be both enabled and limited by birds' evolutionary backgrounds. With massive declines in bird abundances within the last century (Rosenberg 2019), the need to better understand risks and potential adaptability in birds to global change is more critical than ever.

Migration, or large-scale movements of individuals to maximize fitness in seasonal environments, is an essential adaptation that has evolved in thousands of bird species. For example, the longest distance avian traveller is the arctic tern (*Sterna paradisaea*), with an extraordinary 80,000 km annual migration between breeding and wintering grounds (Egevang et al. 2010). When travelling great distances, long-distance migrants rely more on adaptive endogenous cues controlling circannual rhythms, orientation, and physiological changes (Alerstam et al. 2003), compared to short-distance migrants that may rely more upon environmental cues such as temperature and precipitation to adjust timing of migratory behaviours (Rubolini et al. 2007). Studies have shown that long-distance migrants are more

vulnerable to seasonal shifts from climate change (Both et al. 2010), therefore it is important to understand the endogenous cues driving their migrations to identify possible adaptive constraints.

Well-characterized endogenous mechanisms in migration include the circadian and circannual clocks, which control daily and yearly rhythms (Åkesson et al. 2017; Helm et al. 2013). Many studies have focused on the *CLOCK* gene, a candidate gene that has been shown to influence traits correlated with photoperiodic responses such as breeding and moulting phenology in barn swallows (*Hirundo rustica,* Caprioli et al. 2012; Saino et al. 2013). Another known candidate gene linked to circadian rhythms is *ADCYAP1,* associated with migratory restlessness and morphological influences on migration timing in animals such as blackcaps (*Sylvia atricapilla,* Mettler et al. 2015; Mueller et al. 2011). While candidate genes help explain some of the important mechanisms associated with migratory traits, some studies found contrasting results detecting no association between candidate genes and migration timing (Parody-Merino et al. 2019; Lugo Ramos et al. 2017), or inconsistent effect of genes on timing within the same genus (Peterson et al. 2013). The variability across species and limitation of candidate gene studies to small portions of the genome make it difficult to clearly define and quantify the genetics underlying migratory traits.

Phenotypic variation in migration is likely produced by interactions between many genes (Liedvogel & Lundberg 2014), highlighting the importance of identifying novel genes or genomic regions influencing complex traits. Genomics can address this research gap through whole-genome sequencing data containing information on genetic variation at millions of loci distributed throughout an organism's DNA, allowing us to dissect the genomic architecture of important traits. Here, genomic architecture refers to the identity, number, interactions, and

location of genes underlying a phenotypic trait (Hansen 2006). One of the first studies using large-scale sequencing analyses in bird migration was done on populations of willow warblers (*Phylloscopus trochilus*), and identified numerous genes involved in migratory orientation (Lundberg et al. 2013). Genome-wide analyses have identified genes missed in candidate gene studies influencing avian migration (Delmore et al. 2016; Franchini et al 2017; Lundberg et al. 2017; Toews et al. 2019), expanding the search for migratory genes in more organisms to better understand the genetic mechanisms driving migratory phenomena.

In addition, using genomics in conservation research can reveal evolutionary processes such as speciation and adaptation, which can help researchers predict species' abilities to survive in changing environments (Funk et al. 2012; McMahon et al. 2014). Genetic relatedness among individuals and groups can describe population structure and gene flow, providing valuable information for defining conservation units (Funk et al. 2012). For example, low gene flow between differentiated populations can inform population-specific management strategies to maintain genetic diversity, such as in the blue duck (*Hymenolaimus malacorhynchos,* Grosser et al. 2017) and Galápagos petrels (*Pterodroma phaeopygia,* Friesen et al. 2006). Genetic diversity measured between and within populations can have important implications, where low diversity may reduce a species' ability to adapt to changing environments, possibly leading to subpopulation extinction (Wisely et al. 2002). In migratory songbirds specifically, the effects of low genetic diversity are largely unknown. However, these effects may have dire consequences, such as loss of adaptive potential, which for example was predicted in Kirtland's warblers (*Dendroica kirtlandii,* Wilson et al. 2012). Evolution underlies phenotypes observed in modern animals and understanding an organism's evolutionary history is fundamental to anticipating its adaptive capabilities.

As a long-distance Nearctic-Neotropical migrant, the purple martin (*Progne subis*) is an ideal species to study avian genomics due to its large distribution in North American breeding grounds and close association with humans. Because martins rely on artificial housing structures that are easily accessible (Brown & Tarof 2020), researchers can examine nesting behaviours, study breeding biology, and capture numerous individuals to track their movements. By deploying bands and tracking devices such as light-level geolocators, automated radio telemetry (e.g. the Motus Wildlife Tracking system), and archival GPS units, we can gain detailed information about their migratory journeys. Studies using direct-tracking methods have shown that martins travel over 7,000 km from their breeding colonies to roost together in South America in overlapping wintering home ranges (Fraser et al. 2012; Fraser et al. 2017). While they assemble in large mixed flocks in the non-breeding season, adults tend to have a high site fidelity where they return to their previous breeding site (Brown & Tarof 2020; Stutchbury et al. 2009). In its continent-wide breeding distribution, the purple martin exhibits considerable latitudinal variation in migratory timing, providing an incredible system to study this trait. For example, southern-breeding individuals arrive as early as January, while their northern counterparts can arrive as late as June (Neufeld et al. in prep). These southern and northern extremes are interesting because they reflect a latitudinal gradient in phenotypic variation, which may be associated with genetic markers (Johnsen et al. 2007). Like many other aerial insectivores, purple martins have been declining for the past 50 years (Nebel et al. 2010; Sauer et al. 2017), and studies assessing population genetics and architecture of adaptive traits are an important component for conservation action (Supple & Shapiro 2018).

Previous research on purple martin genetics has used mitochondrial DNA (mtDNA) to assess population differentiation (Baker et al. 2008) and characterized microsatellite markers

(Stanley et al. 2011). Baker et al. (2008) examined genetic patterns primarily in the western *P.s. arboricola* populations, showing high genetic diversity maintained by immigrating birds and patterns of low genetic differentiation within subspecies. While mtDNA and microsatellite data are useful tools for studying evolution and population genetics, mtDNA only reflect maternal lineages and both are limited to small portions of the genome. Thus, prior studies of genetic diversity among martins were limited in scope, and aspects of purple martin population genetics remain unknown, such as population structure in the nominate eastern subspecies *P.s subis* and genomic mechanisms influencing its migration patterns.

In my thesis, I assembled a *de novo* purple martin reference genome and generated whole-genome data for individuals breeding across North America. I used these resources to address my overall objective of investigating how migration timing phenotypes can be explained by genomic variation to improve our understanding of adaptive potential in the context of climate change. This research represents the first time genomics and direct migratory timing phenotypes are integrated to examine genomic architecture underlying long-distance migration timing in a migratory bird (Chapter 2). In the second chapter, I also examined purple martin genotypes associated with spring green-up phenology in North America using a continent-wide first bloom dataset. Next, to determine purple martin population structure, I evaluated four widely-distributed breeding populations of *P.s. subis,* and related patterns of differentiation with natal dispersal records (Chapter 3). Results from Chapter 3 provides a background for this species' biogeographic history with possible applications to conservation management. In the concluding chapter, I incorporated thesis results and discussed implications on purple martin adaptability to climate change.

References

Åkesson, S., Ilieva, M., Karagicheva, J., Rakhimberdiev, E., Tomotani, B., and Helm, B. (2017). Timing avian long-distance migration: From internal clock mechanisms to global flights. Phil. Trans. R. Soc. B., 372(1734): 20160252. doi:10.1098/rstb.2016.0252

Alerstam, T., Hedenstrom, A., and Akesson, S. (2003). Long-distance migration: evolution and determinants. Oikos, 103(2): 247-260. doi:10.1034/j.1600-0706.2003.12559.x

Baker, A.J., Greenslade, A.D., Darling, L.M., and Finlay, J.C. (2008). High genetic diversity in the blue-listed British Columbia population of the purple martin maintained by multiple sources of immigrants. Conservation Genetics, 9: 495–505. doi:10.1007/s10592-007-9358-3

Both, C., Van Turnhout, C.A.M., Bijlsma, R.G., Siepel, H., Van Strien, A.J., and Foppen, R.P.B. (2010). Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. Proc. R. Soc. B., 277(1685): 1259-1266. doi:10.1098/rspb.2009.1525

Brown, C.R. and Tarof, S. (2020). Purple Martin (*Progne subis*), version 1.0. In Birds of the World (A.F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. doi:10.2173/bow.purmar.01

Caprioli, M., Ambrosini, R., Boncoraglio, G., Gatti, E., Romano, A., Romano, M., Rubolini, D., Gianfranceschi, L., and Saino, N. (2012). Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. PLoS ONE, 7(4): e35140. doi:10.1371/journal.pone.0035140

Carey, C. (2009). The impacts of climate change on the annual cycles of birds. Phil. Trans. R. Soc. B., 364: 3321–3330. doi:10.1098/rstb.2009.0182

Delmore, K.E., Toews, D.P.L., Germain, R.R., Owens, G.L., and Irwin, D.E. (2016). The Genetics of Seasonal Migration and Plumage Color. Current Biology, 26(16): 2167–2173. doi:10.1016/j.cub.2016.06.015

Egevang, C., Stenhouse, I.J., Phillips, R.A., Petersen, A., Fox, J.W., and Silk, J.R.D. (2010). Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. PNAS, 107(5): 2078–2081. doi:10.1073/pnas.0909493107

Franchini, P., Irisarri, I., Fudickar, A., Schmidt, A., Meyer, A. Wikelski, M., and Partecke, J. (2017). Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. Moleculer Ecology, 26(12): 3204-2316. doi:10.1111/mec.14108

Fraser, K.C., Stutchbury, B.J.M., Silverio, C., Kramer, P.M., Barrow, J., Newstead, D., Mickle, N., Cousens, B.F., Lee, J.C., Morrison, D.M., Shaheen, T., Mammenga, P., Applegate, K., and Tautin, J. (2012). Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. Proc. R. Soc. B., 279(1749): 4901–4906. doi:10.1098/rspb.2012.2207

Fraser, K.C., Shave, A., Savage, A., Ritchie, A., Bell, K., Siegrist, J., Applegate, K., and Pearman, M. (2017). Determining fine-scale migratory connectivity and habitat selection for a migratory songbird by using new GPS technology. Journal of Avian Biology, 48(3): 001-007. doi:10.1111/jav.01091

Friesen, V.L., González, J.A., and Cruz-Delgado, F. (2006). Population genetic structure and conservation of the Galápagos petrel (*Pterodroma phaeopygia*). Conservation Genetics, 7: 105-115. doi:10.1007/s10592-005-8704-6

Funk, C.W., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. (2012). Harnessing genomics for delineating conservation units. Trends Ecol. Evol., 27(9): 489-496. doi:10.1016/j.tree.2012.05.012

Galbraith, A. H., Jones, R., Park, R., Clough, J., Herrod-Julius, H., Harrington, B., and Page, G. (2016). Global Climate Change and Sea Level Rise: Potential Losses of Intertidal Habitat for Shorebirds. Waterbirds, 25(2): 173-183.

Grosser, S., Abdelkrim, J., Wing, J., Robertson, B.C., and Gemmell, N.J. (2017). Strong isolation by distance argues for separate population management of endangered blue duck (*Hymenolaimus malacorhynchos*). Conservation Genetics, 18: 327–341. doi:10.1007/s10592-016-0908-4

Hansen, T.F. (2006). The Evolution of Genetic Architecture. Annu. Rev. Ecol. Evol. Syst., 37(1): 123–157. doi:10.1146/annurev.ecolsys.37.091305.110224

Helm, B., Ben-Shlomo, R., Sheriff, M.J., Hut, R.A., Foster, R., Barnes, B.M., and Dominoni, D. (2013). Annual rhythms that underlie phenology: biological time-keeping meets environmental change. Proc. R. Soc. B., 280(1765): 20130016. doi:10.1098/rspb.2013.0016

Helm, B., Van Doren, B.M., Hoffman, D., and Hoffman, U. (2019). Evolutionary Response to Climate Change in Migratory Pied Flycatchers. Current Biology, 29(21): 3714-3719. doi:10.1016/j.cub.2019.08.072

Hitch, A.T., and Leberg, P.L. (2007). Breeding distributions of North American bird species moving north as a result of climate change. Conservation Biology, 21(2): 534–539. doi:10.1111/j.1523-1739.2006.00609.x

Johnsen, A., Fidler, A.E., Kuhn, S., Carter, K.L., Hoffman, A., Barr, I.R., Biard, C., Charmantier, A., Eens, M., Korsten, P., Sitari, H., Tomiuk, J., and Kempenaers, B. (2007). Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. Molecular Ecology, 16(22): 4867–4880. doi:10.1111/j.1365-294X.2007.03552.x

Jonzén, N., Lindén, A., Ergon, T., Knudsen, E., Vik, J.O., Rubolini, D., Piacentini, D., Brinch, C., Spina, F., Karlsson, L., Stervander, M., Andersson, A., Waldenström, J., Lehikoinen, A., Edvardsen, E., Solvang, R., and Stenseth, N.C. (2006). Rapid advance of spring arrival dates in long-distance migratory birds. Science, 312: 1959–1961. doi:10.1126/science.1126119

Liedvogel, M., and Lundberg, M. (2014). The genetics of animal movement and migration syndromes. Animal Movement Across Scales. Oxford University Press.

Lugo Ramos, J.S., Delmore, K.E., and Liedvogel, M. (2017). Candidate genes for migration do not distinguish migratory and non-migratory birds. J. Comp. Physiol. A., 203: 383–397. doi:10.1007/s00359-017-1184-6

Lundberg, M., Boss, J., Canbäck, B., Liedvogel, M., Larson, K.W., Grahn, M., Åkesson, S., Bensch, S., and Wright, A. (2013). Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler *Phylloscopus trochilus.* BMC Genomics, 14(330). doi:10.1186/1471-2164-14-330

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., and Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evolution Letters, 1(3): 155–168. doi:10.1002/evl3.15

McMahon, B.J., Teeling, E.C., and Höglund, J. (2014). How and why should we implement genomics into conservation? Evolutionary Applications, 7(9): 999–1007. doi:10.1111/eva.12193

Mettler, R., Segelbacher, G., and Schaefer, M.H. (2015). Interactions between a candidate gene for migration (*ADCYAP1*), morphology and sex predict spring arrival in blackcap populations. PLoS ONE, 10(12): e0144587. doi:10.1371/journal.pone.0144587

Montevecchi, W.A., and Myers, R.A. (1997). Centurial and decadal oceanographic influences on changes in northern gannet populations and diets in the north-west Atlantic: Implications for climate change. ICES Journal of Marine Science, 54: 608–614. doi:10.1006/jmsc.1997.0265

Mueller, J.C., Pulido, F., and Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. Proc. R. Soc. B., 278: 2848–2856. doi:10.1098/rspb.2010.2567

Nebel, S., Mills, A., McCracken, J.D., and Taylor, P.D. (2010). Declines of Aerial Insectivores in North American Follow a Geographic Gradient. Avian Conservation and Ecology, 5(2). doi:10.5751/ACE-00391-050201

Neufeld, L.R., Muthukumarana, S., Fischer, J., Ray, J., Siegrist, J., and Fraser, K.C. (2020). Breeding latitude predicts the timing of nest initiation and migration timing around the annual calendar among populations of a declining, migratory songbird purple martin (*Progne subsis*). Unpublished manuscript.

Parody-Merino, Á.M., Battley, P.F., Conklin, J.R., and Fidler, A.E. (2019). No evidence for an association between Clock gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*). Oecologia, 191: 843–859. doi:10.1007/s00442-019-04524-8

Peterson, M.P., Abolins-Abols, M., Atwell, J.W., Rice, R.J., Milá, B., and Ketterson, E.D. (2013). Variation in candidate genes *CLOCK* and *ADCYAP1* does not consistently predict differences in migratory behavior in the songbird genus *Junco*. F1000Research, 2(115). doi:10.12688/f1000research.2-115.v1

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., and Marra, P.P. (2019). Decline of the North American avifauna. Science, 366: 120–124. doi:10.1126/science.aaw1313

Rubolini, D., Møller, A.P., Rainio, K., and Lehikoinen, E. (2007). Intraspecific consistency and geographic variability in temporal trends of spring migration phenology among European bird species. Climate Research, 35: 135–146. doi:10.3354/cr00720

Saino, N., Romano, M., Caprioli, M., Fasola, M., Lardelli, R., Micheloni, P., Scandolara, C., Rubolini, D., and Gianfranceschi, L. (2013). Timing of molt of barn swallows is delayed in a rare *Clock* genotype. PeerJ, 1: e17. doi:10.7717/peerj.17

Sauer, J.R., Pardieck, K.L., Ziolkowski, D.J.Jr., Smith, A.C., Hudson, M.R., Rodriguez, V., Berlanga, H., Niven, D.K., and Link, W.A. (2017). The first 50 years of the North American Breeding Bird Survey. The Condor, 119(3): 576-593. doi:10.1650/CONDOR-17-83.1

Stanley, A.A., Graham, A.M., Wagner, R.H., Morton, E.S., and Schug, M.D. (2011). Characterization of 10 polymorphic microsatellite markers for the purple Characterization of 10 polymorphic microsatellite markers for the purple martin, *Progne subis*, Conservation Genetics Resources, 3: 577-579. doi:10.1007/s12686-011-9408-5

Supple, M.A., and Shapiro, B. (2018). Conservation of biodiversity in the genomics era. Genome Biology 19: 131. doi:10.1186/s13059-018-1520-3

Stutchbury, B.J.M., Hill, J.R., Kramer, P.M., Rush, S.A., and Tarof, S.A. (2009). Sex and age-specific annual survival in a neotropical migratory songbird, the purple martin (Progne subis). The Auk, 126(2): 278-287. doi:10.1525/auk.2009.08038

Toews, D.P.L., Taylor, S.A., Streby, H.M., Kramer, G.R., and Lovette, I.J. (2019). Selection on VPS13A linked to migration in a songbird. PNAS, 116(37): 18272–18274. doi:10.1073/pnas.1909186116

Visser, M.E. (2019). Evolution: Adapting to a Warming World. Current Biology, 29(22): R1189-R1191. doi:10.1016/j.cub.2019.09.062

Visser, M.E. and Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. Nature Ecology and Evolution, 3: 879-885. doi:10.1038/s41559-019-0880-8

Wilson, A.S., Marra, P.P., and Fleischer, R.C. (2012). Temporal patterns of genetic diversity in Kirtland' s warblers (*Dendroica kirtlandii*), the rarest songbird in North America. BMC Ecology, 12(8) doi:10.1186/1472-6785-12-8

Wisely, S.M., Buskirk, S.W., Fleming, M.A., Mcdonald, D.B., and Ostrander, E.A. (2002). Genetic Diversity and Fitness in Black-Footed Ferrets Before and During a Bottleneck. The Journal of Heredity, 93(4): 231–237.

Chapter 2: Genomic variation and regions associated with migration timing in a long-distance migratory songbird

Abstract

The impact of climate change on spring phenology requires adaptation across a wide range of organisms. Long-distance migratory birds are especially vulnerable to changes in spring phenology, as their migration timing may rely predominantly on endogenous mechanisms. However, the genomic architecture of migration timing is largely unknown, thus limiting our ability to understand the internal cues controlling migration. In this study, I investigated the influence of genomic variation and regions associated with migration timing in a long-distance migratory songbird (purple martin, *Progne subis*). I assembled a *de novo* reference genome and integrated whole-genome data with timing phenotypes derived from individuals' direct annual migratory tracks and with the phenology of spring green-up (using first bloom dates). I discovered high heritability for spring and fall migration timing. I also identified significant loci underlying migration timing near multiple genes including *MEF2D* and *CRY1*, which may be related to photoreceptors and circadian rhythms, and *SNX8*, which may play a role in regulating cholesterol. Additionally, I found that timing of first bloom was associated with a high proportion of genomic variance and loci of polygenic effect. These results present new findings in a long-distance migratory bird, providing genomic evidence of high heritability for migration timing and enhancing our understanding of specific genomic regions underlying this trait. Considering the close connection between genomic variation with migration timing and continent-wide spring phenology, my results suggest that regionally-specific adaptations explain

variation in timing and that microevolutionary processes may be required for birds to adapt to shifting seasonality.

Introduction

Climate change is having significant impacts on ecosystems, such as advancing spring phenology in temperate zones and causing shifts in animals' ranges in response to increasing temperatures (Walther et al. 2002), which can have severe consequences for wildlife. For example, migratory animals must synchronize arrival at their breeding grounds with the availability of seasonal resources (Cohen et al. 2018). Organisms unable to adjust to climatic shifts in timing may become mismatched with spring vegetation and peak prey abundances (Mayor et al. 2017; Visser et al. 2012), which may ultimately lead to population declines in many migratory animals (Both et al. 2006; Post & Forchhammer 2008; Saino et al. 2010; Visser & Gienapp 2019). Understanding species' adaptive potential is becoming increasingly important with accelerating rates of global warming leading to greater rates of phenological advances (Post et al. 2018). Some organisms may adapt through microevolutionary changes (Charmantier & Gienapp 2014; Helm et al. 2019; Hoffmann & Sgrò 2011; Walther et al. 2002), however these adjustments are often not sufficient to keep up with the rapid pace of climate change (Visser 2019). While migrants may rely predominantly on endogenous mechanisms (Liedvogel et al. 2011; Pulido 2007), little is known about the underlying genomic mechanisms, limiting our ability to predict species' future responses to changes in timing. Researching these mechanisms linked to migration timing will help us better understand the potential and limitations of species' microevolution.

Decades of work including cross-breeding experiments and quantitative genetic analyses have shown there is a genetic basis to migratory traits, such as migration orientation and propensity (Berthold 1999; Dingle 1991; Helbig 1991; Gwinner 1977). An integral mechanism of migration timing is the internal clock driving biological rhythms and timing of life-cycle stages (Åkesson et al. 2017), which has prompted many migration genetic studies to focus on candidate genes associated with circadian and circannual rhythms, such as *CLOCK* (Caprioli et al. 2012; Costa et al. 2008; Davie et al. 2009; Partch et al. 2014; Saino et al. 2013) and *ADCYAP1* (Mettler et al. 2015; Mueller et al. 2011). While candidate gene analyses provide important findings, results vary across species (Contina et al. 2018; Lugo Ramos et al. 2017; Parody-Merino et al. 2019; Peterson et al. 2013) and are limited to small portions of the genome. Next-generation sequencing (NGS) can help overcome these limitations by surveying the entire genome (Delmore & Liedvogel 2016). Unbiased genome-wide analyses have identified both known candidate genes and previously unspecified genes influencing migratory traits (Baerwald et al. 2016; Delmore et al. 2016; Delmore et al. 2020; Lundberg et al. 2017; Prince et al. 2017; Toews et al. 2019; Zhan et al. 2014), illuminating on the genomic architecture (identity, number, interactions, and location of genes) of migration. Additional work with new species will allow us to answer longstanding questions on the genetics of migration and adaptation more broadly, including if there is a common genetic basis to migration (Liedvogel et al. 2011) and if specific genomic architectures favour the evolution of complex behaviours (Bendesky et al. 2018).

