2018

Migratory patterns and population genetic structure in a declining wetland-dependent songbird

Matthew G. DeSaix
Virginia Commonwealth University

Migratory patterns and population genetic structure in a declining wetland-dependent songbird

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

By

Matthew George DeSaix
Bachelor of Science, Warren Wilson College, 2012

Advised by

Lesley Bulluck, Ph.D.
Department of Biology, Center for Environmental Studies
Virginia Commonwealth University

Rodney Dyer, Ph.D.
Center for Environmental Studies
Virginia Commonwealth University

Catherine Viverette, Ph.D.
Center for Environmental Studies
Virginia Commonwealth University

Virginia Commonwealth University
Richmond, VA
May, 2018

Table of contents

Acknowledgement

I wish to thank my advisors, Lesley Bulluck, Rodney Dyer, and Cathy Viverette for continuous support throughout my graduate studies. Specifically, Lesley has been instrumental in furthering my development as an avian ecologist and provided enthusiastic support for all aspects of my research. I thank Rodney for his guidance in improving myself overall as a scientist (and budding population geneticist), and always keeping me focused on the big picture of my research. I am grateful for Cathy being an exceptionally positive mentor and providing encouragement through all stages of my graduate studies. In addition, I have tremendously enjoyed the adventure of teaching the Avian Ecology course in Panama with Cathy, as well as Ed Crawford, for two winters in a row. I thank my committee member Andrew Eckert for his help with population genetic analyses and lively conversations on topics of evolutionary biology. I thank the VCU Rice River Center for awarding me a research fellowship that supported me through the summer of 2017. I am thankful to all my collaborators in the Prothonotary Warbler Working Group for their help, and notably Matt Johnson for taking me around Audubon's Francis Beidler Forest. I also thank Rupert Hester for our enjoyable times motoring across Holt Lake to visit his prothonotary warbler nest boxes. I thank my partner, Jessie Reese, for her loving support as well as helping me think through a lot of the ideas contained in this research (and editing!). I thank Lindsay Miles for her friendship and patience with my ceaseless inquiries about bioinformatics related issues. I am also grateful to have worked with Jane Remfert, Bonnie Roderique, and Liz Schold, they have been wonderful friends along the way. Finally, I thank my parents, John and Suzanne DeSaix, for their compassion and support that brought me to undertake graduate studies in the first place.

Abstract

Migratory patterns and population genetic structure in a declining wetland-dependent songbird

By Matthew DeSaix, B.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Major Co-advisors:

Lesley Bulluck, Ph.D., Department of Biology, Center for Environmental Studies

Rodney Dyer, Ph.D., Center for Environmental Studies

Catherine Viverette, Ph.D., Center for Environmental Studies

Understanding migratory connectivity is essential for assessing the drivers behind population dynamics and for implementing effective management in migratory species. Genetic markers provide a means to describe migratory connectivity, as well as incorporate population genetic analyses, however genetic markers can be uninformative for species with weak genetic structure. In this study, we evaluate range-wide population genetic structure and migratory connectivity in the prothonotary warbler, *Protonotaria citrea*, a wetland-dependent neotropical migratory songbird, using high-resolution genetic markers. We reveal regional genetic structure between sampling sites in the Mississippi River Valley and the Atlantic Seaboard with overall weak genetic differentiation among populations ($F_{ST} = 0.0051$). By ranking loci by F_{ST} and using subsets of the most differentiated genetic markers (200 – 3000), we identify a maximum

assignment accuracy (89.7% to site, 94.3% to region) using 600 single nucleotide polymorphisms. We assign samples from unknown origin nonbreeding sites to a breeding region, illustrating weak migratory connectivity between prothonotary warbler breeding and nonbreeding grounds. Our results highlight the importance of using high-resolution markers in studies of migratory connectivity with species exhibiting weak genetic structure. Using similar techniques, studies may begin to describe population genetic structure that was previously undocumented, allowing us to infer the migratory patterns of an increasing number of species.

Introduction

Understanding the spatial connections of populations throughout the full annual cycle is critical for the study of the ecology and evolutionary biology of migratory birds, and informing sound management (Sherry and Holmes 1996; Marra et al. 2006; Faaborg et al. 2010). Disturbances in one stage of the annual cycle can have carry-over effects on the population dynamics of subsequent stages (Marra et al. 1998; Sillet et al. 2000; Norris et al. 2004), but it is difficult to assess the influence of these stressors without a comprehensive understanding of the degree of migratory connectivity (Webster et al. 2002), defined as the spatial cohesiveness of populations throughout the annual cycle (Webster and Marra 2004). By incorporating range-wide migratory patterns into ecological models, the drivers behind population trends can begin to be determined and placed into a context of conservation management (Hostetler et al. 2015). For example, recent studies have used information on migratory connectivity to tease apart the influence of factors including nonbreeding season habitat loss and climate change on observed breeding season population trends that differ regionally (Fraser et al. 2012; Rushing et al. 2016; Taylor and Stutchbury 2016; Kramer et al. 2018).

To assess migratory connectivity, populations need to be tracked across large geographic distances between the breeding and nonbreeding grounds, and genetic markers have been instrumental in identifying the migratory patterns of neotropical migratory birds with cross-continental breeding distributions (Kimura et al. 2002; Colbeck et al. 2008; Irwin et al. 2011; Ruegg et al. 2014; Toews et al. 2017). Tracking migratory bird populations with genetic markers requires geographically structured genetic variation on the breeding grounds. Because migratory birds generally exhibit high dispersal capability and and this has been a limitation for migratory

connectivity research because avian species generally exhibit weak population differentiation due to high dispersal (Crochet 2000; Oyler-McCance et al. 2016). The utility of high resolution genetic markers for describing migratory connectivity of species that lack a large breeding range or distinct phylogeographic separation remains to be explored. In the eastern United States, phylogeographical history across taxa is complex and lacks straightforward spatial patterning (Soltis et al. 2006), however results of studies of neotropical migratory bird species in this region consistently show little to no population genetic structure (Ball and Avise 1992; Klein and Brown 1994, Deane et al. 2013; but see Herr et al. 2011). Few studies have attempted to use genetic markers for studying migratory patterns of neotropical migrants with breeding ranges restricted to the eastern United States, but Battey et al. (2017) recently documented strong migratory connectivity across three genetically distinct clusters in the painted bunting (*Passerina ciris*), a species which breeds in two disjunct regions in the southeastern United States.

In comparison to other tracking methods for migratory connectivity (e.g. band recapture, geolocators and GPS, or stable isotope analysis), using genetic markers has the added benefit of providing data for population genetic analyses. For example, the identification of genetically distinct intraspecific conservation units is a critical component of conservation biology and wildlife management (Moritz 1994; Crandall et al. 2000; Haig et al. 2006) and can be incorporated into migratory tracking studies that use genetic markers (e.g. Haché et al. 2017). Furthermore, there is a current need to integrate genomic data into migratory connectivity and ecological population models for migratory birds (Sherry 2018). Recently, Bay et al. (2018) combined population genomics and environmental data to show that North American population declines in yellow warblers are correlated with genomic vulnerability to climate change.

In this study, our aim was to use genome-wide data to describe spatial genetic variation across breeding populations and to determine migratory connectivity in the prothonotary warbler, *Protonotaria citrea*, a neotropical migratory warbler that breeds in the eastern United States. Identifying migratory patterns would help us understand how drivers throughout the annual cycle influence the varied demographic trends across the breeding range of this species. Based on population genetic studies of other migratory New World warblers (*Parulidae*; Clegg et al. 2003; Deane et al. 2013; Lindsay et al. 2008; Irwin et al. 2011, Ruegg et al. 2014), we expected weak to no genetic structure across the breeding range of the prothonotary warbler. In order to fulfill our objective, we addressed the following questions: 1) in a species with a breeding range limited to the eastern United States, is there detectable population genetic structure?; 2) if so can we assign known-origin individuals to their correct population, and what is the optimal number of SNPs to do so?; and 3) where is the breeding origin of individuals sampled from the nonbreeding grounds?

Materials and Methods

Focal species and DNA sampling

The prothonotary warbler (*Protonotaria citrea*) is a wetland-dependent songbird that breeds in the bottomland hardwood forests of the eastern United States (Petit 1997; Figure 1). Due to habitat loss and continuing population declines, the prothonotary warbler is designated a species of conservation priority by Partners in Flight (Rosenberg et al. 2016), as well as a species of "special concern" by U.S. Fish and Wildlife Service (U.S. Fish and Wildlife Service 2008). While the species is estimated to be declining at a rate of 0.17%/year in the past decade (confidence intervals [CI]: -1.43%, 1.13%), trend estimates between the two centers of abundance vary from decreases of -2.02%/year (CI: -4.47%, 0.42%) in Louisiana to increases of 1.80%/year (CI: -0.23%, 4.23%) in North Carolina (Sauer et al. 2017). Prothonotary warbler abundance is unequally distributed across the breeding range with two locales of high concentration centered in southern Louisiana (25.2% of breeding population) and eastern North Carolina (18.9% of breeding population; Partners in Flight Database 2013; Figure S6).

Between 2014 and 2017, we obtained genetic samples from 265 individuals across 17 sites (3 - 29 samples/site, Table 1) distributed throughout the breeding (175 samples) and nonbreeding (90 samples) range (Figure 1). We collected blood samples for each individual using brachial venipuncture and preserved these samples on Whatman FTA cards (Gutierrez-Corchero et al. 2002) and collection was conducted under Virginia Commonwealth University IACUC protocol #AM10230. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA) following a modified lysis protocol for Whatman FTA cards. Two 1.2 mm punches of dried blood were initially incubated in 180 µl Buffer ATL for 10 minutes at

94°C. The sample was briefly allowed to cool and 20 µl Proteinase K Solution were added, and the sample was incubated at 56°C until complete lysis, approximately an hour. Following lysis, our DNA extraction procedure adhered to the manufacturer's protocol for the DNeasy Blood & Tissue Kit and we eluted samples to water.

Library preparation and sequencing

Double digest RADseq (ddRADSeq) was used to produce three multiplexed libraries (Peterson et al. 2012) for sequencing on the Illumina platform following the protocol outlined by Parchman et al. (2012). Genomic DNA was digested with two restriction endonucleases (EcoRI and MseI) and adaptor oligonucleotides, containing 10 base pair (bp) barcodes for the unique identification of individuals, were ligated to the digested fragments. The ligated fragments were amplified using PCR and individuals with unique barcodes were pooled together in sets of 96 samples. Pooled amplified libraries were size selected for fragments in the approximate range of 300-500 bp using gel electrophoresis on 1% agarose gels and purified using QIAquick Gel Extraction Kits (Qiagen, Valencia, CA). Single-end sequencing with one multiplexed library per lane was performed by Novogene Corporation using the Illumina HiSeq 4000 platform. Reads were demultiplexed and trimmed to 60 bp using the process_radtags program in Stacks v.2.0 (Catchen et al. 2013), and trimming length was based on the quality score distribution along the reads. The resulting FASTQ files were processed in the dDocent bioinformatics pipeline (Appendix S1; Puritz et al. 2014), yielding 26,189 single nucleotide polymorphisms (SNPs) that were used as the initial data set for subsequent analyses.

Population structure

We used multiple complementary methods to quantify and describe the extent of genetic population structure across the breeding range to provide insights into the statistical power for assignment of individuals sampled on the wintering grounds. First, we conducted principal component analysis (PCA) of multilocus genotypes to visualize any grouping of the 175 sampled breeding individuals. Initially, the full set of 26,189 SNP genotypes were used (Patterson et al. 2006). Obvious outliers individuals, defined as multilocus genotypes whose PCA coordinates exceeded six standard deviations from the centroids of the first ten principal components, were removed from the dataset. An analysis of variance (ANOVA) was performed on the amount of missing loci data per individual between breeding sites and we excluded sites with high levels of missing data from a subsequent PCA.

In addition to PCA analysis, we also employed the *STRUCTURE* (Pritchard et al. 2000) algorithm, based upon population genetic assumptions of admixture model with correlated allele frequencies and no prior information on sampling location to elucidate broad-scale compartmentalization in the spatial genetic structure of samples from the breeding grounds. We varied the number of groups (K) from 1 to 11 with 5 iterations for each value, and set 10,000 burn-in iterations followed by another 50,000 Markov Chain Monte Carlo (MCMC) steps. We followed the guidelines of Evanno et al. (2005) to identify the most likely number of clusters (K).

