EVOLUTIONARY BIOLOGY OF BIRDS AND THEIR ASSOCIATED MICROBIOTA IN THE MADREAN ARCHIPELAGO SKY ISLANDS

BY

JOSEPH D MANTHEY

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Chairperson: Robert G. Moyle

____

Mark T. Holder

____

John K. Kelly

____

Erik T. Lundquist

____

Mark B. Robbins

Date Defended: 11 December 2015

The dissertation committee for JOSEPH D MANTHEY certifies that this is the approved version of the following dissertation:

EVOLUTIONARY BIOLOGY OF BIRDS AND THEIR ASSOCIATED MICROBIOTA IN THE MADREAN ARCHIPELAGO SKY ISLANDS

____

Chairperson: Robert G. Moyle

Date Approved: 15 December 2015

Abstract

Sky islands are disjunct patches of montane forested habitat in a matrix of desert, grasslands, and scrub. I investigated intraspecific evolutionary biology of two bird species—the White-breasted Nuthatch (*Sitta carolinensis*) and the Brown Creeper (*Certhia americana*)—in the Madrean Archipelago sky islands (Arizona, USA), a biodiversity and evolution hotspot. In addition, I explored patterns of codiversification of these two birds and their associated gut microbial communities. The two bird studies revealed different patterns of diversification within the sky islands. The White-breasted Nuthatch exhibits a pattern of isolation by environment, where genetic differences among populations are related to environmental differences of those localities. In contrast, the Brown Creeper has a strong genetic break between northern and southern populations, with no evidence of gene flow between lineages. When I investigated codiversification of birds and their microbial communities, I found no relationship between host genetic diversity and microbial community alpha diversity, while genetic differentiation between birds was significantly related with beta diversity between microbial communities. This dissertation provides a first step in comparative evolutionary biology of Madrean Archipelago avian taxa, and adds to the knowledge of the factors shaping microbial community diversity in wild animals.

Acknowledgements

First and foremost I want to thank fellow grad students that impacted me on a daily basis while in Lawrence. This was predominantly people involved with KU Ornithology and the KU Ecological Niche Modeling group; thanks for being there. This included, in no particular order, Lukes #1 and #2 (Campillo and Klicka, you guys decide which is which), Pete Hosner, Mike Andersen, Jack Hruska, Jacob Cooper, Carl Oliveros, Robin Jones, Lynnette Dornak, Lindsay Campbell, Erin Saupe, Chris Hensz, Cori Myers, Hannah Owens, Kate Ingenloff, Andrés Lira-Noriega, and Narayani and Vijay Barve. I also want to thank all the herper, bug (entomology), and *Mimulus* folks for all the positive interactions and fun times: Stephen Baca, Mabel Alvarado, Anthony Barley, Laura Breitkreuz, Kin Onn Chan, Kerry Cobb, Jack Colicchio, Allie Fuiten, Jennifer Girón, Jesse Grismer, Carl Hutter, Pat Monnahan, Karen Olson, Marianna Simões, Emmanuel Toussaint, Scott Travers, and Tashi Zaxicuo.

A large portion of my time at KU involved learning and working with others in the NSF IGERT C-CHANGE program. Thanks to Joane Nagel and Natalie Parker for all of the support through the program and getting us involved in many cool experiences. I want to thank all the other trainees for the camaraderie during the program, especially: Brandon Luedtke, Hannah Owens, Laci Gerhart-Barley, Rebecca Crosthwait, Tony Reames, Tony Layzell, Jacob Carter, Lindsay Campbell, and Travis Wentworth. Nothing beats taking a 1 a.m. full light hike across the frigid landscape of Greenland or driving a boat through a flamingo flock.

I want to thank all the people that facilitated all of my work here and helped me become more proficient in fieldwork, labwork, ornithology, and bioinformatics. Thanks to Jaime Keeler, Lori Schlenker, and Aagje Ashe for answering questions and facilitating any

paperwork or other materials needed while in the program. Pat Monnahan, Nick McCool, and John Kelly were immensely helpful in becoming proficient with RAD-seq labwork, and Carl Oliveros helped me out with target capture lab procedures and some bioinformatics issues. When I was first getting into studying microbes Corrie Moreau and Ben Sikes generously gave me advice. I was very grateful to soak up some of the encyclopedic bird knowledge of Pete Hosner and Mike Andersen while they were here in my first few years.

My time at KU wouldn't have been the same without all the excellent travel involved with my time here. The dissertation fieldwork all took place in the summers of 2012 and 2013 in Arizona and California. Thanks to Mark Geiger and Mark Robbins for providing their time to visit some awesome places out west. With the IGERT, we went to Greenland and Mexico in 2012 and 2013. Thanks to all the IGERT trainees and faculty for a great time on those trips, even though I just got my appendix out right before Mexico. I had a great time traveling into the jungle in Honduras in spring 2014 with Oliver Komar, Justin Hite, and Diego Roldán Piña. In 2014, I got to spend a couple months in the Philippines with some great people, and had a few rough experiences as well. Thanks to Kerry Cobb, Luke Campillo, Rafe Brown, Cam Siler, Nick Huron, Joey Brown, Wilson Bulalacao, Jason Bulalacao, Enteng Yngente, Marvic Yngente, Ben Paulson, Gen Morinaga, Jason Fernandez, and any other locals we spent time with. Finally, I recently got to spend some time in Ethiopia in 2015. Thanks to Stephane Boissinot, Xenia Freilich, and Jacobo Reyes-Velasco for some hectic but fun experiences across the Rift Valley.

I would like to thank John Klicka and Sharon Birks at the University of Washington Burke Museum for tissue loans for the *Certhia* project, as well as the state and federal agencies and wildlife officers for their cooperation and help obtaining permits that

contributed to all of this research (U.S. Department of Fish and Wildlife and the Arizona Game and Fish Department).

This research wouldn't have been possible without funding from many sources. These were the American Museum of Natural History Frank M. Chapman Memorial Fund Grant, an American Ornithologists' Union Student Research Grant, KU Biodiversity Institute Panorama Grants, an NSF Doctoral Dissertation Improvement Grant (DEB-1406989) and through the NSF IGERT C-CHANGE Program #0801522 at the University of Kansas, as well as an NSF grant (DEB-1241181) that funded some of my RA semesters.

Two sequencing labs and their personnel made the sequencing in these projects feasible. I'd like to thank Jenny Hackett and Sarah Owens for their help facilitating Illumina sequencing. Sequencing was performed at the COBRE Genome Sequencing Core Laboratory at the University of Kansas, funded through NIH award number P20GM103638, and the Institute for Genomics and Systems Biology Next Generation Sequencing Facility at Argonne National Laboratories.

One of my favorite things over the past few years has been our KU version of "birding." Thanks to Rob Moyle, Luke Campillo, Luke Klicka, and Brian Finley for sitting in the spectrum of weather with me, and commiserating when we don't add any birds to our "list." It was great to be there when Klicka added a Green-winged Teal and a Wood Duck to his life list. Duck hunting is just plain fun. Hopefully we'll get some more before I head to the desert.

I've been blessed with exceptional roommates over the years in Lawrence. Thanks to Ping Li, Carl Oliveros, Carl Hutter, Luke Campillo, and Stephen Baca. Over the past couple years the 1711 house has been especially great in getting the garden and chickens with

Hutter, Baca, and Campillo. Even with the occasional coop murders, it's worth it. I want to particularly thank both Mark Geiger and Luke Campillo, both of whom I spent the most time with in Lawrence. We've been in the field together and to local events together over the four-ish years I've been here. You guys both made it better, thanks.

I want to thank my friends and family that are in different parts of the country and put up with our distance relationship. I've been lucky enough to be visited in Kansas by Woody Walstrom, Drew Dillmann, Kate Noterman, my parents Brad and Rita, and sister Sarah. Thanks to all of you for coming and checking out what Lawrence was all about. Thanks to my friends and family in Wisconsin, South Dakota, and across the rest of the country for being there via email, phone, Skype, or whichever method we used to communicate.

Lastly I want to thank all those who mentored me during my time here at KU. Thanks to Rafe Brown, Rich Glor, Mark Holder, John Kelly, Town Peterson, Andrew Short, and Jorge Soberón for teaching great classes and being a part of a lot of great discussions about science. Thanks to Garth Spellman for continued support long after my tenure in his lab. Thanks to my committee—Mark Holder, John Kelly, Erik Lundquist, Town Peterson, and Mark Robbins—for providing your time, guidance, and support throughout the time I was here. Finally, I'm thankful for the support and freedom my advisor Rob Moyle gave me during my time here. Walking in the door, he let me work on projects that interested me, while also getting me involved in some Asian bird biogeography.

Everyone in my life the past 4.5 years made my time in Lawrence and travels across the world enjoyable, memorable, and a great chapter of my life.

Table of Contents

Title Page	i
Abstract	iii
Acknowledgements	iv
List of Figures	ix
List of Tables	x
List of Appendices	xi
Introduction	1
Chapter 1: Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach	4
Introduction	6
Methods	9
Results	19
Discussion	26
Chapter 2: A genomic investigation of the putative contact zone between divergent Brown Creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation	31
Introduction	33
Methods	37
Results	45
Discussion	50
Chapter 3: Genetic differentiation among birds (Aves: Passeriformes) explains a portion of the beta diversity among their gut microbial communities	55
Introduction	57
Methods	59
Results	65
Discussion	72
Literature Cited	76
Appendix I	86
Appendix II	87
Appendix III	88

List of Figures

Figure 1.1—Sampling map of *Sitta carolinensis*. 10

Figure 1.2—Niche model projections of S. carolinensis in Arizona, USA. 21

Figure 1.3—Regressions of environmental, geographic, and genetic distances. 22

Figure 2.1—Sampling map and STRUCTURE results of *Certhia americana*. 35

Figure 2.2—Example syllables of *C. americana* songs. 44

Figure 2.3—Relationships of genetic differentiation and chromosome size. 48

Figure 2.4—Environmental variable plot and MANOVA of *C. americana* lineages. 49

Figure 3.1—Sampling map of birds and microbial community data. 60

Figure 3.2—Microbial community composition and PCA plots. 66

Figure 3.3—Regressions of beta diversity and geographic and genetic distances. 69

List of Tables

Table 1.1—*S. carolinensis* voucher samples and RAD coverage. 11

Table 1.2—Geographic and environmental distances between localities. 16

Table 1.3—*S. carolinensis* BEDASSLE results. 24

Table 1.4—Loci putatively under selection in *S. carolinensis*. 25

Table 2.1—*C. americana* voucher samples and RAD coverage. 38

Table 2.2—Song repertoires of *C. americana* individuals. 43

Table 2.3—Statistics of *C. americana* SNP datasets. 46

Table 3.1—Sampling locality coordinates of gut microbial communities. 61

Table 3.2—Regression analyses results. 68

Table 3.3—Partial Mantel test results. 70

Table 3.4—Results of STAMP functional metagenome analyses. 71

List of Appendices

Appendix I. Bioclimatic variables used in *S. carolinensis* ecological niche modeling. 86

Appendix II. Outlier loci in *S. carolinensis* selection tests. 87

Appendix III. DAPC results for *C. americana*. 88

Introduction

Sky islands are disjunct patches of montane forested habitat in a lowland matrix of desert, grasslands, and scrub. Sky islands may be centers for diversification, and have been shown to exhibit some of the same island biogeographic patterns (e.g., species area relationship) as oceanic islands. The sky islands of the Madrean Archipelago in the U.S. Southwest and Northwest Mexico are considered a biodiversity and evolution hotspot, are susceptible to habitat changes (e.g., via landscape or climatic forces), and therefore represent an excellent locale to study ecology and evolutionary biology.

In the Madrean Archipelago, pine and songbird diversity is positively related to sky island size, in accordance with the predictions of MacArthur and Wilson's species area relationship. This region of sky islands has also promoted diversification, with speciation among populations on isolated sky islands (e.g., scorpions); alternatively, the matrix of unsuitable habitat between sky islands has acted as a barrier to gene flow in species with better dispersal ability (e.g., black bears), but has not promoted speciation. Investigations of phylogeographic structure and restrictions to gene flow in Madrean Archipelago taxa have uncovered a plethora of patterns; however, little has been done to investigate genetic and geographic patterns in birds. Additionally, no studies have investigated genomic patterns of diversification in Madrean Archipelago taxa.

Just as animals can be isolated on sky islands, animals may act as islands of habitat for symbiotic bacteria. All animals have communities of microorganisms living with them. Biology's "dark matter" has symbiotic roles ranging from beneficial metabolic processes to contributing to disease. In natural systems (e.g., wild animals), little has been done to

investigate codiversification of hosts and their associated microbial communities, especially in vertebrate species.

The goals of this dissertation were twofold. First, I aimed to investigate processes shaping patterns of genetic diversity and differentiation in birds of the Madrean Archipelago sky islands. Second, I looked to explore patterns and processes affecting bacterial community structure in the guts of their avian hosts. This work largely utilized specimen collections, genomics, geographic information systems, and bioinformatics. In three chapters, I used restriction site associated DNA sequencing (RAD-seq) to explore genomic patterns in the White-breasted Nuthatch (*Sitta carolinensis*, Chapter 1) and the Brown Creeper (*Certhia americana*, Chapter 2), and used targeted amplicon sequencing to characterize the birds' gut microbial communities (Chapter 3).

In Chapter 1, I investigated patterns of genetic diversity and differentiation of the White-breasted Nuthatch across the Madrean Archipelago sky islands. Using large panels of single nucleotide polymorphisms (SNPs), I found a lack of any strong phylogeographic structuring. Geography seemed to play little role in structuring genetic differentiation, because there was no pattern of isolation by distance. Conversely, climatic differences between sampling sites explained genetic differentiation among birds, suggesting isolation by environment.

In Chapter 2, I investigated a putative contact zone between Brown Creeper lineages. In previous studies, I found two strongly structured lineages, with a genetic break somewhere in Arizona. Here, I sampled individuals across the sky islands to identify where the split between lineages occurred, identify possible hybridization, and determine levels of gene flow at a small geographic scale. Analysis of thousands of SNPs revealed no

evidence of hybridization, with a sharp geographic break between lineages. Additionally, widespread chromosomal patterns of genetic differentiation were also apparent within the contact zone, where chromosome size was positively related with genetic differentiation among lineages.

In Chapter 3, I explored how host genetic diversity and differentiation affected the hosts' associated gut microbial community alpha and beta diversities. Although host genetic diversity was not associated with microbial community alpha diversity, genetic differentiation between birds was related to beta diversity between microbial communities. Between the Brown Creeper and White-breasted Nuthatch, the microbial communities were significantly different in several metabolic functional categories.

Overall, this dissertation provides a foundation for future comparative analyses of avian evolutionary biology in the Madrean Archipelago. With the two birds studies here, I found vastly different, but equally interesting, patterns of diversification. By utilizing additional aspects of scientific specimens—by collecting gut samples—I was able to explore patterns of coevolution of these birds and their associated microbial communities. The analysis of the avian genomic and microbial community data identified a significant amount of microbial community vertical transmission; continued investigation of codiversification between hosts and their microbial communities in a small geographic scale and a comparative framework would greatly add to our understanding of the factors shaping these symbioses, as well as further our understanding microbiome functional differences in higher taxa.

CHAPTER 1*

Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach

*Manthey, J. D. & R. G. Moyle. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. Molecular Ecology 24: 3628-3638.

Abstract

Understanding landscape processes driving patterns of population genetic differentiation and diversity has been a longstanding focus of ecology and evolutionary biology. Gene flow may be reduced by historical, ecological, or geographic factors resulting in patterns of isolation by distance (IBD) or isolation by environment (IBE). Although IBE has been found in many natural systems, most studies investigating patterns of IBD and IBE in nature have used anonymous neutral genetic markers, precluding inference of selection mechanisms or identification of genes potentially under selection. Using landscape genomics, the simultaneous study of genomic and ecological landscapes, we investigated the processes driving population genetic patterns of White-breasted Nuthatches (*Sitta carolinensis*) in sky islands (montane forest habitat islands) of the Madrean Archipelago. Using more than 4000 single nucleotide polymorphisms and multiple tests to investigate the relationship between genetic differentiation and geographic or ecological distance, we identified IBE, and a lack of IBD, among sky island populations of *S. carolinensis*. Using three tests to identify selection, we identified 79 loci putatively under selection; of these, seven matched CDS regions in the Zebra Finch. The loci under selection were highly associated with climate extremes (maximum temperature of warmest month and minimum precipitation of driest month). These results provide evidence for IBE—disentangled from IBD—in sky island vertebrates and identify potential adaptive genetic variation.

Introduction

Understanding the processes driving patterns of population genetic differentiation and diversity has been a longstanding focus of ecology and evolutionary biology (e.g., Mayr 1942). Historical, geographic, and environmental effects contribute to current patterns among populations that may have differentiated via diverse mechanisms, both adaptive and non-adaptive. Reduced gene flow between populations because of greater geographic distance, or isolation by distance (IBD, Wright 1943) is commonly seen in empirical studies (e.g., Kuchta & Tan 2005; Sharbel et al. 2000). IBD will manifest whenever dispersal is geographically limited. In contrast, gene flow also can be inversely correlated with environmental distance between populations, resulting in adaptive divergence or isolation by environment (IBE, Nosil et al. 2008). IBE may result from various mechanisms, including local adaptation, migration-selection balance, and non-random gene flow and dispersal (Nosil et al. 2008; Edelaar & Bolnick 2012). These processes (IBD and IBE) are not mutually exclusive, and an expected correlation between environmental and geographic distance will produce similar patterns.

Recent meta-analyses (Sexton et al. 2014; Shafer & Wolf 2013; Orsini et al. 2013) have investigated the prevalence of IBD and IBE in natural systems; all identify multiple examples of IBE in the wild, indicating the importance of the processes creating this pattern. Shafer & Wolf (2013) identified ways to improve IBE studies to enhance understanding of adaptation contributing to genetic differentiation in the wild. First, geographic distance and environmental dissimilarity among populations are often correlated. Disentangling the relative effects of geographic distance and environmental dissimilarity between populations on genetic differentiation is paramount to identifying

IBE. A recent study in anoles (17 *Anolis sp.*; Wang et al. 2013) sought to disentangle geographic and environmental factors shaping genetic differentiation among Caribbean populations; although many species showed significant environmental factors shaping population genetic differentiation, the difference was uncoupled from geographic factors in only a single species (*A. chlorocyanus*).

A second issue identified as crucial for informative IBE studies (Shafer & Wolf 2013) is the scope of the genetic data analyzed. Although IBE patterns have been found using neutral markers (e.g., microsatellites or amplified fragment length polymorphisms), it is unclear whether IBE patterns can be identified genome-wide or only in particular genomic regions. Additionally, identifying adaptive genetic variation may be informative for investigating different stages of ecological speciation (Shafer & Wolf 2013). Recently, techniques to acquire thousands of genetic loci across the genome [e.g., restriction site associated DNA (RAD) markers (Miller et al. 2007)] allow researchers to investigate environmental and genomic landscapes simultaneously. The investigation of spatially explicit adaptive genetic diversity and differentiation—landscape genomics (Manel & Holderegger 2013)—has flourished, including new conceptual frameworks for the identification of adaptive genetic variation (Orsini et al. 2013) based on the expected relationships between neutral and non-neutral genetic variation and geographic or environmental distances between populations.

