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

Matthew G. DeSaix
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd
Part of the Integrative Biology Commons, Ornithology Commons, and the Other Ecology and Evolutionary Biology Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/5432

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.
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........................................................................................................ii

Abstract..........................................................................................................................iii

Introduction.....................................................................................................................1

Methods...........................................................................................................................4

Results.............................................................................................................................9

Discussion.........................................................................................................................13

Literature Cited................................................................................................................20

Appendices.......................................................................................................................27

A. Tables..........................................................................................................................27

B. Figures..........................................................................................................................30

C. Supplemental Material...............................................................................................37

Vita.......................................................................................................................................49
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-Corcherro 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 ($F_{ST}/(1 - F_{ST}) \sim \log(\text{Euclidean Distance})$) 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

*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 *Violacea* (*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


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.

<table>
<thead>
<tr>
<th>ID</th>
<th>Site</th>
<th>Region</th>
<th>Latitude</th>
<th>Longitude</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA1</td>
<td>Fort AP Hill</td>
<td>Atlantic</td>
<td>38.15452</td>
<td>-77.32426</td>
<td>20</td>
</tr>
<tr>
<td>VA2</td>
<td>Deep Bottom</td>
<td>Atlantic</td>
<td>37.407349</td>
<td>-77.305381</td>
<td>20</td>
</tr>
<tr>
<td>VA3</td>
<td>Great Dismal Swamp</td>
<td>Atlantic</td>
<td>36.631408</td>
<td>-76.490525</td>
<td>13</td>
</tr>
<tr>
<td>NC1</td>
<td>Holt Lake</td>
<td>Atlantic</td>
<td>35.469092</td>
<td>-78.403068</td>
<td>9</td>
</tr>
<tr>
<td>SC1</td>
<td>Francis Beidler</td>
<td>Atlantic</td>
<td>33.220573</td>
<td>-80.354001</td>
<td>13</td>
</tr>
<tr>
<td>OH1</td>
<td>Hoover Reservoir</td>
<td>Mississippi</td>
<td>40.107202</td>
<td>-82.88641</td>
<td>21</td>
</tr>
<tr>
<td>LA1</td>
<td>Palmetto Island</td>
<td>Mississippi</td>
<td>29.86441</td>
<td>-92.150251</td>
<td>20</td>
</tr>
<tr>
<td>LA2</td>
<td>Bluebonnet Swamp</td>
<td>Mississippi</td>
<td>30.367822</td>
<td>-91.107121</td>
<td>20</td>
</tr>
<tr>
<td>LA3</td>
<td>Barataria</td>
<td>Mississippi</td>
<td>29.783791</td>
<td>-90.115468</td>
<td>14</td>
</tr>
<tr>
<td>AR1</td>
<td>White River</td>
<td>Mississippi</td>
<td>34.358046</td>
<td>-91.090881</td>
<td>20</td>
</tr>
<tr>
<td>WI1</td>
<td>Sugar River</td>
<td>Mississippi</td>
<td>42.530128</td>
<td>-89.32875</td>
<td>8</td>
</tr>
<tr>
<td>PN1</td>
<td>Panama Viejo</td>
<td>Nonbreeding</td>
<td>9.006642</td>
<td>-79.484717</td>
<td>7</td>
</tr>
<tr>
<td>PN2</td>
<td>Juan Diaz</td>
<td>Nonbreeding</td>
<td>9.019719</td>
<td>-79.44465</td>
<td>17</td>
</tr>
<tr>
<td>CO1</td>
<td>Bocas del Atrato</td>
<td>Nonbreeding</td>
<td>8.08918</td>
<td>-76.836956</td>
<td>7</td>
</tr>
<tr>
<td>CO2</td>
<td>Marimonda</td>
<td>Nonbreeding</td>
<td>8.56907</td>
<td>-76.81737</td>
<td>29</td>
</tr>
<tr>
<td>CO3</td>
<td>Cispata</td>
<td>Nonbreeding</td>
<td>9.39281</td>
<td>-75.78397</td>
<td>27</td>
</tr>
<tr>
<td>CO4</td>
<td>Flamencos</td>
<td>Nonbreeding</td>
<td>11.42013</td>
<td>-73.10123</td>
<td>3</td>
</tr>
</tbody>
</table>
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 $p < 0.002$.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Variation (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between region</td>
<td>0.253</td>
<td>0.002</td>
</tr>
<tr>
<td>Between sites within region</td>
<td>0.344</td>
<td>0.002</td>
</tr>
<tr>
<td>Between individuals within sites</td>
<td>14.76</td>
<td>0.002</td>
</tr>
<tr>
<td>Within individuals</td>
<td>84.63</td>
<td>0.001</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Nonbreeding site groupings</th>
<th>Mississippi River Valley (MRV)</th>
<th>Atlantic Seaboard (ATL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN1 &amp; PN2</td>
<td>17 (70.8%)</td>
<td>7 (29.2%)</td>
</tr>
<tr>
<td>CO1 &amp; CO2</td>
<td>29 (80.6%)</td>
<td>7 (19.4%)</td>
</tr>
<tr>
<td>CO3</td>
<td>19 (70.4%)</td>
<td>8 (29.6%)</td>
</tr>
<tr>
<td>CO4</td>
<td>3 (100.0%)</td>
<td>0 (0.0%)</td>
</tr>
</tbody>
</table>
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 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.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Supplemental Material