Population structure was estimated using both multilocus F-statistics and pairwise site differentiation with 95% confidence intervals using the *assigner* package (Gosselin et al. 2016) in *R* (version 3.4.4, R Core Development Team 2018). We conducted a hierarchical analysis of molecular variance (AMOVA) with levels of site and region in the *hierfstat* package (Goudet

2005). Independent of the magnitude of population structure, we also examined the extent to which population differentiation was spatially arranged under a model of Isolation by Distance ($F_{ST}/(1 - F_{ST})$ ~ log(Euclidean Distance)) using a standard Mantel test (1,000 permutations) in the package *adegenet* (Jombart 2008). If regional clustering is revealed, we will investigate isolation by distance with hypothesized regions as well as among all sites.

Population assignment

To test the effect of number of markers on assignment accuracy, we calculated expected genotype frequencies in a training-set with subsets of SNPs (200, 400, 600, 800, 1000, 1200, 1500, 2000, and 3000 SNPs). Only SNPs with both alleles present in every breeding sampling site were used in order to circumvent the issue of calculating genotype frequencies of zero. We ranked loci by single locus F_{ST} estimates from variance components and selected loci with the highest F_{ST} values. Individuals held out were assigned to populations (i.e. breeding sample site) based upon maximum likelihood of multilocus genotype probabilities (Paetkau et al. 1995). If regional structuring is identified, we will consider regional assignment to be based on the region containing the assigned population. By using separate sets of individuals for population allele frequencies and assignment in the hold-out set, we avoided the inherent upward bias in predicted accuracy that occurs when combining classifier and training data (Anderson 2010). Assignment tests were performed using twenty bootstrapped data sets, each containing a 2:1 ratio of training and hold out individuals randomly sampled from each breeding site. The subset of SNPs with the highest assignment accuracy were used in a leave-one-out cross-validation model to describe the variation in assignment success across sites and regions. We examined the effect of sample

size of sites on the population assignment accuracy using a nonlinear regression. All assignment tests were conducted with custom scripts in R which are available through GitHub (https://github.com/mgdesaix/populationAssignment).

Migratory connectivity

As in the assignment methods above, we limited the total number of SNPs to loci that had two alleles in both of the regions, ranked loci by single locus F_{ST} estimates that included regional variance components and selected those with highest F_{ST}. We used the optimal number of loci as determined above to calculate allele frequency by region and assignment was determined by the maximum product of expected genotype frequencies across loci. To estimate mixing of breeding individuals on the nonbreeding sites, we used a chi-square test to compare the proportion of individuals assigned to the regions at each nonbreeding site to an estimate of relative abundance (Partners in Flight Databases 2013) between the regions on the breeding ground. We tested for erroneous regional assignment due to mismatched sample size between the two regions using a chi-square test.

Results

Population structure

The first two principal components explained 0.89% and 0.85% of the genetic variation, respectively, and most individuals were clustered together except for several outlier individuals (Figure S1). After the iterative removal of outliers, the PCA explained 0.88% (PC1) and 0.82% (PC2) of the variation and highlighted genetic clustering at a regional level of individuals from sites along the Atlantic Seaboard (ATL) and Mississippi River Valley (MRV; Figure 2, Figure S2). For subsequent analyses and assignment, ATL and MRV will be referred to as 'regions' and sampling sites will be referred to as 'sites'.

The analysis of variance showed significant differences among the breeding sites based on the proportion of missing loci data from individuals in those sites ($F = 3.69$, $DF = 10$, $p < 0.001$, Figure S2), and SC1 and OH1 were identified as sites that had a significantly higher proportion of missing data in individuals. When we compared this PCA to the PCA without the populations that had a larger number of individuals with missing data (SC1 and OH1), the pattern of clustering was similar. This suggested that the uneven distribution of missing data between breeding sites was not altering the genetic clustering, so we continued subsequent analyses using the full set of individuals and sites.

We ran *STRUCTURE* using all 175 individuals from the 11 sites on the breeding ground, and performed two separate runs using all 26,189 SNPs as well as the number of SNPs we determined to be optimal for population assignment (i.e. 600 SNPs). Using 26,189 SNPs, *STRUCTURE* did not provide an optimization of *K* clusters (Figure S3), which corroborates other studies that have shown *STRUCTURE* to perform poorly compared to other methods when

working with weakly differentiated populations (Waples and Gaggiotti 2006; Chen et al. 2007). For the purposes of assignment we retained *K* of 2 clusters as identified in the PCA.

Overall F_{ST} across populations was 0.0051. Pairwise genetic differentiation, measured with F_{ST}, ranged from 0.00 to 0.011 among the 11 sampling locations. No significant difference in F_{ST} values was found among the following pairs of sampling locations : VA1 & VA3, VA3 & NC1, and NC1 & SC1 (Table S1). Based on an AMOVA, genetic differentiation between regions was significant, though the amount of genetic variation explained by region (F_{CT} = 0.0025, p = 0.002; Table 1) and by site (F_{CT} = 0.0034, p = 0.002) was slight. The results of the mantel test for all breeding sites showed there was a significant positive correlation between genetic and geographic distances (r^2 = 0.40, p = 0.001, Figure S5). When examined within the regions separately, the ATL sites had a non-significant positive association between genetic and geographic distances (r^2 = 0.44, p = 0.26), while the MRV sites had a strong significant positive association between genetic and geographic distances (r^2 = 0.61, p = 0.004).

Population assignment

We considered assignment accuracy at two spatial scales: site and region. In the hold-out validation test, assignment accuracy of individuals to sampling site varied with the number of markers used for assignment (Figure 3) and we identified 600 SNPs as the optimal number of markers. Using the least number of markers tested, 200 SNPs, the median assignment accuracy of individuals to site across training sets was 76.4% (range: 65.5% - 85.5%). Assignment accuracy peaked at 83.6% (range: 76.4% - 87.3%) with the use of 600 SNPs and declined to 50.0% (range: 40.0% - 63.6%) with the maximum number of markers tested, 3000 SNPs. The

maximum threshold of 3000 loci was determined from the 20 training data sets which had a range of 3513 to 3826 available loci. Assignment of individuals to region followed a similar pattern and had the highest assignment accuracy of 94.6% using 600 SNPs (range: 90.9% - 98.2%). We performed the leave-one-out cross-validation assignment with 600 SNPs and the F_{ST} of these loci ranged from 0.072 to 0.42. Compared to the hold-out validation tests, the leave-one-out cross-validation assignment resulted in a modest increase in assignment accuracy by site, 89.7%, and an equal value of regional assignment accuracy, 94.3% (Figure 4; Table S2). Assignment accuracy by population was strongly correlated with the number of individuals sampled at each site ($r^2 = 0.84$, $p = 0.0007$, $F = 20.84$ on 2 and 8 *DF*), and assignment accuracy plateaued when sites had at least 13 individuals (Figure S5).

Migratory connectivity

For the assignment of nonbreeding individuals to an origin on the breeding grounds, we assigned individuals to a general region rather than a specific site for two reasons: 1) it is highly improbable that any sampled individual from the nonbreeding ground originated from one of our sampled sites, 2) we showed that assignment accuracy by site was affected by sample size and because of this we had low assignment accuracy to some of the sites (Table S4). Due to close geographic proximity (<50 km), we combined two of the sites in Panama together as well as two of the sites in Colombia when we conducted the assignment of nonbreeding individuals (Figure 4). Based on these groupings, sample size by site ranged from 3 to 36. There were 19046 loci that had two alleles in individuals from both of the breeding regions, and when reduced to 600 loci with the highest F_{ST} they ranged in F_{ST} from 0.044 to 0.25. Across all nonbreeding sites, 22

samples (24.4% of 90 nonbreeding samples) were assigned to ATL and 68 samples (75.6% of 90 nonbreeding samples) were assigned to MRV. At every nonbreeding site more individuals were assigned to MRV than ATL (Table 2). Using relative abundance data from the Partners in Flight Databases (2013), we determined our defined regions of MRV and ATL to account for 68% and 32%, respectively, of the breeding abundance of prothonotary warblers. A chi-square test of nonbreeding assignment proportions (0.24 and 0.76 from ATL and MRV, respectively) across the nonbreeding site groupings did not reject the null hypothesis of deviation from the relative breeding abundance proportions ($\chi^2 = 4.18$, DF = 4, p = 0.38), indicating that individuals from the two breeding regions were mixing at the nonbreeding sites at a comparable proportion to relative breeding abundance estimates of those two regions. The total proportion of nonbreeding individuals assigned to these two regions was significantly different ($\chi^2 = 8.89$, DF = 1, p = 0.0029) from the proportion of sampled individuals from these two regions (77 individuals or 44% from ATL, 98 individuals or 56% from MRV), indicating that nonbreeding assignment was not erroneously driven by sample size of the assigned region.

Discussion

Regional genetic structure across breeding grounds

Our results provide evidence of weak genetic structure among prothonotary warbler populations across the breeding range, with regional structure between the Mississippi River Valley and Atlantic Seaboard. The weak genetic differentiation within different hierarchical levels (site and region) indicate substantial genetic connectivity among sites within these two regions and/or a recent divergence of large population size (Marko and Hart 2011). We demonstrate that the observed genetic structure in prothonotary warblers is in part explained by a model of isolation by distance and this pattern is observed across the breeding range and sites within MRV, though there was no evidence of isolation by distance solely within ATL sites. The difference in support for isolation by distance between the two regions could be driven by the larger latitudinal separation between sampled sites in MRV (29.8° to 42.5°) than in ATL (33.2° to 38.2°). Interestingly, three of the pairwise genetic differentiation comparisons from closely-spaced ATL populations (VA1 & VA3, VA3 & NC1, NC1 & SC1) were non-significant, however, all sites in MRV with similar geographic separation (< 200 km; LA1, LA2, and LA3) had significant genetic differentiation.

Varied patterns of genetic differentiation at similar scales may be indicative of discrepancies in genetic connectivity due to habitat fragmentation or differing wetland and riparian topography between these areas (Baguette et al. 2012). Despite the generally high rates of dispersal in avian species (Crochet 2000), lower rates of dispersal can occur in forest-dwelling bird that are indisposed to crossing large swaths of open area (Belisle et al. 2001) as well as habitat specialists, such as wetland-dependent species (Haig et al. 1998). If this were the case in

prothonotary warblers, reduced dispersal due to habitat specificity should result in generally higher genetic differentiation between populations than in a non-wetland-dependent migratory species of similar range. In a mature forest-dwelling species of the eastern United States, cerulean warbler (*Setophaga cerulea*), overall population differentiation was estimated at 0.002 and no population genetic structure was found (Deane et al. 2013). The higher genetic differentiation, as well as presence of population genetic structure, in the prothonotary warbler compared to cerulean warbler may in part be related to the wetland preference of prothonotary warblers. In another study, golden-cheeked warblers (*Setophaga chrysoparia*), a habitat-specialist of Ashe-juniper and oak woodlands (Pulich 1976), was shown to have higher levels of genetic differentiation compared to populations of wider-ranging species across similar geographic scales (Lindsay et al. 2008).

Our results show that the use of thousands of SNPs provide a significant measure of breeding-wide genetic structure in the prothonotary warbler, which has not been previously documented in *Parulids* of the eastern United States (Ball and Avise 1992; Klein and Brown 1994; Winker et al. 2000; Deane et al. 2013), nor any neotropical migratory songbird species in the eastern United States apart from the painted bunting (Herr et al. 2011). In the northern part of the prothonotary warbler breeding range, the genetic clusters follow the geographical separation of the range by the Appalachian Mountains; however, we did not have sufficient sampling coverage in the contiguous southern portion of the range to determine precise boundaries of the clusters. In the southeastern United States, the Tombigbee River and Apalachicola River are important topographic features driving genetic discontinuities, primarily in non-avian taxa (Soltis et al. 2006), though Gill et al. (1993) used microsatellites to reveal that

genetic structure in the Carolina chickadee (*Parus caroliniensis*) was partitioned by the Tombigbee discontinuity.