Landscape genomics requires thorough sampling across the genomic landscape as well as the environmental landscape; as such, heterogeneous landscapes provide great potential for investigating adaptive genetic variation across species' ranges. Arizona (U.S. Southwest) has a highly heterogeneous landscape, with many disjunct, isolated mountain

ranges that harbor montane forest separated by lowlands with scrub, grass, or desert. These isolated forests act as islands of habitat, or sky islands, for species that are unable to persist in lowland, non-forested, habitats. The sky islands are located in the transition between temperate and subtropical biomes, increasing the heterogeneity of the landscape.

The White-breasted Nuthatch (*Sitta carolinensis*) is a widespread, North American songbird that inhabits the sky islands of Arizona. Although *S. carolinensis* has phylogeographic structure across North America (Spellman & Klicka 2007; Walstrom et al. 2012), the populations of Arizona are all included in a single clade. In Arizona, the species inhabits oak, pine-oak, and pine woodlands, which form a contiguous tract of habitat through the centre of Arizona along the Mogollon Rim, and many isolated tracts of sky island habitat in the mountain ranges of southern Arizona (Fig. 1.1), also known as the Madrean Archipelago. In the sky islands, the subalpine flora and fauna are elevation-limited by temperature and precipitation extremes (high temperature and low precipitation; Poulos & Camp 2010) along elevational gradients, providing a potential context for selection or biased dispersal that may lead to a pattern of IBE.

Here, using geographic data, ecological data, and thousands of genetic loci obtained using Illumina sequencing, we characterize differentiation among and diversity within populations of *S. carolinensis* in the sky islands of Arizona. With these data, we investigated the following questions and hypotheses:

(1) What patterns are evident in genetic diversity within populations?

H_0: Population genetic diversity is similar across populations.

H_{A1}: Larger sky islands will have higher genetic diversity.

H_{A2}: More suitable environmental localities will have higher genetic diversity.

(2) Does geographic distance or environmental dissimilarity between populations shape patterns of genetic differentiation among populations?

H_0: Population genetic differentiation is structured by limited random dispersal, resulting in a pattern of IBD.

H_A: Habitat differences—and their associated temperature and precipitation differences—will shape genetic differentiation among populations more than geographic distance between populations due to environmental heterogeneity and the fragmented landscape.

Methods

Sampling, laboratory procedures, and SNP dataset creation

Fresh tissue samples of 27 *S. carolinensis* (Rocky Mountain phylogeographic group) were obtained from across the sky islands of Arizona (three from each locality; Table 1.1; Fig. 1.1). The sky islands are islands of montane forest habitat surrounded by lowland desert and scrub. One eastern *S. carolinensis*, and one *S. europea* were used as outgroup samples to confirm that ingroup samples belong to a single clade. Genomic DNA was extracted using a QIAGEN DNeasy blood and tissue extraction kit following manufacturer protocols.

To obtain many anonymous genetic loci, we performed a modified RAD-seq (Miller et al. 2007) protocol. DNA samples were digested with the restriction enzyme NdeI to produce a reduced representation genomic library. Following ligation of custom adapters with attached barcodes for multiplexing (one barcode per individual, minimum two

Figure 1.1. Map of the Madrean Archipelago sky islands of southern Arizona, within the United States of America, and sampling localities for this study. Dark grey areas correspond to montane pine-oak and pine habitats.

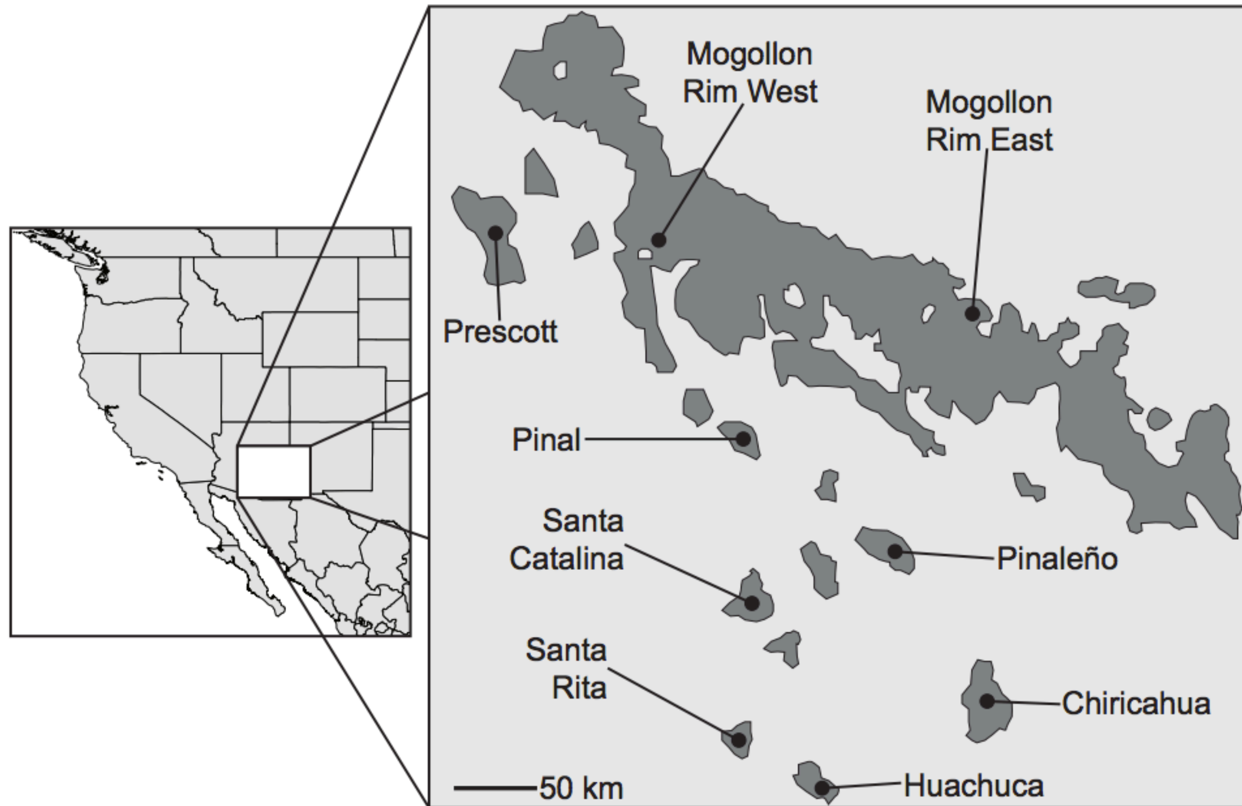


Table 1.1. List of localities, latitude and longitude coordinates, voucher number at the University of Kansas Biodiversity Institute, number of reads for each individual from Illumina sequencing run (# Reads), coverage of each individual from the SNP dataset (Cov.), and mean and standard deviation of sequence reads per locus (Reads Cov.).

Locality	Lat.	Long.	Voucher	# Reads	Cov.	Reads Cov.
Chiricahua	31.78	-109.30	KU 31223	1,025,931	96.3%	88.00 (34.51)
			KU 31225	1,548,897	77.2%	137.40 (54.61)
			MBR 8613	100,932	14.2%	5.82 (3.81)
Huachuca	31.43	-110.29	JDM 054	1,590,374	92.6%	135.30 (49.22)
			MBR 8587	1,237,388	82.6%	33.77 (17.84)
			MBR 8588	684,026	95.2%	47.09 (19.76)
Mogollon East	33.92	-109.27	KU 31224	2,173,135	94.1%	98.44 (40.64)
			MBR 8611	1,010,500	59.5%	16.82 (11.11)
			MBR 8615	583,512	50.9%	10.92 (8.43)
Mogollon West	34.48	-111.42	JDM 056	285,311	75.9%	23.03 (10.74)
			JDM 058	295,361	78.9%	19.52 (11.08)
			MBR 8600	74,929	7.9%	5.38 (2.79)
Pinal	33.29	-110.87	JDM 052	135,912	23.4%	7.94 (3.94)
			JDM 053	125,285	40.5%	10.63 (5.21)
			MBR 8582	1,540,003	88.1%	94.29 (38.33)
Pinaleño	32.69	-109.83	MBR 8598	172,636	46.4%	9.52 (4.62)
			MBR 8599	818,232	91.6%	54.23 (21.39)
			MBR 8605	228,966	62.6%	21.81 (12.78)
Prescott	34.41	-112.42	MBR 8576	987,402	90.5%	41.18 (17.72)
			MBR 8577	377,098	69.3%	23.80 (10.52)
			MBR 8610	2,192,338	93.3%	141.00 (52.30)
Santa Catalina	34.42	-110.73	KU 31220	834,036	95.0%	38.45 (16.10)
			KU 31221	937,297	92.6%	41.31 (20.20)
			KU 31222	126,096	21.7%	7.94 (4.65)
Santa Rita	31.70	-110.88	JDM 059	834,785	92.9%	38.08 (20.26)
			JDM 060	485,778	90.7%	30.79 (13.60)
			MBR 8608	352,386	53.4%	10.44 (6.26)

differences between barcodes), we pooled and purified all samples using AMPure magnetic beads (Agencourt). Size selection of fragments (500 – 600 bp) using a Pippin Prep electrophoresis cassette (Sage Science) further reduced the genomic libraries, followed by another DNA purification step. A brief PCR reaction of the pooled samples was performed in quadruplicate using an initial denaturation period of 98°C for 30 s, 14 cycles of 98°C for 10 s, 64°C for 30 s, and 72°C for 20 s, and a final extension of 72°C for 7 min. The pooled library was tested for quality and quantity of DNA using quantitative PCR and the Agilent Tapestation at the University of Kansas Genome Sequencing Core, followed by sequencing of 100 bp single-end reads on a partial lane of an Illumina HiSeq2500.

Using the STACKS (Catchen et al. 2011) pipeline, we assembled loci de novo from the fastQ files obtained from the Illumina sequencing run. The *process_RADtags* python script included in STACKS was used to assign sequence reads to individuals and remove sequencing reads of poor quality. We used a quality threshold for inclusion as an average phred score of ten in sliding windows of 15 bp. Sequences with possible adapter contamination or lacking the restriction site were removed. After quality control, the *ustacks*, *cstacks*, and *sstacks* modules of STACKS were used with the default settings, but with a modified number of mismatches allowed between individuals from two to five. Finally, we used the populations module of STACKS to create a SNP dataset with the following restrictions: minimum stack depth of five, at least two individuals per sampling locality, minimum minor allele frequency of 0.05, and observed heterozygosity less than 50% (to reduce inclusion of paralogous loci). To ensure that stack depth did not influence our results, we used varying minimum levels (m = 1, 5, 10, 15). This changed the number of loci between datasets, but did not influence levels of genetic differentiation between

populations (all $r > 0.84$, $p < 0.001$). Additionally, to ensure that the number of individuals with low sequencing coverage did not influence the results, we ran STACKS limited to individuals with greater than 50% coverage (Table 1.1; causing exclusion of the Pinal locality). Again, number of loci changed, but differentiation between populations did not ($r = 0.965$, $p < 0.001$). Finally, we ran STACKS including only loci found in 90% of individuals. Differentiation between populations was unaffected ($r = 0.97$, $p < 0.001$). Based on these sensitivity tests, we continued with the original dataset (minimum stack depth of 5 and all individuals). To check for even coverage across the genome, we used a BLAST+ search (Camacho et al. 2009). Here, we BLASTed against the Zebra Finch (*Taeniopygia guttata*) genome, and considered loci a match if they had 70% sequence identity and a maximum e-value of 0.01.

Ecological niche modeling and measuring niche centrality

To assess the current and potential Last Glacial Maximum (LGM) distribution of *Sitta carolinensis*, we created an ecological niche model with the following methodology. We obtained occurrence localities from Global Biodiversity Information Facility (GBIF) and ORNIS, online and curated repositories of museum and herbarium specimens. Occurrence points were visualized and quality-checked in ArcMap v10.1 followed by rarefication to remove points within two km of one another using custom R (R Development Core Team 2012) scripts. Twenty per cent of points were set aside to check model quality. This resulted in 2294 training and 574 testing points.

19 bioclimatic layers were obtained from the WorldClim database (Hijmans et al., 2005; www.worldclim.org). These layers contain worldwide precipitation and temperature

information, including minima, maxima and ranges of values. To reduce using layers with high correlation (R > 0.8; measured in R) across North America, we included only 11 in analyses (Appendix I). Additionally, we obtained bioclimatic layers for the LGM (~21,000 ya), which were produced from model output of the Model for Interdisciplinary Research on Climate (MIROC; Nozawa et al., 2005) and the Community Climate System Model (CCSM3; Collins et al., 2006).

Ecological niche models were created using Maxent (Phillips et al. 2006), which relates environmental variables with species' occurrences to estimate environmental requirements and potential distributional areas. We used a model training area of 500 km buffer around all occurrence points, which we determined as a reasonable region accessible to *S. carolinensis* over evolutionary time (i.e., since the LGM). We set a threshold on output models as inclusion of 95% of training points, allowing 5% omission based on estimated error rates (Peterson et al. 2008). We assessed model performance using a cumulative binomial test, using the data points set aside before model creation; here, the null probability of prediction of occurrence points is the proportion of area predicted in the model across the training area (Peterson et al. 2011).

To characterize each population's position in environmental space, we used the methods of Lira-Noriega & Manthey (2014). Briefly, we extracted 5000 random points from the ecological niche model thresholded region and extracted environmental data from the 19 bioclimatic variables for those points. We transformed this matrix of environmental data using principal component analysis. The mean of scores along the first four principal components, which explained 92% of the variance in the environmental data, was used as an estimate of the species' niche centroid. For each locality, we measured the Euclidean

distance to niche centroid (i.e., we used the environmental values as "coordinates," and measured the distance between them). Additionally, for all pairwise comparisons of populations, we measured Euclidean distance between populations in environmental space. Distances between populations, in both geographic and environmental distances, are shown in Table 1.2. Principal component analysis and measures of environmental distance were all performed in R (R Development Core Team 2012).

To assess habitat size on each sky island and distance to "mainland" (i.e., the Mogollon Rim or Sierra Madre Occidental) we used the Global Land Cover Facility's Tree Cover Continuous Fields dataset, derived from Advanced Very High Resolution Radiometer (AVHRR) satellite imagery (DeFries et al. 2000), to classify forest cover for each sky island. This dataset classifies percent tree cover at 1 km resolution. Using this data, we calculated the area of each sky island (i.e., amount of suitable habitat) as the amount of the region with a minimum of 20% tree cover. To identify the distance to the nearest mainland, we calculated the minimum distance between forested regions in each sky island to forested regions in the mainland using ArcMap v10.1.

Relationships of genetic and ecological data and investigation of selection

To investigate relationships between genetic and ecological characteristics, we performed linear regressions between habitat/ecological conditions and genetic parameter estimates (e.g., F_{ST} and nucleotide diversity), using the following comparisons: (1) current suitable-habitat island size and genetic diversity; (2) niche centrality and genetic diversity; (3) geographic distance between populations and population differentiation (F_{ST}); and (4) environmental distance between populations and population differentiation. The use of F_{ST}

Table 1.2. Geograpic (km) and environmental (PCA-based) distances between populations below and above the diagonal, respectively.

	Prescott	Pinal	Santa Catalina	Santa Rita	Huachuca	Chiricahua	Pinaleño	Mogollon West	Mogollon East
Prescott	-	0.99	1.01	1.58	2.15	1.38	1.44	0.35	1.39
Pinal	189.78	-	1.04	1.36	1.46	1.10	0.95	1.28	2.29
Santa Catalina	271.30	97.65	-	0.62	1.66	1.01	1.61	1.30	1.81
Santa Rita	333.85	176.85	81.32	-	1.33	0.91	1.72	1.88	2.19
Huachuca	386.51	213.94	117.69	63.47	-	0.91	1.20	2.48	2.96
Chiricahua	412.37	223.33	152.38	149.72	101.54	-	0.93	1.70	2.05
Pinalenos	306.96	117.76	89.56	147.96	146.70	112.83	-	1.69	2.37
Mogollon West	92.05	141.77	237.90	313.28	355.25	359.39	247.67	-	1.29
Mogollon East	294.96	163.95	215.20	289.16	292.95	238.04	146.38	207.36	-

with small sample sizes is sufficient for estimating genetic differentiation here because of the inclusion of thousands of genetic markers (Willing et al. 2012). All regressions were performed in R (R Development Core Team 2012).

Given the inherent non-independence of data points in pairwise comparisons, we investigated relationships between genetic differentiation, geographic distance, and ecological distance using two additional methods. First, we performed Partial Mantel tests in R (R Development Core Team 2012), assessing significance with 10,000 permutations. Second, we used BEDASSLE, which models covariance structure in allele frequencies between populations as a decreasing function of ecological and geographic distance (Bradburd et al. 2013). We used BEDASSLE because it simultaneously estimates the relative effects of geographic and environmental distances between populations on genetic differentiation. BEDASSLE uses a Bayesian framework and estimates model parameters with an MCMC algorithm. We ran BEDASSLE—using the beta-binomial model—for 8 million generations and sampling every 100. Performance of the model was evaluated by visualizing MCMC acceptance rates and parameter trace plots as suggested by Bradburd and colleagues (2013). Based on stabilization in trace plots, we discarded the first four million generations as burn-in and used the remaining samples to estimate the relative effect size of ecological versus geographic distance ($\alpha E/\alpha D$). To obtain more interpretable results from BEDASSLE, rather than simply from the PCA-based environmental values, we ran BEDASSLE again with two environmental variables: minimum precipitation of the driest month and maximum temperature of the hottest month, two environmental characteristics important for determining plant communities in the sky islands. Here, we ran BEDASSLE for 10 million generations with the first eight million as burn-in.

To assess potential selection in the SNP dataset, we used three methods. Because of the nature of the dataset (reduced representation library of the genome), it is unlikely that our data included the adaptive genetic variation directly impacting local adaptation, but rather regions of the genome that may be linked with adaptive genetic variation. First, we used BayeScan (Foll & Gaggiotti 2008) to identify diversifying or balancing selection based on allele frequency differences among populations. BayeScan compares the posterior probability of a neutral model with a population-level F_{ST} shared across all loci with the posterior probability of a selection model that incorporates locus-specific F_{ST} estimates to explain differences in allele frequencies among populations. BayeScan was run for 20 pilot runs, with a final run containing a burn-in period of 50,000 iterations followed by an additional 50,000 iterations sampled every ten. We used the default settings for the F_{IS} distribution and prior odds for the neutral model. To determine significance of BayeScan results, we interpreted the log posterior odds ratio using Jeffreys' scale of evidence (Jeffreys 1961); with this scale, a value of one is considered strong evidence for selection.

We used latent fixed mixed modeling (LFMM) to investigate associations between environmental variables (minimum precipitation of driest month and maximum temperature of hottest month) and SNPs while accounting for population structure in the data (Frichot et al. 2013). LFMM uses environmental variables as fixed effects, and population structure is modeled as latent factors (i.e., an inferred variable; Frichot et al. 2013). We ran LFMM with the full range of latent factors (assumed population structure between one and nine distinct populations), using 1000 iterations as burn-in followed by 10,000 iterations to compute LFMM parameters. Significance was applied at an alpha value of 0.001 with a Bonferroni correction for multiple testing.