Early genetic studies were predominantly laboratory-based because it was initially impossible to quantify migratory behaviour in the wild. Because free-living organisms are exposed to a wide variety of abiotic and biotic influences, it is important to measure migratory traits in their natural ecological settings (Helm et al. 2017). Passerines are among the most

widespread taxa and there is a strong genetic basis to their migration, demonstrated by juveniles migrating independently from experienced adults (Alerstam et al. 2003; Liedvogel & Delmore 2018). Only in the beginning of the 21st century were tracking devices such as archival GPS, automated radio-telemetry, and light-level geolocators miniaturized for songbird species (McKinnon & Love 2018). Due to the recent timeframe of developing tracking technology and challenges capturing large numbers of wild birds, it has been difficult for researchers to obtain large sample sizes of migratory data. Purple martins (*Progne subis*) are one of the first two songbird species ever tracked across full migratory journeys (Stutchbury et al. 2009), and we now have one of the most extensive tracking datasets available for a songbird today with over 300 tracks from returning individual purple martins throughout the continent (Neufeld et al. in prep). I will capitalize on this extensive dataset here, matching phenotypic data from light-level geolocators with a new set of genomic resources and analyses to study the genetics of migratory timing.

The purple martin is a Nearctic-Neotropical migrant that travels over 7,000 km between their breeding and wintering sites (Fraser et al. 2012; Fraser et al. 2017). This species displays incredible latitudinal variation in migration timing throughout its North American breeding range. For example, individuals breeding in the southern edge of the range in Florida may arrive as early as mid-January, while their northern counterparts in Alberta may arrive as late as June (Brown & Tarof 2020; Neufeld et al. in prep; PMCA 2020). Despite this variation, there is little genetic differentiation noted between populations (Baker et al. 2008). The combination of the extensive purple martin tracking dataset and the considerable phenotypic variation with limited population differentiation makes it an ideal species for work on the genetics of migratory timing. Additionally, the purple martin is one of many declining aerial insectivores in North America

(Michel et al. 2015; Nebel et al. 2010; Smith et al. 2015) and is susceptible to mismatches in timing with seasonal resources (Fraser et al. 2013), thus it is important to examine potential genetic constraints and adaptability in migration timing in the context of rapidly changing climates.

In this study, I integrated whole-genome sequence data with tracking data from individual purple martins across their annual migrations, representing the first genome-wide association study examining migration timing directly in a songbird. My first objective was to estimate heritability for spring and fall migration timing. Heritability estimates from quantitative genetic analyses may lead to biased estimates (Bérénos et al. 2014; Sardell et al. 2010), and genomic heritability estimates can expand our understanding of the whole-genome contribution to a phenotype (Hu et al. 2014). My second objective was to examine the genomic architecture of migration timing and identify specific regions linked to this trait using traditional association mapping approaches and comparisons of genomic differentiation between birds with extreme phenotypes. For my third objective, I examined genomic associations with first bloom, an ecological variable representing a spring index important for migration timing, to assess how well purple martin genotypes are matched with spring timing and may potentially respond to phenological changes. Genes necessary for determining circadian and circannual rhythms may be important in the martin, but previous work on migration genetics in other songbirds has highlighted the importance of other genes unrelated to these rhythms as well. These studies have also identified genomic regions of large effect (Delmore et al. 2016; Helbig 1991; Lundberg et al. 2017) that are located in areas of reduced recombination. The genome-wide data I am using here will allow me to test the importance of circadian and circannual rhythm genes along with genes showing no previous association with this complex behaviour.

Methods

Field methods

I included samples of *Progne subis subis* from across North America in this study: 89 samples (45 males, 44 females) from 14 different breeding colonies from 2008-2015 (Table A1). I used a second-year female martin from Manitoba, Canada (49°44' N, 97°7' W), for the reference genome. Birds were captured in bird houses while feeding their young, by using drop door traps held with fishing line or by an extendable pole covering the cavity entrance. Up to 150uL of blood was drawn from the bird's brachial vein and then stored in Queen's Lysis buffer (approximately 500-700uL). Light-level geolocators (MK10, MK16) were mounted with leg-loop backpack harnesses and retrieved through recapture in the following years (Fraser et al. 2012; Stutchbury et al. 2009).

Reference genome

DNA for the reference genome was extracted using an *in-situ* agarose plug extraction procedure, then prepared for Pacific BioSciences (PacBio) and 10X Genomics Chromium (10X) libraries for *de novo* assembly (Rhoads & Au 2015; Zheng et al. 2016). After initial assembly with *FALCON* (Chin et al. 2016), I polished the genome using *Arrow* (Chin et al. 2013) and *Pilon* (Walker et al. 2014), and then scaffolded the genome using the *ARKS* and *LINKS* pipeline (Coombe et al. 2018). I evaluated genome quality through multiple programs, including *BUSCO*, which is frequently used as a measure of assembly completeness (Simão et al. 2015), and then mapped scaffolds to chicken (*Gallus gallus*) chromosomes with *SatsumaSynteny* to determine *P.s. subis* scaffold alignment with avian chromosomes (Grabherr et al. 2010). I annotated the final genome with the program *MAKER* (Campbell et al. 2014) using three model species

(chicken, *Gallus gallus*; collared flycatcher, *Ficedula albicollis*; zebra finch, *Taeniopygia guttata*) to identify genes and their locations. See details on genome assembly and annotation in Appendix A.

Resequencing data

I extracted DNA from the remaining birds using the Qiagen DNeasy Blood & Tissue kit (Qiagen 2006) with some procedural modifications to increase DNA yield (details in Appendix A). Samples were sequenced through AgSeq, a skim-based genotyping-by-sequencing (SkimGBS) method that combines Illumina and PerkinElmer technologies to produce low-coverage whole genome data (Bayer et al. 2015; Golicz et al. 2015). Because low coverage results in missing sequence data, single-nucleotide polymorphism (SNP) imputation was performed to fill in missing genotypes based on haplotype blocks (Golicz et al. 2015). Genotype imputation and variant calling were completed after alignment to the reference genome and followed by SNP filtering to remove low-quality sites (details in Appendix A).

Timing data analyses

I analyzed the raw geolocator data (n=89) with R-packages *BAStag* (Wotherspoon et al. 2016) and *GeoLight* (Lisovski & Hahn 2012), using twilight times to estimate daily locations throughout full migratory tracks. I determined the start and end (departure and arrival) dates for both spring and fall migration to represent migration timing phenotypes for each individual. (details on geolocator analyses in Appendix A). Because departure and arrival dates were highly correlated, I focused on spring departure and fall arrival migratory phenotypes to represent the timing of the two seasonal migrations for downstream analyses. I extracted first bloom spring

data from two databases, keeping year consistent with blood sample collection. Through the USA National Phenology Network, first bloom dates consisted of the average for Red Rothomagensis lilac (*Syringa x chinensis*), Arnold Red honeysuckle (*Lonicera tatarica*), and Zabelii honeysuckle (*Lonicera korolkowii*) (USA-NPN 2017). I included data for these species matched to the precise martin colony locations. I selected first bloom dates for the Canadian colonies from NatureWatch Canada (NatureWatch 2014), using Common purple lilac (*Syringa vulgaris*) data closest to martin colony locations to maintain similar plant species for the final dataset (n=79). The overall sample size for first bloom dates was lower than for geolocator data, due to the absence of honeysuckle or lilac data in Alberta in year 2013.

Statistical analysis with migratory phenotypes

I estimated heritability and genomic associations with spring and fall migration timing using *GEMMA* to run Bayesian sparse linear mixed models (BSLMM), which is a polygenic model combining the advantages of standard linear mixed models (LMM) and Bayesian variable selection regression models (BVSR), capable of learning the genetic architecture from the data (Zhou et al. 2013). After standardizing migratory phenotypes through quantile-normalization, I estimated the total variance explained by sparse effects, or alleles of large effects (PGE), and used this to represent the heritability estimate. Next, I used a posterior inclusion probability (PIP) threshold of 0.1 to identify significant SNPs associated with each phenotype (Gompert et al. 2013; Pfeifer et al. 2018). Because BSLMMs do not account for covariates, I also used *GEMMA* to run LMMs to examine isolated effects of SNPs on migration timing while including covariates of sex, colony, year, age, and PC1 on the same dataset (Zhou & Stephens 2013). To obtain values for PC1 (to control for population structure), I conducted a Principal Component Analysis

(PCA) to examine population structure in my set of samples using *smartpca* in the *EIGENSOFT* program (Price et al. 2006). In the LMMs, I determined SNPs above the false discovery rate (FDR) of 0.05 as significant. Using the reference genome annotation, I identified the closest genes within 100 kb to the significant SNPs in the association models and completed a gene ontology (GO) term analysis. Finally, to examine differentiation across the genome related to migration, I estimated Weir and Cockerham fixation indices (F_{ST}) between the 10 earliest and 10 latest spring migrants, representing the extreme phenotypes, in non-overlapping 5 kb windows using *VCFtools* (Danecek et al. 2011). I also estimated F_{ST} between individuals from Alberta and Florida, comparing the most northern and southern populations. Further details on analyses are in Appendix A.

Statistical analysis with first bloom

For mapping genotypes with first bloom data, I ran a latent factor mixed model (LFMM) using R-package *LEA*, which assesses association of allele frequencies to ecological gradients (Frichot & François 2015). Here, I included the same five covariables used in the LMM and used R-package *vqtl* to estimate the proportion of variance explained by genetic effects (Corty & Valder 2018). I determined SNPs above the false discovery rate (FDR) of 0.05 as significant. I also completed a GO term analysis using the same methods described above. See Appendix A for further details.

Results

Reference genome

The final *P. subis* reference genome assembly was 1.17 Gb in length, consisted of 2,896 scaffolds, a N50 scaffold length of 6.13 Mb, and 12,686 genes (Table 2.1). The assembly length was similar to other avian genomes, which are typically between 1.0-1.2 Gb (Ellegren 2013). *BUSCO* analysis revealed that the *P. subis* genome was relatively complete with 91% of avian orthologs detected as complete sequences (89.1% being single-copy and 1.9% being duplicated; Table 2.1), which was in range of other non-model avian genomes (Coelho et al. 2019). The genetic variant dataset for the 89 re-sequenced individuals aligned to the reference genome resulted in 5.9 million SNPs, which I filtered down to 4.6 million SNPs.

Timing data analyses

The light-level geolocator data displayed latitudinal variation as expected, ranging between 89 to 131 days for each timing phenotype (spring and fall departure and arrival). Migration timing dates were earlier in breeding colonies at lower latitudes and later in colonies at higher latitudes (Figure 2.1). For example, spring departure dates ranged from January 10th (Florida) to May 9th (Alberta). The absolute spring migratory tracks are displayed in Figure 2.2, showing weak connectivity with breeding and wintering sites such as observed in Fraser et al. 2012. First bloom dates ranged from January 21st to May 27th, spanning 126 days (Figure A2).

Statistical analysis with migratory phenotypes

A large proportion of variance in both spring and fall was explained by the SNP set through the BSLMMs. Both traits were highly heritable, with median PGE values of 0.78 for

spring (95% CI: 0.01-0.98, SD: 0.29) and 0.85 for fall (95% CI: 0.02-0.99, SD: 0.29) with confidence intervals excluding zero. The wide confidence intervals were likely attributed to the sample size.

BSLMMs identified five SNPs associated with spring departure (PIP 0.11-0.38) and eight SNPs associated with fall arrival (PIP 0.12-0.32). Localized results can be found in Table 2.2 and Figure 2.3. Of prime importance, the two most significant loci (S214:738983 and S86:1071937) were associated with both spring and fall timing, located within 100 kb of genes *MEF2D* and *GDPD2* respectively. In fall arrival, an additional associated SNP was within 100 kb of *CRY1* and another SNP in gene *SNX8*. LMMs did not identify significant SNPs with FDR < 0.05, however, both spring and fall had elevated associations with S214:738983 that were identified in the BSLMM (respective p-values are listed in Table 2.2). I did not find enrichment in the GO-term analysis when combining associations with all phenotypes using all genes within 100 kb of significant SNPs. When analyzing genes associated with SNPs in spring departure separately, even the most significant GO-terms showed little support for gene enrichment (p-value 0.34; Table A3).

F_{ST} across the genome between the earliest and latest migrants revealed regions of high differentiation on chromosome 1 and chromosome 2 (Figure 2.4). The highest F_{ST} (0.6) in chromosome 1 was near the *CCDC59* (coiled-coil domain containing 59) gene, among 13 genes in a 2.5 Mb highly differentiated region. On chromosome 2, the highest F_{ST} (0.45) was near *SLC66A2* (Solute carrier family 66 member 2) gene, among 32 genes also in a 2.5 Mb region. F_{ST} across the genome between Alberta and Florida populations had weaker signals at these regions and lacked clear regions of high differentiation (Figure A5).

Statistical analysis with first bloom

The LFMM identified 63 SNPs associated with first bloom (FDR < 0.05; Figure 2.5). The highest signal was in gene *CEP250* on chromosome 20, however, the overall spread of SNPs across the genome and lack of clear elevation suggest that the ecological association with genotypes may be polygenic (i.e. controlled by many alleles of small effect). The R-package *vqtl* identified a PVE value of 0.54. The GO term analysis using genes within 100 kb of associated SNPs did not reveal any gene enrichment.

Discussion

For the first time in a migratory bird, I paired one of the largest datasets of direct data on migratory timing with an extensive set of genomic resources to study the genetics underlying migration timing. My results demonstrate a strong genetic basis for migration timing, including large proportions of variance explained by genotypes. Additionally, I identified specific genomic loci linked to migratory timing that illuminate new components underlying migration timing and strengthen evidence of a known candidate gene that may be connected to physiological preparedness important for migration. The genomic variation associated with migration timing and spring green-up over a wide latitudinal gradient suggest that microevolutionary changes would be an important component of change in timing across latitude.

Phenotypes that vary across individuals are a result of both environmental and genetic factors, and heritability is represented by the proportion of variance attributed to genetic factors (Visscher et al. 2008). The large heritability estimates in this study (0.78-0.85) demonstrate that migration timing is largely determined by genetics, which is fundamental for evolution to occur. These are the first estimates of heritability derived from genomic data for migratory traits in

songbirds and are higher than previous estimates for migratory traits obtained through quantitative genetic analyses, such as seen in the range of 0.07-0.67 heritability estimates over multiple studies (Pulido & Berthold 2003). The higher heritability estimates in this study are likely due to the whole-genome contribution (Hu et al. 2014) that avoid biases limiting quantitative genetic approaches (Bérénos et al. 2014; Sardell et al. 2010). These estimates are a useful part of understanding potential responses to selection (Charmantier et al. 2016), as animals will need to adapt through selection in microevolutionary changes or more flexibly with individual phenotypic plasticity. While the high genomic variation explaining migratory timing does not preclude phenotypic plasticity, it suggests that changes in timing would likely have to occur through microevolutionary processes. This could constrain migrants' abilities to adapt quickly, as selection on heritable traits are generally slower to change phenotypes than plasticity and may not be sufficient to keep up with rapidly changing climates (Charmantier & Gienapp 2014; Hetem et al. 2014; Visser 2019). While some species may move further northward to match their climate niches, neotropical migrants have not shown shifts in the northern leading edges of their breeding distributions (Rushing et al. 2020), indicating that they will need to adapt to earlier timing. It is important to understand the source of considerable genetic variation in migratory traits (Liedvogel et al. 2011), and future work will inform how influences such as standing genetic variation (presence of more than one allele at a locus in a population) may play a role in facilitating rapid microevolutionary changes (Barrett & Schluter 2008).

The small number of genomic loci significantly associated with spring and fall migration suggest this trait is controlled by a few alleles of large effect. Key regions included signals near genes *MEF2D* (MADS-box domain-containing protein) and *CRY1* (Cryptochrome circadian regulator 1), which may be involved with photoperiod and cryptochromes. Photoperiod and

cryptochromes are known to play a large role in cueing circadian and circannual rhythms and migratory behaviours. There is evidence *MEF2D* is linked to photoreceptors (Andzelm et al. 2015; Nagar et al. 2017; Omori et al. 2015), but this gene has not been documented in birds prior to this study. Many laboratory studies have shown that photoperiod plays a critical role in cueing migration timing, including physiological preparedness and migratory restlessness (Gwinner 1973; Newton 2007; Robart et al. 2018). The genomic signal near *CRY1* supports evidence from previous studies across a broad range of taxa identifying roles of cryptochromes in migration and its function as a circadian photoreceptor (Chang et al. 2019; Fusani et al. 2014; Reppert & de Roode 2018; Zhou et al. 2016). Using photoperiodism as a measure of environmental day length, animals can prepare physiologically for the high energetic costs that migration requires, and time their movements to match seasons throughout the year. Reliance on photoperiod could be a significant constraint on adapting to climate change, since the timing of seasonal resources are shifting but the annual cycle of photoperiod remains unchanged (Wingfield 2008).

Migration is energetically demanding, requiring optimal metabolic and morphological conditions (Braithwaite et al. 2015; Hein et al. 2011). Signals of associated loci in *SNX8* (Sorting nexin 8) and near *GDPD2* (Glycerophosphodiester phosphodiesterase domain containing 2) show new components that may underlie physiological processes for avian migration timing. *SNX8* gene may be associated with membrane lipids and cholesterol regulation (Cullen 2008; Muirhead & Dev 2014), and since birds are known to accumulate fat for long-distance migrations (Araújo et al. 2019), the signal here in the purple martin genome suggest this gene could play a role in migratory birds. Although *GDPD2* is less described in the literature, two studies show it may have a role in bone formation (Corda et al. 2014; Yanaka et al. 2003). Birds have very specialized lightweight bone structure to enable flight (Sullivan et al. 2017), and it is

possible there could be connections to physiological changes in bone development and migration timing if longer-distance travelling birds have different structural characteristics. While this study suggests associations with these genes, further studies could elucidate these mechanisms and its role in migration timing.

Interestingly, the loci associated with spring and fall migration in the association analyses did not show regions of elevated F_{ST}, or differentiation, between early and late migrants. This provided additional support that these regions have specific roles in migration alone, as opposed to being attributed to patterns of variation in other features across the genome (e.g. reduced recombination) or population structure (Cruickshank & Hahn 2014; Noor & Bennett 2009). The overall differentiation between early and late migrants showed regions of elevated F_{ST} in two chromosomes, contrasting with the lack of elevated F_{ST} in the comparison with Alberta and Florida populations, which suggests that these elevated regions are also not related to population structure. Additionally, the gene density in these regions were not lower than the average density across the genome, suggesting factors such as centromeres that would have lower gene density do not contribute to these patterns (Kozubek 2006). Further work is needed to implicate these regions directly with migration.

It is important to investigate connections between spring phenology (e.g. ‘green-up’) with arrival timing in migratory animals, as climate change impacts on phenology can result in asynchrony between animal timing and environmental resources, leading to severe fitness consequences and population declines (Both et al. 2006; Mayor et al. 2017; Visser & Gienapp 2019). Though first bloom dates were correlated with the latitudinal gradient of purple martin spring arrival timing, the environmental model used an ecological variable directly and showed numerous associated loci without emphasis on particular regions, illustrating there may be many

variants of smaller effects involved with matching avian timing to spring phenology. The large proportion of genomic variance associated with first bloom also suggests individuals may be constrained to adjust with increasingly advancing spring timing. Phenological asynchrony with spring green-up highlighted in Mayor et al. (2017) displayed different rates of phenological shifts in migratory species' arrival timing and spring green-up over 12 years, which could be more apparent in species over a longer time series. The first bloom dates in my study spanned eight years, which is a relatively short time span for ecological change. If purple martins are better matched to environmental timing in past decades compared to more recent time, there could be greater mismatch between genomic variation and the fast pace of environmental changes. Additional work is needed to disentangle the association of first bloom with genomic mechanisms, and future studies could also examine environmental variables over a longer time scale.

Ideally, genomic studies would include thousands of samples, such as seen in many human-related association studies (Nishino et al. 2018; Wood et al. 2014). This level of sample size is impossible to obtain in field-based studies examining animals in the wild because of logistical difficulties in capturing a large number of individuals and minimizing disturbance to populations (Steiner et al. 2013). The limitations of the sample size in this study could have missed additional genomic regions associated with migration timing. Additionally, the limited number of samples may have contributed to the lack of linked SNPs showing similar patterns as the significant loci in the association studies (Wang et al. 2011) and in complementary analyses such as the GO-term analysis. Even with limited samples, however, whole-genome sequencing provided high resolution data that captured significant associations in top SNPs of important regions. Many of these SNPs were in intergenic regions, suggesting they occur in regulatory

regions of genes. The three-dimensional structure of the genome can mean these regions occur quite distant from the genes themselves (Williams et al. 2010).

This study presents novel findings on migration timing, opening the door to understanding heritability and components of the genomic architecture of migration timing in long-distance migrants. The strong genomic variation and significant regions associated with purple martin migration timing could have important implications for adaptability in long-distance migrants, suggesting that there may be potential constraints in keeping up with changing climates. This study also reveals some of the specific mechanisms inside the genetic program that may influence responses to photoperiodic changes and physiological cues, which are both important parts of migration, as this feat requires careful timing and high energetic costs and preparations. These findings can have broader implications on a variety of organisms, as many portions of the genome are conserved across other bird species and vertebrates (Burt 2005; Sun et al. 2008), and climate change continues to affect migratory animals all over the world.

Acknowledgements

I am grateful to collaborators K. Applegate, J. Barrow, L. Burgess, L. Chambers, T. Cheskey, P. Clifton, B. Dietrich, J. Fischer, G. Hvenegaard, P. Kramer, P. Mammenga, N. Mickle, M. North, M. Pearman, J. Ray, A. Ritchie, A. Savage, T. Shaheen, J. Siegrist, C. Silverio, and J. Tautin for assistance in collecting samples and deploying geolocators; B. Stutchbury for assistance with both sample collection and organization; K. Jeffries for sharing his laboratory space and equipment for DNA extractions; Alexander Suh National Genomics Infrastructure in Sweden, Science for Life Laboratory, and Texas A&M AgriLife Genomics and Bioinformatics Service for sequencing resources; Texas A&M High Performance Research

Computing and Compute Canada for computing services; C. Garroway for assistance with analyses; M. Thorstensen for assistance with bioinformatics, analyses, and editorial support. This work was supported by funding from the University of Manitoba, Texas A&M University, Natural Sciences and Engineering Research Council Discovery Grant program, Research Manitoba, the Society of Canadian Ornithologists, and Sigma Xi.