Population assignment accuracy

Despite minimal genetic differentiation among sites we assigned breeding-range sampled individuals to site with high accuracy in the hold-out validation model, 83.6%, using only 600 SNPs ranked by F_{ST}. Our results provide further evidence that weakly differentiated markers can be uninformative in population assignment and add noise to the prescribed models, consequently decreasing the overall performance of the model (Benestan et al. 2015). Our assignment accuracy using leave-one-out cross-validation to site (89.7%) and region (94.3%) were comparable to other studies using RADseq data for assignment (Ruegg et al. 2014, Benestan et al. 2015). Ruegg et al. (2014) assigned Wilson's warblers (*Cardellina pusilla*) from across their breeding range to six genetically distinct groups with an average of 88.5% accuracy across groups (range: 80% - 100%). Benestan et al. (2015) correctly assigned American lobsters (*Homarus americanus*) back to sampling site with 80.8% accuracy (range: 56.6% - 95.6%) and to two genetically distinct north and south regions at 93.6% and 94.8%, respectively (Benestan et al. 2015).

The magnitude of genetic differentiation between our sites did not have an effect on assignment accuracy and we successfully assigned individuals to sites that lack significant genetic differentiation from other sites. In turn, assignment accuracy was strongly driven by the sample size of the site being assigned to and our results indicate a need of at least 13 samples per site for assignment accuracy greater than 90%. Our results are similar to other studies using

ddRADseq that have shown at least 34 samples per site may be needed to reach maximum assignment accuracy in populations with weaker genetic differentiation (overall F_{ST} = 0.00011, Benestan et al. 2015) and six to eight samples were sufficient for population genetic diversity estimates in *Violaceae* (*Amphirrhox longifolia*, overall F_{ST} = 0.076, Nazareno et al. 2017). While there is no prescriptive formula for the minimum sample size needed for a study, these results may serve as guidelines for sampling design when considered with general estimates of genetic differentiation, power analyses, and the resolution needed for the specific study.

Weak migratory connectivity

Using 600 genome-wide SNPs, we provide strong evidence of weak migratory connectivity in the core wintering range of the prothonotary warbler. Altogether, the nonbreeding individuals were predominantly assigned to MRV (76%) and this proportion was similarly reflected at each nonbreeding sampling site. Overall, the proportions of assigned individuals to region (MRV and ATL) are comparable to the relative abundance of prothonotary warblers in these two regions on the breeding grounds, signifying the mixing of individuals on the wintering ground. Our nonbreeding sites are concentrated within the center of the nonbreeding range, where we might expect more mixing to occur than on the periphery of the range (Finch et al. 2017). Additional sampling of sites in Central America and Venezuela would provide a more comprehensive depiction of range-wide migratory connectivity in the prothonotary warbler and more research is needed to determine if our results can be extrapolated away from the core of the breeding range. However, recent nonbreeding survey data of prothonotary warblers suggests highest abundance in the core wintering range (Bulluck et al. in

prep), thus providing support for our weak migratory connectivity results being indicative for a majority of prothonotary warbler populations. Our study complements other recent studies that incorporate RADseq data to disentangle patterns of migration in neotropical migratory birds (Ruegg et al. 2014, Battey et al. 2017), and we show that these methods are suitable for species with fine-scale genetic structure that may have been previously undetectable.

Our documentation of migratory connectivity in the prothonotary warbler corroborates the findings of a recent study that used geolocators deployed across the breeding range (Tonra et al. in prep). Geolocator data highlighted the importance of the Magdalena River Valley in central Colombia as the primary region of mixing for wintering prothonotary warblers, however survey data suggest abundance is higher in the coastal mangroves of northern Colombia (Bulluck et al. in prep). Building upon these studies, we show there is population genetic structure between longitudinally separated regions on the breeding grounds and individuals from these genetically distinct regions are mixing on the wintering grounds. In the context of the nonbreeding survey data (Bulluck et al. in prep), our results suggest prothonotary warblers demonstrate a migratory pattern of high nonbreeding mixing, with low nonbreeding range spread (Finch et al. 2017).

Conservation implications

The habitat specificity of the prothonotary warbler makes the species particularly vulnerable to habitat loss at sites throughout the annual cycle (Petit 1997). Habitat loss for prothonotary warblers is of concern because, despite a tapering off of wetland draining, forested wetlands in the eastern United States continue to decline in abundance (Sucik and Marks 2010)

and mangrove forests, the preferred wintering habitat of the prothonotary warbler (Petit 1997), may be completely lost to sea-level rise in the next 100 years (Duke et al. 2007). Our results highlight the importance of coastal mangrove forests in Panama and northern Colombia to migrating prothonotary warblers from across the breeding range. In conjunction with nonbreeding abundance and geolocator data (Tonra et al. *In revision*), our results reveal a high conservation value to the central Colombian coastal mangrove forests for a wide-range of the breeding population of prothonotary warblers. In light of the differing trends for breeding populations, our findings of low migratory connectivity suggest a need to further explore habitat loss and disturbances on the breeding ground as well as along migratory pathways. Population genetic structure of migratory species can be an indication of a migratory divide between the genetically distinct populations (Webster et al. 2002; Rolshausen et al. 2013). If the prothonotary warbler MRV and ATL populations have separate migration routes across the Caribbean Sea and Gulf Coast, then varied stressors along these pathways may drive the different breeding population trends. Further, regional variation in phenology, behavior, and morphology have not been formally assessed in this species and minor genetic variation may have evolutionary and conservation significance.

Conclusion

In this study, we demonstrated the utility of using high-resolution SNP markers for revealing weak genetic structure and provide a baseline for future migratory connectivity studies involving neotropical migratory bird species of the eastern United States. Despite weak genetic differentiation between populations, individuals can be accurately assigned to sites using

relatively few samples (< 20 individuals) and a moderate number of genetic markers (< 1000 SNPs). Using similar techniques, studies may begin to describe population genetic structure that was previously undocumented, allowing us to infer the migratory patterns of an increasing number of species. Our identification, and corroboration (Tonra et al. in prep), of weak migratory connectivity allows us to work toward a full annual cycle model for prothonotary warblers, a crucial step toward the conservation and management of migratory species (Marra et al. 2015). Future studies should continue to expand our knowledge (e.g. stopover site locations and duration) of the full annual cycle of prothonotary warblers and incorporate genomic data into exploring the resiliency of prothonotary warbler populations in regards to climate change and continuing population declines.

Literature Cited

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: Standard methods are upwardly biased. Mol Ecol Resour 10:701–710. doi: 10.1111/j.1755-0998.2010.02846.x

Avise, JC RB (1992) Mitochondrial DNA Phylogeographic Differentiation among Avian Populations and the Evolutionary Significance of Subspecies. Auk 109:626–636

Baguette M, Blanchet S, Legrand D, et al (2013) Individual dispersal, landscape connectivity and ecological networks. Biol Rev 88:310–326. doi: 10.1111/brv.12000

Battey CJ, Linck EB, Epperly KL, et al (2017) A Migratory Divide in the Painted Bunting (Passerina ciris). Am Nat 191:000–000. doi: 10.1086/695439

Bay RA, Harrigan RJ, Underwood VL, et al (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. Science (359) 6371: 83-86. doi: 10.1126/science.aan4380

Bélisle M, Desrochers A, Fortin M (2013) Influence of Forest Cover on the Movements of Forest Birds : A Homing Experiment. 82:1893–1904

Benestan L, Gosselin T, Perrier C, et al (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). Mol Ecol 24:3299–3315. doi: 10.1111/mec.13245

Catchen JM (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi: 10.1111/mec.12354.Stacks

Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study. Mol Ecol Notes 7:747–756. doi: 10.1111/j.1471-8286.2007.01769.x

Clegg SM, Kelly JF, Kimura M, Smith TB (2003) Combining genetic markers and stable isotopes to reveal population connectivity and migration patterns in a neotropical migrant, Wilson's warbler (Wilsonia pusilla). Mol Ecol 12:819–830. doi: 10.1046/j.1365-294X.2003.01757.x

Colbeck GJ, Gibbs HL, Marra PP, et al (2008) Phylogeography of a widespread North American migratory songbird (Setophaga ruticilla). J Hered 99:453–463. doi: 10.1093/jhered/esn025

Crandall KA, Bininda-Emonds ORR, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295. doi: 10.1016/S0169-5347(00)01876-0

Crochet PA (2000) Genetic structure of avian populations - Allozymes revisited. Mol Ecol 9:1463–1469. doi: 10.1046/j.1365-294X.2000.01026.x

Deane PE, McCoy KD, Robertson RJ, et al (2013) Minimal Genetic Structure in the Cerulean Warbler Despite Evidence for Ecological Differentiation among Populations. Condor 115:178–185. doi: 10.1525/cond.2013.120031

Duke NC, Meynecke J-O, Dittmann S, et al (2007) A World Without Mangroves? Science (80-) 317:41b–42b. doi: 10.1126/science.317.5834.41b

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol Ecol 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Faaborg J, Holmes RT, Anders a. D, et al (2010) Conserving migratory landbirds in the New World: Do we know enough? Ecol Appl 20:398–418. doi: 10.1890/09-0397.1

Finch T, Butler S, Franco A, Cresswell W (2017) Low migratory connectivity is common in long-distance migrant birds. J Anim Ecol 38:42–49. doi: 10.1111/1365-2656.12635

Fraser KC, Stutchbury BJM, Silverio C, et al (2012) Continent-wide tracking to determine migratory connectivity and tropical habitat associations of a declining aerial insectivore. Proc R Soc B Biol Sci 279:4901–4906. doi: 10.1098/rspb.2012.2207

Gill FB., Mostrom AM., Mack AL (1993) Speciation in North American Chickadees : I . Patterns of mtDNA Genetic Divergence. Evolution (N Y) 47:195–212

Gosselin, T, Anderson, E. C., Bradbury, I. (2016). assigner: Assignment Analysis with GBS/RAD Data using R. R package version 0.4.1. https://github.com/thierrygosselin/assigner. doi : 10.5281/zenodo.51453

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 2:184–186. doi: 10.1111/j.1471-8278

Gutierréz-Corchero F, Arruga M V, Sanz L, et al (2002) Using FTA (R) cards to store avian blood samples for genetic studies. Their application in sex determination. Mol Ecol Notes 2:75–77. doi: 10.1046/j.1471-8278.2001.00110.x

Haché S, Bayne EM, Villard MA, et al (2017) Phylogeography of a migratory songbird across its Canadian breeding range: Implications for conservation units. Ecol Evol 7:6078–6088. doi: 10.1002/ece3.3170

Haig SM, Beever E a, Chambers SM, et al (2006) Taxonomic Considerations in Listing Subspecies Under the U . S . Endangered Species Act. Conserv Biol 20:1584–1594. doi: 10.1111/j.1523-1739.2006.00530.x

Herr CA, Sykes PW, Klicka J (2011) Phylogeography of a vanishing North American songbird : the Painted Bunting (Passerina ciris). Conserv Genet 1395–1410. doi: 10.1007/s10592-011-0237-6

Haig SM, Mehlman DW, Oring LW (1998) Avian moviments and wetland connectivity in landscape conservation. Conserv Biol 12:749–758

Hostetler JA, Sillett TS, Marra PP (2015) Full-annual-cycle population models for migratory birds. Auk 132:433–449. doi: 10.1642/AUK-14-211.1

Irwin DE, Irwin JH, Smith TB (2011) Genetic variation and seasonal migratory connectivity in Wilson's warblers (Wilsonia pusilla): Species-level differences in nuclear DNA between western and eastern populations. Mol Ecol 20:3102–3115. doi: 10.1111/j.1365-294X.2011.05159.x

Kimura M, Clegg SM, Lovette IJ, et al (2002) Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic-Neotropical warbler (Wilsonia pusilla). Mol Ecol 11:1605–1616. doi: 10.1046/j.1365-294X.2002.01551.x

Klein NK, Brown WM (1994) Intraspecific Molecular Phylogeny in the Yellow Warbler (Dendroica petechia), and Implications for Avian Biogeography in the West Indies. Evolution (N Y) 48:1914. doi: 10.2307/2410517