Lastly, we used Bayenv2 (Günther & Coop 2013) to investigate relationships between population allele frequencies and two environmental variables (minimum precipitation of driest month and maximum temperature of hottest month). Bayenv2 uses a Bayesian model to identify correlations between environmental variables and outlier alleles while accounting for sampling and covariance due to population history (Günther & Coop 2013). Bayes factors for all SNPs (output of Bayenv2) were interpreted using Jeffreys' scale of evidence (Jeffreys 1961), again using a value of one as strong evidence for selection.

Results

Characteristics of sequence data

From the partial Illumina HiSeq2500 lane, we obtained 20,758,546 sequencing reads from 27 individuals. When barcodes and restriction sites were trimmed, this resulted in a total of ~1.89 billion bp. Following quality control (i.e., process_radtags), the number of sequencing reads across individuals was highly variable; with a mean of 768,825 reads and a range of 74,000 to 2.2 million reads (median = 684,026, standard deviation = 620,690). The number of RAD-tags per individual ranged from 5971 to 48,430 (mean = 26,590). This resulted in a total of 6734 loci (2635 polymorphic) with at least two individuals represented per sampling locality. Among polymorphic loci, the mean SNPs per locus was 1.59 (median = 1, standard deviation = 0.89). We found a strong relationship between chromosome size [based on BLAST+ search (Camacho et al. 2009)] and number of polymorphic loci ($R^2 = 0.963$, $p < 0.001$), suggesting our SNPs were spread evenly across

the genome. All pairwise F_{ST} values were low (< 0.09), suggesting current or recent connectivity among populations.

Ecological data

The ecological niche model of *S. carolinensis* (not shown) predicted distributional areas that correspond well with known range limits. Quantitatively, the model showed better predictive ability than null expectations based on a binomial test (510/574 testing points predicted; $p << 0.001$). In Arizona, the niche model generally corresponds with known suitable habitat of the sky islands (Fig. 1.2A). Because of the qualitative and quantitative effectiveness of the niche model's predictability, we were justified to explore correlates of genetic data and ecological data derived from the niche model. When projected to the LGM, the niche model generally showed a wide swath of potentially suitable area across most of the state of Arizona (Fig. 1.2B), implying the potential for widespread connectivity of sky island populations as recently as the LGM.

Genetic diversity could not be explained by suitable habitat size for each population ($p = 0.094$) or distance from a mainland ($p = 0.422$); similarly, distance from niche centroid did not explain genetic diversity of the sky island populations ($p = 0.688$) and was not related to sky island habitat size ($p = 0.104$). Although geographic distance and ecological distance among populations was linked (Fig. 1.3A), regressions between geographic and genetic distance (isolation by distance) were not significant (Fig. 1.3B; $p = 0.623$), whereas the relationship of ecological distance between populations and genetic differentiation (isolation by environment) was significant (Fig. 1.3C; $R^2 = 0.346$, $p < 0.001$). However, because the ecological distance partially encompasses geography (Fig. 1.3A), the ecological

Figure 1.2. Projections of ecological niche models of *S. carolinensis* in Arizona for current conditions (A) and the Last Glacial Maximum (B). In (B), dark grey regions represent areas predicted by one of two climate scenarios and black regions predicted by both climate scenarios.

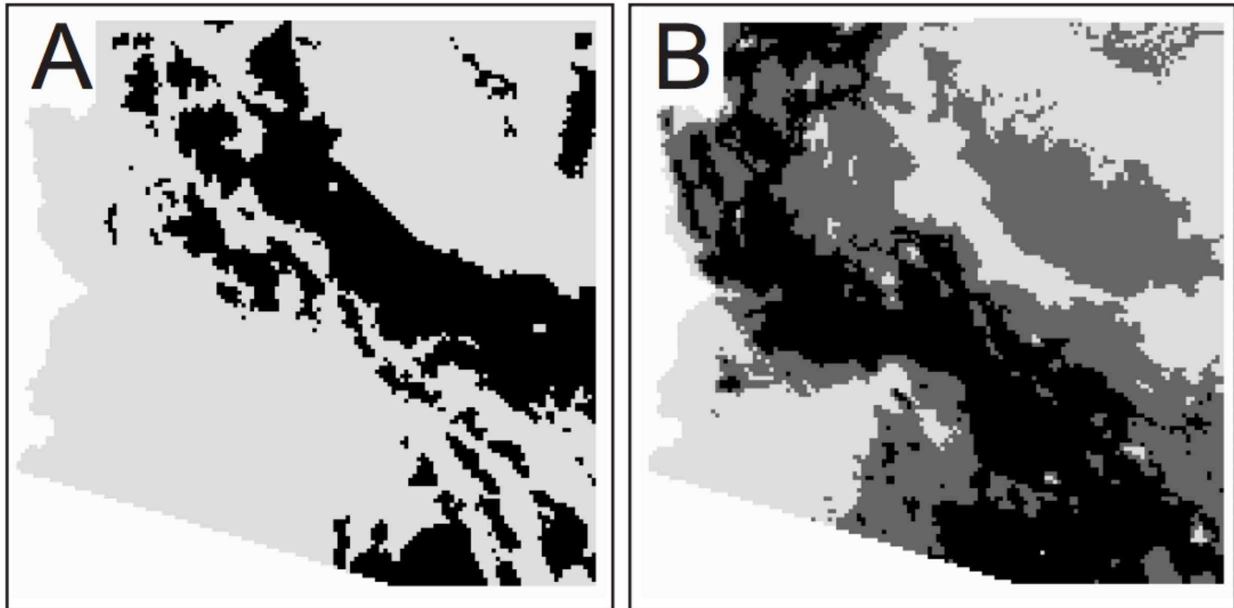
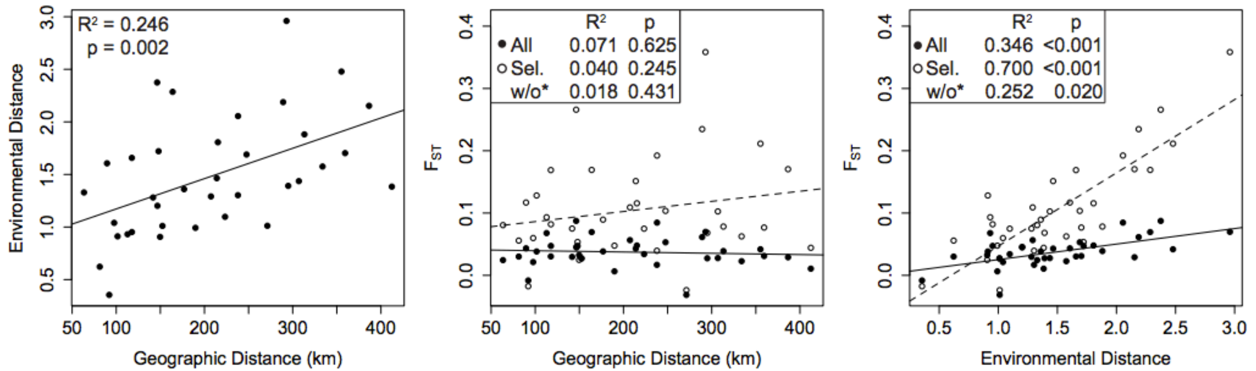


Figure 1.3. Regression of genetic and environmental distance among all population pairwise comparisons (A), regressions of genetic differentiation (F_{ST}) and geographic distance among populations (B), and regressions of genetic differentiation and environmental distance (C; in PCA-transformed units). In (B) and (C), solid points indicate the full (All) dataset and open points only the outlier loci putatively under selection (Sel.). Results (R^2 and p-value) for regressions omitting outlier loci (w/o*) are reported but not plotted.



distance may be interpreted in a biologically more meaningful context as a combination of geographic and environmental distance.

Partial Mantel tests showed no relationship between genetic differentiation and geographic distance between populations ($r = -0.5362$, $p = 0.98$), whereas they identified a strong relationship between genetic differentiation and ecological distance between populations ($r = 0.7283$, $p < 0.001$). Two replicates of BEDASSLE identified no effect of geographic distance, relative to ecological distance, on genetic differentiation ($\alpha E/\alpha D$; Table 1.3). Separate BEDASSLE analyses identified a strong relationship—relative to geographic distance—of minimum precipitation of driest month with genetic differentiation between populations (Table 1.3) and a weaker relationship with maximum temperature of hottest month.

Investigation of selection

Using BayEnv2 and LFMM, we found 36 and 50 outlier SNPs, respectively (Appendix II), including eight in common between analyses (Table 1.4). Among these 78 loci, outgroup individuals contained both major and minor alleles. Because only eight loci overlapped between analyses, we may interpret the other loci as only indicating weak selection, or due to spurious results through chance and which highlight the potential limitations of selection scans using reduced representation genomic sampling. The eight loci in common between analyses BLASTed only to large chromosomes (1-3; Table 1.4), and were not indicative of obvious functional genes associated with climatic conditions. In the program BayeScan, one locus was identified as a potential outlier for selection (BLASTed to

Table 1.3. Results of BEDASSLE runs. Two replicates were run for two datasets. The first compared the relative effect of the PCA-based environmental (E) variable vs. geographic distance (D) on genetic differentiation (S). The second compared the relative effect of maximum temperature of hottest month (T) and minimum precipitation of driest month (P) vs. geographic distance. These can be interpreted as the relative effect of a single degree centigrade or millimeter precipitation vs. a single kilometer distance between populations and the associated genetic differentiation between populations. Results are rounded to nearest integer.

	Mean $\alpha E/\alpha D$	95% Confidence Interval
Run 1 (S)	17,228	9374 – 23,988
Run 2 (S)	19,839	12,841 – 31,456
Run 3 (T)	880	670 – 1140
Run 4 (T)	4500	200 – 1030
Run 3 (P)	9249	6898 – 10,959
Run 4 (P)	16,409	7790 – 23,006

Table 1.4. Outlier loci identified in both LFMM and BayEnv2. For LFMM, results are shown when one or nine latent factors (LF) were used to test associations between environmental variables and SNPs. Environmental variables are temperature of warmest month (TWM) and precipitation of driest month (PDM). BLAST results (to Zebra Finch genome) are shown to chromosome (Chr.) and gene (BLAST).

Locus	LF1 TWM	LF9 TWM	LF1 PDM	LF9 PDM	BayEnv2 TWM	Bayenv2 PDM	Chr.	BLAST
1942			x			x	1	-
4782			x	x		x	1	-
6137	x		x	x		x	1	ephrin-B2
8133			x	x		x	1	uncharacterized protein LOC100190155
14162	x	x	x	x		x	3	Neuroblastoma-amplified gene protein
19820			x	x		x	3	-
22782			x	x		x	-	-
25589			x	x		x	2	-

chromosome 3), although it matched none of the loci identified as outliers with environmental variables.

Of the 79 SNPs identified as putatively under selection, 60 were identified to Zebra Finch chromosomes and seven of these were located in CDS regions (~8.9%), although this was not significantly different than the entire dataset and may be inflated from null expectations because of the methods used for ascertainment of the SNP data. Because we found loci potentially under selection, we again investigated isolation by distance (both geographic and ecological) with F_{ST} values estimated only from loci identified as under selection. Again, regressions between geographic and genetic distance were not significant (Fig. 1.3B; $p = 0.245$), while the relationship between ecological and genetic differentiation was significant (Fig. 1.3C; $R^2 = 0.700$, $p < 0.001$) and more extreme than estimates from all loci (slope of 0.118 compared to 0.025). When loci putatively under selection were removed from the dataset, the regressions' significance did not change (Figs. 1.3B, 1.3C).

Discussion

We identified IBE as the pattern of genetic differentiation in *S. carolinensis* in the Madrean Archipelago sky island region. Of loci putatively under selection (1.2% of loci), ~9% were identified in genic regions (7/60 in CDS; i.e., linked to potential adaptively important genes). In their recent review paper, Orsini and colleagues (2013) summarized the processes driving genetic differentiation in landscape genetics studies. They developed a framework for differentiating between IBD, IBE, and isolation by colonization (e.g., founder effects). In Orsini et al.'s (2013) framework, IBE would be identified if there was no relationship between genetic differentiation and geographic distance between populations,

but a positive relationship between genetic differentiation and environmental dissimilarity between populations. Neutral markers would show a small positive relationship (or small slope in a regression) with markers putatively under selection showing a larger positive relationship (or larger slope) between genetic differentiation and environmental dissimilarity among populations.

Our data (Fig. 1.3) match the patterns associated with IBE identified by Orsini and colleagues (2013) exactly. IBE has been inferred in many species, including vernal grasses (Freeland et al. 2010), fishes (Bond et al. 2014), birds (Smith et al. 2005), mammals (Dudaniec et al. 2013), amphibians (Dudaniec et al. 2012), and invertebrates (Funk et al. 2011, Nosil et al. 2008); however, all of these studies used few markers (tens to hundreds) and no sequence data (all were amplified fragment length polymorphisms or microsatellites), precluding inference of selection mechanisms. Only a study on sticklebacks (Deagle et al. 2011) identified IBE using high volumes of SNP data (1509 SNPs), which can be matched to annotated gene regions of the stickleback genome. The geographic scope of our data is also notable, because in birds, few studies have shown genetic variation associated with environmental characteristics across small geographic regions (e.g., Garroway et al. 2013, Pavlova et al. 2013).

Using three different tests for selection, we identified 79 loci potentially under selection, of which seven were in CDS regions. These genetic markers were all associated with environmental factors that limit distributional ranges on each sky island: maximum temperature of warmest month and minimum precipitation of driest month. These SNPs are located in genic regions, so they represent potential adaptive genetic variation or may be linked to adaptive genetic variation (e.g., other SNPs on same gene). Because the

methods used to identify these outlier loci intrinsically use environmental-SNP associations, they provide possible underlying mechanisms driving divergence among populations. The subalpine flora and fauna inhabiting the sky islands of Arizona are elevation-limited by temperature and precipitation extremes (high temp. and low precip.; Poulos & Camp 2010) in the elevational transect from pine forest to scrub to desert. Additionally, because outgroup genotypes for the loci under selection included major alleles and minor alleles, selection is likely acting on standing genetic variation, although we cannot exclude the possibility of novel mutations. This highlights the effects of large population sizes' standing genetic variation and the potential for adaptation from this variation (Barrett & Schluter 2008).

Potentially adaptive genetic variation associated with temperature and precipitation extremes provides a starting point for understanding the adaptability of species in the Madrean Archipelago sky islands. This process is particularly important in advancing knowledge of species susceptibility to climate change and landscape fragmentation. Recent studies of climate-change sensitive species, the American pika (Henry & Russello 2013) and Atlantic salmon (Bourret et al. 2013), have investigated selection across the genome to identify adaptive genetic variation. Because sky island species are susceptible to environmental changes, they should be investigated for potential adaptive genomic variation to understand future responses to selection pressures caused by environmental changes. In the face of climate change, sky island populations that experience changing temperature regimes will be forced to track climate and habitat, adapt to new conditions, or face extirpation.

Our main result of IBE can be explained by a combination of restricted gene flow due to some level of reproductive isolation among populations and linkage of outlier loci to genomic regions actually under selection. However, since there are presumably high levels of gene flow among populations in the recent evolutionary past (pairwise-F_{ST} values between -0.037 and 0.087), any reproductive barriers among populations must be weak or recently developed. Further evidence of this point is the lack of a relationship between geographic and genetic distance among populations (Fig. 1.3B), suggesting neutral gene flow among populations is not restricted or has only recently ceased.

Alternatively, the results may reflect non-random gene flow among populations, where nuthatches are more often selecting more suitable environments. Because the environmental differences between habitats (and their associated temperature and precipitation differences)—compared to geographic distances between populations—are a large factor in population genetic differentiation (Table 4.1), the birds may be specifically selecting different habitats (e.g., oak vs. pine). Southern Arizona includes the transition between temperate and subtropical habitats, potentially providing an intrinsic habitat gradient allowing for accumulation of genetic differentiation due to selection, migration-selection balance, or non-random dispersal. Non-random gene flow and dispersal may be more likely than reproductive isolation among populations as different clades of White-breasted Nuthatches inhabit different habitat types (Walstrom et al. 2012), which could have been a possible contribution to differentiation among phylogeographic clades.

Because ~9% of putative loci under selection are in CDS regions, it is possible that gene flow has been restricted recently, and local adaptations have accumulated rapidly since the LGM (i.e., environmental selection on a recent evolutionary timescale). Although

no studies have investigated natal dispersal in *S. carolinensis*, a study measuring dispersal in the closely-related *S. europaea* identified a maximum of ~10 km dispersal per generation in a fragmented landscape (Matthysen et al. 1995), suggesting fragmentation since the LGM has likely played a role in limiting gene flow between sky islands, although not limited enough to manifest in a pattern of IBD. Further work should investigate these genomic regions with greater SNP density, also including targeted sequencing to identify potential adaptive protein changes near genomic regions with signatures of selection.

This investigation of *Sitta carolinensis* provided a first genomic assessment of processes driving genetic differentiation among sky island vertebrates and identified IBE, with a lack of IBD. Additionally, this provided a first step in identifying potential adaptive variation in a species inhabiting climate-change sensitive ecosystems.

CHAPTER 2*

A genomic investigation of the putative contact zone between divergent Brown Creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation

* Manthey, J. D., Robbins, M. B. & R. G. Moyle. 2015. A genomic investigation of the putative contact zone between divergent Brown Creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation. Genome doi: 10.1139/gen-2015-0093

Abstract

Sky islands, or montane forest separated by different lowland habitats, are highly fragmented regions that potentially limit gene flow between isolated populations. In the sky islands of the Madrean Archipelago (Arizona, USA), various taxa display different phylogeographic patterns, from unrestricted gene flow among sky islands to complex patterns with multiple distinct lineages. Using genomic-level approaches allows the investigation of differential patterns of gene flow, selection, and genetic differentiation among chromosomes and specific genomic regions between sky island populations. Here, we used thousands of SNPs to investigate the putative contact zone of divergent Brown Creeper (*Certhia americana*) lineages in the Madrean Archipelago sky islands. We found the two lineages to be completely allopatric (during the breeding season) with a lack of hybridization and gene flow between lineages and no genetic structure among sky islands within lineages. Additionally, the two lineages inhabit different climatic and ecosystem conditions and have many local primary song dialects in the southern Arizona mountain ranges. We identified a positive relationship between genetic differentiation and chromosome size, but the sex chromosome (Z) was not found to be an outlier. Differential patterns of genetic differentiation per chromosome may be explained by genetic drift—possibly in conjunction with non-random mating and non-random gene flow—due to variance in recombination rates among chromosomes.