Table of Contents:

Appendix S1) Bioinformatic pipeline and SNP filtering

Figure S1) Supplementary PCAs
Figure S2) Missing data by site
Figure S3) STRUCTURE results
Figure S4) Mantel results
Figure S5) Assignment success by sample size
Figure S6) Breeding Bird Survey relative abundance

Table S1) Pairwise $F_{ST}$
Table S2) Assignment Likelihoods
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) 

![Principal Component Analysis Plot](image1)

B) 

![Principal Component Analysis Plot](image2)
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).
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.

<table>
<thead>
<tr>
<th></th>
<th>VA1</th>
<th>VA2</th>
<th>VA3</th>
<th>NC1</th>
<th>SC1</th>
<th>OH1</th>
<th>LA1</th>
<th>LA2</th>
<th>LA3</th>
<th>AR1</th>
<th>WI1</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA1</td>
<td>-</td>
<td>0.0026</td>
<td>0.0022</td>
<td>0.0040</td>
<td>0.0051</td>
<td>0.0069</td>
<td>0.0066</td>
<td>0.0097</td>
<td>0.0058</td>
<td>0.0064</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>0.0046</td>
<td>0.0049</td>
<td>0.0086</td>
<td>0.0095</td>
<td>0.0097</td>
<td>0.0118</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0120</td>
</tr>
<tr>
<td>VA2</td>
<td>0.0035</td>
<td>-</td>
<td>0 - 0.0006</td>
<td>0.0014</td>
<td>0.00399</td>
<td>0.0064</td>
<td>0.0021</td>
<td>0.0064</td>
<td>0.0113</td>
<td>0.0024</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>0.0054</td>
<td>-</td>
<td>0.0050</td>
<td>-</td>
<td>0.0083</td>
<td>-</td>
<td>0.0042</td>
<td>0.0073</td>
</tr>
<tr>
<td>VA3</td>
<td>0.0036</td>
<td>0</td>
<td>-</td>
<td>0 - 0.0048</td>
<td>0.0007</td>
<td>0.00393</td>
<td>0.00666</td>
<td>0.0066</td>
<td>0.0021</td>
<td>0.00303</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NC1</td>
<td>0.0062</td>
<td>0.0033</td>
<td>0.0017</td>
<td>-</td>
<td>0 - 0.0045</td>
<td>0.0028</td>
<td>0.0041</td>
<td>0.0075</td>
<td>0.0044</td>
<td>0.0056</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SC1</td>
<td>0.0074</td>
<td>0.0033</td>
<td>0.0031</td>
<td>0.0018</td>
<td>-</td>
<td>0.0049</td>
<td>0.0051</td>
<td>0.0049</td>
<td>0.0071</td>
<td>0.0062</td>
<td>0.0069</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>0.0085</td>
<td>0.0086</td>
<td>0.0086</td>
<td>0.0077</td>
<td>0.0096</td>
<td>0.0120</td>
</tr>
<tr>
<td>OH1</td>
<td>0.0085</td>
<td>0.0051</td>
<td>0.0053</td>
<td>0.00515</td>
<td>0.0071</td>
<td>0.0034</td>
<td>0.0034</td>
<td>0.0066</td>
<td>0.0010</td>
<td>0.0028</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0058</td>
<td>0.0091</td>