References

Åkesson, S., Ilieva, M., Karagicheva, J., Rakhimberdiev, E., Tomotani, B., and Helm, B. (2017). Timing avian long-distance migration: From internal clock mechanisms to global flights. Phil. Trans. R. Soc. B., 372: 20160252. doi:10.1098/rstb.2016.0252

Alerstam, T., Hedenstrom, A., and Akesson, S. (2003). Long-distance migration: evolution and determinants. Oikos, 103(2): 247-260. doi:10.1034/j.1600-0706.2003.12559.x

Andzelm, M.M., Cherry, T.J., Harmin, D.A., Boeke, A.C., Lee, C., Hemberg, M., Pawlyk, B., Malik, A.N., Flavell, S.W., Sandberg, M.A., Raviola, E., and Greenberg, M.E. (2015). MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. Neuron, 86(1): 247–263. doi:10.1016/j.neuron.2015.02.038

Araújo, P.M., Viegas, I., Rocha, A.D., Villegas, A., Jones, J.G., Mendonça, L., Ramos, J.A., Masero, J.A., and Alves, J.A. (2019). Understanding how birds rebuild fat stores during migration: insights from an experimental study. Scientific Reports, 9(10065). doi:10.1038/s41598-019-46487-z

Baerwald, M.R., Meek, M.H., Stephens, M.R., Nagarajan, R.P., Goodbla, A.M., Tomalty, K.M.H., Thorgaard, G.H., May, B., and Nichols, K.M. (2016). Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. Molecular Ecology, 25(8): 1785-1800. doi:10.1111/mec.13231

Baker, A.J., Greenslade, A.D., Darling, L.M., and Finlay, J.C. (2008). High genetic diversity in the blue-listed British Columbia population of the purple martin maintained by multiple sources of immigrants. Conservation Genetics, 9: 495–505. doi:10.1007/s10592-007-9358-3

Barret, R.D.H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol., 23(1): 38-44. doi:10.1016/j.tree.2007.09.008

Bayer, P.E., Ruperao, P., Mason, A.S., Stiller, J., Chan, C.K.K., Hayashi, S., Long, Y., Meng, J., Sutton, T., Visendi, P., Varshney, R.K., Batley, J., and Edwards, D. (2015). High-resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. Theor. Appl. Genet., 128: 1039-1047. doi:10.1007/s00122-015-2488-y

Bendesky, A., Kwon, Y.M., Lassance, J.M., Lewarch, C.L., Yao, S., Peterson, B.K., He, M.X., Dulac, C., and Hoekstra, H.E. (2017). The genetic basis of parental care evolution in monogamous mice. Nature, 544: 434–439. doi:10.1038/nature22074

Bérénos, C., Ellis, P.A., Pilkington, J.G., and Pemberton, J.M. (2014). Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. Molecular Ecology, 23: 3434-3451. doi:10.1111/mec.12827

Berthold, P. (1999). A comprehensive theory for the evolution, control and adaptability of avian migration. Ostrich, 70(1): 1–11. doi:10.1080/00306525.1999.9639744

Both, C., Bouwhuis, S., Lessells, C.M., and Visser, M.E. (2006). Climate change and population declines in a long-distance migratory bird. Nature, 441(1): 81–83. doi:10.1038/nature04539

Braithwaite, J.E., Meeuwig, J.J., and Hipsey, M.R. (2015). Optimal migration energetics of humpback whales and the implications of disturbance. Conservation Physiology, 3. doi:10.1093/conphys/conv001

Brown, C.R. and Tarof, S. (2020). Purple Martin (*Progne subis*), version 1.0. In Birds of the World (A.F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. doi:10.2173/bow.purmar.01

Burt, W.D. (2005). Chicken genome: Current status and future opportunities. Genome Research, 15: 1692-1698. doi:10.1101/gr.4141805

Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. Curr Protoc Bioinformatics, 48: 4.11.1-4.11.39 doi:10.1002/0471250953.bi0411s48

Caprioli, M., Ambrosini, R., Boncoraglio, G., Gatti, E., Romano, A., Romano, M., Rubolini, D., Gianfranceschi, L., and Saino, N. (2012). Clock gene variation is associated with breeding phenology and maybe under directional selection in the migratory barn swallow. PLoS ONE, 7(4): e35140. doi:10.1371/journal.pone.0035140

Chang, H., Guo, J.L., Fu, X.W., Wang, M.L., Hou, Y.M., and Wu, K.M. (2019). Molecular characterization and expression profiles of *cryptochrome* genes in a long-distance migrant, *Agrotis segetum* (Lepidoptera: Noctuidae). Journal of Insect Science, 19(1): 1-11. doi:10.1093/jisesa/iey127

Charmantier, A., and Gienapp, P. (2014). Climate change and timing of avian breeding and migration: Evolutionary versus plastic changes. Evolutionary Applications, 7: 15–28. doi:10.1111/eva.12126

Charmantier, A., Doutrelant, C., Dubuc-Messier, G., Fargevieille, A., and Szulkin, M. (2016). Mediterranean blue tits as a case study of local adaptation. Evolutionary Applications, 9(1): 135-152. doi:10.1111/eva.12282

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods, 10(6): 563–569. doi:10.1038/nmeth.2474

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., …Schatz, M.C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nature Methods, 13(12): 1050–1054. doi:10.1038/nmeth.4035

Coelho, L.A., Musher, L.J., and Cracraft, J. (2019). A Multireference-Based Whole Genome Assembly for the Obligate Ant-Following Antbird, *Rhegmatorhina melanosticta* (Thamnophilidae). (2019). Diversity 11(19): 144. doi:10.3390/d11090144

Cohen, J.M., Lajeunesse, M.J., and Rohr, J.R. (2018). A global synthesis of animal phenological responses to climate change. Nature Climate Change, 8: 224–228. doi:10.1038/s41558-018-0067-3

Contina, A., Bridge, E.S., Ross, J.D., Shipley, J.R., and Kelly, J.F. (2018). Examination of *Clock* and *Adcyap1* gene variation in a neotropical migratory passerine. PLoS ONE, 13(1): e0190859. doi:10.1371/journal.pone.0190859

Coombe, L., Zhang, J., Vandervalk, B.P., Chu, J., Jackman, S.D., Birol, I., and Warren, R.L. (2018). ARKS: Chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics, 19(234): 1–10. doi:10.1186/s12859-018-2243-x

Corda, D., Mosca, M.G., Ohshima, N., Grauso, L., Yanaka, N., and Mariggiò, S. (2014). The emerging physiological roles of the glycerophosphodiesterase family. FEBS Journal, 281: 998–1016. doi:10.1111/febs.12699

Corty, R.W., and Valder, W. (2018). vqtl: An R Package for Mean-Variance QTL Mapping. G3: Genes, Genomes, Genetics, 8(12): 3757-3766. doi:10.1534/g3.118.200642

Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23(13): 3133-3157. doi:10.1111/mec.12796

Cullen, P.J. (2008). Endosomal sorting and signaling: an emerging role for sorting nexins. Nat. Rev. Mol. Cell. Bio., 9(7): 574-582. doi:10.1038/nrm2427

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., and 1000 Genomics Project Analysis Group. (2011). The variant call format and VCFtools. Bioinformatics, 27(15): 2156–2158. doi:10.1093/bioinformatics/btr330

Davie, A., Minghetti, M., and Migaud, H. (2009). Seasonal Variations in Clock-Gene Expression in Atlantic Salmon (*Salmo Salar*). Chronobiology International, 26(3): 379–395. doi:10.1080/07420520902820947

Delmore, K.E., and Liedvogel, M. (2016). Investigating factors that generate and maintain variation in migratory orientation: A primer for recent and future work. Frontiers in Behavioral Neuroscience, 10(3): 1-7. doi:10.3389/fnbeh.2016.00003

Delmore, K.E., Toews, D.P.L., Germain, R.R., Owens, G L., and Irwin, D.E. (2016). The Genetics of Seasonal Migration and Plumage Color. Current Biology 26: 2167-2173. doi:10.1016/j.cub.2016.06.015

Delmore, K., Illera, J.C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J.S, Durieux, G., Ishigohoka, J., and Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. ELife, 9:e54462. doi:10.7554/eLife.54462

Dingle, H. (1991). Evolutionary Genetics of Animal Migration. American Zoologist, 31(1): 253-264

Ellegren, H. (2013). The Evolutionary Genomics of Birds. Annu. Rev. Ecol. Evol. Syst., 44: 239-259. doi:10.1146/annurev-ecolsys-110411-160327

Fraser, K.C., Stutchbury, B.J.M., Silverio, C., Kramer, P.M., Barrow, J., Newstead, D., Mickle, N., Cousens, B.F., Lee, J.C., Morrison, D.M., Shaheen, T., Mammenga, P., Applegate, K., and Tautin, J. (2012). Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. Proc. R. Soc. B., 279(1749): 4901–4906. doi:10.1098/rspb.2012.2207

Fraser, K.C., Silverio, C., Kramer, P., Mickle, N., Aeppli, R., and Stutchbury, B.J.M. (2013). A Trans-Hemispheric Migratory Songbird Does Not Advance Spring Schedules or Increase Migration Rate in Response to Record-Setting Temperatures at Breeding Sites. PLoS ONE, 8(5): 1-5. doi:10.1371/journal.pone.0064587

Fraser, K.C., Shave, A., Savage, A., Ritchie, A., Bell, K., Siegrist, J., Ray, J.D., Applegate, K., and Pearman, M. (2017). Determining fine-scale migratory connectivity and habitat selection for a migratory songbird by using new GPS technology. Journal of Avian Biology, 48(3): 001-007. doi:10.1111/jav.01091

Frichot, E., and François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6: 925–929. doi:10.1111/2041-210X.12382

Fusani, L., Bertolucci, C., Frigato, E., and Foà. (2014). Cryptochrome expression in the eye of migratory birds depends on their migratory status. The Journal of Experimental Biology, 217: 918-923. doi:10.1242/jeb.096479

Golicz, A.A., Bayer, P.E., and Edwards, D. (2015). Skim-Based Genotyping by Sequencing. Methods in Molecular Biology, 1245: 257–270. doi:10.1007/978-1-4939-1966-6_19

Gompert, Z., Lucas, L.K., Nice, C.C., and Buerkle, C.A. (2013). Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. Evolution, 67(9): 2498–2514. doi:10.1111/evo.12021

Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: *Satsuma*. Bioinformatics, 26(9): 1145–1151. doi:10.1093/bioinformatics/btq102

Gwinner, E. (1973). Circannual rhythms in birds: their interaction with circadian rhythms and environmental photoperiod. Journal of Reproduction and fertility Supplement., 19: 51-65.

Gwinner, E. (1977). Circannual rhythms in bird migration. Ann. Rev. Ecol. Syst., 8:381-405.

Hein, A.M., Hou, C., and Gillooly, J.F. (2011). Energetic and biomechanical constraints on animal migratoin distance. Ecology letters, 15(2): 104-110. doi:10.1111/j.1461-0248.2011.01714.x

Helbig, A.J. (1991). Inheritance of migratory direction in a bird species: a cross-breeding experiment with SE- and SW-migrating blackcaps (*Sylvia atricapilla*). Behav Ecol Sociobiol., 28: 9–12. doi:10.1007/BF00172133

Helm, B., Visser, M.E., Schwartz, W., Kronfeld-Schor, N., Gerkema, M., Piersma, T., and Bloch, G. (2017). Two sides of a coin: ecological and chronobiological perspectives of timing in the wild. Phil. Trans. R. Soc. B., 372: 20160246. doi:10.1098/rstb.2016.0246

Helm, B., Van Doren, B.M., Hoffman, D., and Hoffman, U. (2019). Evolutionary Response to Climate Change in Migratory Pied Flycatchers. Current Biology, 29(21): 3714-3719. doi:10.1016/j.cub.2019.08.072

Hetem, R.S., Fuller, A., Maloney, S.K., and Mitchell, D. (2014). Responses of large mammals to climate change. Temperature, 1(2): 115-127. doi:10.4161/temp.29651

Hoffmann, A.A., and Sgró, C.M. (2011). Climate change and evolutionary adaptation. Nature, 470: 479–485. doi:10.1038/nature09670

Hu, G., Wang, C., and Da. Y. (2014). Genomic heritability estimation for the early life-history transition related to propensity to migrate in wild rainbow and steelhead trout populations. Ecology and Evolution, 4(8): 1381-1388. doi:10.1002/ece3.1038

Kozubek, S. (2006). Spatial arrangement of the human genome and its possible functional role. Atlas Genet. Cytogenet. Oncol. Haematol., 10(1): 40-55. doi:10.4267/2042/38281

Liedvogel, M., Åkesson, S., and Bensch, S. (2011). The genetics of migration on the move. Trends in Ecology and Evolution, 26(11): 561–569. doi:10.1016/j.tree.2011.07.009

Liedvogel, M. and Delmore, K.E. (2018). (Mirco)evolutionary Changes and the Evolutionary Potential of Bird Migration. In: Tietze D. (eds) Bird Species. Fascinating Life Sciences. doi:10.1007/978-3-319-91689-7_7

Lisovski, S., and Hahn, S. (2012). GeoLight-processing and anslysing light-based geolocator data in R. Methods in Ecology and Evolution, 3: 1055-1059. doi:10.1111/j.2041-210X.2012.00248.x

Lugo Ramos, J.S., Delmore, K.E., and Liedvogel, M. (2017). Candidate genes for migration do not distinguish migratory and non-migratory birds. J. Comp. Physiol. A., 203: 383–397. doi:10.1007/s00359-017-1184-6

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Susanne, Å, and Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evolution Letters, 1(3): 155–168. doi:10.1002/evl3.15

Mayor, S.J., Guralnick, R.P., Tingley, M.W., Otegui, J., Withey, J.C., Elmendorf, S.C., Andrew, M.E., Leyk, S., Pearse, I.S., and Schneider, D.C. (2017). Increasing phenological asynchrony between spring green-up and arrival of migratory birds. Scientific Reports, 7(1902): 1–10.doi:10.1038/s41598-017-02045-z

McKinnon, E.A., and 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. doi:10.1642/auk-17-202.1

Mettler, R., Segelbacher, G., and Schaefer, M.H. (2015). Interactions between a candidate gene for migration (*ADCYAP1*), morphology and sex predict spring arrival in blackcap populations. PLoS ONE, 10(12): e0144587. doi:10.1371/journal.pone.0144587

Michel, N.L., Smith, A.C., Clark, R.G., Morrissey, C.A., and Hobson, K.A. (2015). Differences in spatial synchrony and interspecific concordance inform guild-level population trends for aerial insectivorous birds. Ecography, 39: 774-786. doi:10.1111/ecog.01798

Mueller, J.C., Pulido, F., and Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. Proc. R. Soc. B., 278: 2848–2856. doi:10.1098/rspb.2010.2567

Muirhead, G., and Dev, K.K. (2014). The Expression of Neuronal Sorting Nexin 8 (SNX8) Exacerbates Abnormal Cholesterol Levels. J. Mol. Neurosci., 53: 125-134. doi:10.1007/s12031-013-0209-z

Nagar, S., Noveral, S.M., Trudler, D., Lopez, K.M., McKercher, S.R., Han, X., Yates, J.R., Piña-Crespo, J.C., Nakanishi, N., Satoh, T., Okamoto, S., and Lipton, S.A. (2017). MEF2D haploinsufficiency downregulates the NRF2 pathway and renders photoreceptors susceptible to light-induced oxidative stress. PNAS, 114(20): E4048–E4056. doi:10.1073/pnas.1613067114

NatureWatch Canada. PlantWatch data from May 2013 to May 2014, University of Ottawa, Canada. http://www.naturewatch.ca. Accessed on 17 April 2020.

Nebel, S., Mills, A., McCracken, J.D., and Taylor, P.D. (2010). Declines of Aerial Insectivores in North American Follow a Geographic Gradient. Avian Conservation and Ecology, 5(2). doi:10.5751/ACE-00391-050201

Neufeld, L.R., Muthukumarana, S., Fischer, J., Ray, J., Siegrist, J., and Fraser, K.C. (2020). Breeding latitude predicts the timing of nest initiation and migration timing around the annual calendar among populations of a declining, migratory songbird purple martin (*Progne subsis*). Unpublished manuscript.

Newton, I. (2004). The Migration Ecology of Birds. Academic Press, Waltham, Massachusetts. doi:10.1016/B978-0-12-517367-4.X5000-1

Nishino, J., Ochi, H., Kochi, Y., Tsunoda, T., and Matsui, S. (2018). Sample size for successful genome-wide association study of major depressive disorder. Frontiers in genetics, 28. doi:10.3389/fgene.2018.00227

Noor, M.A.F., and Bennet, S.M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity, 103: 439-444. doi:10.1038/hdy.2009.151

Omori, Y., Kitamura, T., Yoshida, S., Kuwahara, R., Chaya, T., Irie, S., and Furukawa, T. (2015). Mef2d is essential for the maturation and integrity of retinal photoreceptor and bipolar cells. Genes to Cells, 20: 408–426. doi:10.1111/gtc.12233

Parody-Merino, Á.M., Battley, P.F., Conklin, J.R., and Fidler, A.E. (2019). No evidence for an association between *Clock* gene allelic variation and migration timing in a long-distance migratory shorebird (*Limosa lapponica baueri*). Oecologia, 191: 843–859. doi:10.1007/s00442-019-04524-8

Partch, C.L., Green, C.B., and Takahashi, J.S. (2014). Molecular Architecture of the Mammalian Circadian Clock. Trends Cell Biol., 24(2): 90–99. doi:10.1016/j.tcb.2013.07.002

Peterson, M.P., Abolins-Abols, M., Atwell, J.W., Rice, R.J., Milá, B., and Ketterson, E.D. (2013). Variation in candidate genes *CLOCK* and *ADCYAP1* does not consistently predict

differences in migratory behavior in the songbird genus Junco. F1000Research, 2: 115. doi:10.12688/f1000research.2-115.v1

Pfeifer, S.P., Laurent, S., Sousa, V.C., Linnen, C.R., Foll, M., Excoffier, L., Hoekstra, H., and Jensen, J.D. (2018). The evolutionary history of Nebraska deer mice: Local adaptation in the face of strong gene flow. Molecular Biology and Evolution, 35(4): 792–806. doi:10.1093/molbev/msy004

Post, E., and Forchhammer, M.C. (2008). Climate change reduces reproductive success of an Arctic herbivore through trophic mismatch. Phil. Trans. R. Soc. B., 363: 2369–2375. doi:10.1098/rstb.2007.2207

Post, E., Steinman, B.A., and Mann, M.E. (2018). Acceleration of phenological advance and warming with latitude over the past century. Scientific Reports, 8(3927): 1–8. doi:10.1038/s41598-018-22258-0

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics, 38(8): 904-909. doi:10.1038/ng1847

Prince, D.J., O'Rourke, S.M., Thompson, T.Q., Ali, O.A., Lyman, H.S., Saglam, I.K., Hotaling, T.J., Spidle, A.P., and Miller, M.R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science Advances, 3:21603198. doi:10.1126/sciadv.1603198

Pulido, F. (2007). The genetics and evolution of avian migration. BioScience, 57(2): 165-174. doi:10.1641/B570211

Purple Martin Conservation Association (PMCA) Scout Arrival Study. 2020. Available from https://www.purplemartin.org/research/8/scout-arrival-study/ Accessed on 12 July 2020.

Qiagen Inc. 2006. DNeasy Blood & Tissue Kit. Ordered in 2018 from Toronto, Canada

Reppert, S.M., and de Roode, J.C. (2018). Demystifying Monarch butterfly migration. Current Biology, 28: R1009-R1022. doi:10.1016/j.cub.2018.02.067

Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, 13: 278–289. doi:10.1016/j.gpb.2015.08.002

Robart, A.R., McGuire, M.M.K., and Watts, H.E. (2018). Increasing photoperiod stimulates the initiation of spring migratory behaviour and physiology in a facultative migrant, the pine siskin. Royal Society Open Science, 5: 180876. doi:10.1098/rsos.180876

Saino, N., Ambrosini, R., Rubolini, D., Von Hardenberg, J., Provenzale, A., Hüppop, K., Hüppop O., Lehikoinen, A., Lehikoinen, E., Rainio, K., Romano, M., and Sokolov, L. (2011). Climate warming, ecological mismatch at arrival and population decline in migratory birds. Proc. R. Soc. B., 278: 835–842. doi:10.1098/rspb.2010.1778

Saino, N., Romano, M., Caprioli, M., Fasola, M., Lardelli, R., Micheloni, P., Scandolara, C., Rubolini, D., and Gianfranceschi, L. (2013). Timing of molt of barn swallows is delayed in a rare *Clock* genotype. PeerJ, 1: e17. doi:10.7717/peerj.17

Sardell, R.J., Keller, L.F., Arcese, P., Bucher, T., and Reid, J.M. (2010). Comprehensive paternity assignment: genotype, spatial location and social status in song sparrows, *Melospiza Melodia*. Molecular Ecology, 19(19): 4352-4364. doi:10.1111/j.1365-294X.2010.04805.x

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(9): 3210-3212. doi:10.1093/bioinformatics/btv351

Smith, A.C., Hudson, M.A.R., Downes, C.M., and Francis, C.M. (2015). Change Points in the Population Trends of Aerial-Insectivorous Birds in North America: Synchronized in Time across Species and Regions. PLoS ONE, 10(7): e0130768. doi:10.1371/journal.pone.0130768

Steiner, C.C., Putnam, A.S., Hoeck, P.E.A., and Ryder, O.A. (2013). Conservation Genomics of Threatened Animal Species. Annu. Rev. Anim. Biosci., 1: 261-281. doi:10.1146/annurev-animal-031412-103636

Stutchbury, B.J.M., Tarof, S.A., Done, T., Gow, E., Kramer, P.M., Tautin, J., Fox, J.W., and Afanasyev, V. (2009). Tracking long-distance songbird migration by using geolocators. Science, 323: 896. doi:10.1126/science.1166664

Sullivan, T.N., Wang, B., Espinosa, H.D., and Meyers, M.A. (2017). Extreme lightweight structures: avian feathers and bones. Materials today, 20(7): 377-391. doi:10.1016/j.mattod.2017.02.004

Sun, H., Skorgerbø, G., Wang, Z., Liu, W., and Li, Y. (2008). Structural Relationships between Highly Conserved Elements and Genes in Vertebrate Genomes. PLoS ONE, 3(11): e3727. doi:10.1371/journal.pone.0003727

Toews, D.P.L., Taylor, S.A., Streby, H.M., Kramer, G.R., and Lovette, I.J. (2019). Selection on *VPS13A* linked to migration in a songbird. PNAS, 116(37): 18272–18274. doi:10.1073/pnas.1909186116

USA National Phenology Network. (2017). Spring Indices, Current Year - First Bloom - Spring Index. https://www.usanpn.org/data/visualizations. Accessed on 17 April 2020.

Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era- concepts and misconceptions. Nature reviews genetics, 9: 255-266. doi:10.1038/nrg2322

Visser, M.E., te Marvelde, L., and Lof, M.E. (2012) Adaptive phenological mismatches of birds and their food in a warming world. J Ornithol., 153(Suppl 1):S75-S84. doi:10.1007/s10336-011-0770-6

Visser, M.E. (2019). Evolution: Adapting to a Warming World. Current Biology, 29(22): R1189-R1191. doi:10.1016/j.cub.2019.09.062

Visser, M.E., and Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. Nature Ecology & Evolution, 3: 879–885. doi:10.1038/s41559-019-0880-8

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An integrated tool for

comprehensive microbial variant detection and genome assembly improvement. PLoS ONE, 9(11): e112963. doi:10.1371/journal.pone.0112963

Walther, G., Post, E., Convey, P., Menzel, A., Parmesank, C., Beebee, T.J.C., Fromentin, J.M., Hoegh-Guldberg, O., and Bairlein, F. (2002). Ecological response to recent climate change. Nature, 416: 389–395. doi:10.1038/416389a

Wang, L., Jia, P., Wolfinger, R.D., Chen, X., and Zhao, Z. (2011). Gene set analysis of genome-wide association studies: Methodological issues and perspectives. Genomics, 98(1): 1-8. doi:10.1016/j.ygeno.2011.04.006

Williams, A., Spilianakis, C.G., and Flavell, R.A. (2010). Interchromosomal association and gene regulation in *trans*. Trends in Genetics, 26(4): 188–197. doi:10.1016/j.tig.2010.01.007

Wingfield, J.C. (2008). Comparative endocrinology, environment and global change. General and Comparative Endocrinology, 157(3): 207-216. doi:10.1016/j.ygcen.2008.04.017

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., Amin, N., Bunchkovich, M.L., Croteau-Chonka, D.C, … and Frayling, T.M. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nature genetics, 46: 1173-1186. doi:10.1038/ng.3097

Wotherspoon, S., Summer, M., and Lisovski, S. (2016). BAStag: Basic Data Processing for Light Based Geolocation Archival Tags. Version 0.1.3.

Yanaka, N., Imai, Y., Kawai, E., Akatsuka, H., Wakimoto, K., Nogusa, Y., Kato, N., Chiba, H., Kotani, E., Omori, K., and Sakurai, N. (2003). Novel Membrane Protein Containing

Glycerophosphodiester Phosphodiesterase Motif Is Transiently Expressed during Osteoblast Differentiation. Journal of Biological Chemistry, 278(44): 43595–43602. doi:10.1074/jbc.M302867200

Zhan, S., Zhang, W., Niitepõld, K., Hsu, J., Haeger, J.F., Zalucki, M.P., Altizer, A., de Roode, J.C., Reppert, S.M., and Kronforst, M.R. (2014). The genetics of monarch butterfly migration and warning colouration. Nature, 514: 317–321. doi:10.1038/nature13812

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., … Ji, H.P. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nature Biotechnology, 34(3): 303–311. doi:10.1038/nbt.3432

Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genetics, 9(2): e1003264. doi:10.1371/journal.pgen.1003264

Zhou, X., and Stephens, M. (2013). Genome-wide Efficient Mixed Model Analysis for Association Studies. Nature Genetics, 44(7): 821-824. doi:10.1038/ng.2310

Zhou, Z., Peng, P., Chen, J., Wu, X., Wang, Y., and Hong, Y. (2016). Identification of zebrafish magnetoreceptor and cryptochrome homologs. Sci China Life Sci 59(12): 1324-1331. doi:10.1007/s11427-016-0195-x

Table 2.1 Summary statistics of the purple martin draft reference genome assembly, including lengths, BUSCO scores, and number of genes identified in the genome annotation.

Reference genome metrics	
Genome size (bp)	1,165,951,862
Number of scaffolds	2,896
Longest scaffold (bp)	45,082,031
Shortest scaffold (bp)	16,249
Mean scaffold length (bp)	402,608
N50 scaffold length (bp)	6,129,949
L50 scaffold count	44
GC content (%)	43.06
N's per 100 kb	3.99
Complete BUSCOs	7592 (91.0%)
Complete single-copy BUSCOs	7431 (89.1%)
Complete duplicated BUSCOs	161 (1.9%)
Fragmented BUSCOs	118 (1.4%)
Missing BUSCOs	628 (7.6%)
Number of identified genes	12,686

Table 2.2 SNPs with PIP > 0.1 from BSLMMs for spring departure fall arrival phenotypes with respective chromosome number and closest genes within 100 kb; an asterisk (*) notes the SNP is inside the gene. LMM p-values are listed in -log10(p) format.

Spring departure

SNP	Chr	PIP	-log10(p)	Gene	Function
S214:738983	25	0.3751	5.2129	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein
S86:1071937	4	0.3475	2.2315	GDPD2	Glycerophosphodiester phosphodiesterase domain containing 2
S10:25072973	2	0.1385	2.7708	FAM221A	Family with sequence similarity 221 member A

Fall arrival

SNP	Chr	PIP	-log10(p)	Gene	Function
S86:1071937	4	0.3236	2.6375	GDPD2	Glycerophosphodiester phosphodiesterase domain containing 2
S214:738983	25	0.2953	3.7188	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein
S99:528805*	14	0.1521	0.4221	SNX8	Sorting nexin 8
S18:6522504	1	0.1375	2.1805	ADAMTS5	ADAM metallopeptidase with thrombospondin type 1 motif 5
S36:821560	3	0.1370	0.1363	IRAK1BP1	Interleukin 1 receptor associated kinase 1 binding protein 1
S22:13238871	1	0.1370	1.2813	CRY1	Cryptochrome circadian regulator 1
S38:2507638	7	0.1368	1.2370	FASTKD1, LOC101817811	FAST kinase domains 1, Uncharacterized protein

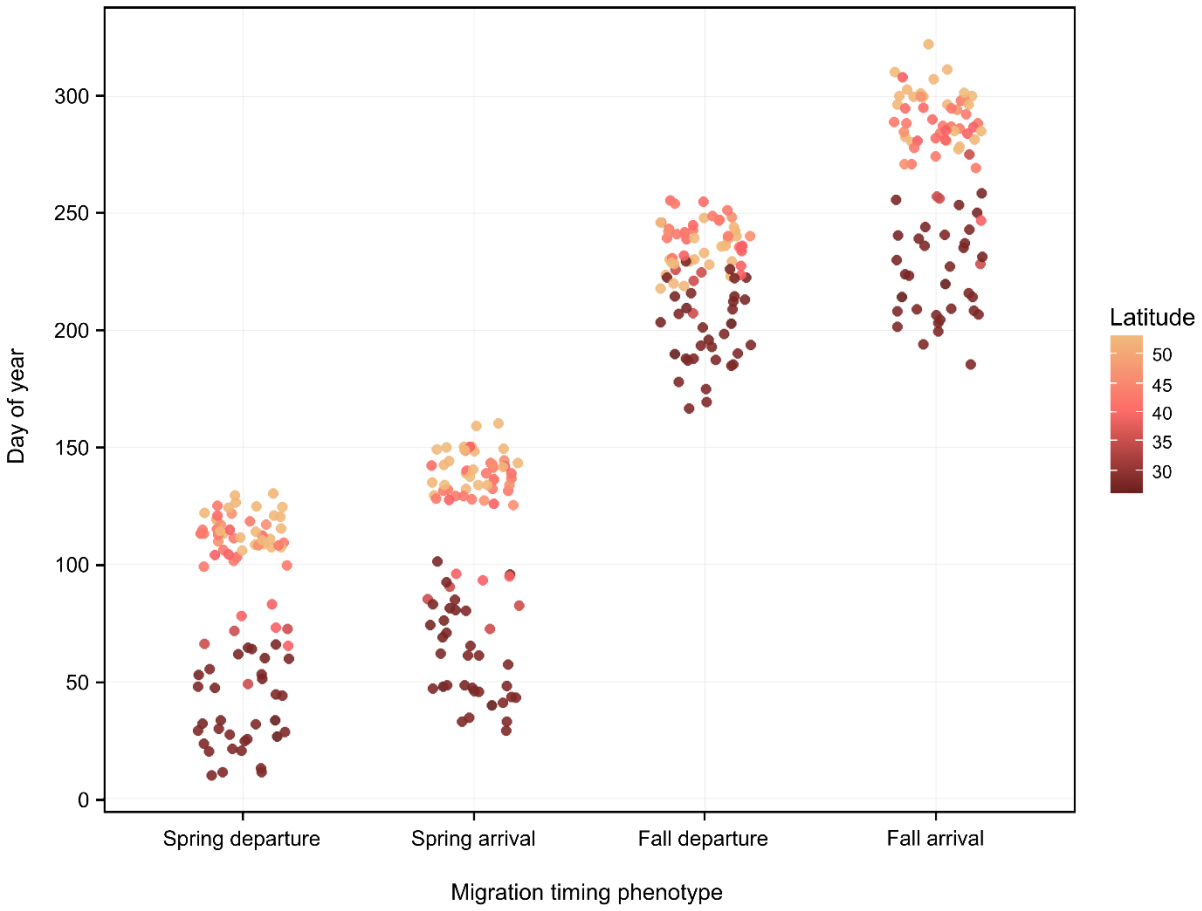

Figure 2.1 Migratory timing phenotypes of 89 geolocator-tracked purple martins, including the start and end of spring and fall migrations in Julian date format for each bird. Latitude is color coded, showing migration timing follows a latitudinal trend, where individuals breeding at higher latitudes exhibit later timing.

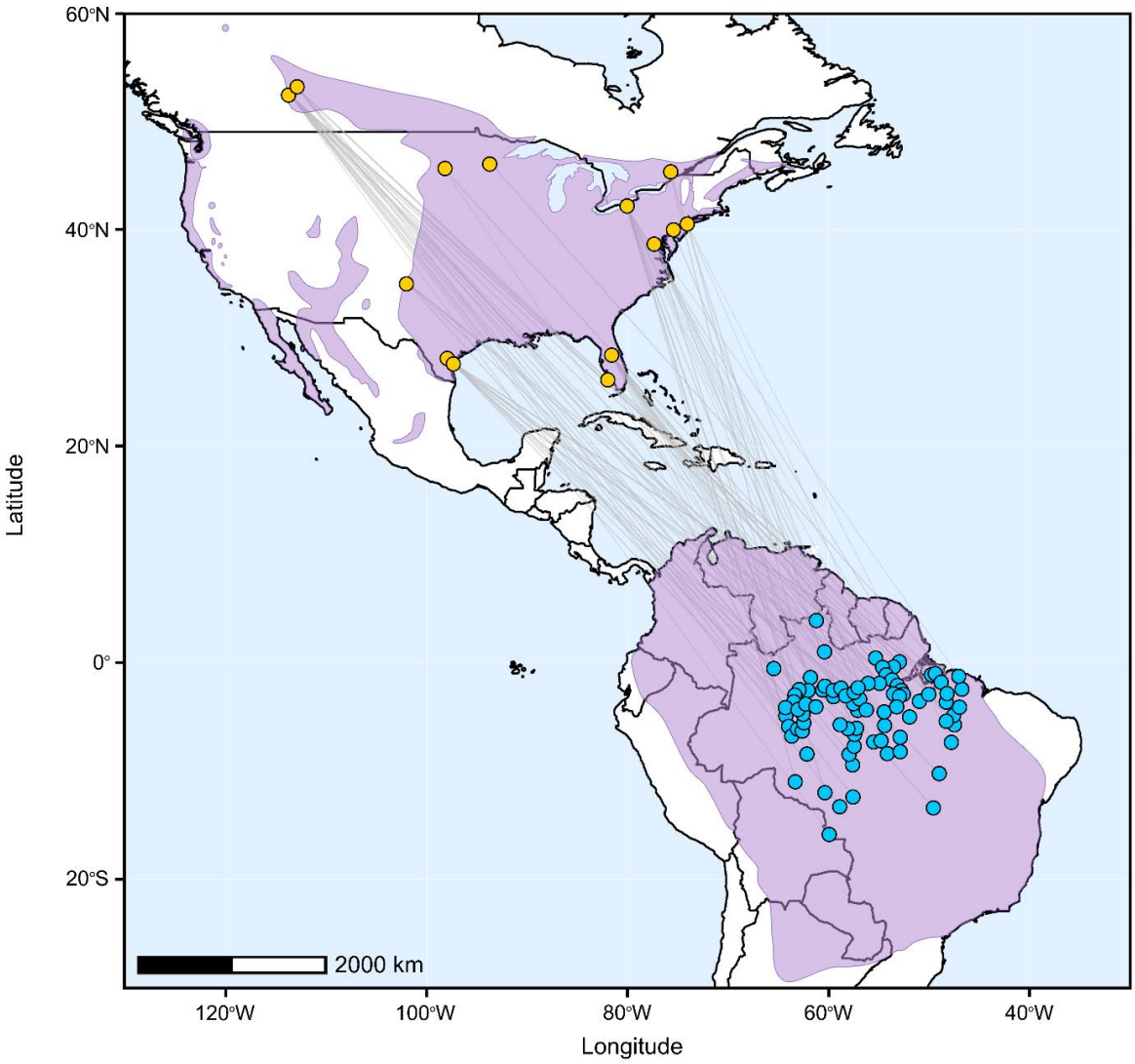

Figure 2.2 Purple martin breeding and wintering distribution, including sampling sites for 89 individuals in breeding range (orange) and their respective wintering destination before spring departure (blue). Grey lines represent the absolute distances of spring migration (ranging between 3,845 – 9,507 km).

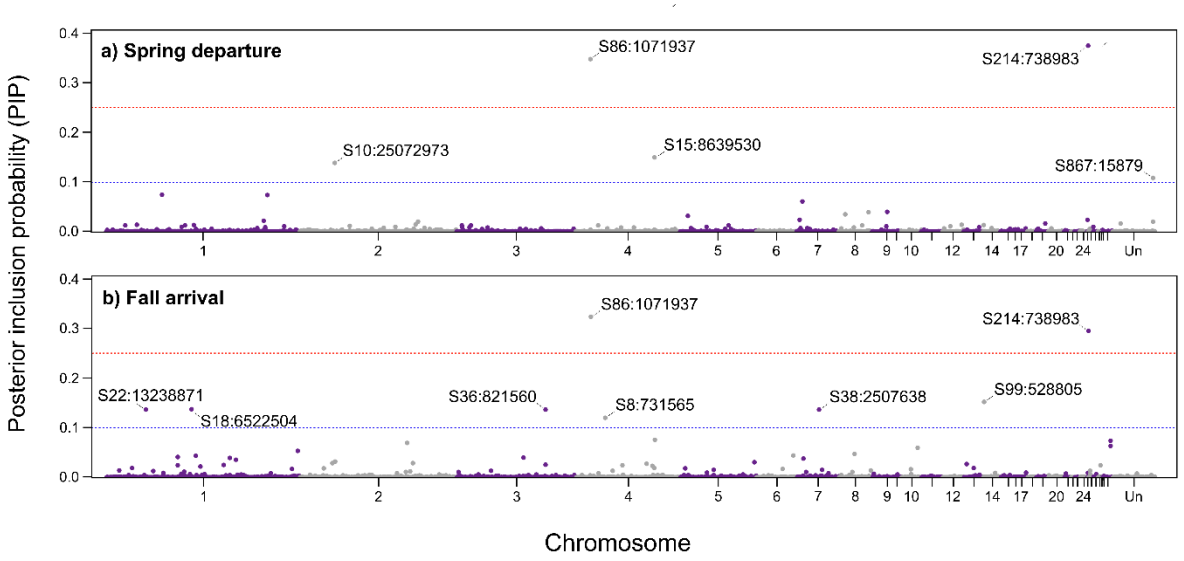

Figure 2.3 Bayesian Sparse Linear Mixed Models (BSLMM) of two migration timing phenotypes: (a) spring departure and (b) fall arrival (n=87). Above the blue threshold line represents PIP > 0.1, and above the red threshold line represents PIP > 0.25. Loci S214:738983 and S86:1071937 are supported in both models.

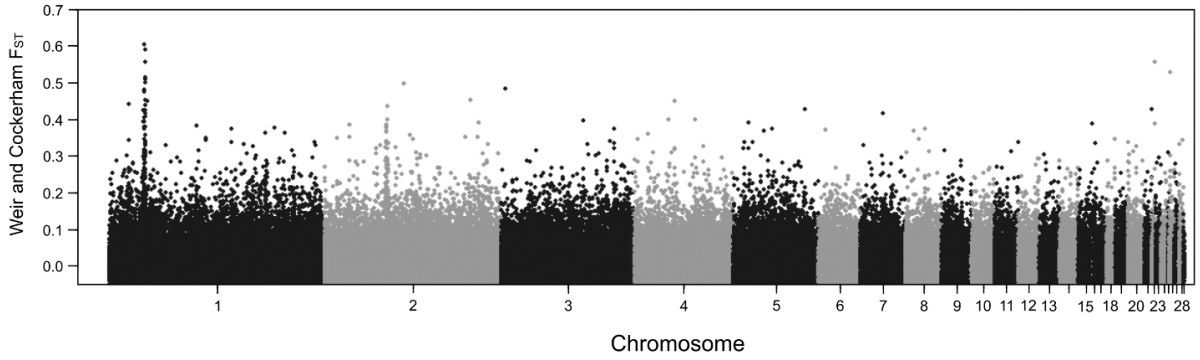

Figure 2.4 F_{ST} across autosomes in genome between the extreme 10 early and 10 late migrants in spring timing, including SNPs in linkage disequilibrium in 5kb non-overlapping windows. Elevation of F_{ST} is present in chromosome 1 and chromosome 2.

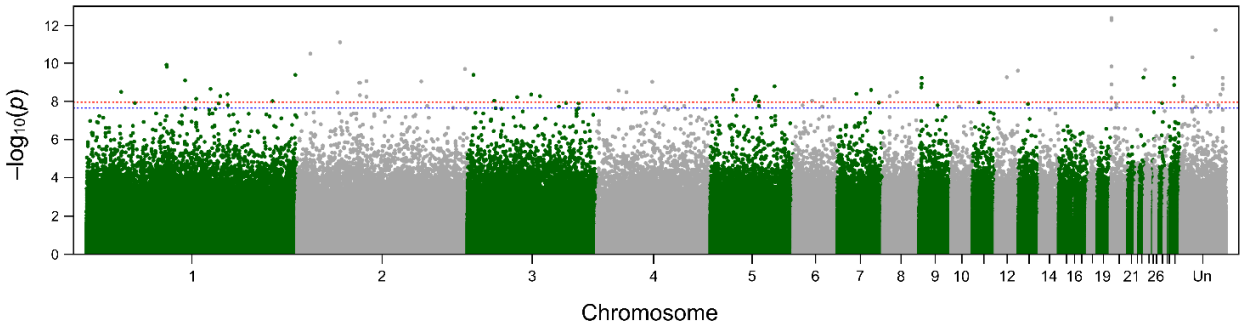

Figure 2.5 Latent Factor Mixed Model (LFMM) using first bloom data as ecological variable with martin genotypes (n=77). Blue threshold line represent FDR 0.1, and red threshold line represents FDR 0.05.

Chapter 3: Genetic differentiation between northern and southern breeding colonies of a declining aerial insectivore, the purple martin (*Progne subis*), and potential constraints on latitudinal natal dispersal

Abstract

Genetic variation is a fundamental component of biodiversity, and studying population structure, gene flow, and differentiation can help guide conservation strategies for many species. Like other aerial insectivores, the purple martin (*Progne subis*) is in decline, and yet very little is known about their genetic diversity and geographic patterns of differentiation and dispersal. In this study, I used next-generation sequencing strategies to fill this knowledge gap, integrating a draft reference genome and whole-genome resequencing data from four North American breeding populations for the most widespread subspecies of purple martin (*P.s. subis*). Additionally, I used continent-wide band encounter records to estimate natal dispersal and relate these patterns to genetic differentiation. I documented population structure across the breeding range and patterns of isolation-by-distance. Estimates of differentiation were ultimately very low (F_{ST}=0.001-0.009), but distinct between the most northern (Alberta) and more southern colonies. Additionally, I found greater longitudinal than latitudinal dispersal distances, consistent with the patterns of population genetic differentiation, signifying potential latitudinal constraints on gene flow are possibly driven by phenological adaptations. These findings demonstrate the importance to conserve northern and southern populations separately to maximize genetic diversity and adaptive potential in the purple martin.

Introduction

Genetic variation is the basis for evolutionary change and has led to the diversity in morphological and behavioural patterns we see in animals today. Understanding patterns of population structure and gene flow across populations is important for anticipating species' adaptive potential and defining units for conservation (Funk et al. 2012). Birds across the world are facing threats such as rapidly changing climates and habitat loss, resulting in steep population declines and putting species at risk for loss of genetic diversity (Rosenberg et al. 2019). Determining conservation units is important for genetic rescue of vulnerable populations and maintaining genetic variation to avoid inbreeding and extinction (Coates et al. 2018). For instance, genetic differentiation patterns observed in the blue duck (*Hymenolaimus malacorhynchos*) led to recommendations for separate management of northern and southern populations for this endangered species (Grosser et al. 2017). Genomic analyses in the willow flycatcher (*Empidonax traillii*) identified a vulnerable subspecies at highest risk of climate-induced extinction, providing direction for targeted recovery work (Ruegg et al. 2018). In another example, low population structure found in regent honeyeaters (*Anthochaera phrygia*) helped guide conservation plans for managing release of captive birds into wild populations to maximize effective population size (Kvistad et al. 2015). These types of conservation actions help maintain genetic diversity of populations and their persistence through changing environments.