Introduction

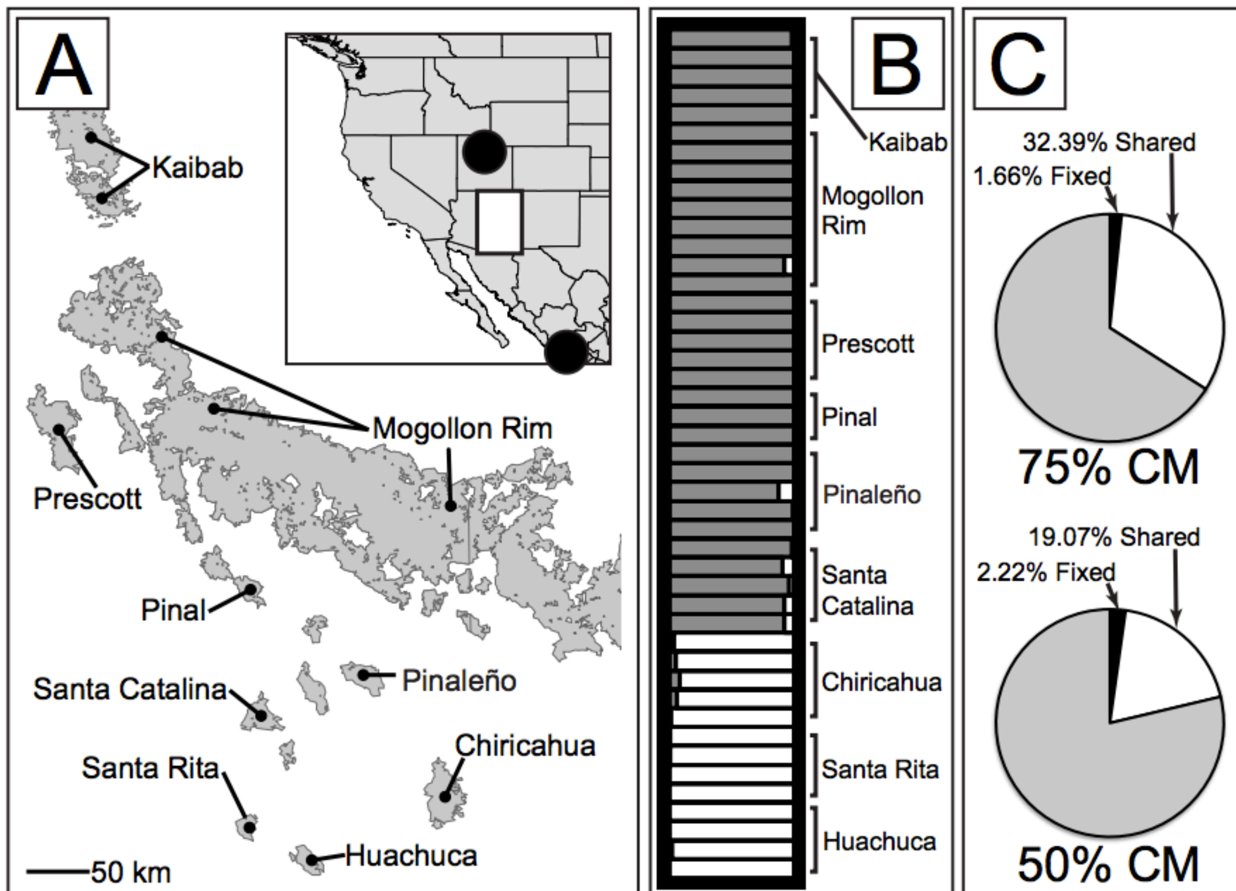
Understanding the processes driving lineage divergence and limiting gene flow across species boundaries has been a longstanding focus of evolutionary biology (Mayr 1942). Highly fragmented landscapes escalate restrictions to gene flow among populations, potentially leading to increased genetic differentiation. The sky islands of the Madrean Archipelago represent one such example of a fragmented landscape for forest species; here, isolated patches of forest (i.e., the sky islands) are separated by desert, grassland, and scrub habitat. In poorly dispersing species, independent lineages have evolved on different sky islands, suggesting little dispersal between populations [e.g., jumping spiders (Masta 2000), bears (Atwood et al. 2011)]. In contrast, species with higher dispersal ability may show fewer restrictions on gene flow among sky islands [e.g., birds (Manthey & Moyle 2015)]. With new methods for obtaining reduced-representation genomic libraries [e.g., restriction-site associated DNA sequencing (RAD-seq); Miller et al. 2007] for many individuals, genetic structure and gene flow between sky islands can now be investigated at the genomic scale. Although researchers have taken a genomic approach to investigate gene flow and selection across species boundaries in many organisms, including—but not limited to—trees (Hersch-Green et al. 2014), mussels (Gardner & Wei 2015), snakes (Schield et al. 2015), fishes (Malek et al. 2012), birds (Lavretsky et al. 2015), and mammals (Janoušek et al. 2012), sky island taxa have received little attention at the genomic level.

Songbirds have high levels of interchromosomal synteny (Kawakami et al. 2014) and the genomic resources of a well-annotated genome (Estrildidae: *Taeniopygia guttata*; Warren et al. 2010). These patterns and tools thus allow the use of high-throughput sequence data of non-model songbirds, with the possibility to identify genomic regions

biased toward a lack of gene flow or increased selective pressures (e.g., sex chromosomes). Biased levels of divergence, lack of gene flow, and greater selective pressures on sex chromosomes have been found in many avian contact zones using large genetic datasets, including flycatchers (*Ficedula*; Ellegren et al. 2012), chickadees (*Poecile*; Taylor et al. 2014), and sparrows (*Passer*; Elgvin et al. 2011), suggesting a common pattern in the Z chromosome's role in speciation. Additionally, this pattern has been found in ducks across populations and species (Lavrestsky et al. 2015).

In contrast, a recent study investigating genomic patterns of differentiation between two divergent lineages of Brown Creeper (*Certhia americana*) did not find biased differentiation of the Z chromosome between lineages, but rather a positive relationship between chromosome size and genetic differentiation (Manthey et al. 2015a). The two divergent [~5% mitochondrial DNA (mtDNA) divergence; Manthey et al. 2011a] lineages come into contact in Arizona, although it is unclear exactly where the contact occurs. Marshall (1956) studied the plumage morphology of *C. americana* in Arizona (USA), where the defining differences between northern and southern forms were coloration of the rump (tawny vs. chestnut, respectively) and coloration of the underparts (white vs. sooty relative to throat color). He found the transition from northern to southern birds between the Santa Catalina Mountains and Huachuca Mountains (Fig. 2.1; his Fig. 11), with intermediate colorations in the Santa Catalina, Rincon, Santa Rita, Chiricahua, and Huachuca ranges. This region includes the transition from temperate to subtropical coniferous forests (Wade et al. 2003), which may act as a mechanism for speciation or biased dispersal within a species. Indeed, the climatic extremes in this region (minimum precipitation of driest month and hottest temperature warmest month) were shown to correspond with genetic

Figure 2.1. (A) Sampling map in Arizona, USA. Gray areas correspond to montane forest. Inset shows locations of parental populations (solid circles) used in this study. (B) STRUCTURE results for the 75% coverage matrix (75% CM) dataset. Each bar represents the probability of population assignment to northern (gray) or southern (white) lineages. All individuals sorted with population assignment values greater than 0.9. (C) Proportion of shared (white), private (gray), and fixed (black) polymorphisms in Arizona between the two lineages [based on genetic structure in part (B)]. (D) Proportion of SNPs fixed between parental populations at each locality in Arizona. Black and white indicate proportion of fixed SNPs for the northern and southern alleles, respectively. Gray indicates the population is polymorphic with both northern and southern alleles.



differentiation in the White-breasted Nuthatch (*Sitta carolinensis*; Manthey & Moyle 2015). The northern and southern forms of *C. americana* also differ in migratory behavior; the southern forms appear to be resident and the northern forms are partially migratory (Phillips et al. 1964). Wintering and migrant specimens of the northern form have been collected in southern Arizona, within the distribution of the southern form (Phillips et al. 1964).

Previous genetic studies (Manthey et al. 2011a,b; Manthey et al. 2015a) of *C. americana* have included sparse sampling from Arizona, only including individuals from the Kaibab National Forest surrounding the Grand Canyon (northern lineage) and the Chiricahua Mountains (southern lineage; Fig. 2.1). The lack of genetic sampling across the putative contact zone, a transition in color pattern across multiple mountain ranges, and chromosomal variation in patterns of genetic differentiation in birds suggest that a genomic-level investigation of *C. americana* in Arizona is needed. Additionally, a genomic-level investigation in a Madrean Archipelago sky island taxon will build upon previous studies to infer comparative population genetic patterns and processes [e.g., birds (Manthey & Moyle 2015), mammals (Fitak et al. 2013, Atwood et al. 2011), invertebrates (Masta 2000, Smith & Farrell 2005), trees (Potter et al. 2013)].

Here, using large numbers of SNPs across the genome, we investigate a putative contact zone between two lineages of *C. americana* and address the following questions and hypotheses:

1. Where is the genetic break between lineages?

   H_0: The genetic break matches morphology, and is clinal across several mountain ranges in Arizona.

H_A: The genetic break is narrow and lineages are allopatric.

2. Within the zone of plumage transition, do the two lineages show a positive relationship between genetic differentiation and chromosome size, as was shown in a broader phylogeographic study?

   H_0: The positive relationship between chromosome size and genetic differentiation remains intact in the contact zone.

   H_A: In the contact zone, this relationship breaks down.

Methods

Sampling, sequencing, and SNP dataset creation

Fresh tissue samples of 45 *C. americana* were obtained across the putative contact zone in the sky islands of Arizona (Fig. 2.1, Table 2.1). The sky islands are islands of montane forest habitat separated by lowland scrub, grassland, and desert habitat. Four samples each of the northern and southern lineages were acquired from Utah, USA (northern) and Jalisco, Mexico (southern) to use as pure parental samples (Table 2.1). A single *C. familiaris* sample was used as an outgroup (Table 2.1). We used a QIAGEN DNeasy blood and tissue kit to extract genomic DNA for each sample.

We performed a modified RAD-seq (Miller et al. 2007) protocol to obtain thousands of genetic loci spread across the genome. We used the restriction enzyme NdeI to digest DNA samples as the first step to create a reduced representation genomic library. Custom adapters with attached barcodes for multiplexing were ligated to digested DNA samples followed by pooling and purification of all samples using an Agencourt AMPure magnetic bead cleanup. To further reduce the genomic libraries, we size-selected fragments in the

Table 2.1. List of samples and their associated RAD-seq coverage. For each individual, the locality (as in Fig. 2.1), collection number [either from University of Washington Burke Museum (UWBM) or University of Kansas Natural History Museum (all others)], percent completeness in the 50% and 75% coverage matrices (C50 and C75, respectively), number of sequencing reads (# Reads), number of RAD-tags, and coverage of sequencing reads for the 50% coverage matrix dataset (RAD Cov.; mean and standard deviation).

Locality	Coll. Number	C50	C75	# Reads	RAD-tags	RAD Cov.
Chiricahua	MBR8595	89.3	95.7	2,193,670	48,677	48.5 (51.1)
Chiricahua	MBR8596	86.3	93.2	1,874,272	47,317	40.7 (41.4)
Chiricahua	JDM057	91.8	96.5	3,102,638	50,786	62.8 (71.8)
Chiricahua	MBR8601	88.2	95.9	1,614,960	34,682	40.6 (44.5)
Chiricahua	KU31226	87.6	94.2	2,468,865	52,090	14.7 (11.5)
Huachuca	MBR8586	89.6	95.7	2,475,852	52,235	52.0 (53.1)
Huachuca	MBR8584	89.2	95.7	2,110,794	40,465	48.9 (51.7)
Huachuca	MBR8585	86.5	92.6	3,582,197	45,483	73.1 (84.6)
Huachuca	KU31227	91.8	97.3	2,344,278	38,860	56.4 (62.4)
Jalisco	UWBM110712	88.1	93.5	3,912,641	65,526	73.9 (84.5)
Jalisco	UWBM110615	92.7	97.7	1,656,810	42,467	39.0 (36.5)
Jalisco	UWBM110621	87.7	93.0	4,114,168	55,404	77.1 (86.2)
Jalisco	UWBM117091	91.3	97.5	2,883,402	39,628	69.5 (78.2)
Santa Ritas	JDM063	88.3	95.2	1,206,553	34,904	30.6 (29.7)
Santa Ritas	JDM064	89.4	95.6	1,831,829	42,147	41.4 (41.7)
Santa Ritas	MBR8609	89.6	95.9	2,997,954	48,142	64.7 (68.6)
Santa Ritas	JDM065	91.5	97.1	1,738,408	42,413	42.1 (40.3)
Kaibab North Rim	JDM002	86.2	93.2	2,037,171	50,198	42.8 (43.4)
Kaibab North Rim	JDM009	91.3	97.4	2,179,302	46,187	50.9 (52.3)
Kaibab North Rim	JDM010	91.1	96.1	3,039,178	52,035	67.6 (64.6)
Kaibab South Rim	JDM061	91.7	97.4	2,467,519	39,978	59.7 (64.5)
Kaibab South Rim	JDM062	90.7	95.7	1,456,471	40,073	35.3 (32.0)
Mogollon Rim	JDM012	90.5	96.1	3,599,547	48,270	79.8 (86.4)
Mogollon Rim	JDM013	92.7	97.9	2,947,298	40,714	71.2 (77.1)
Mogollon Rim	JDM014	89.1	95.6	2,897,455	56,549	55.4 (58.5)
Mogollon Rim	JDM015	79.9	90.5	863,742	26,774	26.4 (24.7)
Mogollon Rim	JDM016	45.6	54.2	155,422	14,710	7.6 (4.3)
Mogollon Rim	JDM017	70.1	82.1	414,123	22,121	14.7 (11.5)
Mogollon Rim	JDM018	63.1	74.9	291,260	20,417	10.6 (7.4)
Mogollon Rim	JDM043	69.1	79.7	879,162	44,515	19.5 (16.3)
Mogollon Rim	JDM045	51.6	57.9	212,617	18,040	7.0 (3.9)
Pinal	MBR8580	90.7	96.6	2,677,802	43,901	59.7 (64.5)
Pinal	MBR8579	91.7	96.9	3,098,164	42,654	72.2 (79.4)
Pinal	MBR8604	89.2	95.3	2,505,339	59,776	50.7 (50.2)
Pinaleño	JDM033	79.6	89.9	686,733	33,435	17.6 (14.6)
Pinaleño	JDM034	91.6	97.0	1,366,277	46,048	32.5 (28.8)
Pinaleño	JDM035	54.1	63.9	650,917	27,576	19.6 (17.4)

Table 2.1. Continued.

Locality	Coll. Number	C50	C75	# Reads	RAD-tags	RAD Cov.
Pinaleño	JDM036	89.1	95.8	978,434	31,435	26.7 (27.0)
Pinaleño	JDM037	56.6	67.4	250,007	18,556	10.1 (6.8)
Prescott	MBR8578	89.1	94.7	3,920,836	58,548	78.3 (87.1)
Prescott	JDM051	94.0	98.7	3,718,143	40,543	90.8 (102.8)
Prescott	JDM055	90.4	97.0	1,346,722	34,398	33.8 (35.4)
Prescott	MBR8602	93.1	98.0	1,902,484	53,109	42.6 (38.8)
Prescott	MBR8603	89.0	94.8	4,669,850	53,519	96.8 (104.4)
Santa Catalina	JDM027	41.1	47.4	122,107	14,648	5.6 (2.5)
Santa Catalina	JDM028	91.5	96.8	2,755,533	45,462	64.6 (65.0)
Santa Catalina	JDM029	70.0	82.7	524,465	22,115	19.7 (16.7)
Santa Catalina	JDM030	93.2	97.8	1,735,413	39,986	42.1 (42.7)
Santa Catalina	JDM032	53.5	64.6	349,605	26,160	10.1 (6.6)
Utah	UWBM114838	68.5	81.5	460,367	36,076	9.8 (6.4)
Utah	UWBM111063	92.6	96.8	2,916,596	42,087	67.5 (72.0)
Utah	UWBM114811	86.8	94.7	644,668	34,290	16.4 (12.6)
Utah	UWBM111062	79.9	92.2	687,820	28,984	18.9 (17.3)
Outgroup - England	KU6761	79.1	84.3	2,085,339	49,129	37.4 (41.3)

range of 500 to 600 bp using a Pippin Prep electrophoresis cassette (Sage Science). Libraries were cleaned again with a bead cleanup, followed by a brief polymerase chain reaction (PCR) of each library in duplicate. PCR conditions were as follows: an initial denaturation period of 98°C for 30 s, 14 cycles of 98°C for 10 s, 64°C for 30 s, and 72°C for 20 s, and a final extension of 72°C for 7 min. PCR reactions were cleaned a final time using a bead cleanup and were tested for DNA quality and quantity using an Agilent Tapestation and quantitative PCR at the University of Kansas Genome Sequencing Core Facility. All libraries were pooled and sequenced on a partial lane (~56%) of an Illumina HiSeq2500 100 bp single-end sequencing run.

We used the STACKS (Catchen et al. 2011) pipeline to assemble loci de novo from fastQ sequence files obtained from the Illumina sequencing run. Sequence reads of poor quality were removed using a quality threshold as an average phred score of ten in sliding windows of 15 bp using the *process_RADtags* python script included in STACKS. Sequences lacking the restriction site were removed. The *ustacks*, *cstacks*, and *sstacks* modules of STACKS were used to assemble catalogs of loci for each individual as well as a catalog of overlapping loci among all individuals (mismatches allowed between individuals = 5). Lastly, the populations module of STACKS was used to create SNP datasets with the following restrictions: minimum stack depth of five, minimum minor allele frequency of 0.05, and observed heterozygosity of 0.5 or less to reduce inclusion of paralogous loci. With these conditions, two SNP datasets were created with different levels of coverage: 50%, and 75% coverage matrices. Here, coverage was determined for three groups: northern parentals, southern parentals, and individuals near the putative contact zone. For loci to be included, the coverage threshold needed to be met in each of these groups. For example, in

the 75% coverage dataset, three northern parentals, three southern parentals, and 34 individuals in the contact zone needed to have data for a particular SNP in order for that SNP to be included.

SNP dataset analyses

We used the BLAST+ utility (Camacho et al. 2009) to match loci in our study to chromosomes in the Zebra Finch (*Taeniopygia guttata*). This is possible due to high levels of synteny in songbirds, but is limited in identifying chromosomal location due to high rates of intrachromosomal recombination in songbirds (Kawakami et al. 2014). Loci were determined a match to the Zebra Finch genome with 70% sequence identity and a maximum e-value of 0.01. We initially tested multiple e-values (0.01, 0.001, 0.0001), but the number of loci matching each chromosome was strongly related between datasets ($R^2 > 0.99$); we therefore proceeded using the e-value less than 0.01 results.

For each chromosome, all matched loci were rerun in STACKS to estimate F_{ST} between northern and southern lineages (based on STRUCTURE results; See RESULTS and Fig. 2.1). Linear regression of chromosome size and F_{ST} was performed in R (R Development Core Team 2011), using chromosomes with a minimum of ten loci.

All analyses were performed for both SNP datasets. To identify any potential admixture or hybridization, we used the program STRUCTURE (Pritchard et al. 2000) with an a priori number of genetic clusters equaling two (K = 2). Initially, we inferred lambda by estimating the K = 1 likelihood and allowing lambda to converge. We used this inferred value of lambda in five subsequent replicates with K = 2, the same F_{ST} assumed in the two groups, and using the admixture model. The burn-in was set as a period of 50,000 steps

followed by 100,000 MCMC iterations. All replicates converged on the same general parameter estimates, and the mean of all replicates was used for reporting assignment of individuals to genetic clusters. Within each of the lineages, STRUCTURE was run for multiple values of K; no additional genetic clustering was identified.

In addition to STRUCTURE analyses, we examined population genetic structure using Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010), implemented in the R package adegenet (Jombart & Ahmed 2011). DAPC transforms the SNP data using principal components analysis and then uses discriminant analysis to identify genetic clusters. Because two lineages were again identified (Appendix III), DAPC was run with the northern and southern lineages separately; here, the most likely number of genetic clusters within each lineage was one.