Kramer GR, Andersen DE, Buehler DA, et al (2018) Population trends in Vermivora warblers are linked to strong migratory connectivity. Proc Natl Acad Sci 201718985. doi: 10.1073/pnas.1718985115

Lindsay DL, Barr KR, Lance RF, et al (2008) Habitat fragmentation and genetic diversity of an endangered, migratory songbird, the golden-cheeked warbler (Dendroica chrysoparia). Mol Ecol 17:2122–2133. doi: 10.1111/j.1365-294X.2008.03673.x

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. Trends Ecol Evol 26:448–456. doi: 10.1016/j.tree.2011.05.007

Marra P, Cohen EB, Loss SR, et al (2015) A call for full annual cycle research in animal ecology. Biol Lett 11:2015.0552. doi: 10.1098/rsbl.2015.0552

Marra PP, Hobson KA, Holmes RT, et al (1998) Linking Winter and Summer Events in a Migratory Bird by Using Stable-Carbon Isotopes. Science (80-) 282:1884–1886

Marra PP, Norris DR, Haig SM, et al (2006) Migratory Connectivity. Conserv Biol Ser - Cambridge 14:455–461. doi: 10.1017/CBO9780511754821.008

Moritz C (1994) Defining "Evolutionarily Significant Units." Trends Ecol Evol 9:373–375. doi: 10.1016/0169-5347(94)90057-4

Nazareno AG, Bemmels JB, Dick CW, Lohmann LG (2017) Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol Ecol Resour 17:1136–1147. doi: 10.1111/1755-0998.12654

Norris DR, Wunder MB, Boulet M (2006) Introduction : The Past and Present of Migratory Connectivity Author (s): Marylène Boulet and D . Ryan Norris Source : Ornithological Monographs , No . 61 , Patterns of Migratory Connectivity in Two Nearctic-Neotropical Songbirds : New Insights from Int. Ornithol Monogr 79–88

Oyler-McCance SJ, Oh KP, Langin KM, Aldridge CL (2016) A field ornithologist's guide to genomics: Practical considerations for ecology and conservation. http://dx.doi.org/101642/AUK-16-491 133:626–648. doi: 10.1642/AUK-16-49.1

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol. Ecol. 4:347–354

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE, 7, e37135. https://doi.org/10.1371/journal.pone. 0037135

Parchman TL, Gompert Z, Mudge J, et al (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol Ecol 21:2991–3005. doi: 10.1111/j.1365-294X.2012.05513.x

Partners in Flight Science Committee Population Estimates Database. (2013). Retrieved April 01, 2017, from http://pif.birdconservancy.org/PopEstimates/Database.aspx#continental

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:2074–2093. doi: 10.1371/journal.pgen.0020190

Petit LJ (1999) Prothonotary Warbler (Protonotaria citrea). Birds North Am Online 26818421:. doi: 10.2173/bna.408

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. doi: 10.1111/j.1471-8286.2007.01758.x

Pulich, W .M. 1976. The Golden-cheeked Warbler: A bioecological study. Texas Parks and Wildlife Department, Austin, Texas.

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431. doi: 10.7717/peerj.431

Rolshausen G, Segelbacher G, Hermes C, et al (2013) Individual differences in migratory behavior shape population genetic structure and microhabitat choice in sympatric blackcaps (Sylvia atricapilla). Ecol Evol 3:4278–4289. doi: 10.1002/ece3.825

Rosenberg K V., Kennedy JA, Dettmers R, et al (2016) Partners in Flight Landbird Conservation Plan: 2016 Revision for Canada and Continental United States. 119 pp

Ruegg KC, Anderson EC, Paxton KL, et al (2014) Mapping migration in a songbird using high-resolution genetic markers. Mol Ecol 23:5726–5739. doi: 10.1111/mec.12977

Rushing CS, Ryder TB, Marra PPv (2016) Quantifying drivers of population dynamics for a migratory bird throughout the annual cycle. Proc R Soc B Biol Sci 283:20152846. doi: 10.1098/rspb.2015.2846

Sauer JR, Niven DK, Hines JE, et al (2017) The North American Breeding Bird Survey, Results and Analysis 1966 - 2015. Laurel, MD

Sherry TW, Holmes RT (1996) Winter habitat quality, population limitation, and conservation of neotropical nearctic migrant birds. Ecology 77:36–48. doi: 10.2307/2265652

Sherry TW (2018) Identifying migratory birds' population bottlenecks in time and space. Proc Natl Acad Sci 201802174. doi: 10.1073/pnas.1802174115

Sillett TS, Holmes RT, Sherry TW (2000) Impacts of a global climate cycle on population dynamics of a migratory songbird. Science (80-) 288:2040–2042. doi: 10.1126/science.288.5473.2040

Soltis DE, Morris AB, McLachlan JS, et al (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293. doi: 10.1111/j.1365-294X.2006.03061.x

Sucik MT, Marks E (2015) The status and recent trends of wetlands in the United States. U.S. Department of Agriculture. Summary Report: 2012 National Resources Inventory, Natural Resources Conservation Service, Washington, D.C.

Taylor CM, Stutchbury BJM (2016) Effects of breeding versus winter habitat loss and fragmentation on the population dynamics of a migratory songbird. Ecol Appl 26:424–437. doi: 10.1890/14-1410

Toews DPL, Brelsford A, Grossen C, et al (2016) Genomic variation across the Yellow-rumped Warbler species complex. Auk 133:698–717. doi: 10.1642/AUK-16-61.1

U.S. Fish and Wildlife Service (2008) Birds of Conservation Concern 2008. Report 85. doi: 10.1177/0146167211402094

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–1439. doi: 10.1111/j.1365-294X.2006.02890.x

Webster MS, Marra PP, Haig SM, et al (2002) Links between worlds: Unraveling migratory connectivity. Trends Ecol Evol 17:76–83. doi: 10.1016/S0169-5347(01)02380-1

Webster MS, Marra PP (2004) Birds of Two Worlds—The Ecology and Evolution of Migration. The Johns Hopkins University Press, Baltimore, Maryland.

Winker K, Graves GR, Braun MJ (2000) Genetic Differentiation among Populations of a Migratory Songbird : Limnothlypis swainsonii. J Avian Biol 31:319–328

Appendix A - Tables

Table 1: Sampling site locations of prothonotary warblers. The classification of region is as follows: 'Atlantic' identifies breeding sites in Atlantic Seaboard and 'Mississippi' identifies breeding sites in the Mississippi River Valley. The label 'nonbreeding' is used for Panama and Colombia sites. The last column, n, provides the sample size from each location.

ID	Site	Region	Latitude	Longitude	n
VA1	Fort AP Hill	Atlantic	38.15452	-77.32426	20
VA2	Deep Bottom	Atlantic	37.407349	-77.305381	20
VA3	Great Dismal Swamp	Atlantic	36.631408	-76.490525	13
NC1	Holt Lake	Atlantic	35.469092	-78.403068	9
SC1	Francis Beidler	Atlantic	33.220573	-80.354001	13
OH1	Hoover Reservoir	Mississippi	40.107202	-82.88641	21
LA1	Palmetto Island	Mississippi	29.86441	-92.150251	20
LA2	Bluebonnet Swamp	Mississippi	30.367822	-91.107121	20
LA3	Barataria	Mississippi	29.783791	-90.115468	14
AR1	White River	Mississippi	34.358046	-91.090881	20
WI1	Sugar River	Mississippi	42.530128	-89.32875	8
PN1	Panama Viejo	Nonbreeding	9.006642	-79.484717	7
PN2	Juan Diaz	Nonbreeding	9.019719	-79.44465	17
CO1	Bocas del Atrato	Nonbreeding	8.08918	-76.836956	7
CO2	Marimonda	Nonbreeding	8.56907	-76.81737	29
CO3	Cispata	Nonbreeding	9.39281	-75.78397	27
CO4	Flamencos	Nonbreeding	11.42013	-73.10123	3

Table 2: Results from the analysis of molecular variance. The designation of region was based on the genetic clusters shown in the PCA plot (i.e. MRV and ATL), while sites correspond to all of the breeding sites. All results were significant at $p < 0.002$.

Levels	Variation (%)	p-value
Between region	0.253	0.002
Between sites within region	0.344	0.002
Between individuals within sites	14.76	0.002
Within individuals	84.63	0.001

Table 3: Assignment of individuals from the nonbreeding sites. Values are given for the number of individuals assigned from each nonbreeding site grouping to the corresponding region, and percentage of assigned individuals to that region from a particular grouping is listed in parentheses.

Nonbreeding site groupings	Mississippi River Valley (MRV)	Atlantic Seaboard (ATL)
PN1 & PN2	17 (70.8%)	7 (29.2%)
CO1 & CO2	29 (80.6%)	7 (19.4%)
CO3	19 (70.4%)	8 (29.6%)
CO4	3 (100.0%)	0 (0.0%)

Appendix B - Figures

Figure 1: Range map of prothonotary warbler with all sampling locations shown and labeled by site ID. Site ID uses the first two letters of the state (breeding sites) or country (nonbreeding sites) followed by a number.

Figure 2: Results from population genetic structure analysis using PCA. Genotypes of individuals contained 26,189 SNPs and individual outliers were removed that had principal component values of more than 6 standard deviations on the first ten principal components; individuals are colored by region with blue representing individuals from Mississippi River Valley and yellow representing individuals from the Atlantic Seaboard. The inset map has the sampling locations with the same color scheme.

Figure 3: Population assignment with hold-out model.. In both plots, the y-axis corresponds to the proportion of individuals that were correctly assigned to known origin site (upper plot) or region (lower plot) across 20 bootstrapped training sets in the hold-out validation model. The x-axis is categorically divided by the number of SNPs used to assign hold-out individuals to sites and regions. The trajectory of assignment success followed the same pattern in both plots, with maximum assignment accuracy ranging from 600 to 1200 SNPs.

Figure 4: Population assignment with leave-one-out cross-validation model. Values in the circles are the percentage of assigned individuals from the sampled population (y-axis) to the inferred population (x-axis). Blue circles along the diagonal correspond to individuals that were

correctly assigned by site and red circles represent individuals that were incorrectly assigned to a site. Any circles within one of the gray shaded areas represents individuals that were correctly assigned to region, with correct assignment to Mississippi River Valley being in the top-right and Atlantic Seaboard in the bottom left of the plot. Compared to the hold-out validation tests, the leave-one-out cross-validation assignment resulted in a modest increase in assignment accuracy by site, 89.7%, and an equal value of regional assignment accuracy, 94.3%.

Figure 5: Assignment of individuals from non-breeding sites. Top-left map shows breeding sampling locations with genetic clusters by color: MRV in blue and ATL in yellow. Lower map shows nonbreeding sampling locations and pie charts correspond to groupings of nonbreeding sampling sites with value indicating number of individuals assigned to the MRV or ATL.

Figure 1.