Spatial patterns in populations change over time and can influence a species' evolutionary trajectory. Habitat stability in the tropics has allowed more time for genetic drift and local adaptation in tropical species, broadly leading to greater speciation and population differentiation rates compared to birds in temperate areas (Harvey et al. 2017). In recent

evolutionary history, Pleistocene glacial cycles have impacted diversity and distributions of widespread species in temperate regions (Avise & Walker 1998; Holder et al. 1999). Population contractions persisting in refugia during glacial maxima followed by rapid range expansions can result in low levels of differentiation within regions (Colbeck et al. 2008; Milá et al. 2000; Milot et al. 2000; Ruegg & Smith 2002). Many phylogenetic studies revealed significant east to west differentiation stemming from two or more refugia during glaciation divided by the Rocky Mountains or Great Plains (Kimura et al. 2002; Lovette et al. 2004; Peters et al. 2005), such as seen in the yellow warbler (*Dendroica petechia,* Boulet & Gibbs 2006; Milot et al. 2000) and the Swainson's thrush (*Catharus ustulatus,* Ruegg & Smith 2002). Other species like the American redstart (*Setophaga ruticilla*) do not exhibit this pattern, instead they appear to have originated from a single refugia in eastern North America (Colbeck et al. 2008). In addition to geographical barriers, genetic differentiation can be influenced by isolation-by-distance, a pattern of decreased relatedness with increasing geographical distance (Malécot 1975; Wright 1942). Isolation-by-distance relationships can reveal patterns of dispersal and gene flow on contemporary timescales (Aguillon et al. 2017; Gibbs et al. 2000), which have important implications for population connectivity and genetic variation that we see in wildlife populations today (Grosser et al. 2017).

The purple martin (*Progne subis*) is one of the most widely distributed species of migratory songbirds in North America, occurring on both sides of the Rocky Mountains across three subspecies. Genetic analyses have been conducted on two of the subspecies, showing significant differentiation between the eastern *P.s. subis* subspecies and the western *P.s. arboricola* subspecies, which likely split in the second half of the Pleistocene (Baker et al. 2008). Baker et al. (2008) used mitochondrial haplotypes from eight sites in *P.s. arboricola* and *P.s. subis* subspecies to address genetic diversity in a vulnerable British Columbia population, and

found a high level of gene flow within western populations but with some mixing via rare migrants from the east. Although *P.s. subis* is the most widespread purple martin subspecies, with colonies ranging from Alberta to Florida, no previous study has examined genetic patterns covering both southern and northern populations together. *P.s. subis* sample sites in Baker et al. (2008) were limited to the northern portion of the range, and a study using microsatellite markers in this subspecies was limited to one population in Virginia (Stanley et al. 2011). The purple martin is among many aerial insectivores that have been declining more steeply than other passerine groups over the past 50 years (NABCI-Canada 2019; Nebel et al. 2010; Sauer et al. 2017), with some colonies declining more rapidly than others (Tautin et al. 2009). Thus, it is important to understand the genetic patterns in the nominate subspecies to identify distinct groups that may warrant population-specific conservation strategies.

I used next-generation sequencing to examine patterns of population structure and differentiation across several breeding colonies of the eastern *P.s. subis* subspecies. Purple martins are long-distance migrants and individuals from the *P.s. subis* subspecies likely originated from the same eastern refugia (Baker et al. 2008), so I expected the overall population differentiation to be relatively low and to recover a pattern of isolation-by-distance. However, because purple martin adults exhibit high breeding site fidelity (Brown & Tarof 2020; Stutchbury et al. 2009), I also expected to find some population structure across the breeding range. Very little is known about natal dispersal in most songbirds because juveniles are difficult to track and generally have higher mortality than adults (McKim-Louder et al. 2013; Sullivan 1989). However, collective banding of purple martins over many years has supplied a large dataset of encounter records, which provided a unique opportunity to examine patterns of natal dispersal. Since natal dispersal can have significant impacts on gene flow between populations, I

also compared patterns of population differentiation with those of natal dispersal across the range.

Methods

Sampling and genotyping

This study included samples of *P.s. subis* from across North America: 71 adults (35 males, 36 females) from six different breeding colonies representing Alberta, Pennsylvania, Texas, and Florida populations (Table 3.1), and one sample from Manitoba for the reference genome (described in Chapter 2). Birds were captured in bird houses while feeding their young, by using drop door traps held with fishing line or by an extendable pole covering the cavity entrance. Up to 150uL of blood was drawn from the bird's brachial vein and then stored in Queen's Lysis buffer. After DNA extraction and sequencing, I assembled and annotated the purple martin draft reference genome – the backbone to genotype the samples in this study. The genetic variants for the 71 samples were called from the reference in the form of single-nucleotide polymorphisms (SNP). I filtered the SNPs to remove low-quality sites, sex-linked scaffolds, and SNPs in linkage disequilibrium to create a final dataset to use in analyses. Details on methodology for extraction, sequencing, assembly, annotation, genotyping, and filtering are described in Chapter 2 and Appendix A.

Population structure and differentiation

I conducted a Principal Component Analysis (PCA) to examine population structure using *smartpca* in the *EIGENSOFT* program (Price et al. 2006). In addition, I used *ADMIXTURE* to estimate genetic ancestry (Alexander et al. 2009). For population differentiation, I estimated

Weir and Cockerham fixation indices (F_{ST}) for each population pair (Weir and Cockerham 1984) using *VCFtools* (Danecek et al. 2011). To further evaluate patterns of isolation-by-distance, I completed a Mantel test through R-package *ade4* (Dray & Dufour 2007) to assess correlation between F_{ST} and colony differences in distances. Further details on analyses are in Appendix B.

Natal dispersal and genomic differentiation

To determine how gene flow relates to natal dispersal, I compared estimates of natal dispersal with correlations between genetic differentiation across latitude and longitude. Through the North American Bird Banding Program (NABBP 2018), I obtained encounter records of purple martins across North America between 1922 – 2018 (n=3,928), including both western and eastern groups but largely consisting of *P.s. subis* individuals (n=3,631). I filtered the dataset to include only natal breeding dispersal records, using the following criteria: records between May to July (to exclude migratory encounters), initial banding age no older than hatch-year, and encounters at least one year following banding year (see Appendix B for details on filtering criteria). Using initial banding and encounter locations, I calculated dispersal distances using R-package *geosphere*, which incorporated the spherical coordinate system (Hijmans et al. 2014). After excluding individuals that did not disperse, I categorized records as longitudinal or latitudinal dispersers based on the proportion of movements in cardinal directions (south-north and west-east) using trigonometric calculations between banding and encounter locations (see Appendix B for details). Next, I used population differentiation data determined from the previous section to complete two additional Mantel tests examining correlations between F_{ST} and colony differences in latitude and longitude.

Results

Population structure and differentiation

The PCA clustered individuals from each of the four populations of *P.s. subis* separately, with a clear distinction between Alberta and the remaining three populations (Figure 3.1b). Although 71 individuals were used for this study, I removed four outliers from the Florida population that displayed extremely disproportionate variation in both PC axes here and in subsequent analyses (Figure B1; more details in Appendix B). Results from *ADMIXTURE* further supported the distinction of the Alberta population. Although assuming one ancestral population (K=1) presented the lowest CV error of 0.809, the admixture results using K=2 (CV error=0.850) are consistent with the results from the PCA and indicate introgression from the cluster limited to Alberta is also present in Pennsylvania, Texas, and Florida (Figure B2). Population differentiation was low between all populations, with F_{ST} values ranging from 0.0011 to 0.0090 (Figure 3.2a). Similar to results from the PCA, the highest F_{ST} estimates came from pairwise comparisons with Alberta. I recovered a clear picture of isolation-by-distance across these populations (Figure 3.2b), with a correlation coefficient of 0.894 (Mantel test, p-value 0.038) between F_{ST} and colony distances.

Natal dispersal and genomic differentiation

I filtered the purple martin encounter banding records from 3,928 to 2,060 individuals to include only data pertaining to natal dispersal (individuals banded as nestlings or juveniles and encountered again in a later breeding season). In the filtering process, I removed one outlier, an encounter found dead outside the breeding range along the Rocky Mountains. On average, juveniles dispersed 64 km, ranging from 0 to 2,644 km. Most juveniles (70.9%) dispersed 50 km

or less from their natal site, whereas only 1.6% dispersed farther than 500 km (Table 3.2, Figure 3.3a). Among these top dispersers (n=33), 29 travelled longitudinally with an average of 909 km (26,353 km total; SD 482 km) and only four travelled latitudinally with an average distance of 523 km (2,093 km total; SD 191 km) shown in Figure 3.3b, displaying greater longitudinal movements in natal dispersal. This result was consistent with population differentiation patterns, where Mantel tests revealed correlations of 0.678 (p-value 0.122) between F_{ST} and longitude, and 0.734 (p-value 0.079) between F_{ST} and latitude, indicating higher genetic differentiation by shifting latitude and greater gene flow across longitudes.

Discussion

I examined populations of *P.s. subis* across most of their breeding range using next-generation sequencing and continent-wide banding records. Overall levels of differentiation were quite low and the principal components themselves accounted for very little variance, suggesting gene flow is high across the subspecies. However, I documented evidence of some population structure, mostly between Alberta and southern populations, that is likely related in large part to geographic distance and high site fidelity exhibited in purple martins (Brown & Tarof 2020; Stutchbury et al. 2009). This suggestion was also supported by a pattern of isolation-by-distance. I discuss these findings in light of conservation concerns for purple martins, and their geographic history and patterns of natal dispersal.

A species' adaptive potential is largely determined by population sizes and their genetic variance (Bourne et al. 2014), which are essential to conserve and maintain in the face of environmental and anthropogenic impacts. The purple martin is one of the steeply declining aerial insectivores in North America (Nebel et al. 2010), vulnerable to many factors such as

changes in insect abundances, contaminants, habitat loss, and phenological shifts from warming temperatures (Spiller & Dettmers 2019). Genetic differentiation among populations, in contrast to panmictic groups (Kvistad et al. 2015), has implications for classification decisions in conservation listings, such as prescribed in studies for vulnerable species (Friesen et al. 2006; Grosser et al. 2017). My results show that northern (Canada) and southern (United States) *P.s. subis* breeding colonies exhibit the largest differentiation, indicating that these groups should be classified as distinct conservation units, with separate population management for conserving this species' genetic diversity. Specifically, populations in Alberta and Florida that represent the extremes of the breeding range display the strongest genetic differentiation, suggesting these should be prioritized. The observed genetic variation may include alleles related to behaviours that have geographical or latitudinal constraints, such as migration timing, which are important to maintain for providing necessary genetic variation for selection to act on in response to environmental changes (Bourne et al. 2014; Lande & Shannon 1996).

Large geographic events shape evolutionary histories and impact present-day diversity (Lovette 2005). The split between *P.s. subis* and *P.s. arboricola* is suggested to have occurred during the Pleistocene (Baker et al. 2008). My results expand on the genetic patterns in purple martins throughout the largest portion of their range, displaying isolation-by-distance and overall low genetic differentiation within *P.s. subis*. These are similar to patterns observed within their western counterpart *P.s. arboriocla* (Baker et al. 2008), and together these studies support the presence of both eastern and western refugium during the Pleistocene and post-glacial range expansions northward. The purple martin is among many widespread North American migratory birds whose evolutionary history shows presence of two glacial refugia, separated by geographical barriers such as the Great Plains and Rocky Mountains (Milot et al. 2000; Peters et

al. 2005; Ruegg & Smith 2002). This likely led to the observed differences in distribution, behaviours, and morphology in purple martin subspecies (Brown & Tarof 2020). For example, *P.s. arboricola* breeds in the western edges of California through British Columbia and is larger compared to *P.s. subis*, and *P.s. subis* is more dominantly colonial during breeding (Brown & Tarof 2020). Although there is still little known about the genetic history of the third subspecies, *P.s. hesperia*, breeding in south-western United States and Mexico (Brown & Tarof 2020), the use of natural cacti cavities and morphological similarities with *P.s. arboricola* (Brown & Tarof 2020) could suggest that this subspecies may not have stemmed from the same eastern refugia as *P.s. subis*.

In species where adults exhibit high site fidelity, it is especially important to address natal dispersal because it can significantly impact gene flow between populations. My results show that on average, young purple martins disperse short distances from their natal site when returning to the breeding range, which is consistent with the general pattern of natal dispersal in terrestrial birds (Sutherland et al. 2000). However, about a third of the juveniles dispersed more than 50 km, with several travelling over 2,000 km, displaying a large range in dispersal movements. Although purple martins from widespread breeding populations overwinter in mixed flocks (Fraser et al. 2012; Fraser et al. 2017), juveniles returning to the breeding grounds as second-year adults arrive later than experienced breeders who may outcompete them in claiming breeding territories (Brown & Tarof 2020). This could cause some juveniles to disperse farther to find necessary breeding habitat.

Among the dispersing individuals, my results show a higher quantity of longitudinal dispersals (moving on the east-west scale), which could indicate latitudinal constraints on movements and gene flow. When comparing this pattern with the genomics portion of this study,

it is interesting to observe similar support for greater longitudinal movements from correlations between F_{ST} with latitude and longitude. Although the total number of populations in the genetic differentiation analyses may limit the power to thoroughly assess patterns of directional gene flow, the distribution of sampled populations covers a large portion of the breeding range. Even with high gene flow overall, the higher correlation with differentiation and latitude (compared to longitude) in conjunction with greater longitudinal natal dispersal could indicate that there are limitations on latitudinal movements. These limitations could be related to constraints on spring phenology. It is understood that plant phenology, such as first bloom, generally follows a latitudinal trend across North America (Gerst et al. 2020), which is also an indicator of timing in seasonal resources that migratory birds need during the breeding season (Lehikoinen et al. 2019). Purple martins are known to follow a latitudinal gradient in spring arrival timing where southern colonies arrive earlier than northern ones (Brown & Tarof 2020) and potentially rely largely on endogenous cues tied with spring phenology.

This study sets up the genetic background for the eastern *P.s. subis* subspecies, providing data on four key populations distributed throughout the breeding range and mechanisms driving differentiation between northern and southern colonies. The genetic distinction between Alberta and the remaining populations demonstrate that northern and southern groups represent different conservation units and require separate management to capture the majority of *P.s. subis* genetic diversity. Although the rapid range expansion after refugia during the Pleistocene glaciation is recent in evolutionary time, this widespread species must now adapt quickly to global environmental changes. The latitudinal constraints on natal dispersal and differentiation further support the need to conserve northern and southern populations separately, which will help maximize the purple martin's adaptive potential.

Acknowledgements

I am grateful to collaborators J. Barrow, L. Chambers, P. Clifton, J. Fischer, G. Hvenegaard, P. Kramer, M. Pearman, A. Ritchie, A. Savage, J. Siegrist, C. Silverio, and J. Tautin for assistance in collecting samples; B. Stutchbury for assistance with both sample collection and organization; K. Jeffries for sharing his laboratory space and equipment for DNA extractions; Alexander Suh, National Genomics Infrastructure in Sweden, Science for Life Laboratory, and Texas A&M AgriLife Genomics and Bioinformatics Service for sequencing resources; Texas A&M High Performance Research Computing and Compute Canada for computing services; North American Bird Banding Program for providing encounter records; M. Thorstensen for assistance with bioinformatics, analyses, and editorial support. This work was supported by funding from the University of Manitoba, Texas A&M University, Natural Sciences and Engineering Research Council Discovery Grant program, Research Manitoba, the Society of Canadian Ornithologists, and Sigma Xi.

References

Aguillon, S.M., Fitzpatrick, J.W., Bowman, R., Schoech, S.J., Clark, A.G., Coop, G., and Chen, N. (2017). Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLoS Genetics, 13(8): e1006911. doi:10.1371/journal.pgen.1006911

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

Avise, J.C., and Walker, D. (1998). Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. R. Soc. Lond. B., 265: 457-463. doi:10.1098/rspb.1998.0317

Baker, A.J., Greenslade, A.D., Darling, L.M., and Finlay, J.C. (2008). High genetic diversity in the blue-listed British Columbia population of the purple martin maintained by multiple sources of immigrants. Conservation Genetics, 9: 495–505. doi:10.1007/s10592-007-9358-3

Boulet, M., and Gibbs, L. (2006). Lineage origin and expansion of a Neotropical migrant songbird after recent glaciation events. Molecular Ecology, 15(9): 2505-2525. doi:10.1111/j.1365-294X.2006.02956.x

Bourne, E.C., Bocedi, G., Travis, J.M.J., Pakeman, R.J., Brooker, R.W., and Schiffers, K. (2014). Between migration load and evolutionary rescue: dispersal, adaptation and the response of spatially structured populations to environmental change. Proc. R. Soc. B., 281: 20132795. doi:10.1098/rspb.2013.2795

Brown, C.R. and Tarof, S. (2020). Purple Martin (*Progne subis*), version 1.0. In Birds of the World (A.F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. doi:10.2173/bow.purmar.01

Coates, D.J., Byrne, M., and Moritz, C. (2018). Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. Frontiers in Ecology and Evolution, 6(165): 1–13. doi:10.3389/fevo.2018.00165

Colbeck, G.J., Gibbs, H.L., Marra, P.P., Hobson, K., and Webster, M.S. (2008). Phylogeography of a widespread North American migratory songbird (*Setophaga ruticilla*). Journal of Heredity, 99(5): 453–463. doi:10.1093/jhered/esn025

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

Dray, S. and Dufour, A.B. (2007). The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software, 22(4): 1–20. doi:10.18637/jss.v022.i04

Fraser, K.C., Stutchbury, B.J.M., Silverio, C., Kramer, P.M., Barrow, J., Newstead, D., Mickle, N., Cousens, B.F., Lee, J.C., Morrison, D.M., Shaheen, T., Mammenga, P., Applegate, K., and Tautin, J. (2012). Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. Proc. R. Soc. B., 279(1749): 4901–4906. doi:10.1098/rspb.2012.2207

Fraser, K.C., Shave, A., Savage, A., Ritchie, A., Bell, K., Siegrist, J., Applegate, K., and Pearman, M. (2017). Determining fine-scale migratory connectivity and habitat selection for a migratory songbird by using new GPS technology. Journal of Avian Biology, 48(3): 001-007. doi:10.1111/jav.01091

Friesen, V.L., González, J.A., and Cruz-Delgado, F. (2006). Population genetic structure and conservation of the Galápagos petrel (*Pterodroma phaeopygia*). Conservation Genetics, 7: 105-115. doi:10.1007/s10592-005-8704-6

Funk, C.W., McKay, J.K., Hohenlohe, P.A., and Allendorf, F.W. (2012). Harnessing genomics for delineating conservation units. Trends Ecol. Evol., 27(9): 489-496. doi:10.1016/j.tree.2012.05.012

Gerst, K.L., Crimmins, T.M., Posthumus, E.E., Rosemartin, A.H., and Schwartz, M.D. (2020). How well do the spring indices predict phenological activity across plant species? International Journal of Biometeorology, 64: 889–901. doi:10.1007/s00484-020-01879-z

Gibbs, H.L, Dawson, R.J.G., and Hobson, K.A. (2000) Limited differentiation in microsatellite DNA variation among northern populations of the yellow warbler: evidence for male-biased gene flow? Molecular Ecology, 9: 2137-2147

Grosser, S., Abdelkrim, J., Wing, J., Robertson, B.C., and Gemmell, N.J. (2017). Strong isolation by distance argues for separate population management of endangered blue duck (*Hymenolaimus malacorhynchos*). Conservation Genetics, 18: 327–341. doi:10.1007/s10592-016-0908-4

Harvey, M.G., Seeholzer, G.F., Smith, B.T., Rabosky, D.L., Cuervo, A.M., and Brumfield, R.T. (2017). Positive association between population genetic differentiation and speciation rates in New World birds. PNAS, 114(24): 6328–6333. doi:10.1073/pnas.1617397114

Hijmans, R.J., Williams, E., and Vennes, C. (2014). Geosphere: spherical trigonometry. R package version 1.3-11

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

Kimura, M., Clegg, S.M., Lovette, I.J., Holder, K.R., Girman, D.J., Milá, B., Wade, P., and Smith, T.B. (2002). Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (*Wilsonia pusilla*). Molecular Ecology, 11(9): 1605-1616. doi:10.1046/j.1365-294X.2002.01551.x

Kvistad, L., Ingwersen, D., Pavlova, A., Bull, J.K., and Sunnucks, P. (2015). Very low population structure in a highly mobile and wide-ranging endangered bird species. PLoS ONE 10(12): 1–20. doi:10.1371/journal.pone.0143746

Lande, R. and Shannon, S. (1996). The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment. Evolution, 50(1): 434–437.

Lehikoinen, A., Lindén, A., Karlsson, M., Andersson, A., Crewe, T.L., Dunn, E.H., Gregory, G., Karlsson, L., …Tjørnløv, R S. (2019). Phenology of the avian spring migratory passage in Europe and North America: Asymmetric advancement in time and increase in duration. Ecological Indicators, 101: 985–991. doi:10.1016/j.ecolind.2019.01.083

Lovette, I.J., Clegg, S.M., and Smith, T.B. (2004). Limited utility of mtDNA markers for determining connectivity among breeding and overwintering locations in three neotropical migrant birds. Conservation Biology, 18(1): 156-166. doi:10.1111/j.1523-1739.2004.00239.x

Lovette, I.J. (2005). Glacial cycles and the tempo of avian speciation. Trends in Ecology and Evolution, 20(2): 57–59. doi:10.1016/j.tree.2004.11.011

Malécot, G. (1975). Heterozygosity and relationship in regularly subdivided populations. Theoretical Population Biology, 8(2): 212–241. doi:10.1016/0040-5809(75)90033-7

McKim-Louder, M.I., Hoover, J.P., Benson, T.J., and Schelsky, W.M. (2013). Juvenile Survival in a Neotropical Migratory Songbird Is Lower than Expected. PLoS ONE, 8(2): e56059. doi:10.1371/journal.pone.0056059

Milá, B., Girman, D.J., Kimura, M., and Smith, T.B. (2000). Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proc. R. Soc. Lond. B., 267: 1033-1040. doi:10.1098/rspb.2000.1107

Milot, E., Gibbs, H.L., and Hobson, K.A. (2000). Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Molecular Ecology, 9: 667–681. doi:10.1046/j.1365-294X.2000.00897.x

NABCI Canada. (2019). The State of Canada's Birds, 2019. Environment and Climate Change Canada, Ottawa, Canada.

Nebel, S., Mills, A., McCracken, J.D., and Taylor, P.D. (2010). Declines of Aerial Insectivores in North American Follow a Geographic Gradient. Avian Conservation and Ecology, 5(2). doi:10.5751/ACE-00391-050201

North American Bird Banding Program. (2018). Patuxent Wildlife Research Center.