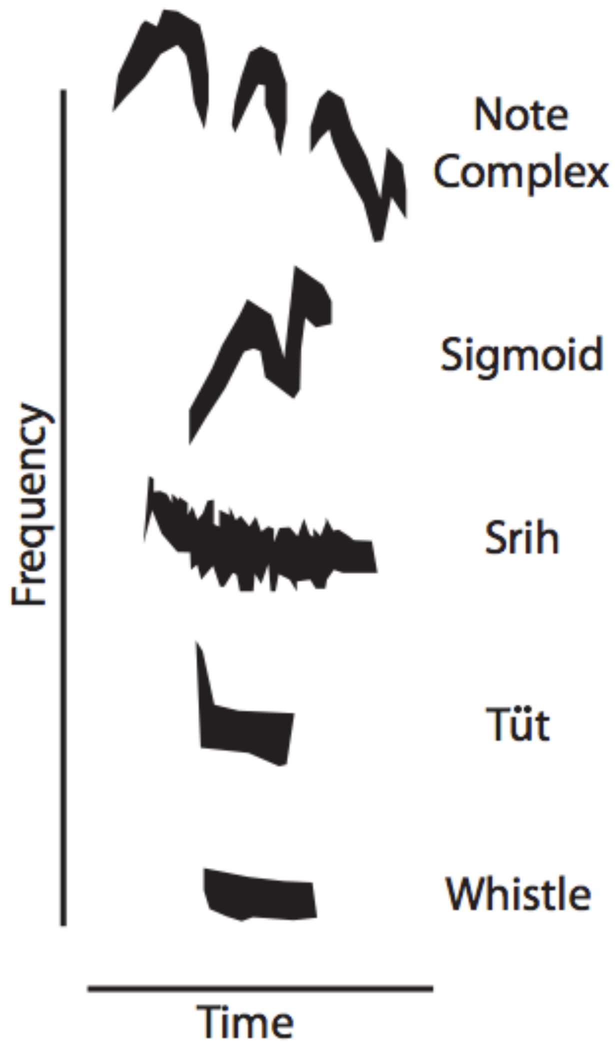
Song and environmental characteristics

During June 2013, song recordings were collected from multiple localities in the contact zone (Table 2.2), although not enough sampling to perform an extensive quantitative analysis. Here, our goal was to find potentially identifying characteristics between the song of northern and southern lineages. We used the terminology of syllables used by Baptista and Krebs (2000), who studied *C. americana* song dialects in California, USA. The song repertoire of *C. americana* generally consists of five types of syllables (as in Baptista and Krebs 2000; Fig. 2.2): (1) srih—rapidly frequency modulated note, (2) note complex—group of notes forming an obvious cohesive unit, (3) sigmoid—note shaped like a sideways letter "S" on the spectrogram, (4) tüt—note that rapidly drops in frequency and then remains sustained, and (5) whistle—a frequency sustained note. We used the Raven

Table 2.2. Song repertoires of *C. americana* near the contact zone. Based on limited sample sizes, songs did not noticeably differ in length or frequency range between lineages and varied in the number of syllables from three to five. All syllable types are as in Figure 2.2. In the Santa Rita Mountains, birds appear to have a unique syllable, where the end of a sigmoid is rapidly frequency modulated (similar to a srih syllable). Recording number refers to Macaulay Library reference number.

Locality	#	1	2	3	4	5
Prescott	202873	Srih	Long Sigmoid	Note Complex	-	-
Prescott	202882	Srih	Long Sigmoid	Note Complex	-	-
Prescott	202884	Srih	Sigmoid	Whistle	Note Complex	-
Prescott	202888	Srih	Whistle	Whistle	Note Complex	-
Santa Catalina	203222	Srih	Sigmoid	Sigmoid	Whistle	-
Santa Catalina	203223	Srih	Sigmoid	Sigmoid	Whistle	-
Santa Rita	203232	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	Whistle
Santa Rita	203233	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	Whistle
Santa Rita	203234	Srih	Srih-Sigmoid	Sigmoid	Sigmoid	-
Huachuca	203253	Srih	Sigmoid	Tüt	Tüt	Tüt
Pinaleño	203292	Srih	Whistle	Long Sigmoid	Whistle	-
Pinaleño	203298	Srih	Whistle	Long Sigmoid	Whistle	-
Mog. Rim (East)	203628	Srih	Sigmoid	Whistle	Note Complex	-

Figure 2.2. Definition of syllables used in song analysis. Song vocabulary follows Baptista and Krebs 2000. Frequency of each note is not on the same scale (i.e., the Sigmoid is not at an innately higher frequency than a Srih).



Pro v1.4 (Cornell Lab of Ornithology) software to audibly and visually inspect songs. Measurements were taken of the following song attributes: total length, frequency range, number of syllables, and types of syllables. All songs are deposited in the Macaulay Library (Table 2.2).

Because another species (*Sitta carolinensis*) was shown to have genetic structure based on temperature and precipitation extremes in the Madrean Archipelago sky islands of Arizona, USA (Manthey & Moyle 2015), we obtained environmental data (minimum precipitation of driest month and maximum temperature of warmest month) from worldclim.org (Hijmans et al. 2005) to test this possibility in *C. americana*. We obtained all specimen records of *C. americana* around the putative contact zone (Arizona, USA, Sonora, Mexico, and Chihuahua, Mexico) from VertNet, an online repository of museum-vouchered specimen data. Specimens were restricted to breeding months (late May-early August) to prevent inclusion of migrants. To assess if northern and southern localities (based on genetic results in Fig. 2.1) significantly differed in environmental characteristics, we performed a multivariate analysis of variance (MANOVA) in R (R Development Core Team 2012).

Results

Number of reads and RAD-tags per individual was highly variable (Table 2.1). The 50% and 75% coverage matrices had 15,531 and 7,995 loci, with ~2.8 SNPs per locus on average (Table 2.3). STRUCTURE analyses, for both the 50% and 75% coverage matrices, identified a strong and well-defined split between northern and southern populations (Fig. 2.1B). All individuals sorted to their respective lineages with greater than 90% probability

Table 2.3. Statistics for the 50% and 75% coverage matrices (CM). Number of loci, single nucleotide polymorphisms (SNPs), SNPs per locus (PL; mean and standard deviation), SNPs per locus in the contact zone (PL CZ), and percent missing data in each dataset (Miss.).

Dataset	Loci	SNPs	SNPs PL	SNPs PL CZ	Miss.
50% CM	15,531	44,400	2.86 (1.59)	2.65 (1.50)	17.11%
75% CM	7,995	22,700	2.84 (1.53)	2.62 (1.44)	10.08%

assignment (Fig. 2.1B), indicating a lack of hybridization between lineages. However, if hybridization rarely occurs, our limited sampling may not be sufficient to identify hybrids. Within each lineage, the most likely number of genetic clusters was one, suggesting a lack of fine-scale spatial structure. Similar to STRUCTURE analyses, DAPC identified the same individuals sorting to two genetic clusters and a lack of genetic structure within each of the lineages (Appendix III).

The parental populations (Utah and Mexico) had 2296 and 766 fixed differences for the 50% and 75% coverage matrices, respectively. Note that these numbers may be inflated due to small sample sizes. Most SNPs fixed between parental populations were similarly fixed in sampled localities of Arizona (Fig. 2.1D). In the contact zone specifically, most SNPs were private alleles (SNPs found in only one lineage), with more than 1.5% of SNPs fixed between lineages (Fig. 2.1C), even between populations as close as 70 km.

The number of loci per chromosome was highly related to chromosome size ($R^2 > 0.98$), suggesting our data were spread evenly across the genome (based on BLAST+ results). When F_{ST} was computed per chromosome across the putative contact zone between northern and southern lineages, genetic differentiation was positively related to chromosome size (Fig. 2.3A). Additionally, the pattern was similar between contact zone samples and parental samples (Fig. 2.3B), although the scale of F_{ST} was different between groups, likely due to different sample sizes.

Northern and southern localities near the contact zone differed significantly in environmental characteristics (Fig. 2.4), likely a reflection of the latitudinal difference between lineage distributions. The geographic break between lineages is at the split between temperate and subtropical coniferous forests (Wade et al. 2003), however, and

Figure 2.3. (A) Relationship of genetic differentiation (F_{ST}) between lineages in the contact zone and chromosome size for the 50% and 75% coverage SNP matrices (CM). (B) Comparison of chromosomal patterns of genetic differentiation between lineages in the contact zone and in parental populations. Arrows indicate the Z chromosome.

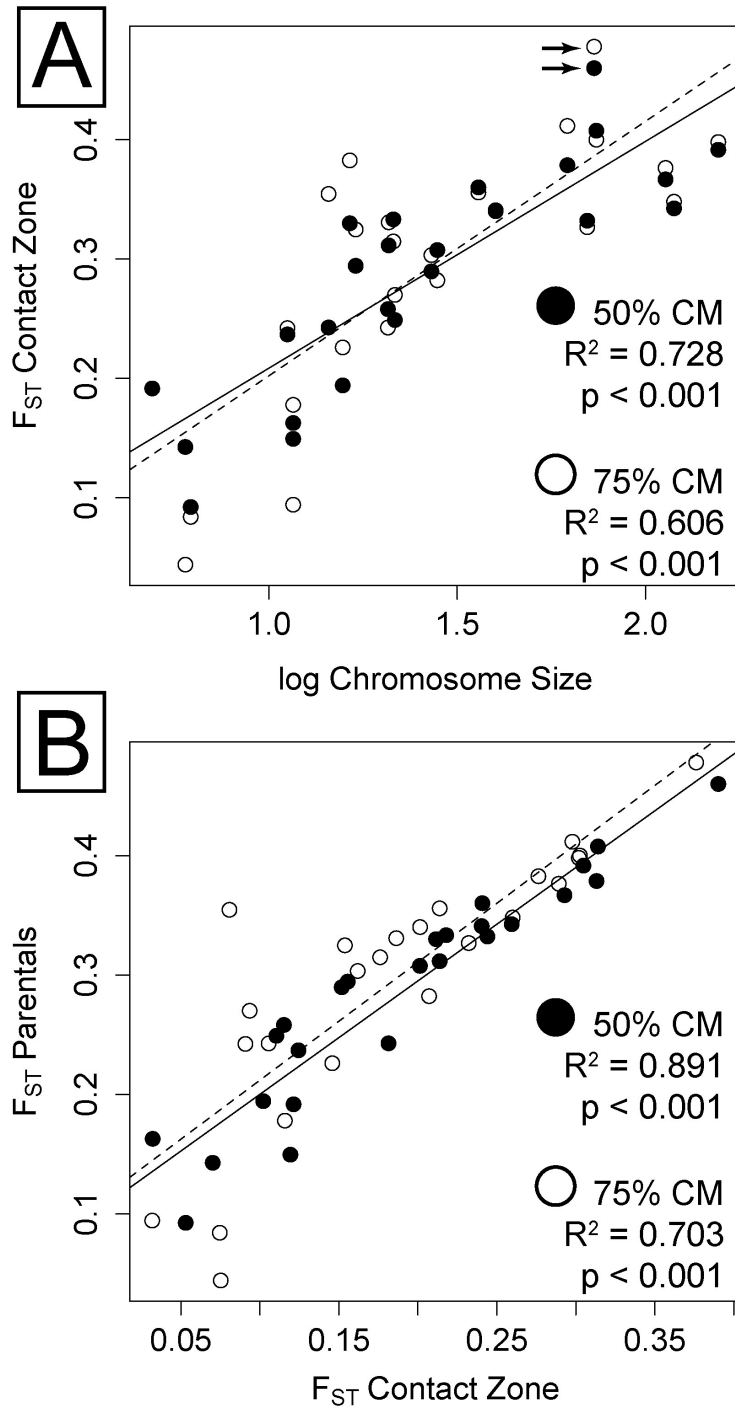
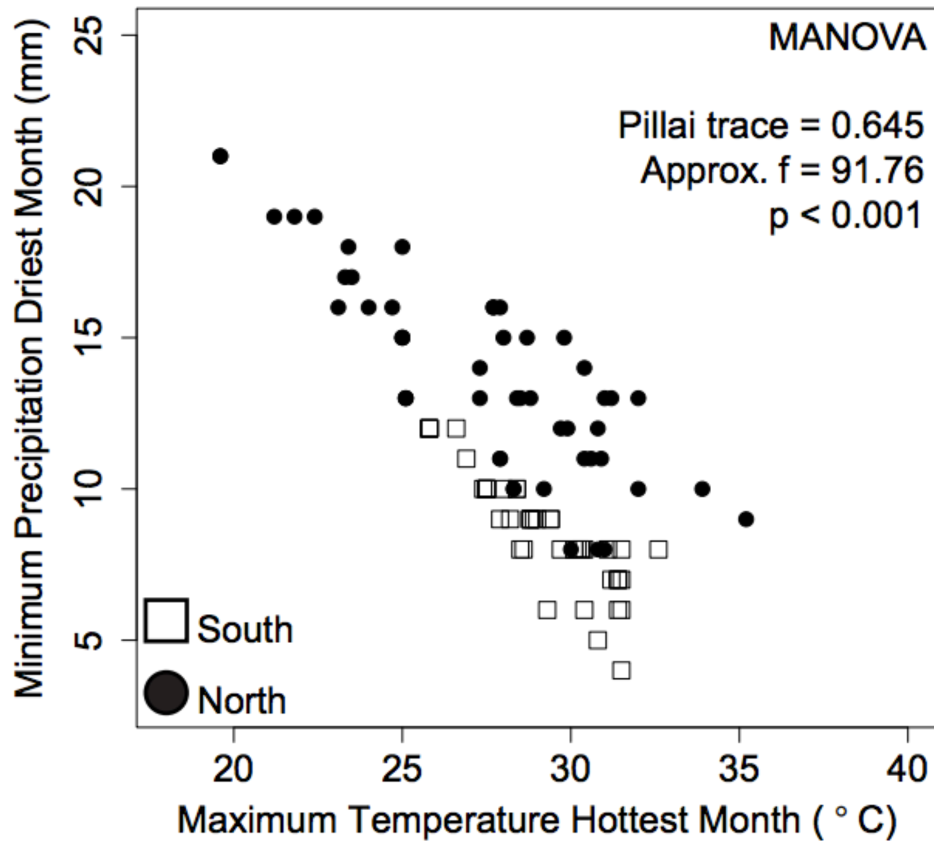


Figure 2.4. Scatter plot of environmental variables (minimum precipitation of driest month and maximum temperature of warmest month) for both lineages and results of a multivariate analysis of variance (MANOVA) between lineages of these two environmental variables.



may represent a real difference in climate regimes inhabited by each lineage. In this region, the northern lineage inhabits forests dominated by Ponderosa Pine (*Pinus ponderosa*); in contrast, the southern lineage populations are in the northernmost Madrean forest localities, where *Pinus* species are interspersed with Madrean oak species (Genus: *Quercus*). Further south (northern Mexico), the sky island habitat of *C. americana* becomes dominated by mixed oak communities.

No differences between lineages were apparent in song length, frequency range, or song structure (Table 2.2). All individuals' songs began with the Srih note, ranged between 1.12 and 1.53 seconds in duration, and had similar frequency ranges (3.0 – 8.2 kilohertz). Differences in song mostly existed among sampling localities; all but one locality (Santa Rita Mountains) with multiple individuals had the same beginning and ending syllables (Table 2.2). In the Santa Rita Mountains, the first four syllables were always the same, with the only difference being whether the song ended with a whistle. These results suggest that each locality has developed a specific local dialect, independent of genetic lineage, although more song recordings are needed to test this.

Discussion

High throughput sequencing to investigate genetic structure and gene flow

The use of high-throughput sequencing has been critical in identifying chromosomal levels of differentiation between species at contact zones (e.g., Ellegren et al. 2012, Taylor et al. 2014), usually identifying higher relative effects (e.g., divergence, lack of introgression, selection) of sex chromosomes compared to autosomes. Studies with thousands of SNPs have been used to provide genomic scale resolution in contact zone studies compared to

traditional studies using mtDNA or microsatellites alone. Large panels of SNPs provide an increasingly robust method to diagnose hybridization or lack thereof. Here, using more than 40,000 SNPs we identified a lack of hybridization and a sharp geographic break between two genetically distinct lineages (Fig. 2.1). Interestingly, this pattern is in contrast to Marshall's (1956) evidence for a more gradual cline across several mountain ranges based on morphology. This may be caused by multiple factors; we may have not sampled the genetic variation that contributes to morphological variation (e.g., *Sporophila* seedeaters, Campagna et al. 2015). Alternatively, morphological differences could be due to an environment by genotype interaction; here, a cline in environmental conditions (i.e., latitudinal cline) may contribute to a cline in morphological characters.

The observed genetic pattern is not unique to *C. americana*, however. The sky islands between the Rocky Mountains and Sierra Madre Occidental span the transition between temperate and subtropical coniferous forests (Wade et al. 2003). Here, at the southern range limits of the Rocky Mountains, higher elevations are dominated by Ponderosa Pine. This transitions to the northern range limits of the Sierra Madre Occidental, where higher elevations are dominated by mixed pine (Genus *Pinus*) and Madrean oak species (Genus *Quercus*). As such, many montane songbird species (> 50) have their northern or southern range limits in this region. The specific geographic break identified here in *C. americana* is evident in three other bird species, including the southern range limit of the Mountain Chickadee (*Poecile gambeli*) and the northern range limit of two species: Mexican Chickadee (*Poecile sclateri*), and the montane range of the Eastern Bluebird (*Sialis sialis*). Between *C. americana* lineages, divergence likely occurred during the Pleistocene (Manthey et al. 2011b); however, during the Pleistocene glacial cycles there

is little evidence for increased distributional overlap among lineages during glacial maxima based on ecological niche modeling (Manthey et al. 2014), suggesting mechanisms other than climate cycles have restricted gene flow among lineages.

Chromosomal patterns and divergence hypotheses

In the contact zone, we found a positive relationship between genetic differentiation and chromosome size (Fig. 2.3A). The Z chromosome had the highest F_{ST} between lineages (arrows in Fig. 2.3A) and one of the narrower cline widths (Fig. 2.3C), although it did not appear to be an outlier based on relationships with chromosome size. Many studies of hybridizing bird species have found elevated genetic differentiation on the Z chromosome relative to autosomes (e.g., Carling et al. 2010; Storchová et al. 2010) when using intron sequence data and small numbers of loci. Indeed, in this study system (*C. americana*), elevated Z chromosome differentiation has been shown in analyses of introns alone (Manthey & Spellman 2014). This discrepancy suggests that using only intron sequences biases inference of chromosome wide patterns of differentiation. However, a genome-level analysis of *Ficedula* flycatchers still estimated the Z chromosome to be an order of magnitude more differentiated on average than autosomes (Ellegren et al. 2012), suggesting small-scale genetic sequencing may be biased in only some cases.

An alternative hypothesis to explain this pattern may be that only recent speciation events exhibit elevated sex chromosome differentiation. Relative to *Ficedula* flycatchers (Ellegren et al. 2012) and *Passerina* buntings (Carling & Brumfield 2008), the two *Certhia* lineages are more divergent based on mitochondrial and nuclear sequence data (Manthey et al. 2011a,b, 2015a). Under a time-since-speciation scenario, biased genomic

introgression (i.e., less on sex chromosomes) may have ceased much earlier in *Certhia* due to lack of interbreeding between lineages.