<td>0.0058</td>
<td>0.0010</td>
<td>0.0052</td>
<td>0.0041</td>
</tr>
<tr>
<td>LA1</td>
<td>0.0077</td>
<td>0.0033</td>
<td>0.0050</td>
<td>0.0060</td>
<td>0.0066</td>
<td>0.0046</td>
<td>-</td>
<td>0.0047</td>
<td>0.0008</td>
<td>0.0018</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>0.0091</td>
<td>0.0043</td>
<td>0.0037</td>
<td>0.0058</td>
</tr>
<tr>
<td>LA2</td>
<td>0.0107</td>
<td>0.0074</td>
<td>0.0079</td>
<td>0.0091</td>
<td>0.0092</td>
<td>0.0078</td>
<td>0.0059</td>
<td>-</td>
<td>0.0037</td>
<td>0.0048</td>
<td>0.0069</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>0.0070</td>
<td>0.0006</td>
<td>0.0066</td>
<td>0.0110</td>
</tr>
<tr>
<td>LA3</td>
<td>0.0071</td>
<td>0.0023</td>
<td>0.0039</td>
<td>0.0066</td>
<td>0.0059</td>
<td>0.0078</td>
<td>0.0020</td>
<td>0.0050</td>
<td>-</td>
<td>0.0007</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>0.0029</td>
<td>0.0035</td>
<td>0.0077</td>
</tr>
<tr>
<td>AR1</td>
<td>0.0074</td>
<td>0.0032</td>
<td>0.0045</td>
<td>0.0073</td>
<td>0.0081</td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0057</td>
<td>0.0018</td>
<td>0.0035</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0029</td>
<td>0.0072</td>
</tr>
<tr>
<td>WI1</td>
<td>0.0101</td>
<td>0.0052</td>
<td>0.0067</td>
<td>0.0107</td>
<td>0.0093</td>
<td>0.0040</td>
<td>0.0039</td>
<td>0.0089</td>
<td>0.0055</td>
<td>0.0053</td>
<td>-</td>
</tr>
</tbody>
</table>
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 multilocus genotype probabilities. The numerator of the likelihood ratio is the likelihood of the assigned region.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Nonbreeding Site</th>
<th>Assigned Region</th>
<th>Likelihood to ATL</th>
<th>Likelihood to MRV</th>
<th>Likelihood Ratio</th>
<th>Log Likelihood Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PN1_01</td>
<td>PN1</td>
<td>ATL</td>
<td>1.47E-170</td>
<td>2.28E-175</td>
<td>64400</td>
<td>4.81</td>
</tr>
<tr>
<td>PN1_02</td>
<td>PN1</td>
<td>MRV</td>
<td>2.66E-171</td>
<td>1.28E-161</td>
<td>4.81E+09</td>
<td>9.68</td>
</tr>
<tr>
<td>PN1_03</td>
<td>PN1</td>
<td>MRV</td>
<td>8.25E-166</td>
<td>2.54E-155</td>
<td>3.07E+10</td>
<td>10.5</td>
</tr>
<tr>
<td>PN1_04</td>
<td>PN1</td>
<td>MRV</td>
<td>3.58E-192</td>
<td>5.42E-175</td>
<td>1.51E+17</td>
<td>17.2</td>
</tr>
<tr>
<td>PN1_05</td>
<td>PN1</td>
<td>MRV</td>
<td>1.87E-179</td>
<td>1.21E-162</td>
<td>6.48E+16</td>
<td>16.8</td>
</tr>
<tr>
<td>PN1_06</td>
<td>PN1</td>
<td>MRV</td>
<td>1.92E-187</td>
<td>1.32E-182</td>
<td>68900</td>
<td>4.84</td>
</tr>
<tr>
<td>PN1_07</td>