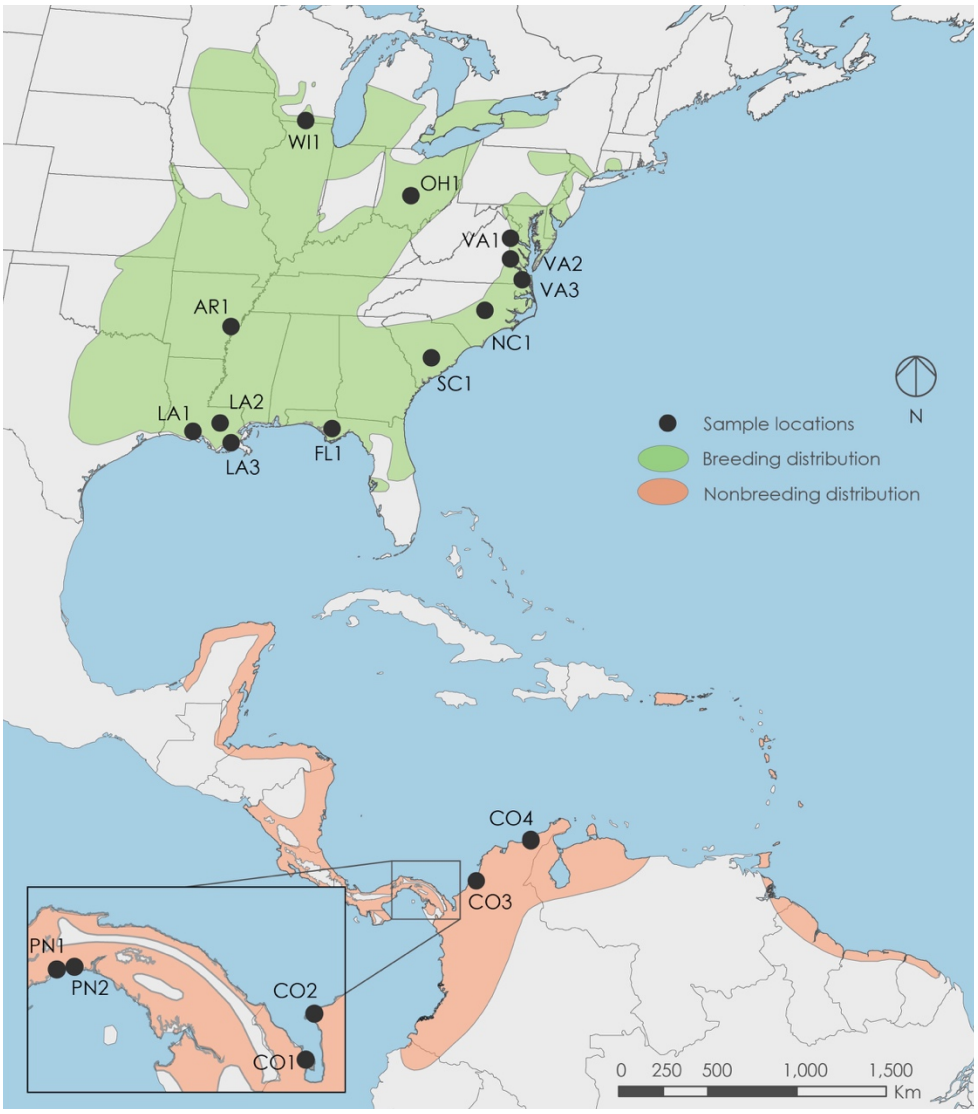

WI1
OH1
VA1
VA2
VA3
AR1
NC1
SC1
LA1
LA2
LA3
FL1
Sample locations
Breeding distribution
Nonbreeding distribution
N
CO4
CO3
PN1
PN2
CO2
CO1
0 250 500 1,000 1,500
Km

Figure 2.

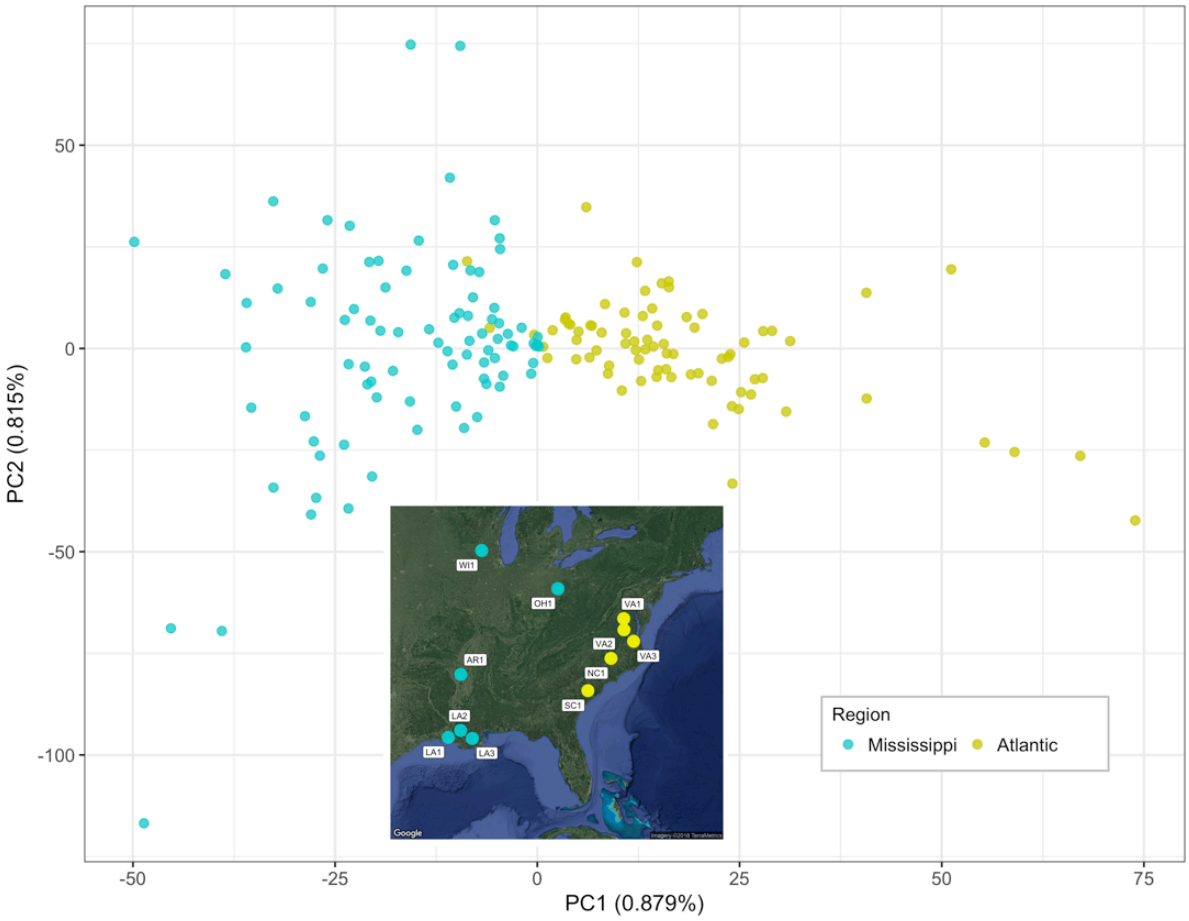

PC2 (0.815%)
50
0
-50
-100
-50
-25
0
25
50
75
PC1 (0.879%)
WI1
OH1
VA1
VA2
VA3
AR1
NC1
SC1
LA2
LA1
LA3
Google
Region
Mississippi
Atlantic

Figure 3.

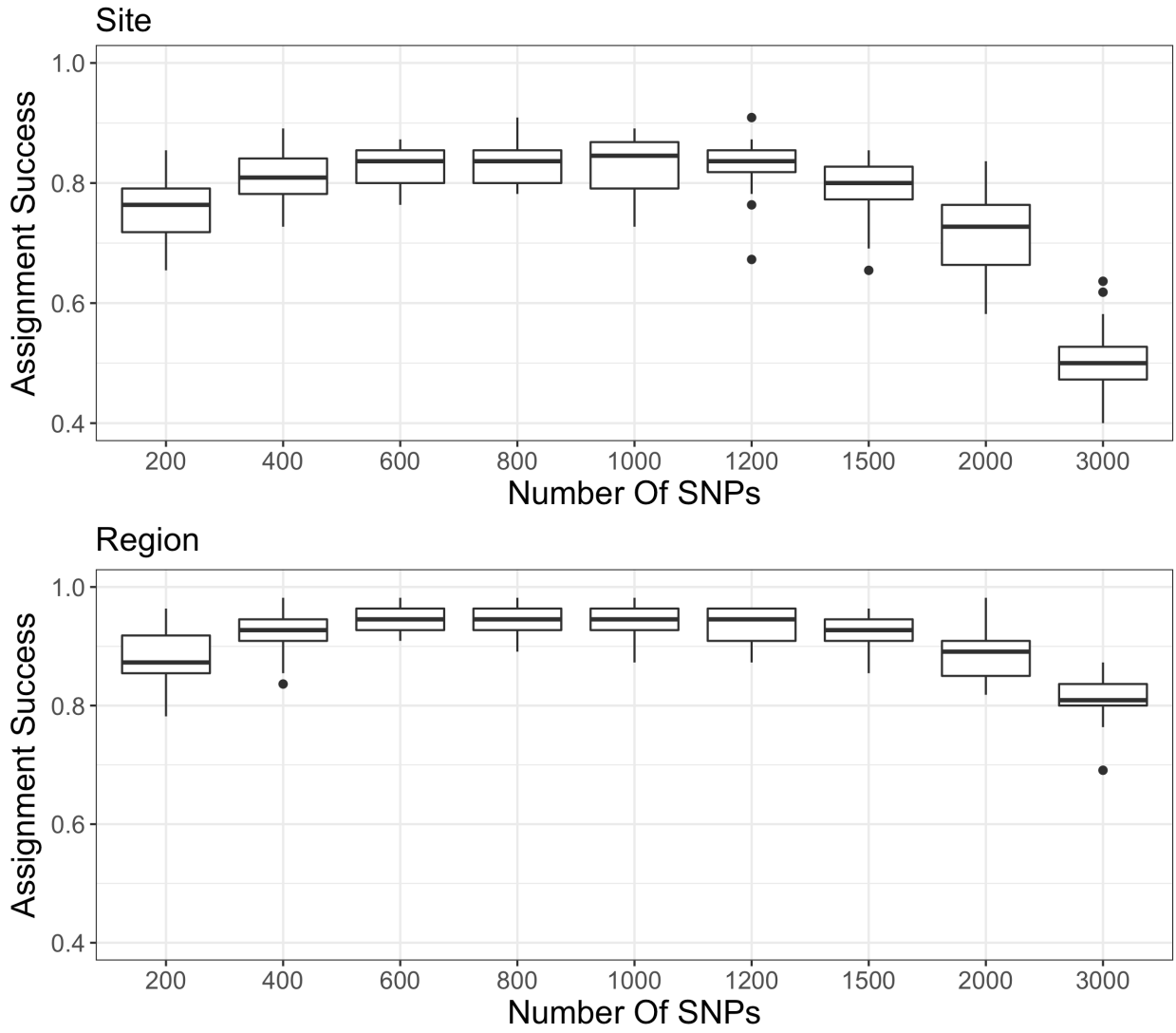

Site
Assignment Success
1.0
0.8
0.6
0.4
200
400
600
800
1000
1200
1500
2000
3000
Number Of SNPs
Region
Assignment Success
1.0
0.8
0.6
0.4
200
400
600
800
1000
1200
1500
2000
3000
Number Of SNPs

Figure 4.