Peters, J.L., Gretes, W., and Omland, K.E. (2005). Late Pleistocene divergence between eastern and western populations of wood ducks (*Aix sponsa*) inferred by the "isolation with migration" coalescent method. Molecular Ecology, 14: 3407–3418. doi:10.1111/j.1365-294X.2005.02618.x

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8): 904–909. doi:10.1038/ng1847

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., and Marra, P.P. (2019). Decline of the North American avifauna. Science, 366: 120–124. doi:10.1126/science.aaw1313

Ruegg, K.C. and Smith, T.B. (2002). Not as the crow flies: A historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). Proc. R. Soc. Lond. B., 269: 1375–1381. doi:10.1098/rspb.2002.2032

Ruegg, K., Bay, R.A., Anderson, E.C., Saracco, J.F., Harrigan, R.J., Whitfield, M., Paxton, E.H., and Smith, T.B. (2018). Ecological genomics predicts climate vulnerability in an endangered southwestern songbird. Ecology Letters, 21: 1085–1096. doi:10.1111/ele.12977

Rushing, C.S., Royle, J.A., Ziolkowski, D.J., and Pardieck, K.L. (2020). Migratory behavior and winter geography drive differential range shifts of eastern birds in response to recent climate change. PNAS, 117(23): 12897-12903. doi:10.1073/pnas.2000299117

Sauer, J.R., Pardieck, K.L., Ziolkowski, D.J., Smith, A.C., Hudson, M.A.R., Rodriguez, V., Berlanga, H., Niven, D.K., and Link, W.A. (2017). The first 50 years of the North American Breeding Bird Survey. The Condor, 119: 576-593. doi:10.1650/CONDOR-17-83.1

Spiller, K.J. and Dettmers, R. (2019). Evidence for multiple drivers of aerial insectivore declines in North America. The Condor, 121: 1–13. doi:10.1093/condor/duz010

Stanley, A.A., Graham, A.M., Wagner, R.H., Morton, E.S., and Schug, M.D. (2011). Characterization of 10 polymorphic microsatellite markers for the purple Characterization of 10 polymorphic microsatellite markers for the purple martin, *Progne subis*, Conservation Genetics Resources, 3: 577-579. doi:10.1007/s12686-011-9408-5

Stutchbury, B.J.M., Hill, J.R., Kramer, P.M., Rush, S.A., and Tarof, S.A. (2009). Sex and age-specific annual survival in a neotropical migratory songbird, the purple martin (Progne subis). The Auk, 126(2): 278-287. doi:10.1525/auk.2009.08038

Sullivan, K.A. (1989). Predation and Starvation: Age-Specific Mortality in Juvenile Juncos (*Junco phaenotus*). Journal of Animal Ecology, 58(1): 275-286.

Sutherland, G.D., Harestad, A.S., Price, K., and Lertzman, K.P. (2000). Scaling of natal dispersal distances in terrestrial birds and mammals. Conservation Ecology, 4(1).

Tautin, J., Cousens, B., Kostka, K., Kotska, S., and Airola, D.A. (2008). Addressing regional declines in Purple Martin populations. Partners in Flight Conference: Tundra to Tropics, 82–87.

Wright, S. (1942). Isolation by Distance. Genetics, 28: 114–138.

Table 3.1 Site locations for purple martin samples (n=71) across six breeding colonies, collectively representing four populations.

Population	Colony	Latitude	Longitude	N	Year(s)
Alberta	Camrose	53.011	-112.864	5	2012-2014
Alberta	Lacombe	52.391	-113.612	17	2012-2014
Florida	Bay Lake	28.360	-81.588	23	2013-2015
Pennsylvania	Erie	42.115	-80.145	14	2008-2013
Texas	Corpus Christi	27.680	-97.407	4	2013
Texas	Sandia	27.997	-97.879	8	2013

Table 3.2 Number of purple martin natal dispersals in breeding distribution in a given distance interval, with proportion of observed dispersal distance from a total of 2060 records.

Distance	Number	Percent (%)
$0 \leq 20$ km	1051	51.0%
$20 \leq 50$ km	410	19.9%
$50 \leq 100$ km	289	14.0%
$100 \leq 200$ km	144	7.0%
$200 \leq 500$ km	133	6.5%
$500 \leq 1{,}000$ km	27	1.3%
> 1,000 km	6	0.3%

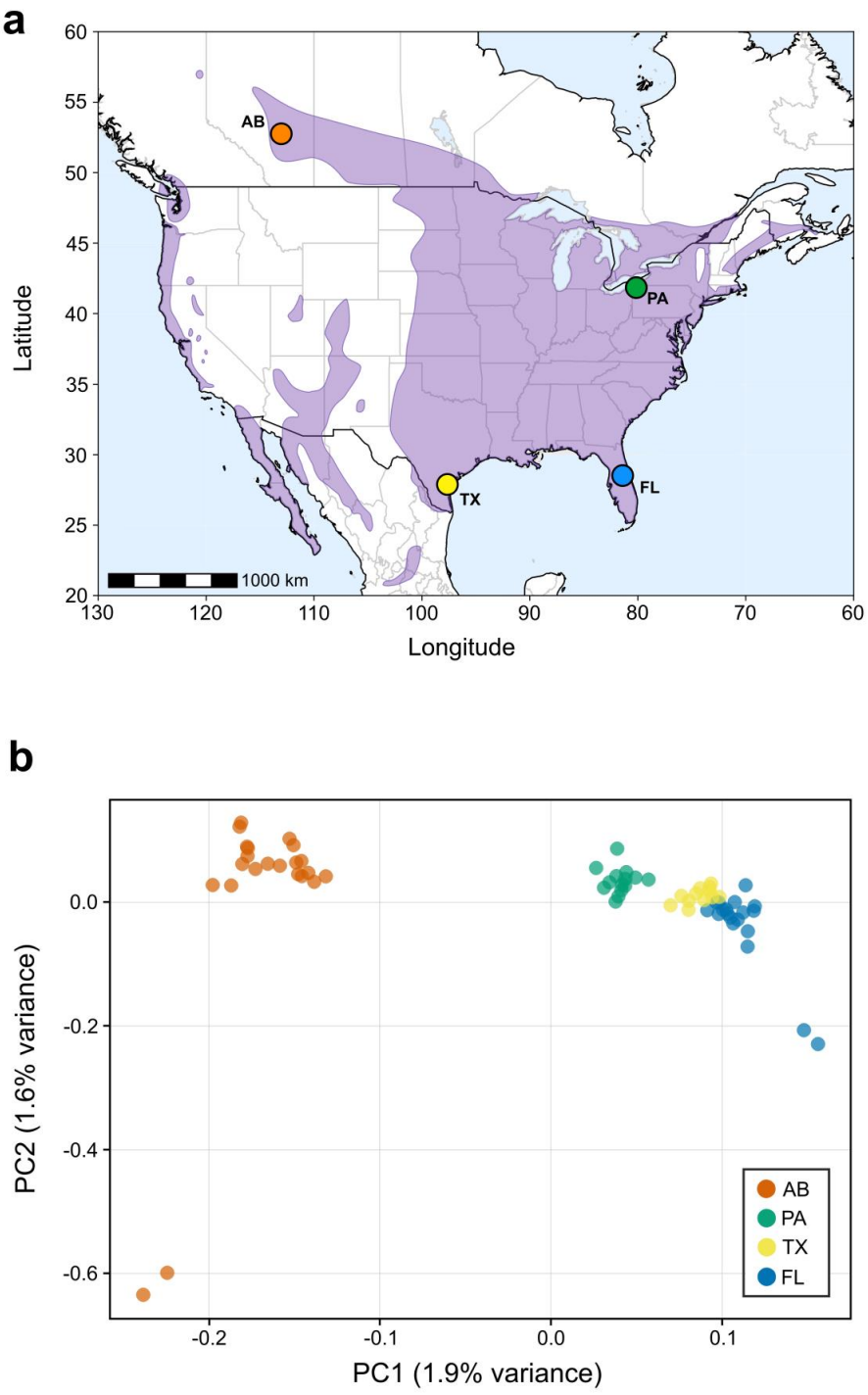

Figure 3.1 (a) Breeding distribution and four sampling sites for purple martins in study. (b) Results from a principal component analysis (n=67), displaying population structure (AB=Alberta, PA=Pennsylvania, TX=Texas, FL=Florida).

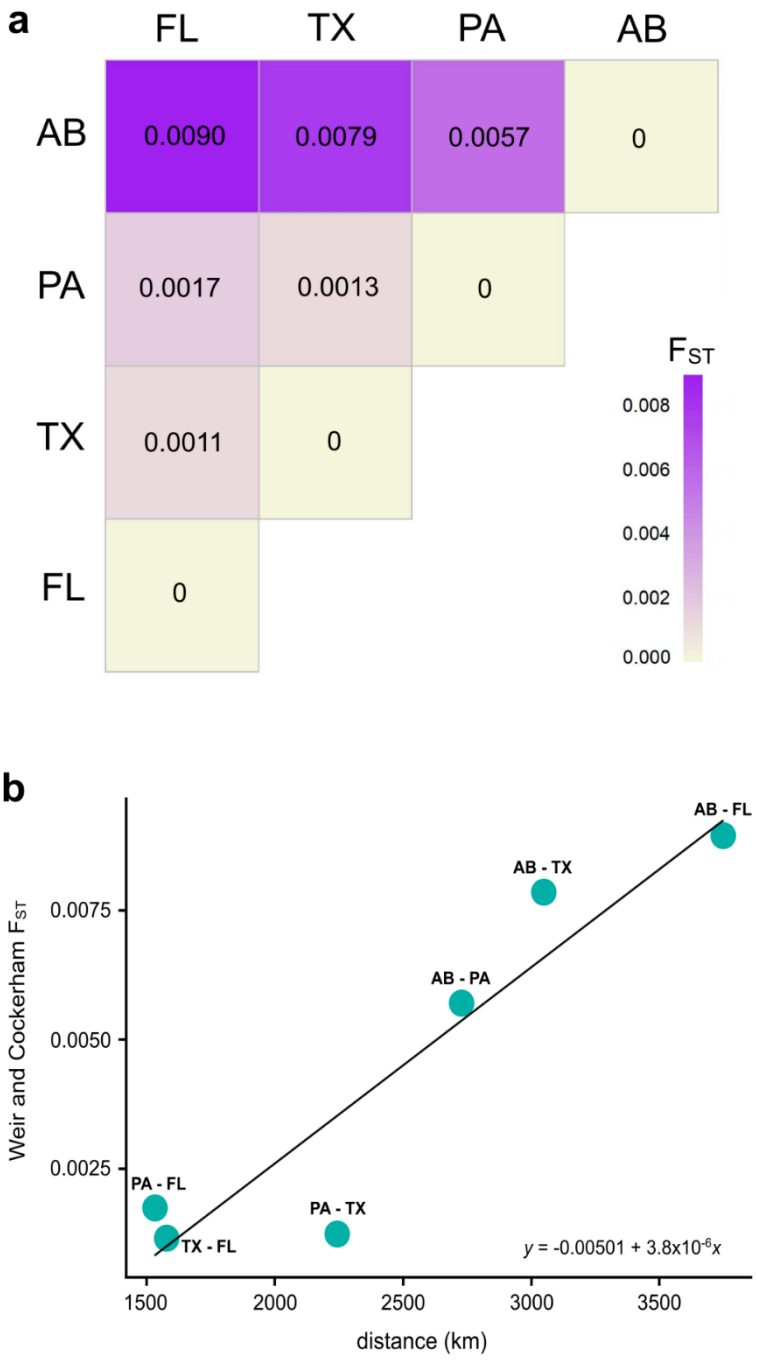

Figure 3.2 (a) Heatmap showing F_{ST} estimated between Alberta, Pennsylvania, Texas, and Florida, displaying low population differentiation (n=67). (b) Same estimates showing a positive relationship with geographic distance between population pairs, following a pattern of isolation by distance.

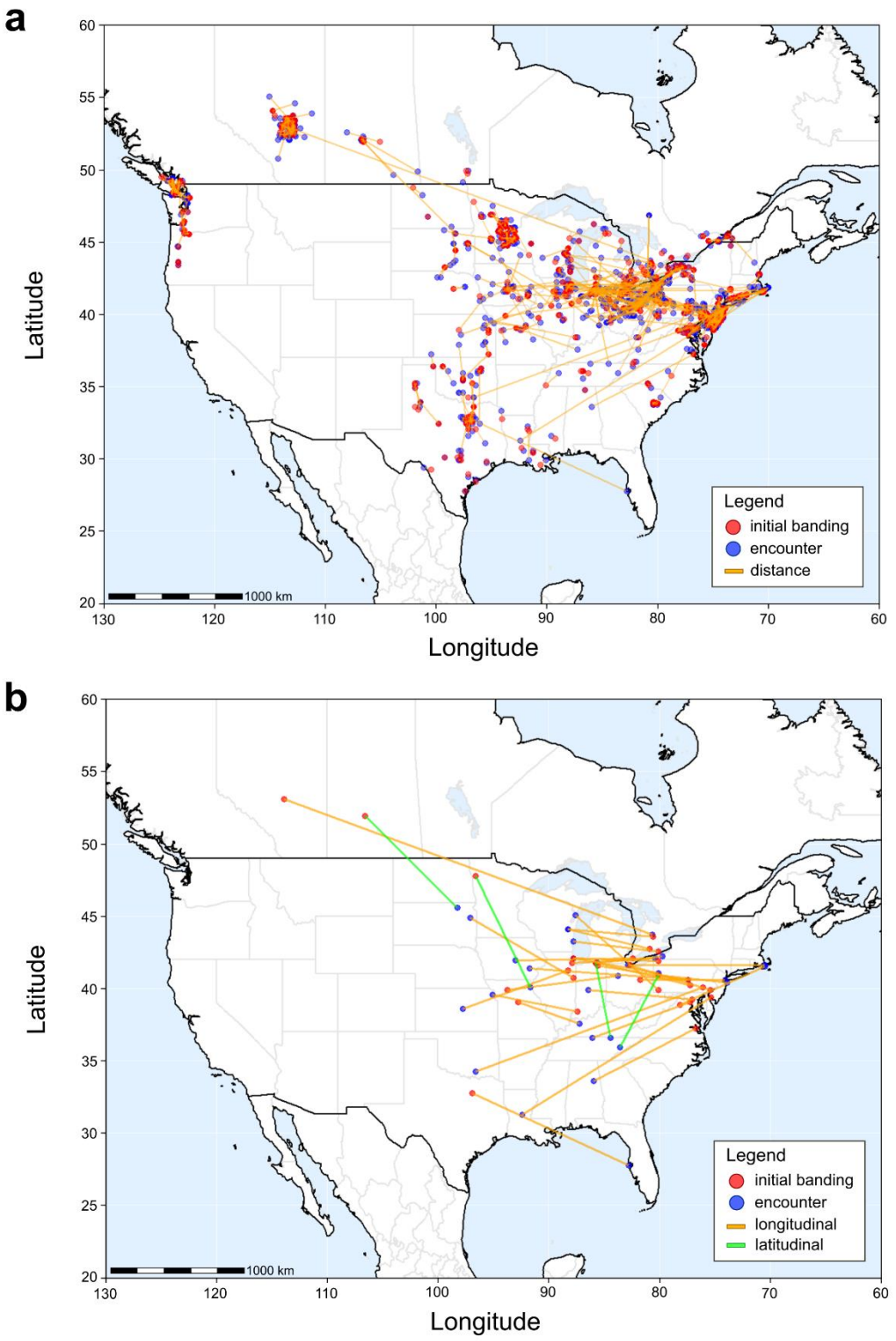

Figure 3.3 (a) Natal dispersal across 2060 band records, with birth location marked in red and breeding location as adult in blue. (b) Dispersals of 500 km or more (n=33), displaying larger longitudinal than latitudinal movements.

Chapter 4: Thesis conclusion

It is important to conserve the incredible biodiversity present on our planet today, as every species has a connected role in the health of ecosystems. Climate change is one of many factors threatening animal populations, and its unprecedented pace urges us to understand species' adaptive potential and implement conservation action. The impacts of advancing spring phenology on many migratory birds can lead to seasonal mismatches and severe consequences to populations (Visser & Gienapp 2019). While some species can adjust migration timing through phenotypic plasticity (Charmantier et al. 2008; Pulido 2007), or microevolution (Helm et al. 2019; Jonzén et al. 2006), the degree of adaptability can vary greatly across organisms and might not be sufficient to keep up with climate change (Visser 2019).

In this thesis, I integrated genomic data, migratory tracks, first bloom dates, and dispersal data, to examine the genomic architecture of migration timing and population structure and patterns of differentiation in a widespread migratory songbird, the purple martin (*Progne subis*). The purple martin is among many declining migratory insectivores (Michel et al. 2015; Nebel et al. 2010; Smith et al. 2015) vulnerable to phenological mismatches (Fraser et al. 2013). Individual variability observed in purple martin migration timing demonstrates a degree of flexibility to seasonal phenology (Fraser et al. 2019), however, it is unknown if this variation is sufficient to keep up with climate change. My results suggest that there is a strong genetic component in migration timing and limitations on gene flow between northern and southern breeding colonies, implicating that microevolutionary changes may be required for this species to adapt to rapid climate change.

In my second chapter, I capitalized on the large existing light-level geolocator dataset and examined the genomic associations with migration timing phenotypes and spring phenology. For

the first time in a long-distance migratory songbird, I connected genomics with spring and fall migratory timing phenotypes and found high heritability in migration timing described from strong genomic variation, which was further supported by complementary analyses with correlated first bloom dates. Through association mapping, I identified loci within and near multiple genes correlated with spring and fall migrations and related to important physiological mechanisms such as photoreception and energetic preparations involved in migration timing. In addition to identifying a region near a known candidate gene (*CRY1*) that plays a role in cryptochromes and the circadian clock (Fusani et al. 2014: Liedvogel & Mouritsen 2010), I also identified a significant region near *MEF2D*, a gene with a role in photoreception (Andzelm 2015; Omori 2015) has not been previously identified in bird migration. Genomic regions potentially related to cholesterol regulation (*SNX8*, Muirhead & Dev 2014) and bone formation (*GDPD2*, Yanaka et al. 2003) also show novel insights into genomic determinants related to migration timing.

In my third chapter, I assessed population structure and differentiation in four *P.s. subis* populations across the North American breeding range and related natal dispersal to genomic patterns. The Alberta population was significantly distinct from the other three populations, indicating that northern and southern purple martin populations require separate conservation management to maintain their genetic diversity. Overall genetic differentiation was low and followed patterns of isolation-by-distance, supporting origins from an eastern refugia during the Pleistocene, which also observed in many other widespread temperate species (Milot et al. 2000; Ruegg & Smith 2002). Dispersal patterns displayed greater longitudinal movements, indicating that dispersal between breeding colonies may have latitudinal constraints. This matches the

observed patterns of population differentiation, where correlations of differentiation were higher with latitude, indicating more gene flow across longitudes.

Evidence for strong genomic variation and associated regions underlying purple martin migration timing has broad implications on migratory animals, revealing high heritability for migration timing and important components that may be associated with this trait. Migration is an essential part of the purple martin life cycle, and they are vulnerable to shifts in seasonal resource availability (Fraser et al. 2013). By guiding conservation classifications and management of genetically differentiated populations, we can help maintain genetic diversity in this species and maximize adaptive potential in the face of threats from climate change. Because the future holds a great amount of uncertainty for purple martins and many other organisms, continued research on key migratory traits and population genomics is important for further understanding adaptive responses to climate change.

References

Andzelm, M.M., Cherry, T.J., Harmin, D.A., Boeke, A.C., Lee, C., Hemberg, M., Pawlyk, B., Malik, A.N., Flavell, S.W., Sandberg, M.A., Raviola, E., and Greenberg, M.E. (2015). MEF2D drives photoreceptor development through a genome-wide competition for tissue-specific enhancers. Neuron, 86(1): 247–263. doi:10.1016/j.neuron.2015.02.038

Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E.B., and Sheldon, B.C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. Science, 320(5877): 800-803. doi:10.1126/science.1157174

Fraser, K.C., Silverio, C., Kramer, P., Mickle, N., Aeppli, R., and Stutchbury, B.J.M. (2013). A Trans-Hemispheric Migratory Songbird Does Not Advance Spring Schedules or Increase Migration Rate in Response to Record-Setting Temperatures at Breeding Sites. PLoS ONE, 8(5): 1-5. doi:10.1371/journal.pone.0064587

Fraser, K.C., Shave, A., de Greef, E., Siegrist, J., and Garroway, C.J. (2019). Individual variability in migration timing can explain long-term population-level advances in a songbird. Front. Ecol. Evol., 7(324). doi:10.3389/fevo.2019.00324

Fusani, L., Bertolucci, C., Frigato, E., and Foà. (2014). Cryptochrome expression in the eye of migratory birds depends on their migratory status. The Journal of Experimental Biology, 217: 918-923. doi:10.1242/jeb.096479

Helm, B., Van Doren, B.M., Hoffman, D., and Hoffman, U. (2019). Evolutionary Response to Climate Change in Migratory Pied Flycatchers. Current Biology, 29(21): 3714-3719. doi:10.1016/j.cub.2019.08.072

Jonzén, N., Lindén, A., Ergon, T., Knudsen, E., Vik, J.O., Rubolini, D., Piacentini, D., Brinch, C., Spina, F., Karlsson, L., Stervander, M., Andersson, A., Waldenström, J., Lehikoinen, A., Edvardsen, E., Solvang, R., and Stenseth, N.C. (2006). Rapid advance of spring arrival dates in long-distance migratory birds. Science, 312: 1959–1961. doi:10.1126/science.1126119

Liedvogel, M., and Mouritsen, H. (2010). Cryptochromes- a potential magnetoreceptor: what do we know and what do we want to know? J. R. Soc. Interface, 7: S147-S162.

doi:10.1098/rsif.2009.0411.focus

Michel, N.L., Smith, A.C., Clark, R.G., Morrissey, C.A., and Hobson, K.A. (2015). Differences in spatial synchrony and interspecific concordance inform guild-level population trends for aerial insectivorous birds. Ecography, 39: 774-786. doi:10.1111/ecog.01798

Milot, E., Gibbs, H.L., and Hobson, K.A. (2000). Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Molecular Ecology, 9: 667–681. doi:10.1046/j.1365-294X.2000.00897.x

Muirhead, G., and Dev, K.K. (2014). The Expression of Neuronal Sorting Nexin 8 (SNX8) Exacerbates Abnormal Cholesterol Levels. J. Mol. Neurosci., 53: 125-134. doi:10.1007/s12031-013-0209-z

Nebel, S., Mills, A., McCracken, J.D., and Taylor, P.D. (2010). Declines of Aerial Insectivores in North American Follow a Geographic Gradient. Avian Conservation and Ecology, 5(2). doi:10.5751/ACE-00391-050201

Omori, Y., Kitamura, T., Yoshida, S., Kuwahara, R., Chaya, T., Irie, S., and Furukawa, T. (2015). Mef2d is essential for the maturation and integrity of retinal photoreceptor and bipolar cells. Genes to Cells, 20: 408–426. doi:10.1111/gtc.12233

Pulido, F. (2007). Phenotypic changes in spring arrival: evolution, phenotypic plasticity, effects of weather and condition. Climate Research, 35: 5-23. doi:10.3354/cr00711

Ruegg, K.C. and Smith, T.B. (2002). Not as the crow flies: A historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). Proc. R. Soc. Lond. B., 269: 1375–1381. doi:10.1098/rspb.2002.2032

Smith, A.C., Hudson, M.A.R., Downes, C.M., and Francis, C.M. (2015). Change Points in the Population Trends of Aerial-Insectivorous Birds in North America: Synchronized in Time across Species and Regions. PLoS ONE, 10(7): e0130768. doi:10.1371/journal.pone.0130768

Visser, M.E. (2019). Evolution: Adapting to a Warming World. Current Biology, 29(22): R1189-R1191. doi:10.1016/j.cub.2019.09.062

Visser, M.E. and Gienapp, P. (2019). Evolutionary and demographic consequences of phenological mismatches. Nature Ecology and Evolution, 3: 879-885. doi:10.1038/s41559-019-0880-8

Yanaka, N., Imai, Y., Kawai, E., Akatsuka, H., Wakimoto, K., Nogusa, Y., Kato, N., Chiba, H., Kotani, E., Omori, K., and Sakurai, N. (2003). Novel Membrane Protein Containing Glycerophosphodiester Phosphodiesterase Motif Is Transiently Expressed during Osteoblast Differentiation. Journal of Biological Chemistry, 278(44): 43595–43602. doi:10.1074/jbc.M302867200

APPENDIX A: Chapter 2 supplemental materials

Methods

Reference genome

Genome assembly

The Science for Life Laboratory in Sweden extracted DNA from one individual using an *in-situ* agarose plug extraction procedure. They prepared Pacific BioSciences (PacBio) libraries to produce long reads for *de novo* assembly (Rhoads & Au 2015), and 10X Genomics Chromium (10X) libraries for filling gaps in the *de novo* assembly (Zheng et al. 2016). The initial genome assembly was assembled with the PacBio sequences using *FALCON* algorithms with SMRT Link v7.0.0 (Chin et al. 2016). I polished the genome using two rounds of the program *ArrowGrid* v0.6.0 with the PacBio subreads bam files (Chin et al. 2013; Koren et al. 2017). To prepare the next polishing step, I aligned 10X fastq files to the arrow-corrected genome using *BWA* v0.7.17 (Li & Durbin 2009) and sorted the resulting bam file with *SAMtools* v1.9 (Li et al. 2009). Next, I polished the genome with *Pilon* v1.23 to correct bases, fix mis-assemblies, and fill gaps (Walker et al. 2014). I scaffolded the genome with the *ARKS* v1.0.4 (Coombe et al. 2018), using interleaved 10X reads put together with *LongRanger* v2.2.2 (Marks et al. 2019). I checked the genome for duplicated scaffolds using dedupe.sh in *BBMap* v38.44 (Bushnell 2014). To look for possible contaminants, I used the nucleotide database from the National Center for Biotechnology Information (NCBI 2020) and the program *BLAST+* v2.9 (Camacho et al. 2009) to identify any foreign DNA. The final genome was evaluated with *QUAST* v5.0.2 (Gurevich et al. 2013), *Assemblathon 2* (Bradnam et al. 2013), and *BUSCO* v4.0.5 with the aves dataset (aves_odb10) and augustus species set to chicken (Seppey et al. 2019).