<td>PN1</td>
<td>MRV</td>
<td>1.48E-179</td>
<td>1.55E-164</td>
<td>1.05E+15</td>
<td>15</td>
</tr>
<tr>
<td>PN2_01</td>
<td>PN2</td>
<td>ATL</td>
<td>3.39E-184</td>
<td>4.90E-187</td>
<td>690</td>
<td>2.84</td>
</tr>
<tr>
<td>PN2_02</td>
<td>PN2</td>
<td>ATL</td>
<td>3.15E-173</td>
<td>1.50E-177</td>
<td>21000</td>
<td>4.32</td>
</tr>
<tr>
<td>PN2_03</td>
<td>PN2</td>
<td>MRV</td>
<td>4.93E-159</td>
<td>1.32E-158</td>
<td>2.69</td>
<td>0.429</td>
</tr>
<tr>
<td>PN2_04</td>
<td>PN2</td>
<td>ATL</td>
<td>1.55E-161</td>
<td>1.93E-163</td>
<td>80.4</td>
<td>1.91</td>
</tr>
<tr>
<td>PN2_05</td>
<td>PN2</td>
<td>ATL</td>
<td>8.17E-190</td>
<td>1.19E-191</td>
<td>68.6</td>
<td>1.84</td>
</tr>
<tr>
<td>PN2_06</td>
<td>PN2</td>
<td>MRV</td>
<td>2.47E-159</td>
<td>5.43E-156</td>
<td>2200</td>
<td>3.34</td>
</tr>
<tr>
<td>PN2_07</td>
<td>PN2</td>
<td>MRV</td>
<td>7.56E-171</td>
<td>1.73E-161</td>
<td>2.28E+09</td>
<td>9.36</td>
</tr>
<tr>
<td>PN2_08</td>
<td>PN2</td>
<td>ATL</td>
<td>7.40E-135</td>
<td>4.83E-135</td>
<td>1.53</td>
<td>0.186</td>
</tr>
<tr>
<td>PN2_09</td>
<td>PN2</td>
<td>MRV</td>
<td>1.16E-162</td>
<td>2.02E-146</td>
<td>1.74E+16</td>
<td>16.2</td>
</tr>
<tr>
<td>PN2_10</td>
<td>PN2</td>
<td>MRV</td>
<td>2.77E-148</td>
<td>6.85E-146</td>
<td>247</td>
<td>2.39</td>
</tr>
<tr>
<td>PN2_11</td>
<td>PN2</td>
<td>MRV</td>
<td>2.57E-155</td>
<td>7.98E-151</td>
<td>31000</td>
<td>4.49</td>
</tr>
<tr>
<td>PN2_12</td>
<td>PN2</td>
<td>ATL</td>
<td>2.47E-94</td>
<td>1.70E-94</td>
<td>1.46</td>
<td>0.164</td>
</tr>
<tr>
<td>PN2_13</td>
<td>PN2</td>
<td>MRV</td>
<td>2.36E-178</td>
<td>6.57E-178</td>
<td>2.79</td>
<td>0.445</td>
</tr>
<tr>
<td>PN2_14</td>
<td>PN2</td>
<td>MRV</td>
<td>2.16E-174</td>
<td>3.64E-162</td>
<td>1.69E+12</td>
<td>12.2</td>
</tr>
<tr>
<td>PN2_15</td>
<td>PN2</td>
<td>MRV</td>
<td>4.01E-168</td>
<td>8.11E-156</td>
<td>2.02E+12</td>
<td>12.3</td>
</tr>
<tr>
<td>PN2_16</td>
<td>PN2</td>
<td>MRV</td>
<td>2.04E-178</td>
<td>2.43E-165</td>
<td>1.19E+13</td>
<td>13.1</td>
</tr>
<tr>
<td>PN2_17</td>
<td>PN2</td>
<td>MRV</td>
<td>1.43E-182</td>
<td>1.45E-164</td>
<td>1.01E+18</td>
<td>18</td>
</tr>
<tr>
<td>CO3_01</td>
<td>CO3</td>
<td>MRV</td>
<td>2.77E-165</td>
<td>2.02E-160</td>
<td>72900</td>
<td>4.86</td>
</tr>
<tr>
<td>CO3_02</td>
<td>CO3</td>
<td>ATL</td>
<td>1.52E-123</td>
<td>1.61E-125</td>
<td>94</td>
<td>1.97</td>
</tr>
<tr>
<td>CO3_03</td>
<td>CO3</td>
<td>ATL</td>
<td>8.44E-150</td>
<td>6.82E-158</td>
<td>1.24E+08</td>
<td>8.09</td>
</tr>
<tr>
<td>CO3_04</td>
<td>CO3</td>
<td>ATL</td>
<td>5.65E-123</td>
<td>1.52E-127</td>