A final hypothesis is that sex chromosome evolution is similar to autosomal evolution when genetic mechanisms involved with postzygotic isolation between taxa do not occur [e.g., large-Z effect or Haldane's rule (Coyne & Orr 1989)]. In this scenario, a lack of hybridization prevents increased differentiation on the Z to manifest. Here, prezygotic isolation would have needed to occur throughout the divergence process. Behavioral differences between populations could cause divergence on chromosomes via genetic drift and non-random mating. This possibility may also be the isolation mechanism between *Poecile gambeli* and *P. sclateri*, sister species (Harris et al. 2013) that do not hybridize (McCarthy 2006), occupy the same habitat, and have the identical genetic break shown here in *C. americana*.

Some behavioral characteristics potentially minimizing interbreeding between lineages include non-random gene flow, non-random mating, and natal philopatry. Non-random gene flow may lead to population differentiation with or without the presence of fitness variation (Edelaar & Bolnick 2012). Variation in dispersal ability, individual behavior, and habitat preference genes alone can lead to biased gene flow and subsequent genetic differentiation through time (Edelaar & Bolnick 2012). Short natal dispersal distance (Davis 1978) and natal philopatry (Cramp & Perrins 1993) in creepers could increase the effects of non-random gene flow. Because northern and southern forms of *C. americana* at least overlap seasonally (Phillips et al. 1964) and there is still a distinct genetic break between lineages (Fig. 2.1) suggests that behavioral characteristics may minimize gene flow among lineages.

Multiple lines of evidence suggest non-random mating, with or without other sources of non-random gene flow, may be causing the patterns observed in *C. americana*. First, local dialect formation appears to be common, as has been shown in California, USA (Baptista & Krebs 2006) and now in Arizona, USA (Table 2.2). Song variation has been suggested as a cue for finding locally-adapted mates and thus promoting co-divergence of song and genetics (Slabbekoorn & Smith 2002), but has also been shown to differ widely among populations without speciation events [e.g., White-crowned Sparrows (*Zonotrichia leucophrys*), Marler & Tamura 1962]. Additionally, each lineage is associated with significantly different climates (Fig. 2.4)—and associated ecosystems—a potential context for non-random dispersal and gene flow.

Lastly, the genomic patterns observed here (Figs. 2.1 & 2.3)—namely chromosome-differentiation relationships—are similar between local (i.e., in the contact zone) and highly disjunct (e.g., eastern USA and southern Mexico; Manthey et al. 2015a) populations. The same patterns between geographically and ecologically disparate populations suggest selection is not the driving force of differentiation, as it would be unlikely for such widespread populations to be undergoing the same selective pressures (Manthey et al. 2015a). Rather, genetic drift—in conjunction with non-random mating and possibly non-random gene flow—is causing the differential patterns of genetic differentiation observed between lineages (Fig. 2.3). Genetic drift alone could cause divergence patterns correlated to chromosome size because of differential average recombination; recombination rates scale with chromosome size due to meiotic crossover requirements (Lynch 2007).

CHAPTER 3*

Genetic differentiation among birds (Aves: Passeriformes) explains a portion of the beta diversity among their gut microbial communities

* Manthey, J. D., & R. G. Moyle. Genetic differentiation among birds (Aves: Passeriformes) explains a portion of the beta diversity among their gut microbial communities.

Abstract

All animals have evolved in the presence of microorganisms, with symbioses ranging from negative influences in disease to beneficial roles in host nutrient synthesis. The mutually obligate relationship of the host and associated microbiota suggests evolution in one would have an impact on the other, yet this remains inadequately explored in nature. Here, we used gut microbiota samples from populations of two songbird species—*Sitta carolinensis* and *Certhia americana*—across multiple geographic localities in the Madrean Archipelago sky islands of Arizona, USA to explore coevolution between hosts and their gut microbial communities. We explored relationships between host genetic diversity and associated microbial community alpha diversity, as well as between host genetic differentiation, distance between sampling localities, and microbial community beta diversity. We found no relationship between host genetic diversity and microbial alpha diversity. We identified no appreciable effect of geography on microbiota community assembly, but found a significant proportion of the microbial community beta diversity to be explained by genetic differentiation among hosts. Additionally, we found significant differences in predicted proportions of metabolic functional units in the microbiomes of *C. americana* and *S. carolinensis*. These results suggest a lack of horizontal transfer of bacterial OTUs among localities, as well as a significant proportion of the microbiota vertically inherited.

Introduction

All animals have evolved in the presence of microorganisms. Animals have several distinct communities of symbiotic microorganisms (Turnbaugh et al. 2007), many representing mutually obligate relationships (e.g., the gut microbial community, Xu & Gordon 2003). The role of the microbiome in animals ranges from negative influences in disease (Kingsley & Baumler 2000) to mutually beneficial roles in development (Yin et al. 2009) or nutrient synthesis (Mai & Baer 2008). Despite some negative interactions, the mutually obligate relationship between hosts and their microbial communities suggests evolution in one will have an impact on the other and may work in both directions.

In natural systems, coevolutionary patterns of microbial communities and their hosts have been investigated across geography (e.g., Hird et al. 2014) and phylogeny (e.g., Banks et al. 2009, Hird 2013). Many of the relationships between microbial communities and their hosts have been described in model organisms (e.g., livestock species such as chicken; Lan et al. 2005), but studies investigating assembly of these microbial communities in natural systems have yielded ambiguous results. For example, some studies find some geographic structuring of microbial communities within species (Hird et al. 2014), whereas others find some phylogenetic signal within microbial communities (Banks et al. 2009). To better investigate patterns of microbial community assembly in natural systems, as well as investigate patterns of coevolution between microbial communities and their hosts, studies should include aspects of both geography and phylogeny. Sampling multiple species across many geographic localities would enable the identification of the relative effects of geography, phylogeny, ecology, and stochasticity on the formation of vertebrate microbial communities.

Two songbirds, the Brown Creeper (*Certhia americana*) and White-breasted Nuthatch (*Sitta carolinensis*), have similar ecological niches in the sky islands of the Madrean Archipelago in southern Arizona. Despite their shared affinities for preening invertebrates on tree bark, they have very different patterns of genetic structure in this region (Fig. 3.1). The Brown Creeper has a strong genetic break between lineages (dotted line in Fig. 3.1; Manthey et al. 2015b), with individuals on either side of the break genetically pure for their respectively lineages. Alternatively, the White-breasted Nuthatch exhibits a pattern of isolation by environment (Manthey & Moyle 2015) where individuals in more similar environmental conditions, regardless of geographic location, are more genetically similar. The observed dissimilar phylogeographic patterns across the same sampling regime between species provides a comparative framework to assess the relative contributions of geography and phylogeny to vertebrate microbial community assembly. Because of the similar ecological niche of the two species (invertebrate preeners), the effect of each species' ecological niche on microbial community assembly is minimized between species.

Here, we utilize restriction-site associated DNA sequence (RAD-seq; Miller et al. 2007) data of the host organisms in conjunction with targeted sequencing of the V4 variable region of the 16S ribosomal subunit of the microbiotic communities for each host individual. Using these two datasets, we assessed the following questions and hypotheses:

1) Is microbial community assembly determined by geography, phylogeny, or a combination?

    $H1_0$) No genetic or geographic structuring of microbial communities.

$H1_{A1}$) Differentiation among microbial communities is correlated with genetic differentiation among host birds.

$H1_{A2}$) Differentiation among microbial communities is correlated with geographic distance between sampling sites of host birds.

$H1_{A3}$) A combination of geographic and genetic factors best explain microbial community differentiation between individual hosts.

2) Is genetic diversity within host individuals correlated with microbial community alpha diversity?

$H2_0$) Genetic diversity in hosts is correlated with intestinal microbial alpha diversity.

$H2_A$) No relationship between host genetic diversity and intestinal microbial alpha diversity.

Methods

Bird genetic data and bioinformatics

Avian genetic data were retrieved from previous studies investigating landscape genomics of the White-breasted Nuthatch (Manthey & Moyle 2015) and contact zone genomics of the Brown Creeper (Manthey et al. 2015b). The two studies used identical methods in obtaining RAD-seq data for each taxon; therefore, the datasets could be utilized together. Sampling localities of both taxa were nearly identical (Fig. 3.1; Table 3.1). Additionally, both species forage on tree bark for invertebrates, reducing a possible contributing factor for variance in gut microbial communities.

Figure 3.1. Sampling localities of *Certhia americana* and *Sitta carolinensis* individuals and associated microbiota in the Madrean Archipelago of the Southwest USA (Arizona). Shaded gray regions indicate montane forest habitat.

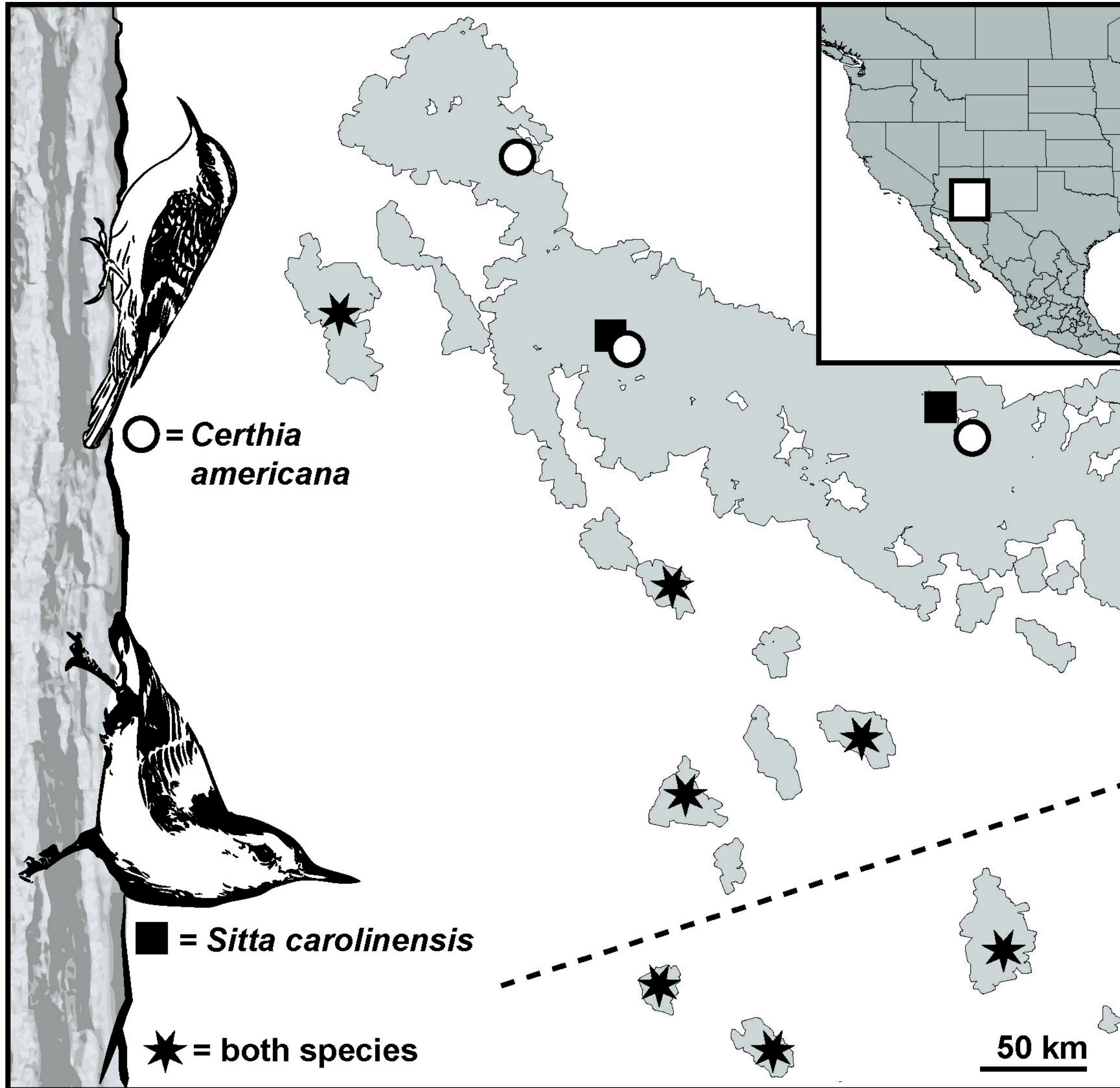


Table 3.1. Sampling localities (approximate lat./long.) and number of individuals for which RAD-seq. and gut microbiota sequencing was performed for each species.

Locality	Latitude	Longitude	# *Certhia*	# *Sitta*
Mogollon Rim West 1	34.97	-111.54	4	-
Mogollon Rim West 2	34.48	-111.42	-	2
Mogollon Rim West 3	34.45	-111.38	2	-
Prescott	34.41	-112.42	5	2
Mogollon Rim East 1	33.92	-109.27	-	3
Mogollon Rim East 2	33.81	-109.16	2	-
Pinal	33.29	-110.87	3	3
Pinaleño	32.69	-109.83	5	3
Santa Catalina	32.42	-110.73	5	2
Chiricahua	31.78	-109.30	4	3
Santa Rita	31.70	-110.88	2	3
Huachuca	31.43	-110.29	2	3

We used the STACKS (Catchen et al. 2011) pipeline to assemble a single nucleotide polymorphism (SNP) library using default settings in the *ustacks*, *cstacks*, and *sstacks* modules with the exception of mismatches allowed between stacks when building the catalogue of loci (in *cstacks*). We initially tested multiple values for the allowed number of mismatches (n = 1-10) between stacks to build the catalogue of loci (e.g., Mastretta-Yanes et al. 2015). We completed the rest of the STACKS pipeline and calculated population genetic statistics [nucleotide diversity (π) and genetic differentiation (F_{ST})] of *S. carolinensis* and each lineage of *C. americana*. Based on plateauing of these statistics when the number of mismatches (n) allowed ranged between six and eight, we included the middle value (i.e., n = 7) for downstream analysis. Initial sequence quality control was performed in previous studies (Manthey et al. 2015b, Manthey & Moyle 2015). The *populations* module of STACKS (Cathen et al. 2011) was used to create SNP libraries with the following settings: minimum stack depth of five, minimum minor allele frequency of 0.05, observed heterozygosity less than 0.5 (to reduce inclusion of paralogous loci), a minimum of 50% of individuals from each taxonomic group (*S. carolinensis* and each lineage of *C. americana*), and a minimum of one individual for each species at each locality. Variants of minimum stack depth (m = 5, 10, 15) did not change genetic differentiation among lineages. With this final SNP dataset, downstream analyses were performed on individual and locality levels. Genetic diversity [observed heterozygosity (H_0)] and genetic differentiation (F_{ST}) were calculated in STACKS. In the individual-based dataset, F_{ST} was highly related with (log) pairwise differences between individuals ($R^2 = 0.984$, $p << 0.001$); we therefore continued with the F_{ST} values for analyses.

Gut microbial communities and bioinformatics

Intestinal microbiota samples included full intestinal contents of the duodenum, jejunum, and ileum of each bird, extracted from each individual within two hours of collection and immediately frozen. MO BIO Laboratories (Carlsbad, California) PowerSoil DNA Isolation Kits were used to extract DNA from intestinal samples following manufacturer protocols. The V4 variable region of the 16S ribosomal RNA gene (~250 bp) was sequenced in all samples on a single Illumina MiSeq 151x151 paired-end sequencing run at the Institute for Genomics and Systems Biology Next Generation Sequencing Facility at Argonne National Laboratories. Recent studies have shown indistinguishable conclusions of studies based on the V4 region versus the entire gene sequence of 16S rDNA (Caporaso et al. 2012).

Paired-end sequences were quality-checked and assembled using QIIME (Caporaso et al. 2010), an open-source software package for analysis of microbial communities. 100% sequence identity in overlap between paired reads was required for retention. Sequences with unassigned bases (N calls) or lacking specified barcode sequences were removed. Reads were assigned to operational taxonomic units (OTUs) using the open-reference OTU picking process invoked in QIIME. First, sequences were aligned to the Greengenes database (McDonald et al. 2011) by utilizing uclust (Edgar 2010). All reads that were not clustered using the reference collection were clustered de novo. Final OTU picking was performed at two thresholds (95% and 97% sequence identity) with all subsequent analyses performed on both datasets. All OTUs belonging to Cyanobacteria were filtered, as they likely represent ingested plant material (Waite & Taylor 2014).

For alpha and beta diversity metrics, we created OTU datasets based on microbial sequences from individuals (i.e., each bird) and localities (i.e., all birds from one sampling

site). Microbial community similarity between samples was measured using UniFrac distance (Lozupone & Knight 2005), a phylogeny-based distance metric of community similarity. In addition to the standard UniFrac distance, we also measured the weighted UniFrac (Lozupone et al. 2007), which utilizes OTU abundances to calculate distance. Alpha diversity of microbial communities was measured using phylogeny-based (Faith et al. 2006) and count-based (e.g. OTU richness) metrics as implemented in the QIIME package. All estimates of alpha diversity were based on ten replicates of rarefication to 17,000 and 68,000 reads for individual and locality-based datasets, respectively; this ensured comparable coverage across all samples. Lastly, we used principal coordinate analyses, as implemented in QIIME, to visualize variation in microbial community distance (i.e., Unifrac distance) matrices.

Co-analysis of avian and microbial data

We used regression and correlation analyses in R (R Development Core Team 2013) to identify factors shaping microbial communities. First, we investigated the effects of genetic distance and geographic distance on microbial community differentiation using multiple regression. To look at the effect of genetic or geographic distance on beta diversity, while controlling for the other factor, we used partial Mantel tests. Finally, we explored the effect of host genetic diversity on microbial community alpha diversity (i.e., variation in alpha diversity explained by observed heterozygosity).

To investigate potential functional differences of gut microbial communities between avian taxa, we used PICRUSt (Langille et al. 2013). In PICRUSt, we used the 97% clustering OTU table from all individuals, filtered to only include known taxa from the

Greengenes database, normalized by 16S gene copy number for each taxon, predicted functions for the metagenome, and categorized the predictions into KEGG pathway categories. Using Statistical Analysis of Metagenomic Profiles (STAMP) v2.1.3 (Parks et al. 2014), we investigated differences in relative abundance of functional categories using analysis of variance (ANOVA), followed by a post-hoc Games-Howell test. We corrected for multiple testing using Benjamini-Hochberg FDR (Benjamini and Hochberg 1995).