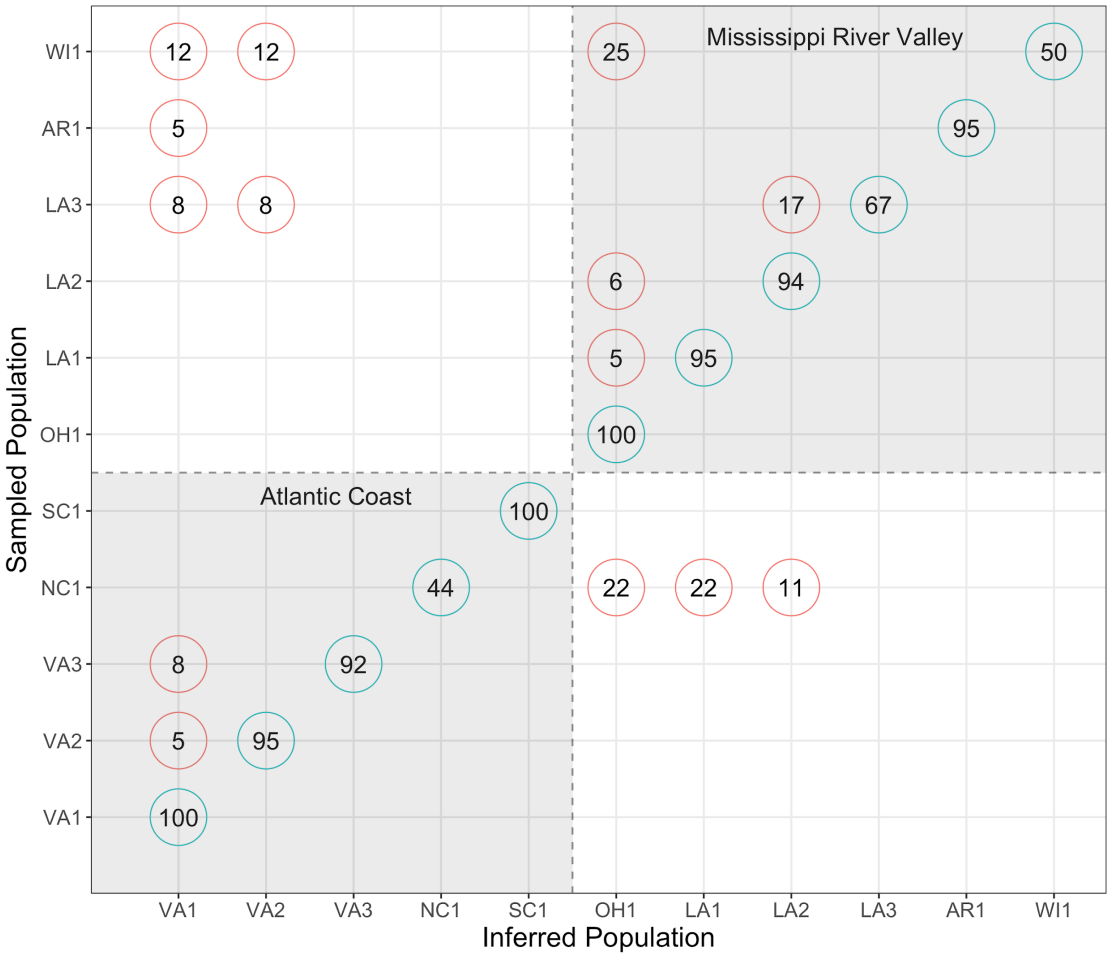
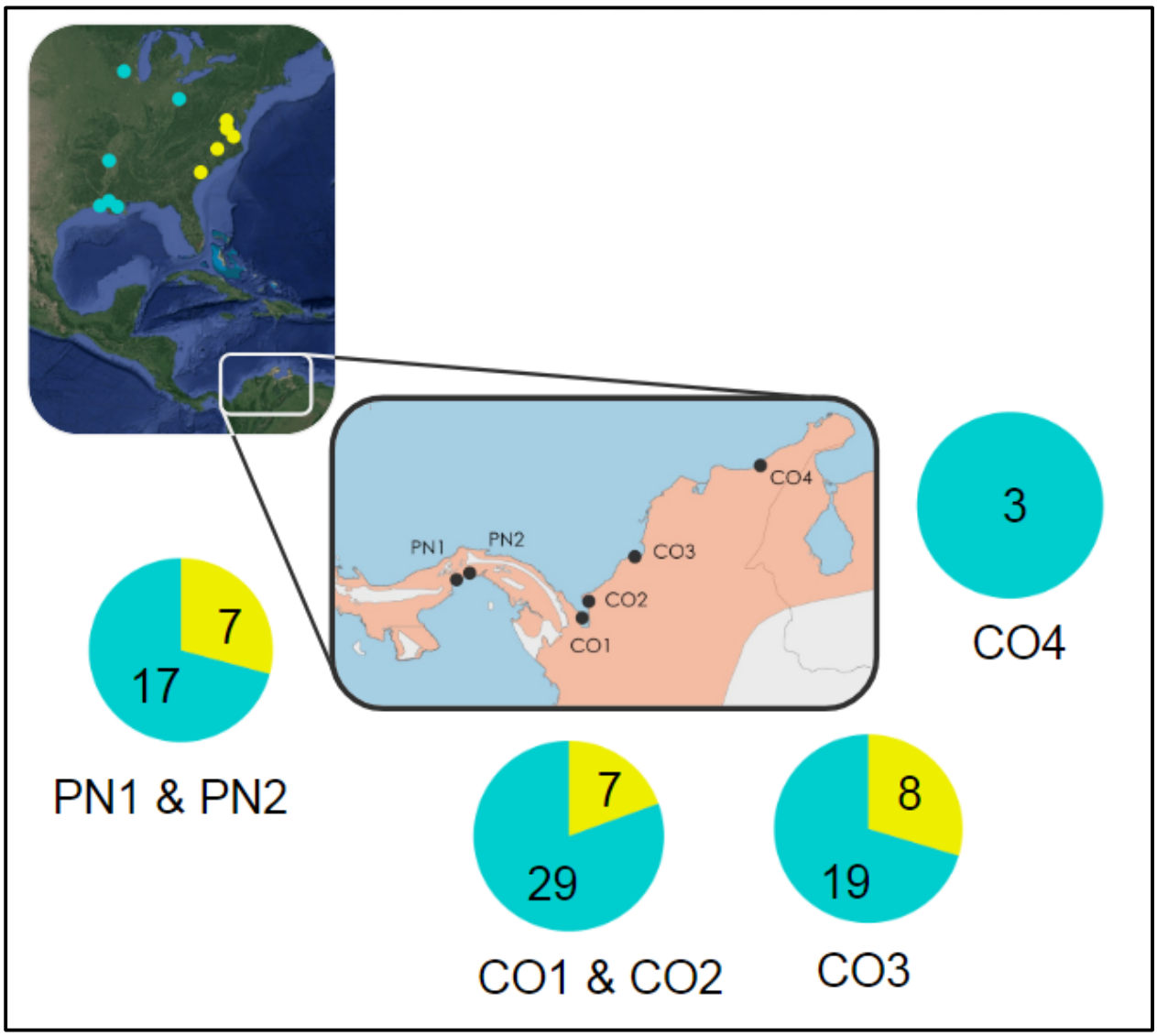

Mississippi River Valley
Atlantic Coast
Sampled Population
Inferred Population
WI1
AR1
LA3
LA2
LA1
OH1
SC1
NC1
VA3
VA2
VA1
12
12
5
8
8
8
5
100
100
44
92
95
25
50
95
17
67
6
94
5
95
100
22
22
11

Figure 5.



Supplemental Material

Table of Contents:

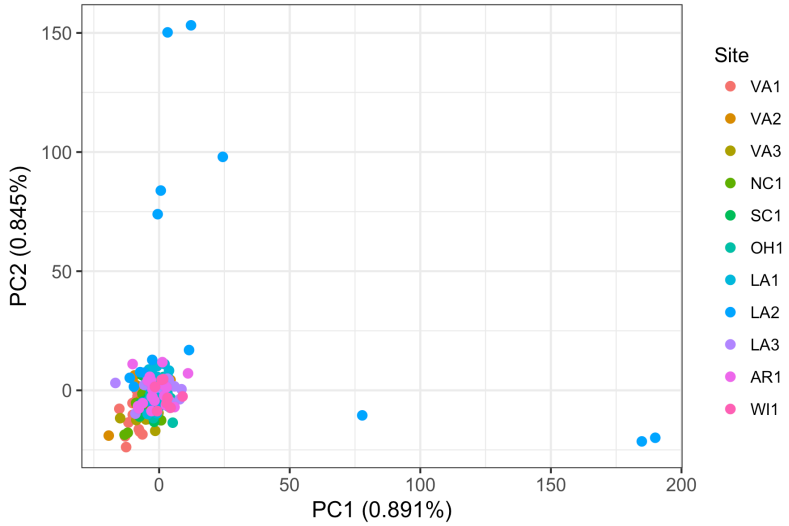
Appendix S1: Bioinformatic Pipeline and SNP filtering

The demultiplexed FASTQ files were processed in the dDocent bioinformatics pipeline (Puritz, Hollenbeck, & Gold, 2014) with minimal filtering thresholds of 3x coverage per loci and presence of loci in at least 10 individuals, which yielded 145260 putative single nucleotide polymorphisms (SNPs). For the resulting variant call format (VCF) files, we used VCFtools (Danecek et al. 2011) to select for biallelic SNPs with less than 50% missing data and remove any indels, which retained 114632 loci. Using a custom python script, we further selected for SNPs that were between -0.5 and 0.5 F_{IS}, greater than 0.01 minor allele frequency, and had a PHRED quality score greater than 20, resulting in 41328 SNPs. We reduced the SNP data set to 1 SNP per RAD tag to avoid any issues with linkage disequilibrium and this resulted in a set of 26189 SNPs that were used as the initial data set for subsequent analyses.

Figure S1: Principal Components Analysis

The plot of the initial PCA with all 26,189 SNPs (A) showed several outlier individuals at greater than 6 standard deviations from the mean of the principal component score across the first 10 principal components. After four iterations of removing outlier individuals and reperforming PCA, 162 individuals remained (B) and no sites had a disproportionate amount of outlier individuals that were removed.

A)



B)

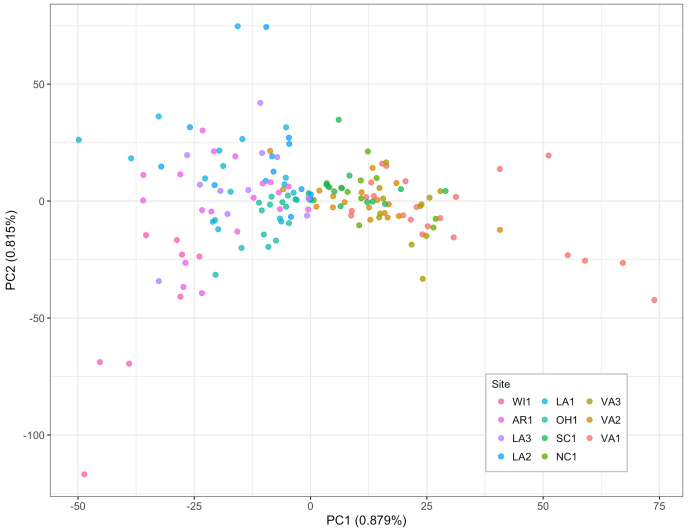


Figure S2: Missing loci data across sites

The sites OH1 (n = 21) and SC1 (n = 13) had a larger median proportion of missing loci data within their individuals and a larger range of missing data than the other sites. PCA was conducted without these two sites but the resulting plots showed the same trends as the original PCA (Figure S1), thus, all 11 breeding sites and 175 individuals were retained for subsequent analyses.

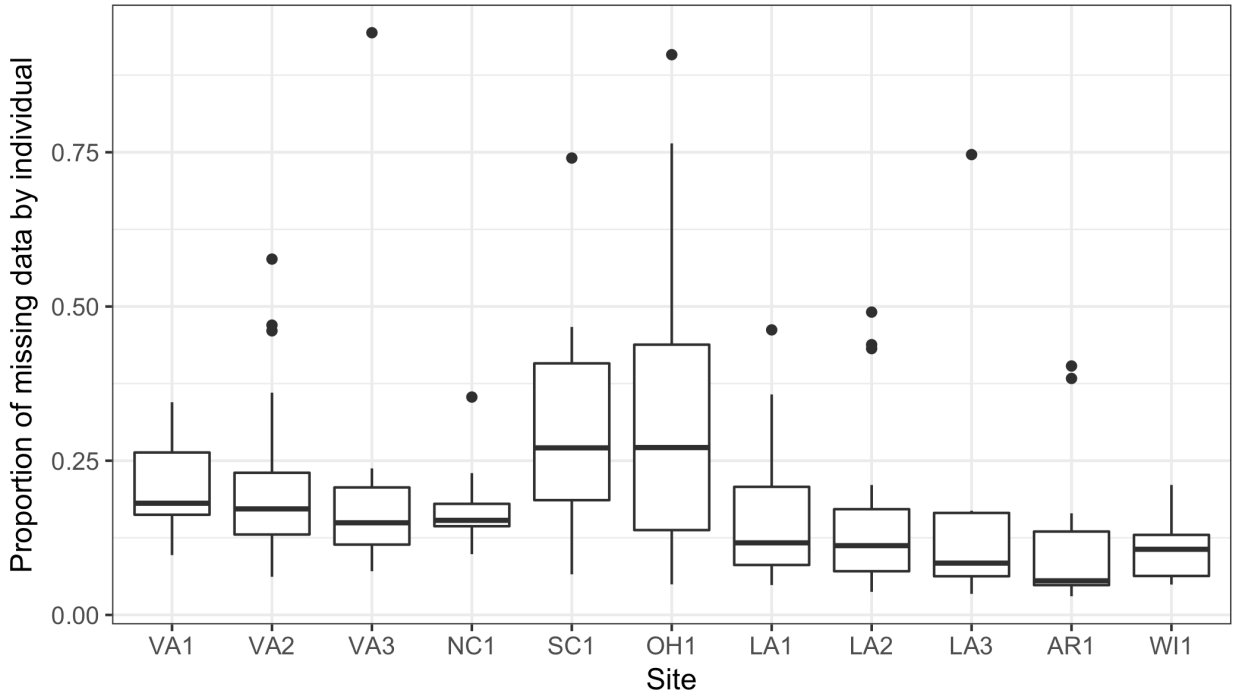


Figure S3: STRUCTURE plots

Plots from Evanno (2015) method using 26,189 SNPs. The mean likelihood of number of K clusters does not provide support for more than one cluster.