<td>37200</td>
<td>4.57</td>
</tr>
<tr>
<td>CO3_05</td>
<td>CO3</td>
<td>ATL</td>
<td>4.83E-103</td>
<td>9.05E-104</td>
<td>5.34</td>
<td>0.728</td>
</tr>
<tr>
<td>CO3_06</td>
<td>CO3</td>
<td>MRV</td>
<td>3.05E-99</td>
<td>1.74E-98</td>
<td>5.71</td>
<td>0.757</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>CO3_07</td>
<td>CO3</td>
<td>ATL</td>
<td>4.17E-141</td>
<td>4.03E-144</td>
<td>1030</td>
<td>3.01</td>
</tr>
<tr>
<td>CO3_08</td>
<td>CO3</td>
<td>MRV</td>
<td>8.96E-113</td>
<td>2.07E-110</td>
<td>231</td>
<td>2.36</td>
</tr>
<tr>
<td>CO3_09</td>
<td>CO3</td>
<td>MRV</td>
<td>2.86E-118</td>
<td>1.83E-104</td>
<td>6.37E+13</td>
<td>13.8</td>
</tr>
<tr>
<td>CO3_10</td>
<td>CO3</td>
<td>MRV</td>
<td>4.24E-105</td>
<td>2.72E-97</td>
<td>6.40E+07</td>
<td>7.81</td>
</tr>
<tr>
<td>CO3_11</td>
<td>CO3</td>
<td>MRV</td>
<td>8.14E-127</td>
<td>2.89E-111</td>
<td>3.55E+15</td>
<td>15.6</td>
</tr>
<tr>
<td>CO3_12</td>
<td>CO3</td>
<td>MRV</td>
<td>5.98E-141</td>
<td>5.15E-130</td>
<td>8.61E+10</td>
<td>10.9</td>
</tr>
<tr>
<td>CO3_13</td>
<td>CO3</td>
<td>MRV</td>
<td>1.74E-171</td>
<td>1.99E-169</td>
<td>114</td>
<td>2.06</td>
</tr>
<tr>
<td>CO3_14</td>
<td>CO3</td>
<td>MRV</td>
<td>1.38E-171</td>
<td>1.90E-152</td>
<td>1.37E+19</td>
<td>19.1</td>
</tr>
<tr>
<td>CO3_15</td>
<td>CO3</td>
<td>MRV</td>
<td>1.75E-172</td>
<td>3.10E-163</td>
<td>1.77E+09</td>
<td>9.25</td>
</tr>
<tr>
<td>CO3_16</td>
<td>CO3</td>
<td>MRV</td>
<td>1.20E-170</td>
<td>3.04E-158</td>
<td>2.54E+12</td>
<td>12.4</td>
</tr>
<tr>
<td>CO3_17</td>
<td>CO3</td>
<td>ATL</td>
<td>1.57E-131</td>
<td>3.09E-133</td>
<td>50.8</td>
<td>1.71</td>
</tr>
<tr>
<td>CO3_18</td>
<td>CO3</td>
<td>ATL</td>
<td>1.04E-95</td>
<td>6.24E-105</td>
<td>1.67E+09</td>
<td>9.22</td>
</tr>
<tr>
<td>CO3_19</td>
<td>CO3</td>
<td>ATL</td>
<td>7.25E-48</td>
<td>1.29E-52</td>
<td>562000</td>
<td>4.75</td>
</tr>
<tr>
<td>CO3_20</td>
<td>CO3</td>
<td>MRV</td>
<td>1.09E-116</td>
<td>1.54E-114</td>
<td>141</td>
<td>2.15</td>
</tr>
<tr>
<td>CO3_21</td>
<td>CO3</td>
<td>MRV</td>
<td>9.58E-145</td>
<td>2.59E-134</td>
<td>2.71E+10</td>
<td>10.4</td>
</tr>
<tr>
<td>CO3_22</td>
<td>CO3</td>
<td>MRV</td>
<td>1.45E-92</td>
<td>3.04E-79</td>
<td>2.09E+13</td>
<td>13.3</td>
</tr>
<tr>
<td>CO3_23</td>
<td>CO3</td>
<td>MRV</td>
<td>7.68E-132</td>
<td>1.70E-131</td>
<td>2.21</td>
<td>0.344</td>
</tr>
<tr>
<td>CO3_24</td>
<td>CO3</td>
<td>MRV</td>
<td>7.51E-162</td>
<td>6.12E-157</td>
<td>81400</td>
<td>4.91</td>
</tr>
<tr>
<td>CO3_25</td>
<td>CO3</td>
<td>MRV</td>
<td>3.11E-173</td>
<td>6.46E-166</td>