Genome annotation

I annotated the genome using four rounds of the program *MAKER* v2.31.10 (Campbell et al. 2014). The first round included *RepeatMasker* and *Exonerate* using protein data from three model species, the chicken (*Gallus gallus*), collared flycatcher (*Ficedula albicollis*), and zebra finch (*Taeniopygia guttata*) obtained from *Ensembl* database (Cunningham et al. 2019). A hidden Markov model (HMM) was created using the outputs from the first round, and then included to in the second round to train *SNAP* using computational gene prediction. The third round used the updated HMM model from the second round and included the *Augustus* chicken model. The fourth and final round had the same parameters as the third, but additionally filtered out genes of annotation edit distance (AED) score over 0.5. After completing the *MAKER* rounds, the annotation was aligned with protein data from the *uniprot* database (Bateman 2019) using *BLAST+* v2.9 (Camacho et al. 2009) and protein functions using *InterProScan* v5.4 (Jones et al. 2014).

Resequencing data

DNA extractions

The following modifications were made from the Qiagen DNeasy Blood & Tissue kit for the 89 blood samples to maximize DNA yield, starting with pipetting between 150-250uL of the sample mixture containing blood and Queen's Lysis buffer to add to 20uL proteinase K, and adjusting the total volume to about 300uL with phosphate-buffered saline in the first step. During incubation, instead of using a thermomixer for ten minutes, all samples were incubated in a dry bath block heater (at 56°C) for 1-8 hours, and then taken out for vortexing every 30-60 minutes. After the three instructed wash steps with the spin columns, an additional dry spin was

performed to fully remove residual ethanol and other liquids before elution. In the final extraction step, all DNA samples were eluted in 100-200uL of warmed (56°C) ultrapure distilled water instead of using the provided buffer AE in the extraction kit. Each sample was then checked for quality, using a NanoDrop OneC Spectrophotometer to examine DNA concentrations and absorbance ratios for purity (minimum concentration 25ng/uL, and minimum 260/280 and 260/230 ratios of 1.8). Then, the samples were tested on a gel electrophoresis machine to confirm intact DNA fragments. Finally, each sample was loaded on a well-skirted PCR plate in a randomized order to avoid lane bias during sequencing and diluted with ultrapure distilled water to reach a concentration of 25ng/uL in a volume of 60uL.

Genotype imputation

In the AgSeq pipeline conducted at Texas A&M AgriLife Research, the data was first filtered and trimmed using *Trimmomatic* v0.38 (Bolger et al. 2014) then aligned to the reference genome using *Bowtie2* v2.3.4.2 (Langmead & Salzberg 2012), and sorted, realigned, and filtered by MQ>5 by *SAMTools* v1.9 and *PICARD* v.2.18.4 (Broad Institute 2019). The variants were called using *GATK* v3.5 using HapltotypeCaller (Poplin et al. 2017), and missing genotypes were imputed with *Beagle* v.4.1 (Browning & Browning 2016) and filtered to remove genotype probabilities below 0.9. The imputation process used information from surrounding genotypes and relied on linkage disequilibrium structure and haplotype blocks (Golicz et al. 2015).

SNP filtering

Initial filtering quality thresholds included SNPs with quality score (QUAL) > 20 and mapping quality (MQ) > 20 using *vcflib* v1 (Garrison 2016). Using *VCFtools* v.0.1.16 (Danecek

et al. 2011) I removed sites with indels, and then filtered the SNPs to a dataset with minor allele frequency (MAF) > 0.05 (5%), max-missing genotype score of 0.8, matching Hardy-Weinberg equilibrium, and containing no tri-allelic sites. I removed scaffolds with abnormally high read depths (4 or higher) from the AgSeq sequences, as they are likely representing repetitive regions. In addition to Z and W-linked scaffolds identified from mapping the genome to a chicken genome with *SatsumaSynteny* (Grabherr et al. 2010), I also used F_{ST} values through *VCFtools* (Danecek et al. 2011) between males and female samples to recognize additional sex-linked scaffolds. These filters were used to conservatively identify autosomes in both males and females. For the principal component analysis specifically (to determine PC1 for a covariable), I used *PLINK* v1.9 (Purcell et al. 2007) to remove SNPs in linkage disequilibrium (LD-pruning).

Genome-wide association studies

Light-level geolocator analysis

The purple martin's behaviour of foraging and residing in open areas made light-level geolocators ideal for capturing sunrises and sunsets with minimal shading. The timing of these twilights estimated the daily locations of the birds during the entire migratory tracks, using the midpoint of rise-set events for determining longitudes and day length for estimating latitudes (Hill 1994). To define sunrises and sunsets, I used the *preprocessLight* function in the R-package *BAStag* (Wotherspoon et al. 2016). Here, the light-intensity threshold was set at 32 to systematically separate day and night, and false twilights or outliers due to shading or light pollution were removed through manual inspection. In addition, the sudden onset of extreme light fluctuations in spring indicated the bird entering and exiting nest cavities, signifying the spring arrival date. Next, I estimated daily coordinates using the twilight data with the *coord*

function in R-package *GeoLight* (Lisovki & Hahn 2012). Here, the calibration for the sun elevation angle was calculated using two weeks of light data at the end of the breeding season, where the bird was known to stay in their respective colony before departing on fall migration. Since there is little variation in day lengths around the spring and fall equinox, the latitudes around the equinox were removed using a tol of 0.13 (Lisovki & Hahn 2012) as these would result in inaccurate locations. I determined spring and fall departure and arrival dates from the location coordinates based on sudden shifts in latitude and longitude, and confirmed with the *changeLight* function, which defined residency periods throughout the year.

Association Mapping

For the BSLMM, the SNP dataset was converted from a vcf file to a binary plink format (bim, bed, fam files) using *PLINK* v1.9 (Purcell et al. 2007). Migratory phenotypes (spring departure, spring arrival, fall departure, fall arrival) were quantile normalized using R-package *preprocressCore* (Bolstad 2020) to standardize the phenotypic data for comparisons. Prior to running the BSLMM, I calculated a relatedness matrix through *GEMMA* and then ran four BSLMM models for each phenotype in *GEMMA* using Markov Chain Monte Carlo (MCMC) (Zhou et al. 2013) and combined the hyp and param outputs files for analyses. The posterior inclusion probabilities (PIP) is the probability that the SNP is associated with the phenotypic variation (Gompert et al. 2013), and I plotted the resulting PIPs using R-package *qqman* (Turner 2014) after ordering SNPs by chromosome. Because BSLMMs do not take in covariates, I ran univariate LMMs for each migratory phenotype using *GEMMA* (Zhou & Stephens 2013), including sex, colony, year, age, and PC1 as covariates. This model assumed every genetic variant affects the phenotypes, where effect sizes are normally distributed. Using the original

bam files, I checked areas around each significant SNP identified in the BSLMMs and LMMs to ensure the raw sequences around the identified regions were in place and did not contain a large portion of missing sites.

I ran the LFMM with data in lfmm and geno formats using R-package *LEA* (Frichot & François 2015) and a latent factor K=1. Models were tested with K=2 and 3, however K=1 had the best fit due to the number ancestral groups in the dataset. I used the default settings with 10,000 iterations and 5,000 burnin periods to compute LFMM parameters for all SNPs. This model was based on MCMC, and p-values were re-adjusted using fisher-stouffer method on combined z-scores of three runs and restricted for FDR > 0.05.

I used R-package *goseq* (Young et al. 2010) to look for gene enrichment associated with gene ontology terms using all genes within 100 kb to significant SNPs.

Analyses were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing and Compute Canada, and all additional R-based analyses were completed in R version 3.6.1 (R Core Team 2019).

References

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research, 47: D506–D515. doi:10.1093/nar/gky1049

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, 30(15): 2114-2120. doi:10.1093/bioinformatics/btu170

Bolsatd, B. (2020). preprocessCore: A collection of pre-processing functions. R package version 1.50. https://github.com/bmbolstad/preprocessCore

Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Cahpuis, G., Chikhi, R., … Korf, I.F. (2013). Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. GigaScience, 2(10). doi:10.1186/2047-217X-2-10

Broad Institute, GitHub Repository. (2019). Picard Toolkit. http://broadinstitute.github.io/picard/

Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of Reference Samples. American Journal of Human Genetics, 98: 116–126. doi:10.1016/j.ajhg.2015.11.020

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. United States.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10(421): 1–9. doi:10.1186/1471-2105-10-421

Campbell, M.S., Holt, C., Moore, B., and Yandell, M. (2014). Genome Annotation and Curation Using MAKER and MAKER-P. Current Protocols in Bioinformatics, 48: 4.11.1-4.11.39. doi:10.1002/0471250953.bi0411s48

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E., Turner, S.W., and Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods, 10(6): 563–569. doi:10.1038/nmeth.2474

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M.,

Luo, C., Ecker, J.R., Cantu, D., Rank, D.R., and Schatz, M.C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nature Methods, 13(12): 1050–1054. doi:10.1038/nmeth.4035

Coombe, L., Zhang, J., Vandervalk, B.P., Chu, J., Jackman, S.D., Birol, I., and Warren, R.L. (2018). ARKS: Chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics, 19(234): 1–10. doi:10.1186/s12859-018-2243-x

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I M., Bennet, R., Bhai, J., Billis, K., Boddu, S., Cummins, C., Davidson, C., Dodiya, K.J., Gall, A., Girón, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O., Janacek, S.H., … Flicek, P. (2019). Ensembl 2019. Nucleic Acids Research, 47: D745–D751. doi:10.1093/nar/gky1113

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., and 1000 Genomics Project Analysis Group. (2011). The variant call format and VCFtools. Bioinformatics, 27(15): 2156–2158. doi:10.1093/bioinformatics/btr330

Frichot, E., and François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6: 925–929. doi:10.1111/2041-210X.12382

Garrison, E. (2016). Vcflib, a simple C++ library for parsing and manipulating VCF files. https://github.com/vcflib/vcflib

Golicz, A.A., Bayer, P.E., and Edwards, D. (2015). Skim-Based Genotyping by Sequencing. Methods in Molecular Biology, 1245: 257–270. doi:10.1007/978-1-4939-1966-6_19

Gompert, Z., Lucas, L.K., Nice, C.C., and Buerkle, C.A. (2013). Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. Evolution, 67(9): 2498–2514. doi:10.1111/evo.12021

Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: *Satsuma*. Bioinformatics, 26(9): 1145–1151. doi:10.1093/bioinformatics/btq102

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. Bioinformatics, 29(8): 1072–1075. doi:10.1093/bioinformatics/btt086

Hill, R.D. (1994). Theory of geolocation by light levels. In B.J.L. Boeuf, & R.M. Laws (Ed.), Elephant seals: Populatoin ecology, behaviour and phyiology (pp. 227-236). Berekeley, CA: University of California Press.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, F., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., and Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics, 30(9): 1236–1240. doi:10.1093/bioinformatics/btu031

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat

separation. Genome Research, 27(5): 722–736. doi:10.1101/gr.215087.116

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4): 357–359. doi:10.1038/nmeth.1923

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

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

Lisovski, S., and Hahn, S. (2012). GeoLight-processing and anslysing light-based geolocator data in R. Methods in Ecology and Evolution, 3: 1055-1059. doi:10.1111/j.2041-210X.2012.00248.x

Marks, P., Garcia, S., Barrio, A.M., Belhocine, K., Bernate, J., Bharadwaj, R., Bjornson, K., Catalanotti, C., Delaney, J., Fehr, A., Fiddes, I.T., Galcin, B., Heaton, H., Herschleb, J., Hindson, C., … Church, D.M. (2019). Resolving the full spectrum of human genome variation using Linked-Reads. Genome Research, 29(4): 635–645. doi:10.1101/gr.234443.118

National Center for Biotechnology Information (NCBI) (2020). Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., … Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of

samples. BioRxiv, 201178. doi:10.1101/201178

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Maller, B.D., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Americna Journal of Human Genetics, 81(3): 559-575. doi:1086/519795

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statisical Computing, Vienna, Austria.

Rhoads, A., and Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, 13: 278–289. doi:10.1016/j.gpb.2015.08.002

Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M. (eds) Gene Prediction. Methods in Molecular Biology 1962. Humana, New York, NY.

Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. biorXiv. doi:10.1101/005165

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE, 9(11): e112963.doi:10.1371/journal.pone.0112963

Wotherspoon, S., Summer, M., and Lisovski, S. (2016). BAStag: Basic Data Processing for Light Based Geolocation Archival Tags. Version 0.1.3.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology, 11: R14. doi:10.1186/gb-2010-11-2-r14

Zheng, G.X.Y., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., … Ji, H.P. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nature Biotechnology, 34(3): 303–311. doi:10.1038/nbt.3432

Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genetics, 9(2): e1003264. doi:10.1371/journal.pgen.1003264

Zhou, X., and Stephens, M. (2013). Genome-wide Efficient Mixed Model Analysis for Association Studies. Nature Genetics, 44(7): 821-824. doi:10.1038/ng.2310

Tables and figures

Table A1 Site locations for geolocator-tracked purple martins (n=89) across 14 breeding colonies in North America.

Province/State	Colony	Latitude	Longitude	N	Year(s)
Alberta	Camrose	53.011	-112.864	5	2012-2014
Alberta	Lacombe	52.391	-113.612	17	2012-2014
Florida	Bay Lake	28.360	-81.588	19	2013-2015
Florida	Naples	26.149	-81.746	2	2013
Minnesota	Millie Lacs	46.146	-93.724	1	2011
New Jersey	Middletown	40.390	-74.001	3	2011
Ontario	Ottawa	45.351	-75.827	2	2014
Pennsylvania	Erie	42.115	-80.145	14	2008-2013
Pennsylvania	Malvern	40.036	-75.514	1	2013
South Dakota	Columbia	45.598	-98.310	2	2011-2013
Texas	Amarillo	35.040	-101.933	4	2013
Texas	Corpus Christi	27.680	-97.407	4	2013
Texas	Sandia	27.997	-97.879	8	2013
Virginia	Woodbridge	38.613	-77.263	7	2010-2013

Table A2 SNPs with PIP > 0.1 from BSLMMs with spring departure, spring arrival, fall departure, and fall arrival phenotypes with respective chromosome number. LMM p-values are listed in -log10(p) format. Genes closest to the SNP are listed with distance (bp).

Spring departure

SNP	Chr	PIP	-log10(p)	Gene	Function	Distance (bp)
S214:738983	25	0.3751	5.2129	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein	88,046 727
S86:1071937	4	0.3475	2.2315	GDPD2, HTR2C	Glycerophosphodiester phosphodiesterase domain containing 2, 5-hydroxytryptamine receptor 2C	45,244 243,812
S15:8639530	4	0.1496	0.4548	WWC2, DCTD	WW and C2 domain containing 2, dCMP deaminase	208,186 116,554
S10:25072973	2	0.1385	2.7708	FAM221A, NPY	Family with sequence similarity 221 member A, Neuropeptide Y	1,580 126,713
S867:15879	Un	0.1085	0.4431	NA	Similar to Olfactory receptor	13,542 7,977

Spring arrival

SNP	Chr	PIP	-log10(p)	Gene	Function	Distance (bp)
S214:738983	25	0.5972	6.3671	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein	88,046 727
S15:8639530	4	0.5171	0.7556	WWC2 DCTD	WW and C2 domain containing 2, dCMP deaminase	208,186 116,554
S144:528865	2	0.1533	1.0996	VSTM2A	V-set and transmembrane domain containing 2A	267,260

Fall departure

SNP	Chr	PIP	-log10(p)	Gene	Function	Distance (bp)
S214:738983	25	0.6324	3.3972	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein	88,046 727
S10:21203894	2	0.6314	2.1724	SCIN, AGMO	Scinderin, Alkylglycerol monooxygenase	758,330 425,402
S48:6684786	1	0.2284	1.6308	COX11, STXBP4	Cytochrome c oxidase copper chaperone COX11, Syntaxin binding protein 4	6,737 2,418
S2:1666231	1	0.1172	1.0458	NA	Protein unknown	In
S44:6323892	1	0.1150	0.6133	ETNK1, BCAT1	Ethanolamine kinase 1, Branched-chain-amino-acid aminotransferase	54,746 878,407

Fall arrival						
SNP	Chr	PIP	-log10(p)	Gene	Function	Distance (bp)
S86:1071937	4	0.3236	2.6375	GDPD2, HTR2C	Glycerophosphodiester phosphodiesterase domain containing 2, 5-hydroxytryptamine receptor 2C	45,244 243,812
S214:738983	25	0.2953	3.7188	MEF2D, RHBG	MADS-box domain-containing protein, Ammonia transporter protein	88,046 727
S99:528805	14	0.1521	0.4221	SNX8	Sorting nexin 8	in
S18:6522504	1	0.1375	2.1805	N6AMT1, ADAMTS5	N-6 adenine-specific DNA methyltransferase 1, ADAM metallopeptidase with thrombospondin type 1 motif 5	500,042 23,085
S36:821560	3	0.1370	0.1363	IRAK1BP1, HTR1B	Interleukin 1 receptor associated kinase 1 binding protein 1, 5-hydroxytryptamine receptor 1B	2,598 459,355
S22:13238871	1	0.1370	1.2813	BTBD11, CRY1	BTB domain containing 11, Cryptochrome circadian regulator 1	184,587 26,978
S38:2507638	7	0.1368	1.2370	FASTKD1, LOC101817811	FAST kinase domains 1, Uncharacterized protein	9,712 2,031
S8:731565	4	0.1198	0.9355	AGA	Aspartylglucosaminidase	946,873

Table A3 The top four most significant results form gene ontology enrichment analyses, with adjusted p-values below 0.6, focusing on all genes within 100 kb of SNPs associated with spring departure. The number of differentially expressed (DE) genes is listed for each go-term.