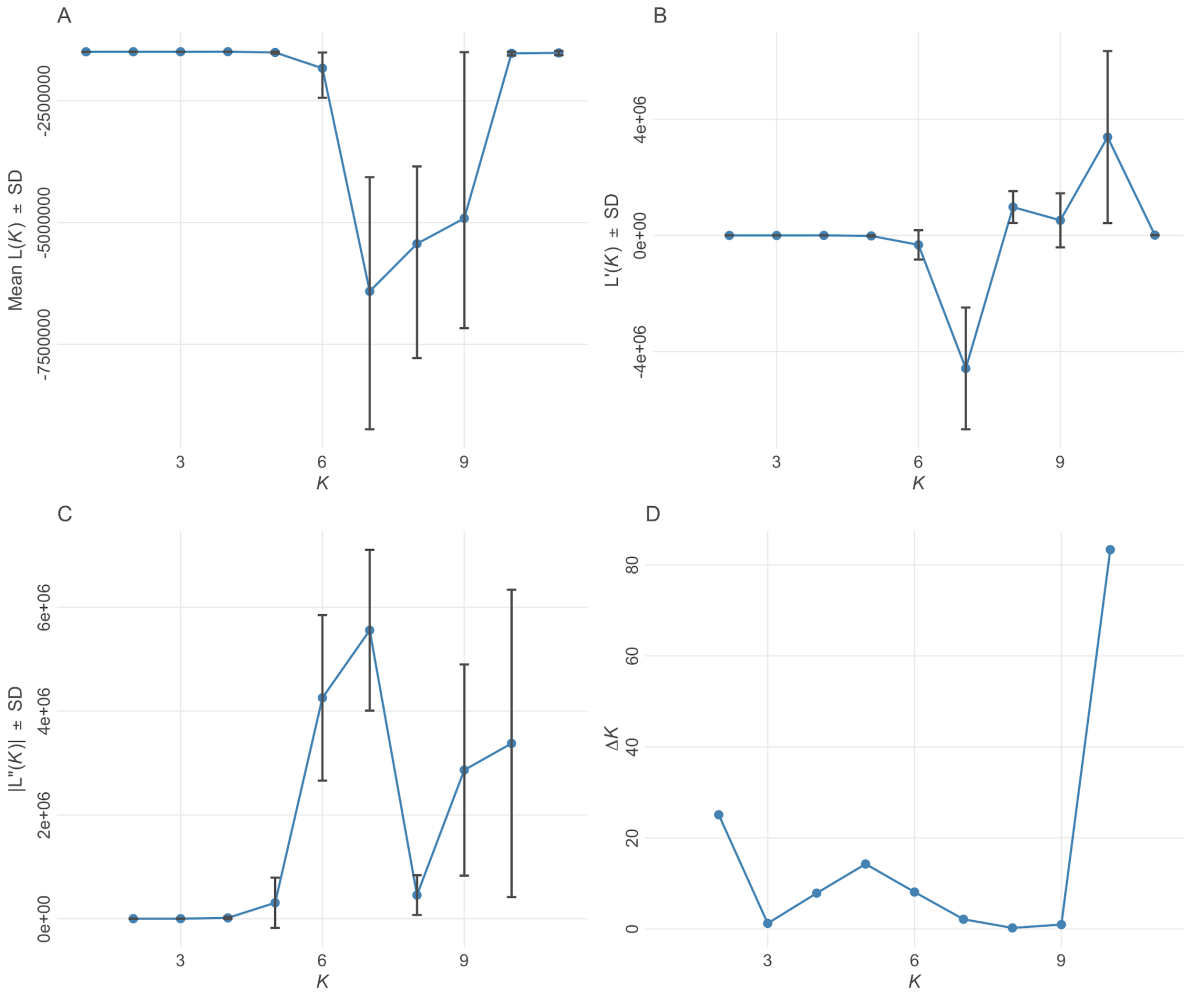


Figure S4: Isolation by distance

A kernel density plot of genetic distance by the geographic distance (log_{10} transformed) revealed a center near the maximum of geographic distance on the x-axis. However, near the maximum values of geographic distance there is a wide-range of genetic distance values. The lower genetic distance values with a large geographic distance represent comparisons of widely separated sites in the MRV that are genetically similar. The high genetic distance values with large geographic distances are the between-site comparisons that cross diagonally between the two regions (e.g. WI1 to SC1).

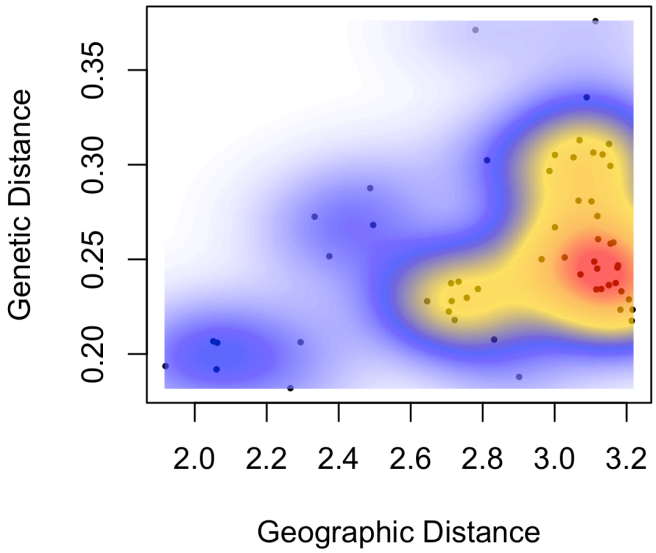


Figure S5: Assignment success by sample size

In the leave-one-out cross-validation model, the assignment success of known origin individuals to breeding site ranged from 44% to 100%. All sites represented by more than 12 samples had a greater than 90% assignment success of their individuals. These data had a significant positive correlation ($r^2 = 0.84$, $p = 0.0007$, $F = 20.84$ on 2 and 8 *DF*).

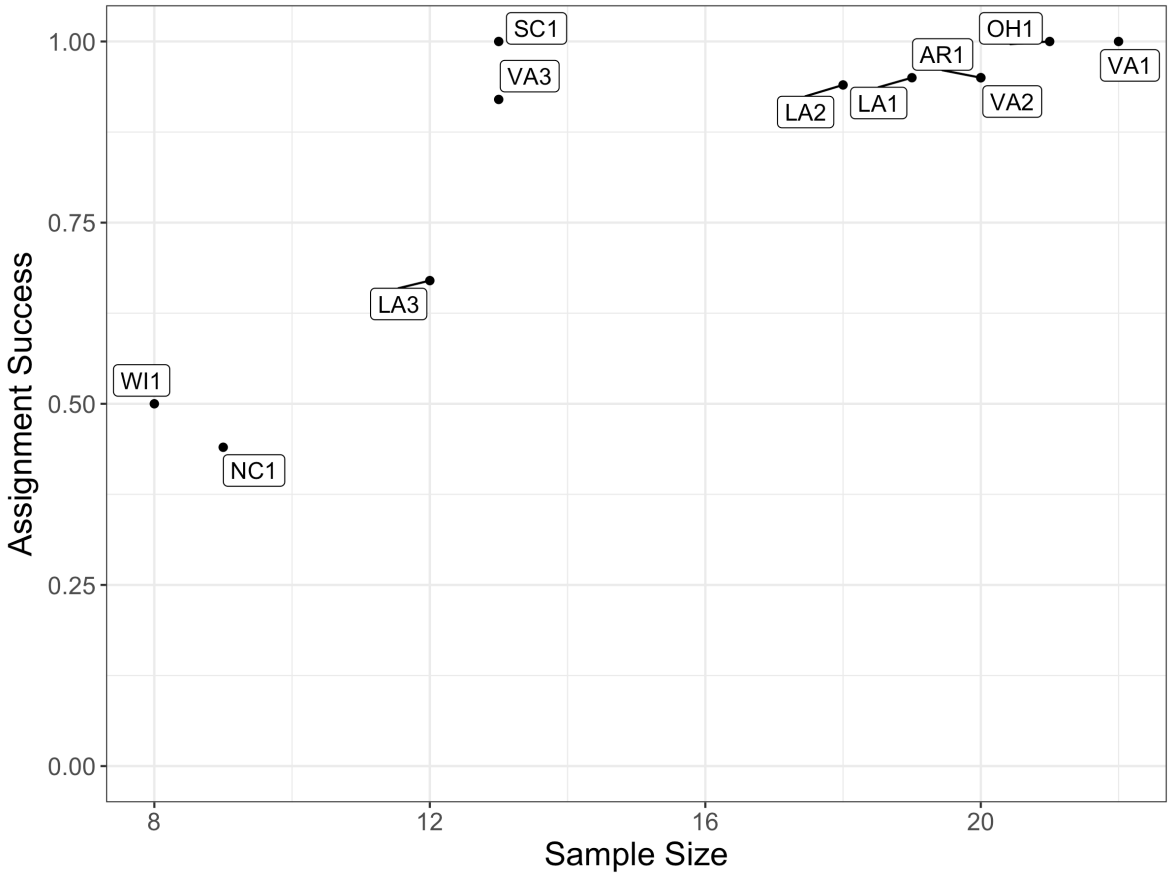


Figure S6: Relative breeding abundance provided by Breeding Bird Survey Data (Sauer et al. 2017).

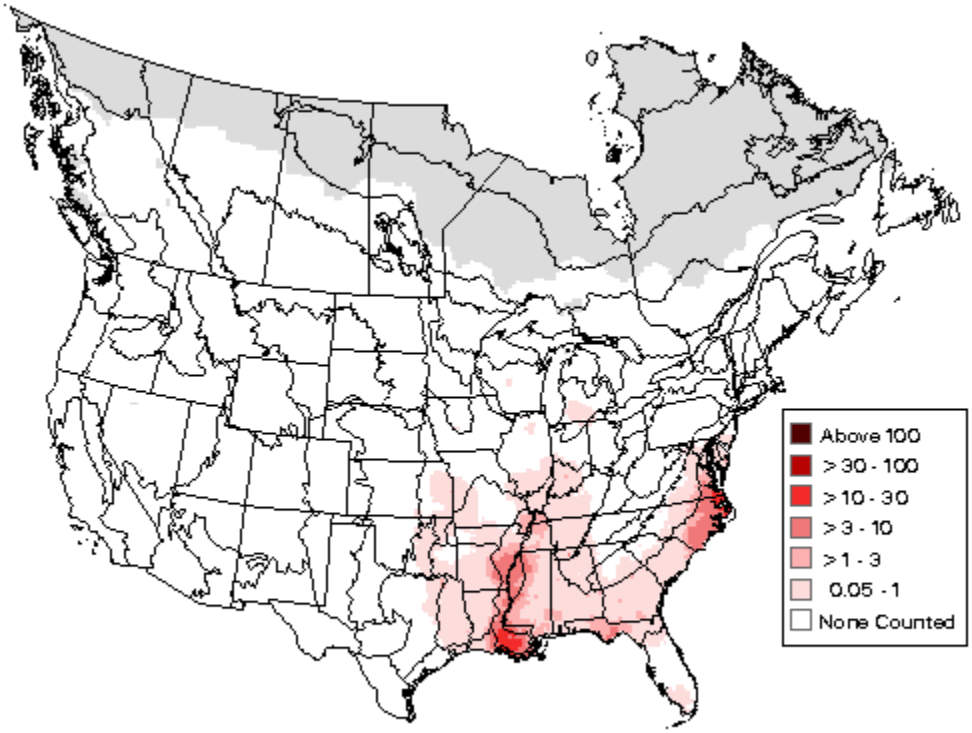


Table S1: Pairwise F_{ST}

Pairwise F_{ST} values from sites across the breeding range. Values below the diagonal are Weir & Cockerham (1984) F_{ST} values calculated from 100 bootstrap iterations with the assigner package (Gosselin et al. 2016). Values above the diagonal are bootstrapped confidence intervals of the 0.025 and 0.975 quantiles.