<td>2070000</td>
<td>7.32</td>
</tr>
<tr>
<td>CO3_26</td>
<td>CO3</td>
<td>MRV</td>
<td>4.65E-155</td>
<td>4.03E-152</td>
<td>867</td>
<td>2.94</td>
</tr>
<tr>
<td>CO3_27</td>
<td>CO3</td>
<td>MRV</td>
<td>4.41E-152</td>
<td>5.51E-141</td>
<td>1.25E+11</td>
<td>11.1</td>
</tr>
<tr>
<td>CO2_01</td>
<td>CO2</td>
<td>MRV</td>
<td>1.22E-167</td>
<td>8.14E-163</td>
<td>66800</td>
<td>4.82</td>
</tr>
<tr>
<td>CO2_02</td>
<td>CO2</td>
<td>MRV</td>
<td>5.76E-162</td>
<td>5.46E-152</td>
<td>9.48E+09</td>
<td>9.98</td>
</tr>
<tr>
<td>CO2_03</td>
<td>CO2</td>
<td>MRV</td>
<td>2.79E-175</td>
<td>1.64E-144</td>
<td>5.86E+30</td>
<td>30.8</td>
</tr>
<tr>
<td>CO2_04</td>
<td>CO2</td>
<td>MRV</td>
<td>5.37E-132</td>
<td>1.30E-123</td>
<td>2.42E+08</td>
<td>8.38</td>
</tr>
<tr>
<td>CO2_05</td>
<td>CO2</td>
<td>MRV</td>
<td>7.44E-170</td>
<td>5.26E-153</td>
<td>7.07E+16</td>
<td>16.8</td>
</tr>
<tr>
<td>CO2_06</td>
<td>CO2</td>
<td>MRV</td>
<td>2.28E-182</td>
<td>1.01E-166</td>
<td>4.44E+15</td>
<td>15.6</td>
</tr>
<tr>
<td>CO2_07</td>
<td>CO2</td>
<td>MRV</td>
<td>1.56E-157</td>
<td>1.19E-143</td>
<td>7.60E+13</td>
<td>13.9</td>
</tr>
<tr>
<td>CO2_08</td>
<td>CO2</td>
<td>MRV</td>
<td>3.73E-198</td>
<td>1.91E-191</td>
<td>5120000</td>
<td>6.71</td>
</tr>
<tr>
<td>CO2_09</td>
<td>CO2</td>
<td>MRV</td>
<td>3.43E-102</td>
<td>2.80E-96</td>
<td>816000</td>
<td>5.91</td>
</tr>
<tr>
<td>CO2_10</td>
<td>CO2</td>
<td>ATL</td>
<td>4.44E-100</td>
<td>3.80E-103</td>
<td>1170</td>
<td>3.07</td>
</tr>
<tr>
<td>CO2_11</td>
<td>CO2</td>
<td>MRV</td>
<td>1.82E-147</td>
<td>1.27E-133</td>
<td>6.97E+13</td>
<td>13.8</td>
</tr>
<tr>
<td>CO2_12</td>
<td>CO2</td>
<td>ATL</td>
<td>0.0129</td>
<td>0.00121</td>
<td>10.6</td>
<td>1.03</td>
</tr>
<tr>
<td>CO2_13</td>
<td>CO2</td>
<td>MRV</td>
<td>1.82E-142</td>
<td>6.65E-132</td>
<td>3.65E+10</td>
<td>10.6</td>
</tr>
<tr>
<td>CO2_14</td>
<td>CO2</td>
<td>MRV</td>
<td>3.88E-117</td>
<td>9.34E-115</td>
<td>241</td>
<td>2.38</td>
</tr>
<tr>
<td>CO2_15</td>
<td>CO2</td>
<td>MRV</td>
<td>1.23E-154</td>
<td>9.01E-143</td>
<td>7.30E+11</td>
<td>11.9</td>
</tr>
<tr>
<td>CO2_16</td>
<td>CO2</td>
<td>MRV</td>
<td>7.85E-113</td>
<td>5.46E-109</td>
<td>6950</td>
<td>3.84</td>
</tr>
<tr>
<td>CO2_17</td>
<td>CO2</td>
<td>MRV</td>
<td>1.85E-161</td>
<td>3.17E-160</td>
<td>17.1</td>
<td>1.23</td>
</tr>
<tr>
<td>CO2_18</td>
<td>CO2</td>
<td>MRV</td>
<td>1.45E-132</td>
<td>5.75E-130</td>
<td>395</td>
<td>2.6</td>
</tr>
<tr>
<td>CO2_19</td>
<td>CO2</td>
<td>ATL</td>
<td>3.28E-83</td>
<td>1.89E-83</td>
<td>1.74</td>
<td>0.24</td>
</tr>
</tbody>
</table>