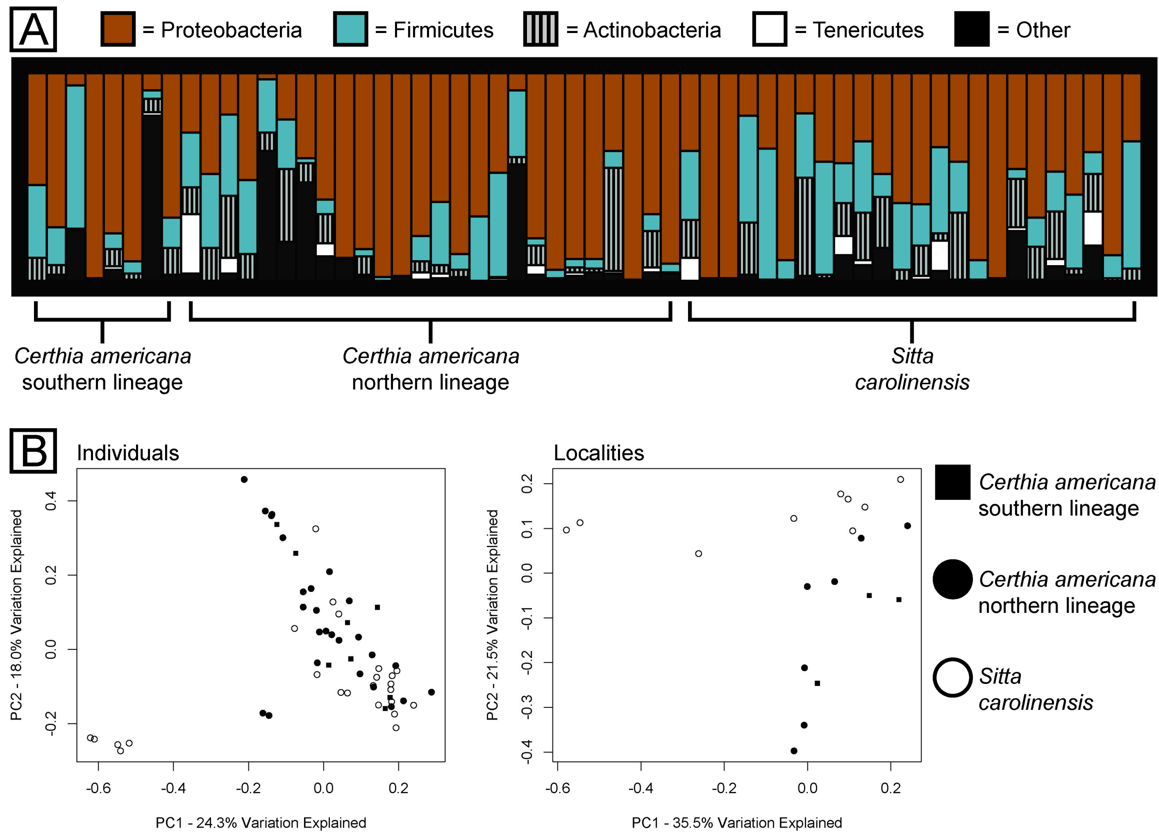
Results

Characteristics of datasets

Following quality control, our microbial sequence dataset contained 7,880,262 merged pairs of sequences, with a median length of 253 bp. The number of sequences per individual was highly variable (mean = 131,338 reads, sd = 85,606). Across all individuals, there was a mean of 207 (sd = 113) and 193 (sd = 104) OTUs identified for the 97% and 95% clustering datasets, respectively. Three phyla were represented in all individuals (Fig. 3.2A): Actinobacteria (10.3% mean abundance), Firmicutes (19.9%), and Proteobacteria (59.8%). The bacterial genera *Corynebacterium*, *Enterococcus*, *Rickettsiella*, and *Staphylococcus* were present in all individuals. In total, 8089 and 5506 OTUs were identified for the 97% and 95% clustering datasets, respectively.

The final RAD-seq dataset contained 696 loci and 4123 SNPs. In the individual-based dataset, observed heterozygosity ranged between 0.057% and 0.200%, while F_{ST} ranged between -0.027 and 0.979. The locality-based dataset had a range of observed heterozygosity between 0.099% and 0.185%, and an F_{ST} range between -0.028 and 0.943.

Figure 3.2. (A) Composition of microbial communities in *Sitta* and *Certhia* gut samples. All phyla represented by < 2% of overall microbiota samples were pooled into the "Other" category. (B) Principal coordinate analyses of the individual and locality based microbial communities (based on 97% clustering threshold).



Co-analysis of microbial and avian datasets

Principal coordinates analyses did not indicate clear taxonomic clustering using individual-based information, while we found clear separation between *Certhia* and *Sitta* samples using the locality-based data (Fig. 3.2B). Microbial community alpha diversity and host genetic diversity showed no clear relationship in the individual or locality-based datasets (Table 3.2).

Genetic differentiation among birds and beta diversity of their associated microbial communities was significantly correlated for most comparisons (Table 3.2, Fig. 3.3; based on multiple regression), with the exception of the locality-based datasets using unweighted Unifrac distances (Table 3.2). In the multiple regressions, geographic distance among samples had no impact on differentiation among microbial communities in most comparisons (Table 3.2, Fig. 3.3). When one factor was controlled for using partial Mantel tests, genetic differentiation again showed a significant relationship with differences in microbial communities (Table 3.3), but there was no apparent relationship between geographic distance and microbial beta diversity.

STAMP analyses revealed no differences between metagenomic profiles of the northern and southern lineages of *C. americana*. In contrast, six functional categories of metagenomic profiles differed significantly between *Certhia* and *Sitta* (Table 3.4). Of the six features that differed between taxa, metabolism-related categories were overrepresented (binomial test $p = 0.016$).

Table 3.2. Summary of regression analyses. At top, the relationship between alpha diversity [either phylogenetic diversity (PD) or OTU richness (# OTUs)] and host genetic diversity [observed heterozygosity (H_0)]. At bottom, results of multiple regression investigating the relationships between beta diversity (weighted or unweighted Unifrac distance), genetic differentiation (Gen.), and geographic distance (Geo.). Results are shown for the 97% (OTU97) and 95% (OTU95) clustering thresholds, and for datasets based on individuals (I) or localities (L).

Alpha Diversity	**H_0 R^2**	**H_0 p**	
OTU97 PD (L)	0.032	0.463	
OTU97 # OTUs (L)	0.091	0.209	
OTU95 PD (L)	0.033	0.460	
OTU95 # OTUs (L)	0.094	0.201	
OTU97 PD (I)	0.039	0.136	
OTU97 # OTUs (I)	0.040	0.130	
OTU95 PD (I)	0.035	0.158	
OTU95 # OTUs (I)	0.041	0.126	
Beta Diversity	**Gen. + Geo. Dist. R^2**	**Gen. p**	**Geo. p**
OTU97 Unifrac (Unweighted) (L)	0.010	0.249	0.531
OTU97 Unifrac (Weighted) (L)	0.066	***0.001***	0.366
OTU95 Unifrac (Unweighted) (L)	0.010	0.371	0.311
OTU95 Unifrac (Weighted) (L)	0.086	***<0.001***	0.844
OTU97 Unifrac (Unweighted) (I)	0.004	***0.010***	0.744
OTU97 Unifrac (Weighted) (I)	0.018	***<0.001***	0.382
OTU95 Unifrac (Unweighted) (I)	0.010	***0.014***	***0.002***
OTU95 Unifrac (Weighted) (I)	0.021	***<0.001***	0.738

Figure 3.3. Results of simple linear regressions for individual (A) and locality (B) based datasets (based on 97% clustering threshold). Also shown are the results of partial Mantel tests (second line of values over each plot), including the Mantel statistic (r) and associated p-value (but also see Table 3.3).

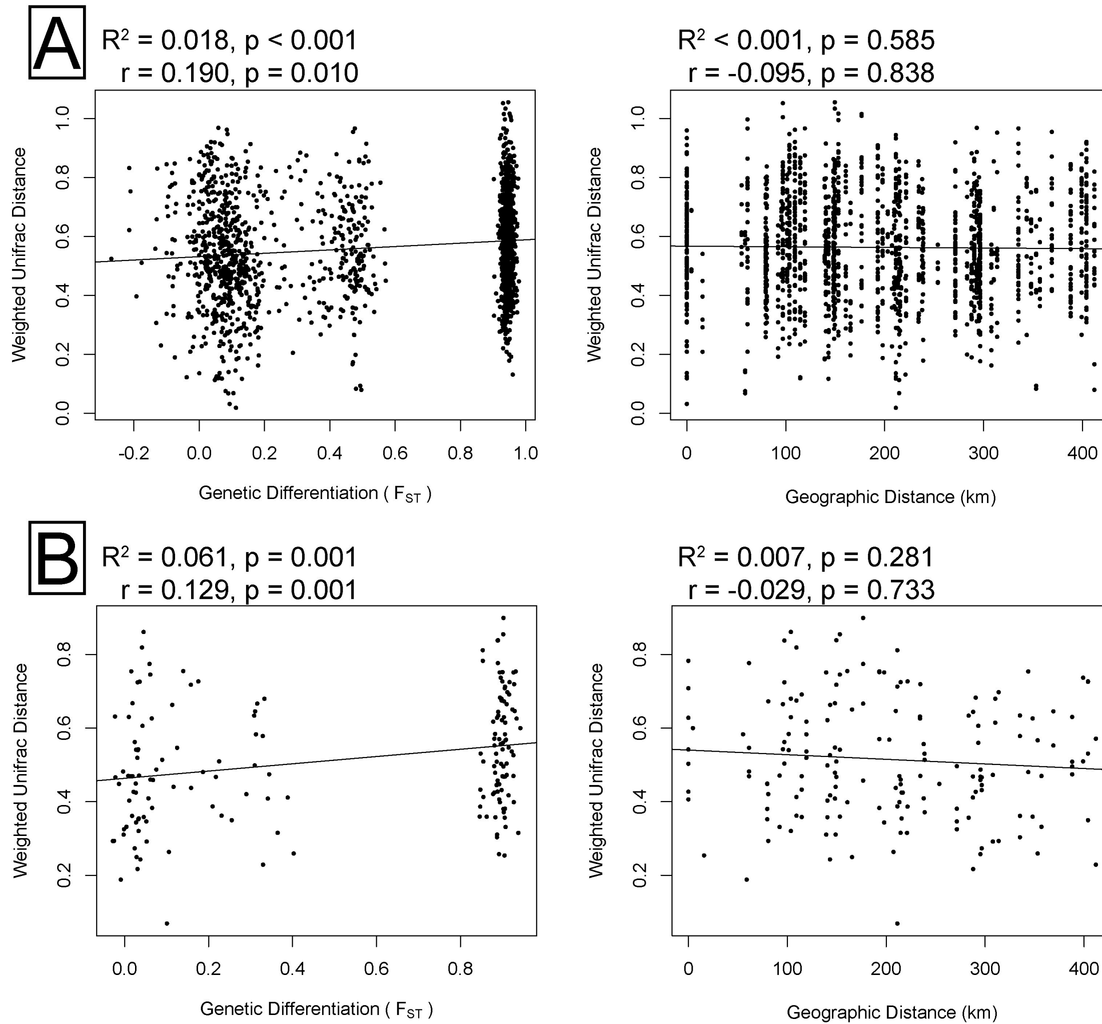


Table 3.3. Partial Mantel tests investigating the relationship between beta diversity (weighted or unweighted Unifrac distance) and genetic differentiation [while controlling for geographic distance; Gen. Dist. (Geo. Dist.)] or geographic distance (while controlling for genetic differentiation). Mantel test statistic (r) and associated significance level (p) based on 1000 permutations are reported. Results are shown for the 97% (OTU97) and 95% (OTU95) clustering thresholds, and for datasets based on individuals (I) or localities (L).

Beta Diversity	**Gen. Dist. (Geo. Dist.)**	**Geo. Dist. (Gen. Dist.)**
OTU97 Unifrac (Unweighted) (L)	r = 0.087, p = 0.141	r = 0.041, p = 0.330
OTU97 Unifrac (Weighted) (L)	***r = 0.190, p = 0.010***	r = -0.095, p = 0.838
OTU95 Unifrac (Unweighted) (L)	r = 0.053, p = 0.232	r = 0.066, p = 0.234
OTU95 Unifrac (Weighted) (L)	***r = 0.239, p = 0.007***	r = -0.052, p = 0.711
OTU97 Unifrac (Unweighted) (I)	***r = 0.060, p = 0.020***	r = 0.000, p = 0.464
OTU97 Unifrac (Weighted) (I)	***r = 0.129, p = 0.001***	r = -0.029, p = 0.733
OTU95 Unifrac (Unweighted) (I)	***r = 0.060, p = 0.024***	r = 0.068, p = 0.063
OTU95 Unifrac (Weighted) (I)	***r = 0.139, p = 0.001***	r = -0.001, p = 0.500

Table 3.4. Results of statistical analysis of metagenomic profiles (STAMP) analyses. No categories differed between *Certhia* lineages, while 6 of 36 differed between *Certhia* and *Sitta*. The predicted metagenome percentage (MG %) and multiple testing corrected significance values are reported.

Functional Category	*Certhia* MG %	*Sitta* MG %	Corrected p-value
Amino Acid Metabolism	0.25	0.14	<<0.001
Biosynthesis of other Secondary Metabolites	11.65	10.17	<<0.001
Glycan Biosynthesis and Metabolism	1.02	0.77	0.034
Metabolism of other Amino Acids	2.43	2.95	<<0.001
Signaling Molecules and Interaction	2.24	1.87	<<0.001
Xenobiotics Biodegradation and Metabolism	4.21	2.57	<<0.001

Discussion

Here, we characterized the gut microbiota of two songbirds of the Madrean Archipelago sky islands of the U.S. Southwest. We found that small (~0.5% to 9%, Table 3.2) but significant proportions of the microbial community beta diversity could be explained by genetic differentiation among avian hosts. Between taxonomic groups (i.e., *Certhia* and *Sitta*), significant differences existed in proportions of functional KEGG modules grouped by biochemical pathways related to metabolism.

Composition of gut microbiota

This study adds to the growing body of literature characterizing avian microbiota (for a review see Waite and Taylor 2014). Across multiple studies, five phyla have been shown to persist at high levels in the avian gut: Proteobacteria, Firmicutes, Actinobacteria, Tenericutes, and Bacteroidetes (Waite and Taylor 2014). Here, we found a similar pattern, with a high prevalence of all the aforementioned phyla except Bacteroidetes (Fig. 3.2A). Previous work has shown that Bacteroidetes comprises a large proportion of mammals' (Nelson et al. 2013a) and non-Passeriformes bird species' (e.g., chicken, penguins, turkey, ratites; Waite and Taylor 2014) microbiota. This study, and a recent study across multiple Passeriformes species (Hird et al. 2014), identified a smaller role for Bacteroidetes in the songbird microbiota. Interestingly, four bacterial genera were ubiquitous throughout all samples, including *Corynebacterium*, *Enterococcus*, *Rickettsiella*, and *Staphylococcus*. This prevalence may be because of the pervasive nature of these bacterial genera in nature and animals. For example, *Enterococcus* sp. have been found in a majority of animal and environmental microbiota surveys (Kühn et al. 2003). In contrast to the ubiquitous

bacteria, a large proportion (~48% based on 97% clustering threshold) of bacterial OTUs were private to a single individual. In both mammals (Ley et al. 2008) and birds (Banks et al. 2009), microbiota studies have often found a majority of OTUs to be specific to a particular host. These highly contrasting patterns suggest that the ability to colonize hosts (i.e., transferability of specific microbiota) differs greatly among bacterial OTUs.

Based on predicted functionality of metagenomic content, the *Certhia* and *Sitta* gut microbiome was largely (> 50%) metabolism related. No statistically different metagenomic functional groups were identified between the two *Certhia* lineages, but numerous differences identified between *Certhia* and *Sitta* were related to metabolism (Table 3.4). This result is similar to the findings of the meta-analysis of Waite and Taylor (2014), in which many functional differences were found among host species.

Factors shaping the gut microbiota

In natural systems, the vertebrate gut microbiota has been shown to vary with ecology (e.g., diet; Waite and Taylor 2014; Ley et al. 2008), age (Godoy-Vitorino et al. 2010, van Dongen et al. 2013), geography (Lucas and Heeb 2005, Hird et al. 2014), and genetics (Ochman et al. 2010, Banks et al. 2009, Dewar et al. 2013). In birds, the majority of studies have used categorical analyses to investigate factors shaping the gut microbiota (e.g., Waite and Taylor 2014). Here, incorporating genome-wide genetic data for hosts in conjunction with 16S data from microbial communities, we investigated the relationship between host genetics and microbial community structure in a multivariate framework.

Our first question attempted to explain microbial community diversity in relation to host genetic diversity; we found no relationship between the two variables (Table 3.2).

This result is echoed by an experiment in mice (*Mus musculus*, Kreisinger et al. 2014), in which inbred lines showed no significant difference in alpha diversity compared to wild-caught mice. In leopard seals (*Hydrurga leptonyx*), captive individuals have been shown to harbor a higher alpha diversity of gut microbiota relative to their wild counterparts (Nelson et al. 2013b). These results go against null expectations; in wild populations, more opportunities should exist for horizontal transfer of microbiota than in captive populations. Similarly, hosts with higher genetic diversity have likely had exposure to more divergent populations and associated microbiota in previous generations.

The aim of our second question and associated hypotheses was to investigate the relationships between beta diversity of microbial communities, genetic differentiation of hosts, and distance between sampling localities. Here, we found small but significant amounts of variation in beta diversity explained by genetic differentiation among hosts (Fig. 3.3, Tables 3.2 & 3.3). In contrast, geographic distance between sampling sites had no appreciable relationship with beta diversity. These relationships were simultaneously investigated between individuals of the same lineage, between different clades of the same species, and between species. Notably, the results were congruent when the microbiota were investigated at two levels: within individuals and within sampling localities of the same species (Fig. 3.3, Tables 3.2 & 3.3).

In other studies of avian gut microbiota community composition, the role of geography has varied; while some species, such as Hoatzin (*Ophisthocomus hoazin*; Godoy-Vitorino et al. 2012) and Brown-headed Cowbird (*Molothrus ater*; Hird et al. 2014), have shown differences between sampling localities, other species' gut microbial communities showed no effects of geographic sampling locality [e.g., Adélie penguins (*Pygoscelis*

adeliae); Banks et al. 2009]. Like geography, genetics has been shown to play a role in avian microbiota assembly in some taxa, both within (e.g., Adélie penguins; Banks et al. 2009) and among species (e.g., multiple penguin species; Dewar et al. 2013). In contrast, species assignment did not explain microbiota community assemblages across multiple passerine birds (Hird et al. 2014).

The varied relationships of geography and genetics with avian gut microbial community composition suggest mixed levels of dispersal and inheritance effects on different bird species' microbiota. Substantial dispersal of hosts among localities would be required to manifest a signal of isolation by distance in gut microbial communities. With a lack of significant dispersal, opportunities would not exist for lateral transfer of bacterial OTUs among host localities. In contrast, too much dispersal of hosts, and associated lateral transfer of bacterial OTUs, could lose any geographic signal and cause a large proportion of the microbiota to be ubiquitous among localities. The effects of genetics on microbial community structure depend on the proportion of those communities passed between parents and offspring. Differential inheritance among taxa would lead to mixed signals of the relationships between microbial community beta diversity and genetic differentiation among birds. Here, we found a lack of geographic signal, suggesting limited lateral transfer of microbiota among disjunct sampling localities. In contrast, we found a significant relationship between genetic differentiation and microbial community differences, indicating a partial role of inheritance on microbiota composition in *C. americana* and *S. carolinensis*.

Literature Cited

Atwood, T.C., J. K. Young, J. P. Beckmann, S. W. Breck, J. Fike, O. E. Rhodes, and K. D. Bristow. 2011. Modeling connectivity of black bears in a desert sky island archipelago. Biological Conservation 144: 2851–2862.

Banks, J. C., S. C. Cary, and I. D. Hogg. 2009. The phylogeography of Adelie penguin faecal flora. Environmental Microbiology 11: 577–588.

Baptista, L.F., and R. Krebs. 2000. Vocalizations and relationships of Brown Creepers *Certhia americana*: a taxonomic mystery. Ibis 142: 457–465.

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

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57: 289–300.