	VA1	VA2	VA3	NC1	SC1	OH1	LA1	LA2	LA3	AR1	WI1
VA1	-	0.0026 - 0.0046	0.0022 - 0.0049	0.0040 - 0.0086	0.0051 - 0.0095	0.0069 - 0.0097	0.0066 - 0.0089	0.0097 - 0.0118	0.0058 - 0.0086	0.0064 - 0.0086	0.0077 - 0.0120
VA2	0.0035	-	0 - 0.0006	0.0016 - 0.0054	0.0014 - 0.0050	0.00399 - 0.0064	0.0021- 0.0043	0.0064 - 0.0083	0.0013 - 0.0036	0.0024 - 0.0042	0.0033 - 0.0073
VA3	0.0036	0	-	0 - 0.0048	0.0007 - 0.0049	0.00393 - 0.00681	0.00354 - 0.00666	0.0066 - 0.0095	0.0021 - 0.0054	0.00303 - 0.0056	0.0046 - 0.0089
NC1	0.0062	0.0033	0.0017	-	0 - 0.0045	0.0028 - 0.0068	0.0041 - 0.0080	0.0075 - 0.0110	0.0044 - 0.0089	0.0056 - 0.0090	0.0086 - 0.0136
SC1	0.0074	0.0033	0.0031	0.0018	-	0.0049 - 0.0089	0.0051 - 0.0086	0.0071 - 0.0111	0.0037 - 0.0077	0.0062 - 0.0096	0.0069 - 0.0120
OH1	0.0085	0.0051	0.0053	0.00515	0.0071	-	0.0034 - 0.0058	0.0066 - 0.0091	0.0010 - 0.0043	0.0028 - 0.0052	0.0001 - 0.0041
LA1	0.0077	0.0033	0.0050	0.0060	0.0066	0.0046	-	0.0047 - 0.0070	0.0008 - 0.0032	0.0018 - 0.0037	0.0024 - 0.0058
LA2	0.0107	0.0074	0.0079	0.0091	0.0092	0.0078	0.0059	-	0.0037 - 0.006	0.0048 - 0.0066	0.0069 - 0.0110
LA3	0.0071	0.0023	0.0039	0.0066	0.0059	0.0078	0.0020	0.0050	-	0.0007 - 0.0029	0.0035 - 0.0077
AR1	0.0074	0.0032	0.0045	0.0073	0.0081	0.0026	0.0027	0.0057	0.0018	-	0.00356 - 0.0072
WI1	0.0101	0.0052	0.0067	0.0107	0.0093	0.0040	0.0039	0.0089	0.0055	0.0053	-

Table S2: Assignment Likelihood

Results from the leave-one-out cross-validation model. Likelihood of assignment to each region is listed as the product of of multilocus genotype probabilities. The numerator of the likelihood ratio is the likelihood of the assigned region.

Individual	Nonbreeding Site	Assigned Region	Likelihood to ATL	Likelihood to MRV	Likelihood Ratio	Log Likelihood Ratio
PN1_01	PN1	ATL	1.47E-170	2.28E-175	64400	4.81
PN1_02	PN1	MRV	2.66E-171	1.28E-161	4.81E+09	9.68
PN1_03	PN1	MRV	8.25E-166	2.54E-155	3.07E+10	10.5
PN1_04	PN1	MRV	3.58E-192	5.42E-175	1.51E+17	17.2
PN1_05	PN1	MRV	1.87E-179	1.21E-162	6.48E+16	16.8
PN1_06	PN1	MRV	1.92E-187	1.32E-182	68900	4.84
PN1_07	PN1	MRV	1.48E-179	1.55E-164	1.05E+15	15
PN2_01	PN2	ATL	3.39E-184	4.90E-187	690	2.84
PN2_02	PN2	ATL	3.15E-173	1.50E-177	21000	4.32
PN2_03	PN2	MRV	4.93E-159	1.32E-158	2.69	0.429
PN2_04	PN2	ATL	1.55E-161	1.93E-163	80.4	1.91
PN2_05	PN2	ATL	8.17E-190	1.19E-191	68.6	1.84
PN2_06	PN2	MRV	2.47E-159	5.43E-156	2200	3.34
PN2_07	PN2	MRV	7.56E-171	1.73E-161	2.28E+09	9.36
PN2_08	PN2	ATL	7.40E-135	4.83E-135	1.53	0.186
PN2_09	PN2	MRV	1.16E-162	2.02E-146	1.74E+16	16.2
PN2_10	PN2	MRV	2.77E-148	6.85E-146	247	2.39
PN2_11	PN2	MRV	2.57E-155	7.98E-151	31000	4.49
PN2_12	PN2	ATL	2.47E-94	1.70E-94	1.46	0.164
PN2_13	PN2	MRV	2.36E-178	6.57E-178	2.79	0.445
PN2_14	PN2	MRV	2.16E-174	3.64E-162	1.69E+12	12.2
PN2_15	PN2	MRV	4.01E-168	8.11E-156	2.02E+12	12.3
PN2_16	PN2	MRV	2.04E-178	2.43E-165	1.19E+13	13.1
PN2_17	PN2	MRV	1.43E-182	1.45E-164	1.01E+18	18
CO3_01	CO3	MRV	2.77E-165	2.02E-160	72900	4.86
CO3_02	CO3	ATL	1.52E-123	1.61E-125	94	1.97
CO3_03	CO3	ATL	8.44E-150	6.82E-158	1.24E+08	8.09
CO3_04	CO3	ATL	5.65E-123	1.52E-127	37200	4.57
CO3_05	CO3	ATL	4.83E-103	9.05E-104	5.34	0.728

CO3_06	CO3	MRV	3.05E-99	1.74E-98	5.71	0.757
CO3_07	CO3	ATL	4.17E-141	4.03E-144	1030	3.01
CO3_08	CO3	MRV	8.96E-113	2.07E-110	231	2.36
CO3_09	CO3	MRV	2.86E-118	1.83E-104	6.37E+13	13.8
CO3_10	CO3	MRV	4.24E-105	2.72E-97	6.40E+07	7.81
CO3_11	CO3	MRV	8.14E-127	2.89E-111	3.55E+15	15.6
CO3_12	CO3	MRV	5.98E-141	5.15E-130	8.61E+10	10.9
CO3_13	CO3	MRV	1.74E-171	1.99E-169	114	2.06
CO3_14	CO3	MRV	1.38E-171	1.90E-152	1.37E+19	19.1
CO3_15	CO3	MRV	1.75E-172	3.10E-163	1.77E+09	9.25
CO3_16	CO3	MRV	1.20E-170	3.04E-158	2.54E+12	12.4
CO3_17	CO3	ATL	1.57E-131	3.09E-133	50.8	1.71
CO3_18	CO3	ATL	1.04E-95	6.24E-105	1.67E+09	9.22
CO3_19	CO3	ATL	7.25E-48	1.29E-52	56200	4.75
CO3_20	CO3	MRV	1.09E-116	1.54E-114	141	2.15
CO3_21	CO3	MRV	9.58E-145	2.59E-134	2.71E+10	10.4
CO3_22	CO3	MRV	1.45E-92	3.04E-79	2.09E+13	13.3
CO3_23	CO3	MRV	7.68E-132	1.70E-131	2.21	0.344
CO3_24	CO3	MRV	7.51E-162	6.12E-157	81400	4.91
CO3_25	CO3	MRV	3.11E-173	6.46E-166	20700000	7.32
CO3_26	CO3	MRV	4.65E-155	4.03E-152	867	2.94
CO3_27	CO3	MRV	4.41E-152	5.51E-141	1.25E+11	11.1
CO2_01	CO2	MRV	1.22E-167	8.14E-163	66800	4.82
CO2_02	CO2	MRV	5.76E-162	5.46E-152	9.48E+09	9.98
CO2_03	CO2	MRV	2.79E-175	1.64E-144	5.86E+30	30.8
CO2_04	CO2	MRV	5.37E-132	1.30E-123	2.42E+08	8.38
CO2_05	CO2	MRV	7.44E-170	5.26E-153	7.07E+16	16.8
CO2_06	CO2	MRV	2.28E-182	1.01E-166	4.44E+15	15.6
CO2_07	CO2	MRV	1.56E-157	1.19E-143	7.60E+13	13.9
CO2_08	CO2	MRV	3.73E-198	1.91E-191	5120000	6.71
CO2_09	CO2	MRV	3.43E-102	2.80E-96	816000	5.91
CO2_10	CO2	ATL	4.44E-100	3.80E-103	1170	3.07
CO2_11	CO2	MRV	1.82E-147	1.27E-133	6.97E+13	13.8
CO2_12	CO2	ATL	0.0129	0.00121	10.6	1.03
CO2_13	CO2	MRV	1.82E-142	6.65E-132	3.65E+10	10.6
CO2_14	CO2	MRV	3.88E-117	9.34E-115	241	2.38
CO2_15	CO2	MRV	1.23E-154	9.01E-143	7.30E+11	11.9
CO2_16	CO2	MRV	7.85E-113	5.46E-109	6950	3.84
CO2_17	CO2	MRV	1.85E-161	3.17E-160	17.1	1.23
CO2_18	CO2	MRV	1.45E-132	5.75E-130	395	2.6
CO2_19	CO2	ATL	3.28E-83	1.89E-83	1.74	0.24

CO2_20	CO2	MRV	2.66E-111	3.47E-105	1310000	6.12
CO2_21	CO2	MRV	9.49E-132	1.44E-124	15200000	7.18
CO2_22	CO2	MRV	0.663	0.85	1.28	0.108
CO2_23	CO2	ATL	2.17E-159	2.56E-161	85	1.93
CO2_24	CO2	ATL	6.97E-157	3.57E-157	1.95	0.291
CO2_25	CO2	MRV	1.70E-144	1.06E-141	623	2.79
CO2_26	CO2	MRV	7.76E-119	5.39E-115	6950	3.84
CO2_27	CO2	MRV	2.91E-89	3.78E-87	130	2.11
CO2_28	CO2	ATL	2.76E-84	1.67E-89	166000	5.22
CO2_29	CO2	MRV	6.18E-191	2.45E-180	3.96E+10	10.6
CO1_01	CO1	MRV	4.12E-104	8.49E-93	2.06E+11	11.3
CO1_02	CO1	MRV	2.51E-143	8.00E-142	31.8	1.5
CO1_03	CO1	MRV	3.09E-155	1.44E-143	4.67E+11	11.7
CO1_04	CO1	ATL	1.37E-121	1.45E-125	9460	3.98
CO1_05	CO1	MRV	1.85E-151	1.68E-149	90.9	1.96
CO1_06	CO1	MRV	9.72E-165	6.15E-161	6330	3.8
CO1_07	CO1	MRV	1.99E-162	1.16E-140	5.81E+21	21.8
CO4_01	CO4	MRV	1.75E-162	5.86E-154	3.35E+08	8.53
CO4_02	CO4	MRV	3.49E-152	8.17E-147	234000	5.37
CO4_03	CO4	MRV	4.99E-171	5.96E-156	1.19E+15	15.1

Vita

Matthew George DeSaix graduated with a B.S. in Environmental Studies and a concentration in Sustainable Agriculture from Warren Wilson College in 2012. Afterward, he spent several years gallivanting/conducting avian field research across the Americas, in places such as New Mexico, Peru, California, Hawai'i, and Ecuador, for organizations and universities. His avian research projects focused on breeding demography and tropical avian ecology. In 2015, Matthew had a brief hiatus from field work and assisted Dr. Luke Tornabene in the Laboratory of Analytical Biology at the Smithsonian Natural History Museum to conduct DNA barcoding of dwarf-gobies. During his graduate studies, he traveled twice to Panama to teach the Panama Avian Field Ecology course, and he taught Introductory Biology Laboratory and Ornithology Laboratory courses as well. While at VCU, he also completed the Graduate Certificate in Geographic Information Systems through the L. Douglas Wilder School of Government and Public Affairs at VCU.