GO term	Description	Adjusted p-value	DE
GO:0005540	hyaluronic acid binding	0.3416	2
GO:0007155	cell adhesion	0.3416	2
GO:0008519	ammonium transmembrane transporter activity	0.5124	1
GO:0015696	ammonium transport	0.5124	1

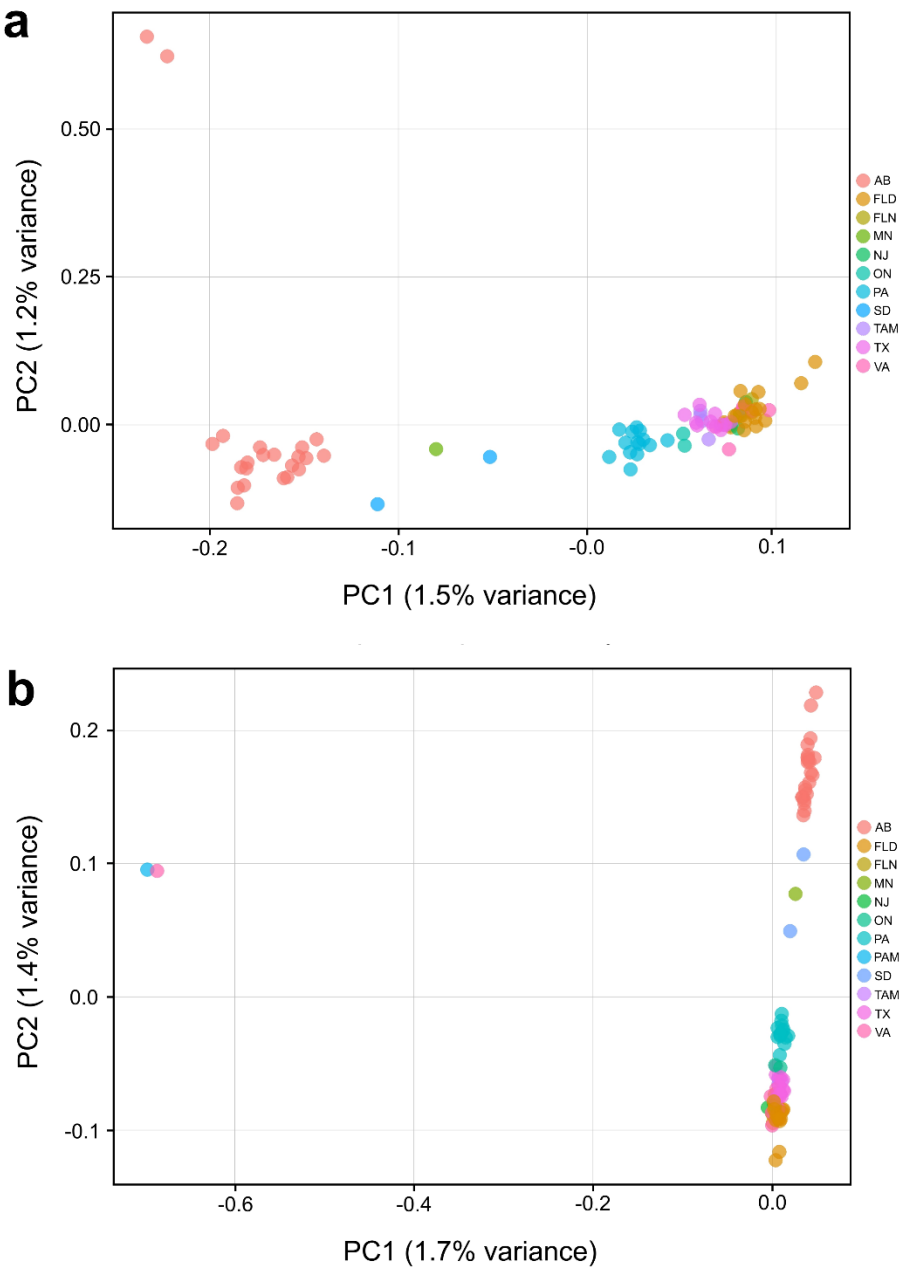

Figure A1 Results from principal component analyses: (a) n=87 displaying population structure across sample sites along PC1, (b) n=89 displaying population structure including two outlier individuals (from TAM and VA) that were removed for GWAS analyses. AB=Alberta, FLD=Florida (Disney), FLN=Florida (Naples), MN=Minnesota, NJ=New Jersey, ON=Ontario, PA=Pennsylvania, PAM=Pennsylvania (Malvern), SD=South Dakota, TAM=Texas (Amarillo), TX=Texas (Corpus Christi, Sandia), VA=Virginia.

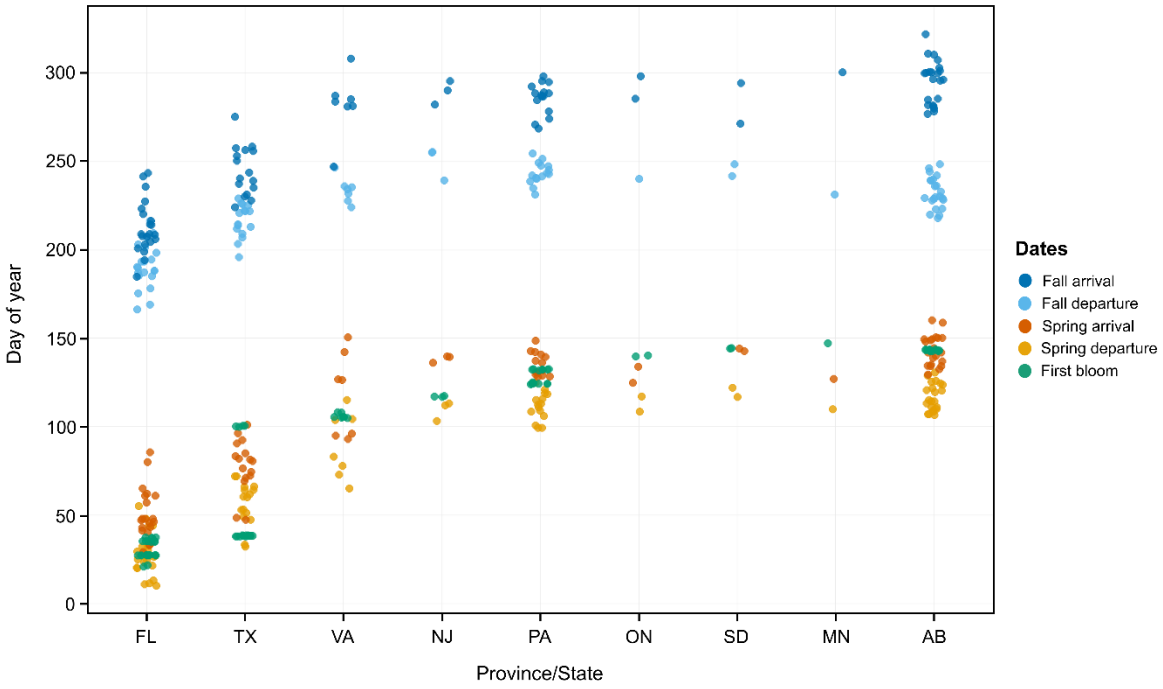

Figure A2 Migratory timing phenotypes and first bloom data for 89 geolocator-tracked purple martins, divided by province/state and ordered by latitude. Dates include the start and end of spring and fall migrations in Julian date format for each bird and the first bloom date for their respective year and site location.

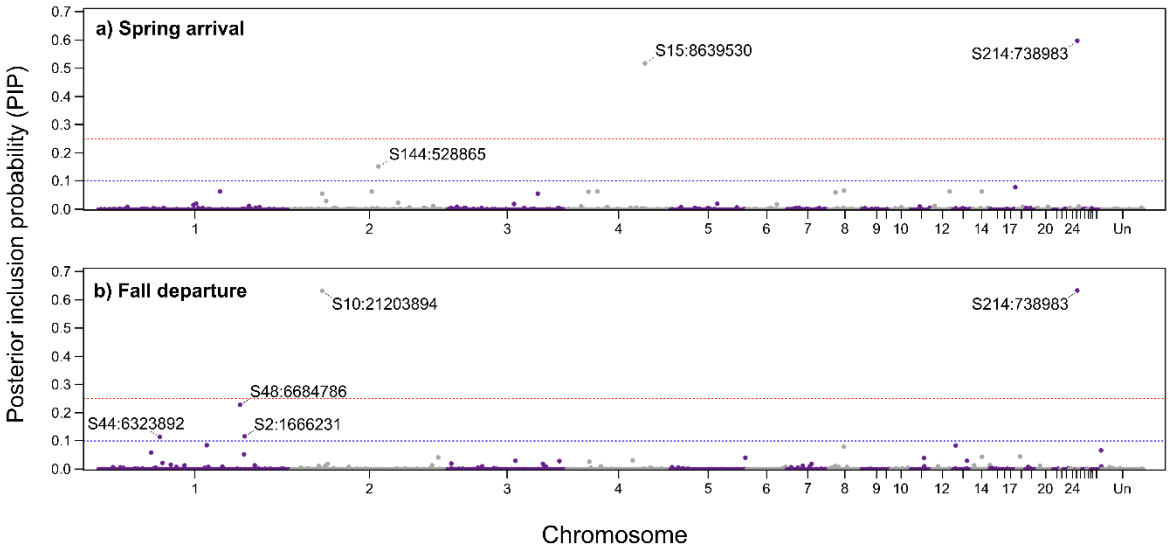

Figure A3 Bayesian Sparse Linear Mixed Models (BSLMM) of two migration timing phenotypes: (a) spring arrival and (b) fall departure (n=87). Above the blue threshold line represents PIP > 0.1, and above the red threshold line represents PIP > 0.25.

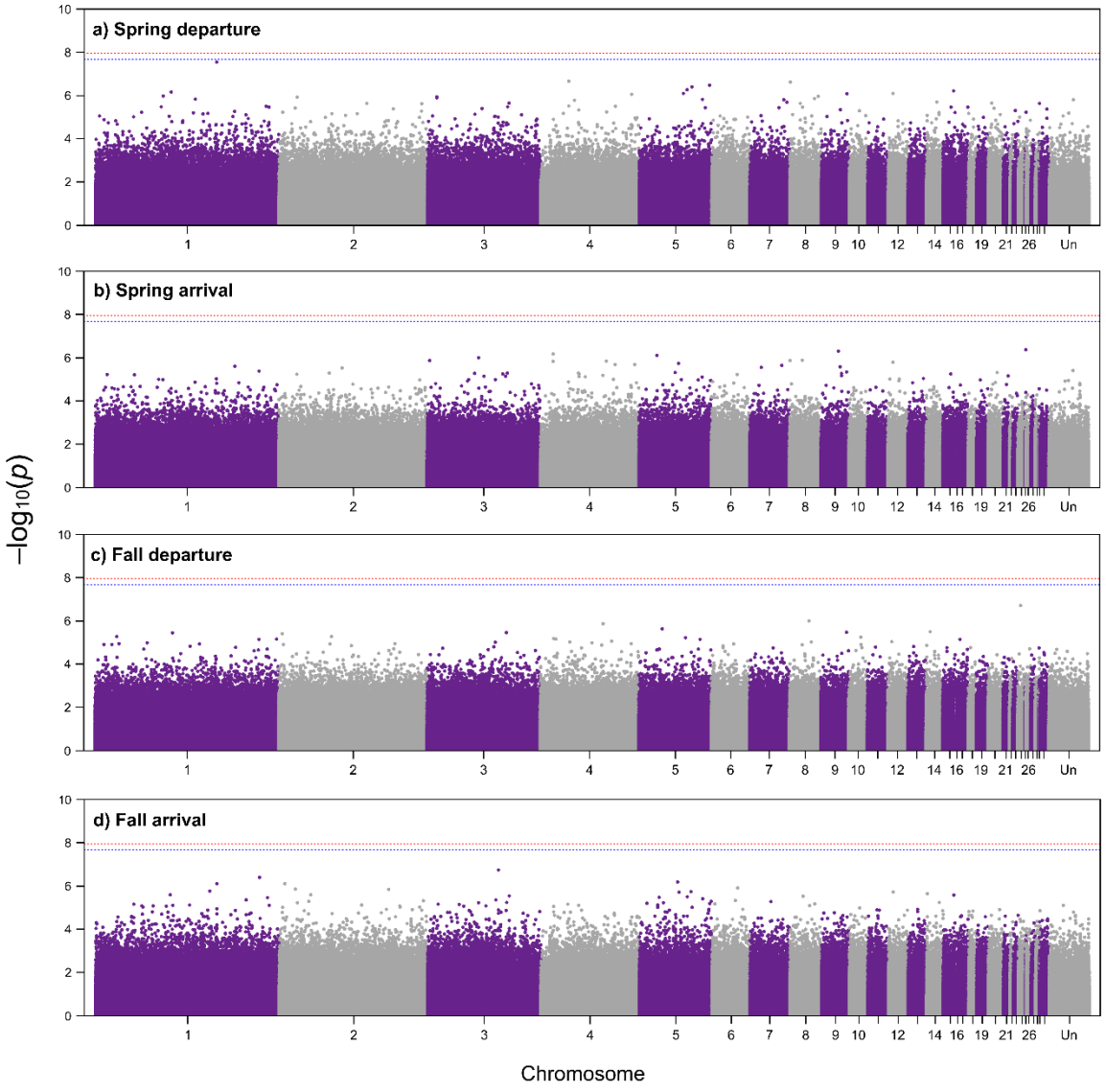

Figure A4 Linear Mixed Models (LMM) of four migration timing phenotypes: (a) spring departure, (b) spring arrival, (c) fall departure, and (d) fall arrival (n=87). Sex, colony, age, year, and PC1 are included as covariates. P-values are from Wald likelihood ratio tests. Above the blue threshold line represents false discovery rate (FDR) > 0.1, and above the red threshold line represents FDR > 0.05.

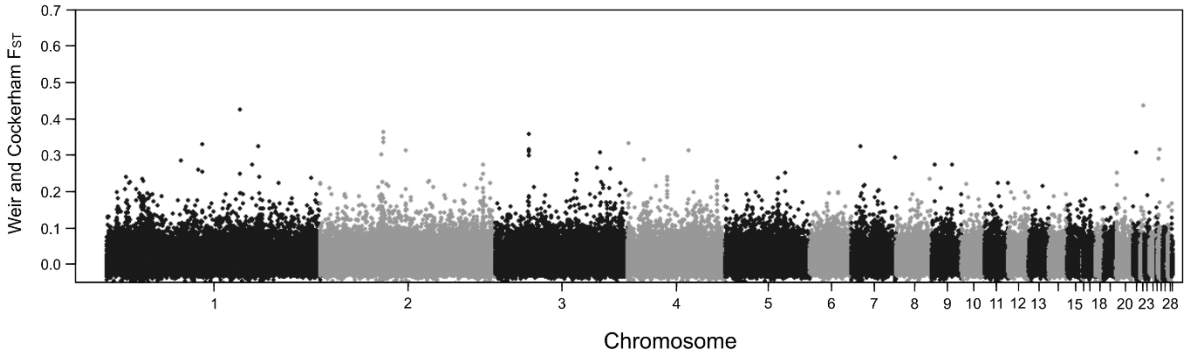

Figure A5 F_{ST} across autosomes in genome between individuals from Alberta and Florida, representing the northern and southern edges of the breeding distribution. SNPs are in linkage disequilibrium in 5kb non-overlapping windows.

APPENDIX B: Chapter 3 supplemental materials

Methods

Population structure & differentiation

PCA outliers

I used *PLINK* v1.9 (Purcell et al. 2007) to convert the SNP dataset from vcf to bed, bim, and fam files. Next, I ran *smpartca* using *EIGENSOFT* v7.2.1 (Price et al. 2006). In the PCA with 71 samples, four outliers from Florida heavily impacted the PCA on both PC1 and PC2 (Figure B1). I used additional programs to calculate PCA, including *PLINK* (Purcell et al. 2007) and R-package *pcadapt* (Privé et al. 2020), however all runs had similar results. I also checked the sample information for any potential abnormalities, but the four outliers were among other Florida samples of the same years and sexes. The order of samples on the PCR plate for sequencing did not show any patterns that could have affected these four samples. I also checked the genotype coverage for these individuals and they were consistent with the rest of the samples, making it difficult to pinpoint causes of significant variation in SNPs displayed in the PCA.

F_{ST} and mantel tests

All pairwise fixation indices were calculated using *VCFtools* v0.1.16 (Danecek et al. 2011) using 5kb non-overlapping windows. After manually calculating distance, latitude, and longitude between each population pair and formatting them into separate matrices, I used *mantl.rtest* function in R-package *ade4* (Dray & Dufour 2007) to run three Mantel tests for calculating correlation of F_{ST} with distance, latitude, and longitude using 9999 random permutations for each run.

Natal dispersal

Filtering banding records

I started with 3,928 banding encounter records obtained through the North American Bird Banding Program Bird Banding Office. First, I filtered out individuals that were initially captured as adults (any age older than HY), in order to include only juveniles in the dataset. Next, I filtered out inexact banding or encounter dates that did not specify month. I also filtered out duplicate encounters for the same individual (keeping the first encounter as an adult) as well as single encounters that were observed in the same year as initial banding. Finally, I filtered out data between months of August through April, in order to ensure exlusion of records encountered during migration. Although some populations are known to arrive at breeding grounds earlier than May (i.e. Florida), I did not observe more breeding dispersal records from the southern breeding populations when adjusting the filter criteria to include earlier months. Instead, adjusting for earlier months only added data outside of North America, which were assumed to be birds on migration. Thus, I kept the filter including only months of May through July to confidently use records of birds in breeding territories.

Quantifying latitudinal and longitudinal movements

After calculating dispersal distances for each individual using *distm* function in R-packge *geosphere* v1.5.10 (Hijmans et al. 2014), I narrowed the dataset down to individuals with dispersal distance greater than 0 km (n=1,725) in order to examine longitudinal and latitudinal dispersal patterns. Continuing with *geosphere* to incorporate the spherical coordinate system, I calculated movements in cardinal directions for each individual using a combined coordinate (initial banding longitude and encounter latitude) as a reference point for calculating distances

trigonomically. North-south distances were estimated by distance between initial banding coordinates and banding longitude & encounter latitude, and east-west distances were calculated by distance between encounter coordinates and banding longitude & encounter latitude. Inidividuals with ratios of east-west distance over north-south that were greater than one were categorized as longitudinal dispersers, and ratios below one were considered latitudinal dispsersers. Total distances of longitudinal dispersers were 94,815 km (n=845) and latitudinal dispersers were 68,653 km (n=880). Because these included short dispersals, I also examined patterns among disperers travelling a minimum of 100 km (n=310) and 500 km (n=33). For the minimum of 100 km, longitudinal dispersers travelled 63,430 km total (n=191) and latitudinal dispersers travelled 26,022 km total (n=119). Using a minimum threshold of 500 km dispersal distances, I found an even greater difference with longitudinal dispersers travelling 26,353 km (n=29) and latitudinal dispersers travelling 2,093 km (n=4).

Analyses were conducted with the advanced computing resources provided by Texas A&M High Performance Research Computing and Compute Canada, and all additional R-based analyses were completed in R version 3.6.1 (R Core Team 2019).

References

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., and 1000 Genomics Project Analysis Group. (2011). The variant call format and VCFtools. Bioinformatics, 27(15): 2156–2158. doi:10.1093/bioinformatics/btr330

Dray, S. and Dufour, A.B. (2007). The ade4 package: Implementing the duality diagram for ecologists. Journal of Statistical Software, 22(4): 1–20. doi:10.18637/jss.v022.i04

Hijmans, R.J., Williams, E., and Vennes, C. (2014). Geosphere: spherical trigonometry. R package version 1.3-11

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8): 904–909. doi:10.1038/ng184

Privé, F., Luu, K., Vilhjálmsson, B.J., and Blum, M.G. (2020) Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. Molecular Biology and Evolution, 37(7): 2153-2154. doi:10.1093/molbev/msaa053

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Maller, B.D., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Americna Journal of Human Genetics, 81(3): 559-575. doi:1086/519795

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statisical Computing, Vienna, Austria.

Table and figures

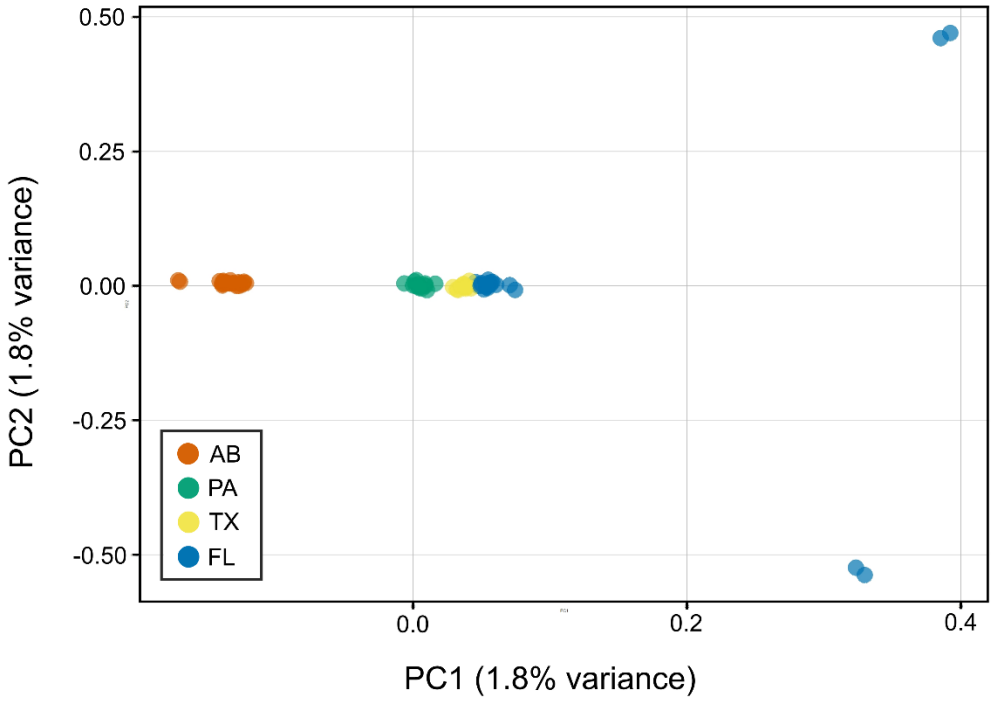

Figure B1 Results from a principal component analysis (n=71), displaying population structure (AB=Alberta, PA=Pennsylvania, TX=Texas, FL=Florida) with four outliers from Florida impacting both PC1 and PC2.

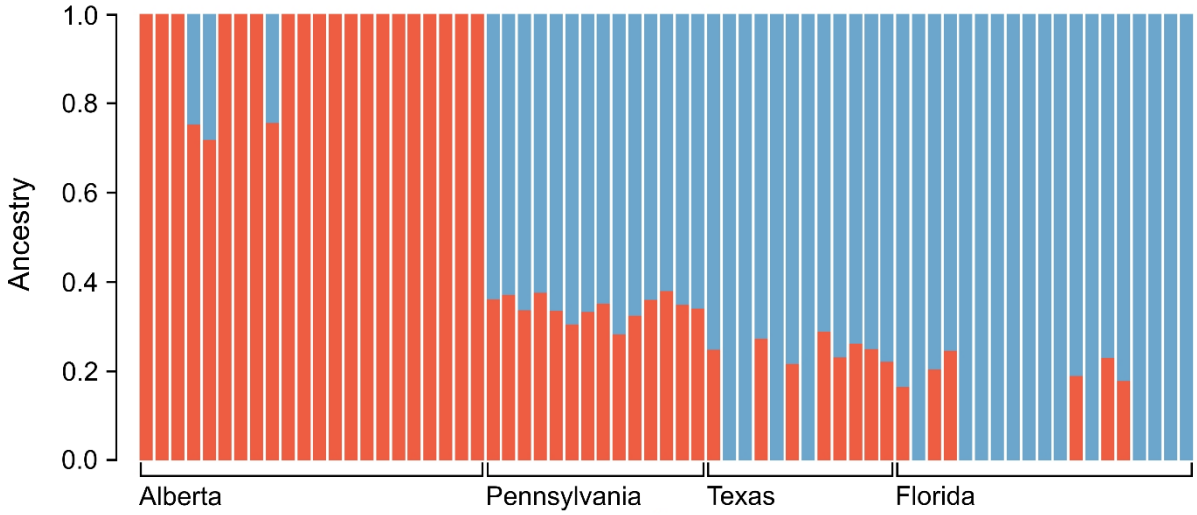

Figure B2 Results from Admixture assuming two ancestral populations (K=2). Populations on x-axis are order by decreasing latitude (Alberta, Pennsylvania, Texas, and Florida), and each individual genotype is represented by one vertical bar.