Bond, M. H., P. A. Crane, W. A. Larson, and T. P. Quinn. 2014. Is isolation by adaptation driving genetic divergence among proximate Dolly Varden char populations?. Ecology and Evolution 4: 2515–2532.

Bourret, V., M. Dionne, M. P. Kent, S. Lien, and L. Bernatchez. 2013. Landscape genomics in Atlantic Salmon (*Salmo salar*): searching for gene–environment interactions driving local adaptation. Evolution 67: 3469–3487.

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67: 3258–3273.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

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

Caporaso, J.G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7: 335–336.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6: 1621–1624.

Carling, M.D., and R. T. Brumfield. 2008. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in *Passerina* buntings. Genetics 178: 363–377.

Carling, M.D., I. J. Lovette, and R. T. Brumfield. 2010. Historical divergence and gene flow: coalescent analyses of mitochondrial, autosomal and sex-linked loci in *Passerina* buntings. Evolution 64: 1762–1772.

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics 1: 171–182.

Catchen, J. M., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool for population genomics. Molecular Ecology 22: 3124–3140.

Collins, W. D., C. M. Bitz, M. L. Blackmon, G. B. Bonan, C. S. Bretherton, J. A. Carton, et al. 2006 The community climate system model version 3 (CCSM3). Journal of Climatology 19: 2122–2143.

Cramp, S., and C. M. Perrins. 1993. The Birds of the Western Palearctic: Flycatchers to Shrikes, vol. 7. Oxford University Press, Oxford, UK.

Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. In Speciation and its consequences (eds. Otte, D. & Endler, J.), Sinauer, Sunderland, MA, USA.

Davis, C. M. 1978. A nesting study of the Brown Creeper. Living Bird 17: 237–263.

Deagle, B. E., F. C. Jones, Y. F. Chan, D. M. Absher, D. M. Kingsley, and T. E. Reimchen. 2011. Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. Proceedings of the Royal Society B: Biological Sciences: rspb20111552.

DeFries, R., M. Hansen, J. R. G. Townshend, A. C. Janetos, and T. R. Loveland. 2000. A new global 1 km data set of percent tree cover derived from remote sensing. Global Change Biology 6: 247–254.

Dewar, M. L., J. P. Arnould, P. Dann, P. Trathan, R. Groscolas, and S. Smith. 2013. Interspecific variations in the gastrointestinal microbiota in penguins. MicrobiologyOpen 2: 195–204.

Dudaniec, R. Y., J. R. Rhodes, W. J. Worthington, M. Lyons, K. E. Lee, C. A. McAlpine, and F. N. Carrick. 2013. Using multilevel models to identify drivers of landscape-genetic structure among management areas. Molecular Ecology 22: 3752–3765.

Dudaniec, R. Y., S. F. Spear, J. S. Richardson, and A. Storfer. 2012. Current and historical drivers of landscape genetic structure differ in core and peripheral salamander populations. PLoS One 7: e36769.

Edelaar, P., and D. I. Bolnick. 2012. Non-random gene flow: an underappreciated force in evolution and ecology. Trends in Ecology and Evolution 27: 659–665.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461.

Elgvin, T. O., J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Saether, et al. 2011. Hybrid speciation in sparrows II: a role for sex chromosomes? Molecular Ecology 20: 3823–3837.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491: 756–760.

Faith, D. P. and A. M. Baker. 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online 2: 121.

Fitak, R. R., J. L. Koprowski, and M. Culver. 2013. Severe reduction in genetic variation in a montane isolate: the endangered Mount Graham red squirrel (*Tamiasciurus hudsonicus grahamensis*). Conservation Genetics 14: 1233–1241.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977–993.

Freeland, J. R., P. Biss, K. F. Conrad, and J. Silvertown. 2010. Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow. Journal of Evolutionary Biology 23: 776–782.

Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Molecular Biology and Evolution 30: 1687–1699.

Funk, D. J., S. P. Egan, and P. Nosil. 2011. Isolation by adaptation in *Nechlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. Molecular Ecology 20: 4671–4682.

Gardner, J. P. A., and K. J. Wei. 2014. The genetic architecture of hybridisation between two lineages of greenshell mussels. Heredity 114: 344–355.

Garroway, C. J., R. Radersma, I. Sepil, A. W. Santure, I. Cauwer, J. Slate, and B. Sheldon. 2013. Fine-scale genetic structure in a wild bird population: the role of limited dispersal and environmentally based selection as causal factors. Evolution 67: 3488–3500.

Godoy-Vitorino, F., K. C. Goldfarb, E. L. Brodie, M. A. García-Amado, F. Michelangeli, and M. G. Domínguez-Bello. 2010. Developmental microbial ecology of the crop of the folivorous hoatzin. The ISME Journal 4: 611–620.

Godoy-Vitorino, F., S. J. Leal, W. A. Díaz, J. Rosales, K. C. Goldfarb, and M. A. García-Amado. 2012. Differences in crop bacterial community structure between hoatzins from different geographical locations. Research in Microbiology 163: 211–220.

Günther, T., and G. Coop. 2013. Robust identification of local adaptation from allele frequencies. Genetics 195: 205–220.

Harris, R. B., M. D. Carling, and I. J. Lovette. 2014. The influence of sampling design on species tree inference: a new relationship for the new world chickadees (Aves: *Poecile*). Evolution 68: 501–513.

Henry, P., M. A. Russello. 2013. Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. Ecology and Evolution 3: 3906–3917.

Hersch-Green, E. I., G. J. Allan, and T. G. Whitham. 2014. Genetic analysis of admixture and patterns of introgression in foundation cottonwood trees (Salicaceae) in southwestern Colorado, USA. Tree Genetics and Genomes 10: 527–539.

Hird, S. M. 2013. Novel computational tools and utilization of the gut microbiota for phylogeographic inference (Doctoral dissertation, Louisiana State University).

Hird, S. M., B. C. Carstens, S. W. Cardiff, D. L. Dittmann, and R. T. Brumfield. 2014. Sampling locality is more detectable than taxonomy or ecology in the gut microbiota of the brood-parasitic Brown-headed Cowbird (*Molothrus ater*). PeerJ 2: e321.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965–1975.

Janoušek, V., L. Wang, L. Luzynski, P. Dufková, M. M. Vyskočilová, M. W. Nachman, et al. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. Molecular Ecology 21: 3032–3047.

Jeffreys, H. 1961. The Theory of Probability. Oxford University Press, Oxford.

Jombart, T., and I. Ahmed. 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27: 3070–3071.

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

Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, et al. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology 23: 4035–4058.

Kingsley, R. A., and A. J. Baumler. 2000. Host adaptation and the emergence of infectious disease: the Salmonella paradigm. Molecular Microbiology 36: 1006–1014.

Kreisinger, J., D. Čížková, J. Vohánka, and J. Piálek. 2014. Gastrointestinal microbiota of wild and inbred individuals of two house mouse subspecies assessed using high-throughput parallel pyrosequencing. Molecular Ecology 23: 5048–5060.

Kuchta, S. R., and A. M. Tan. 2005. Isolation by distance and post-glacial range expansion in the rough-skinned newt, *Taricha granulosa*. Molecular Ecology 14: 225–244.

Kühn, I., A. Iversen, L. G. Burman, B. Olsson-Liljequist, A. Franklin, M. Finn, et al. 2003. Comparison of enterococcal populations in animals, humans, and the environment-a European study. International Journal of Food Microbiology 88: 133–145.

Lan, Y., M. W. A. Verstegen, S. Tamminga, and B. A. Williams. 2005. The role of the commensal gut microbial community in broiler chickens. World's Poultry Science Journal 61: 95–104.

Langille, M. G., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, et al. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31: 814–821.

Lavretsky, P., J. M. DaCosta, B. E. Hernández-Baños, A. Engilis, M. D. Sorenson, and J. L. Peters. 2015. Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. Molecular Ecology 24: 5364-5378.

Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, et al. 2008. Evolution of mammals and their gut microbes. Science 320: 1647–1651.

Lira-Noriega, A., and J.D. Manthey. 2014. Relationship of genetic diversity and niche centrality: A survey and analysis. Evolution 68: 1082–1093.

Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71: 8228–8235.

Lozupone, C. A., M. Hamady, S. T. Kelley, and R. Knight. 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology 73: 1576–1585.

Lucas, F. S., and P. Heeb. 2005. Environmental factors shape cloacal bacterial assemblages in great tit *Parus major* and blue tit *P. caeruleus* nestlings. Journal of Avian Biology 36: 510–516.

Lynch, M. 2007. The Origins of Genome Architecture. Sinauer Associates, Sunderland, Massachusetts.

Mai, V., and D. J. Baer. 2010. Understanding the extent and sources of variation in gut microbiota studies: a prerequisite for establishing associations with disease. Diversity 2: 1085–1096.

Malek, T. B., J. W. Boughman, I. Dworkin, and C. L. Peichel. 2012. Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. Molecular Ecology 21: 5265–5279.

Manel, S., and R. Holderegger. 2013. Ten years of landscape genetics. Trends in Ecology and Evolution 28: 614–621.

Manthey, J.D., and R. G. Moyle. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. Molecular Ecology 24: 3628–3638.

Manthey, J. D., and G. M. Spellman. 2014. Increased differentiation and reduced gene flow in sex chromosomes relative to autosomes between lineages of the brown creeper *Certhia americana*. Journal of Avian Biology 45: 149–156.

Manthey, J. D., J. Klicka, and G. M. Spellman. 2011a. Cryptic diversity in a widespread North American songbird: phylogeography of the Brown Creeper (*Certhia americana*). Molecular Phylogenetics and Evolution 58: 502–512.

Manthey, J. D., J. Klicka, and G. M. Spellman. 2011b. Isolation-driven divergence: speciation in a widespread North American songbird (Aves: Certhiidae). Molecular Ecology 20: 4371–4384.

Manthey, J. D., J. Klicka, and G. M. Spellman. 2014. Effects of climate change on the evolution of Brown Creeper (*Certhia americana*) lineages. Auk 131: 559–570.

Manthey, J. D., J. Klicka, and G. M. Spellman. 2015a. Chromosomal patterns of diversity and differentiation in creepers: a next-gen phylogeographic investigation of *Certhia americana*. Heredity 115: 165–172.

Manthey, J. D., M. B. Robbins, and R. G. Moyle. 2015b. A genomic investigation of the putative contact zone between divergent Brown Creeper (*Certhia americana*) lineages: chromosomal patterns of genetic differentiation. Genome doi: 10.1139/gen-2015-0093.

Marler, P., and M. Tamura. 1962. Song “dialects” in three populations of White-crowned Sparrows. Condor 64: 368–377.

Marshall, J. T. 1956. Summer birds of the Rincon Mountains, Saguaro National Monument, Arizona. Condor 58: 81–97.

Masta, S. E. 2000. Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): recent vicariance of sky island populations? Evolution 54: 1699–1711.

Mastretta-Yanes, A., N. Arrigo, N. Alvarez, T. H. Jorgensen, D. Piñero, and B. C. Emerson. 2015. Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. Molecular Ecology Resources 15: 28–41.

Matthysen, E., F. Adriaensen, A. A. Dhondt. 1995. Dispersal distances of nuthatches, Sitta europaea, in a highly fragmented forest habitat. Oikos 72: 375–381.

Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Harvard University Press, Cambridge, MA, USA.

McCarthy, E. M. 2006. Handbook of avian hybrids of the world. Oxford University Press, New York, NY, USA.

McDonald, D., M. N. Price, J. Goodrich, E. P. Nawrocki, T. Z. DeSantis, A. Probst, et al. 2011. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME Journal 6: 610–618.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 17: 240–248.

Nelson, T. M., T. L. Rogers, and M. V. Brown. 2013a. The gut bacterial community of mammals from marine and terrestrial habitats. PloS One 8: e83655.

Nelson, T. M., T. L. Rogers, A. R. Carlini, and M. V. Brown. 2013b. Diet and phylogeny shape the gut microbiota of Antarctic seals: a comparison of wild and captive animals. Environmental Microbiology 15: 1132–1145.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogenous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. Evolution 62: 316–336.

Nozawa, T., T. Nagashima, H. Shiogama, and S. A. Crooks. 2005. Detecting natural influence on surface air temperature change in the early twentieth century. Geophysicial Research Letters 32: L20719.

Ochman, H., M. Worobey, C. H. Kuo, J. B. N. Ndjango, M. Peeters, B. H. Hahn, and P. Hugenholtz. 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biology 8: 2638.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. Meester. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Molecular Ecology 22: 5983–5999.

Parks, D. H., G. W. Tyson, P. Hugenholtz, and R. G. Beiko. 2014. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 30: 3123–3124.

Peterson, A. T., M. Papeş, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling 213: 63-72.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. Ecological niches and geographic distributions. Princeton University Press, Princeton.

Phillips, A. J., J. Marshall, and G. Monson. 1964. The Birds of Arizona. University of Arizona Press, Tucson, AZ.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modelling of species geographic distributions. Ecological Modelling 190: 231–259.

Potter, K. M., V. D. Hipkins, M. F. Mahalovich, and R. E. Mean. 2013. Mitochondrial DNA haplotype distribution patterns in *Pinus ponderosa* (Pinaceae): range-wide evolutionary history and implications for conservation. American Journal of Botany 100: 1562–1579.

Poulos, H. M., and A. E. Camp. 2010. Topographic influences on vegetation mosaics and tree diversity in the Chihuahuan Desert Borderlands. Ecology 9: 1140–1151.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

R Development Core Team. 2012. R: A Language for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schield, D. R., D. C. Card, R. H. Adams, T. Jezkova, J. Reyes-Velcasco, F. N. Proctor, et al. 2015. Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). Molecular Phylogenetics and Evolution 83: 213–223.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution 68: 1–15.

Sharbel, T. F., B. Haubold, and T. Mitchell-Olds. 2000. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. Molecular Ecology 9: 2109–2118.

Shafer, A., and J. B. Wolf. 2013. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. Ecology Letters 16: 940–950.

Slabbekoorn, H., and T. B. Smith. 2002. Bird song, ecology, and speciation. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 357: 493–503.

Smith, C. I., and B. D. Farrell. 2005. Phylogeography of the longhorn cactus beetle *Moneilema appressum* LeConte (Coleoptera: Cerambycidae): was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? Molecular Ecology 14: 3049–3065.

Smith, T. B., R. Calsbeek, R. K. Wayne, K. H. Holder, D. Pires, and C. Bardeleben. 2005. Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. Journal of Evolutionary Biology 18: 257–268.

Spellman, G. M., and J. Klicka. 2007. Phylogeography of the white-breasted nuthatch (*Sitta carolinensis*): diversification in North American pine and oak woodlands. Molecular Ecology 16: 1729–1740.

Storchová, R., J. Reif, and M. W. Nachman. 2010. Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. Evolution 64: 456–471.

Taylor, S. A., R. L. Curry, T. A. White, V. Ferretti, and I. J. Lovette. 2014. Spatiotemporally consistent genomic signatures of reproductive isolation in a moving hybrid zone. Evolution 68: 3066–3081.

Turnbaugh, P. J., R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, and J. I. Gordon. 2007 The human microbiome project. Nature 449: 804–810.

van Dongen, W. F., J. White, H. B. Brandl, Y. Moodley, T. Merkling, S. Leclaire, et al. 2013. Age-related differences in the cloacal microbiota of a wild bird species. BMC Ecology 13: 11.

Wade, T. G., K. H. Riitters, J. D. Wickham, and K. B. Jones. 2003. Distribution and causes of global forest fragmentation. Conservation Ecology 7.

Waite, D. W., and M. W. Taylor. 2014. Characterizing the avian gut microbiota: membership, driving influences, and potential function. Frontiers in Microbiology 5: 223.

Walstrom, V. W., J. Klicka, and G. M. Spellman. 2012. Speciation in the White-breasted Nuthatch (*Sitta carolinensis*): a multilocus perspective. Molecular Ecology 21: 907–920.

Wang, I. J., R. E. Glor, and J. B. Losos. 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. Ecology Letters 16: 175–182.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, et al. 2010. The genome of a songbird. Nature 464: 757–762.

Willing, E. M., C. Dreyer, and C. van Oosterhout. 2012. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. PLoS One 7: e42649.

Wright, S. 1943. Isolation by distance. Genetics 28: 114–138.

Xu, J., and J. I. Gordon. 2003. Honor thy symbionts. Proceedings of the National Academy of Sciences 100: 10452–10459.

Yin, Y., F. Lei, L. Zhu, S. Li, Z. Wu, R. Zhang, G. F. Gao, B. Zhu, and X. Wang. 2009. Exposure of different bacterial inocula to newborn chicken affects gut microbiota development and ileum gene expression. The ISME Journal 4: 367–376.

Appendix I

Bioclimatic layers and their descriptions, used for ecological niche modeling in *Sitta carolinensis*.

Layer	Description
Bio1	Annual mean temperature
Bio2	Monthly temperature range
Bio4	Temperature seasonality
Bio5	Max temperature warmest month
Bio6	Min temperature coldest month
Bio9	Mean temperature driest quarter
Bio12	Annual precipitation
Bio15	Precipitation seasonality
Bio17	Precipitation driest quarter
Bio18	Precipitation warmest quarter
Bio19	Precipitation coldest quarter

Appendix II.

Outlier loci identified in LFMM and BayEnv2 in *Sitta carolinensis*. For LFMM, results are shown when one or nine latent factors (LF) were used to test associations between environmental variables and SNPs. Environmental variables are temperature of warmest month (TWM) and precipitation of driest month (PDM).

Locus	K1 TWM	K9 TWM	K1 PDM	K9 PDM	BayEnv2 TWM	Bayenv2 PDM	both
266	x						
599	x	x	x	x			
1709						x	
1942			x			x	x
1971						x	
2015	x	x	x	x			
2076			x	x			
2245	x	x					
2322	x	x	x				
2946						x	
3905	x	x					
3926						x	
4167			x	x			
4307			x	x			
4782			x	x		x	x
4836	x		x	x			
4928			x	x			
5290	x	x	x	x			
5586			x				
5619	x	x					
5970							
5976	x		x				
6137	x		x	x		x	x
6377						x	
6657						x	
7347	x	x	x				
8133			x	x		x	x
8375						x	
8425					x	x	
8984	x						
9065						x	
9414						x	
9477						x	
9599						x	
9779	x	x					
10315	x		x	x			

10922			x	x				
11718	x	x	x	x				
12309							x	
12436	x	x	x	x				
12531							x	
12673			x	x				
14016			x					
14162	x	x	x	x			x	x
15430			x	x				
16207							x	
16278							x	
16535	x	x						
16614							x	
16829	x	x						
17546	x	x	x	x				
17564								
17616			x					
19820			x	x			x	x
20075							x	
20578							x	
21626			x	x				
21956	x	x						
22693	x	x						
22720							x	
22782			x	x			x	x
23035	x	x						
23512							x	
23692							x	
24511	x	x	x	x				
24569							x	
25589			x	x			x	x
26003	x	x	x					
26292	x	x						
26398			x					
26758							x	
26778	x	x	x	x				
28408			x					
28722							x	
29026							x	
29142							x	
29349							x	
30021	x	x	x	x				

Appendix III.

Certhia americana DAPC Results. Bayesian Information Criterion plot indicating the most likely number of genetic clusters for the 50% and 75% coverage datasets, on the left and right respectively. The lowest value (in both cases 2) indicates the assumed best number of genetic clusters. Below these, the assignment of the individuals from each locality to genetic cluster 1 or 2.