47
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2_20</td>
<td>CO2</td>
<td>MRV</td>
<td>2.66E-111</td>
<td>3.47E-105</td>
<td>1310000</td>
</tr>
<tr>
<td>CO2_21</td>
<td>CO2</td>
<td>MRV</td>
<td>9.49E-132</td>
<td>1.44E-124</td>
<td>15200000</td>
</tr>
<tr>
<td>CO2_22</td>
<td>CO2</td>
<td>MRV</td>
<td>0.663</td>
<td>0.85</td>
<td>1.28</td>
</tr>
<tr>
<td>CO2_23</td>
<td>CO2</td>
<td>ATL</td>
<td>2.17E-159</td>
<td>2.56E-161</td>
<td>85</td>
</tr>
<tr>
<td>CO2_24</td>
<td>CO2</td>
<td>ATL</td>
<td>6.97E-157</td>
<td>3.57E-157</td>
<td>1.95</td>
</tr>
<tr>
<td>CO2_25</td>
<td>CO2</td>
<td>MRV</td>
<td>1.70E-144</td>
<td>1.06E-141</td>
<td>623</td>
</tr>
<tr>
<td>CO2_26</td>
<td>CO2</td>
<td>MRV</td>
<td>7.76E-119</td>
<td>5.39E-115</td>
<td>6950</td>
</tr>
<tr>
<td>CO2_27</td>
<td>CO2</td>
<td>MRV</td>
<td>2.91E-89</td>
<td>3.78E-87</td>
<td>130</td>
</tr>
<tr>
<td>CO2_28</td>
<td>CO2</td>
<td>ATL</td>
<td>2.76E-84</td>
<td>1.67E-89</td>
<td>166000</td>
</tr>
<tr>
<td>CO2_29</td>
<td>CO2</td>
<td>MRV</td>
<td>6.18E-191</td>
<td>2.45E-180</td>
<td>3.96E+10</td>
</tr>
<tr>
<td>CO1_01</td>
<td>CO1</td>
<td>MRV</td>
<td>4.12E-104</td>
<td>8.49E-93</td>
<td>2.06E+11</td>
</tr>
<tr>
<td>CO1_02</td>
<td>CO1</td>
<td>MRV</td>
<td>2.51E-143</td>
<td>8.00E-142</td>
<td>31.8</td>
</tr>
<tr>
<td>CO1_03</td>
<td>CO1</td>
<td>MRV</td>
<td>3.09E-155</td>
<td>1.44E-143</td>
<td>4.67E+11</td>
</tr>
<tr>
<td>CO1_04</td>
<td>CO1</td>
<td>ATL</td>
<td>1.37E-121</td>
<td>1.45E-125</td>
<td>9460</td>
</tr>
<tr>
<td>CO1_05</td>
<td>CO1</td>
<td>MRV</td>
<td>1.85E-151</td>
<td>1.68E-149</td>
<td>90.9</td>
</tr>
<tr>
<td>CO1_06</td>
<td>CO1</td>
<td>MRV</td>
<td>9.72E-165</td>
<td>6.15E-161</td>
<td>6330</td>
</tr>
<tr>
<td>CO1_07</td>
<td>CO1</td>
<td>MRV</td>
<td>1.99E-162</td>
<td>1.16E-140</td>
<td>5.81E+21</td>
</tr>
<tr>
<td>CO4_01</td>
<td>CO4</td>
<td>MRV</td>
<td>1.75E-162</td>
<td>5.86E-154</td>
<td>3.35E+08</td>
</tr>
<tr>
<td>CO4_02</td>
<td>CO4</td>
<td>MRV</td>
<td>3.49E-152</td>
<td>8.17E-147</td>
<td>234000</td>
</tr>
<tr>
<td>CO4_03</td>
<td>CO4</td>
<td>MRV</td>
<td>4.99E-171</td>
<td>5.96E-156</td>
<td>1.19E+15</td>
</tr>
</tbody>
</table>
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.