

ORIGINAL ARTICLE OPEN ACCESS

Comparative Whole Genome Phylogeography Reveals Genetic Distinctiveness of Appalachian Populations of Boreal Songbirds

Abigail A. Kimmitt^{1,2}  | Teresa M. Pegan^{1,3} | Kristen S. Wacker¹ | Andrew W. Jones^{4,5} | Benjamin M. Winger¹

¹Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan, USA | ²Department of Biology, Hofstra University, Hempstead, New York, USA | ³Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA | ⁴Department of Ornithology, Cleveland Museum of Natural History, Cleveland, Ohio, USA | ⁵Spring Island Trust, Okatie, South Carolina, USA

Correspondence: Benjamin M. Winger (wingerb@umich.edu)**Received:** 17 June 2025 | **Revised:** 17 August 2025 | **Accepted:** 12 September 2025**Funding:** This material is based upon work supported by the National Science Foundation under Grant No. 2146950 to B.M.W.**Keywords:** comparative population genomics | genetic diversity | molecular parallelism | seasonal migration

ABSTRACT

Intraspecific genetic diversity across a species' geographic range is relevant to adaptive potential and long-term population persistence, and identifying genetically distinct groups within species can direct management decisions focused on conserving species-level genetic diversity. Comparative phylogeography using whole genome techniques allows for investigation of whether co-distributed species exhibit shared spatial genetic differentiation at fine spatial scales, thereby facilitating a comparative approach to both landscape and conservation genetics. By sequencing over 900 low-coverage whole genomes, we evaluated the concordance of genetic structure and diversity from 12 co-occurring species of migratory birds whose breeding ranges span adjacent North American ecogeographic regions: the vast boreal forest belt and the temperate montane Appalachian forests. We detected concordant phylogeographic patterns in 11 of 12 species wherein populations from the southern Appalachians were genetically distinct from boreal belt populations. Our results reveal that small populations persisting in the southern Appalachian Mountains consistently harbor genetic diversity that is subtly distinct from much larger, widespread boreal populations of the same species. However, in most species, levels of standing genetic diversity were not significantly different between Appalachian and boreal populations despite the drastic difference in geographic extent of these populations. We found no evidence for shared signatures of selection across the genome, suggesting that the concordance of spatial genetic structure across species emerges from species-specific patterns of molecular divergence across the genome rather than parallel patterns of selection. Conservation of the Appalachian ecosystem would likely support maintenance of distinct genetic diversity in several migratory avian species with widespread distributions.

1 | Introduction

As a “place-based” field, comparative phylogeography investigates the geographic context of genetic variation and population structure in co-occurring species (Avice et al. 1987; Edwards et al. 2022; Knowles 2009). The development of cost-effective whole genome techniques has created a new era of

comparative phylogeography wherein researchers can evaluate spatial genetic differentiation at high resolution among many co-distributed species, thereby facilitating landscape and conservation genetics at a comparative scale (Grueber 2015; McGaughan et al. 2022; Rissler 2016). Here, we use whole genome sequencing to investigate the concordance of genetic structure and diversity patterns in co-occurring migratory

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Evolutionary Applications* published by John Wiley & Sons Ltd.

birds whose breeding ranges span adjacent ecogeographic regions: the vast boreal forest belt that extends longitudinally across northern North America and a connected narrow peninsula of montane habitat that runs south along the Appalachian Mountains (Figure 1). Specifically, we evaluated 900 genomes from 12 bird species to test whether Appalachian populations are genetically distinct from the considerably larger boreal belt populations, whether they have consistent population differences in genetic diversity, and whether parallel patterns of molecular evolution underlie population differentiation in these co-distributed species.

Many boreal plants and animals have broad ranges that extend across northern North America, with relict populations persisting further south in similar habitats provided by montane temperate forests (Campbell et al. 2021; Wiken et al. 2011). The southern and central Appalachian Mountains are at a lower latitude than the boreal forest belt and cover a much smaller geographic area, but support mixed deciduous-coniferous montane forest that is suitable for many boreal forest species of birds (Kimmitt et al. 2024a; Leppold and Mulvihill 2011; Miller et al. 2024) and other taxa (Campbell et al. 2021; Mitchell et al. 1997) (Figure 1). Phylogeographic structure between boreal and Appalachian populations could be driven by historic refugial species distributions or the unique ecology that distinguishes the southern and central Appalachians from the northern boreal forest belt (Omernik and Griffith 2014), with differences in biogeographic history, climate, and forest structure (Cogbill and White 1991; Lyman and Edwards 2022; Wiken et al. 2011). However, species with high dispersal ability are expected to exhibit low genetic differentiation (Claramunt et al. 2012; Medina et al. 2018), such

that we might expect that migratory birds would exhibit panmixia or subtle patterns of isolation by distance, rather than discrete differentiation between the boreal and Appalachian ecoregions. A recent population genetic analysis of a migratory boreal avian assemblage of 35 co-distributed species found that species vary in their spatial population genetic patterns across the North American boreal forest belt, ranging from panmixia to significant isolation by distance (Pegan et al. 2025b). However, the degree of genetic divergence between widespread boreal and geographically restricted Appalachian populations of migratory birds remains understudied at a comparative scale. Here, by incorporating additional sampling for 12 species from Pegan et al. (2025b) whose ranges extend south into the Appalachian montane forests, we evaluate whether Appalachian populations exhibit variable or shared phylogeographic patterns in relation to the geographically more expansive boreal forest belt.

The effects of habitat loss and climate change on montane taxa are pressing conservation issues globally (Freeman et al. 2018, 2021), as well as locally in the Appalachian region (Lewis et al. 2023; Milanovich et al. 2010; Potter et al. 2010). Populations of boreal birds in the southern peripheries of their ranges, including Appalachia, are likely at higher extirpation risk than boreal belt populations due to the smaller geographic area paired with acute anthropogenic pressure (Leppold and Mulvihill 2011; Lituma et al. 2021; Mitchell et al. 1999; Ralston and DeLuca 2020). A net loss of approximately 2.5 billion migratory North American birds since 1970 (Rosenberg et al. 2019), and climate-induced range shifts forcing birds to higher latitudes (McCaslin and Heath 2020) further highlight the need to understand genetic diversity in southern

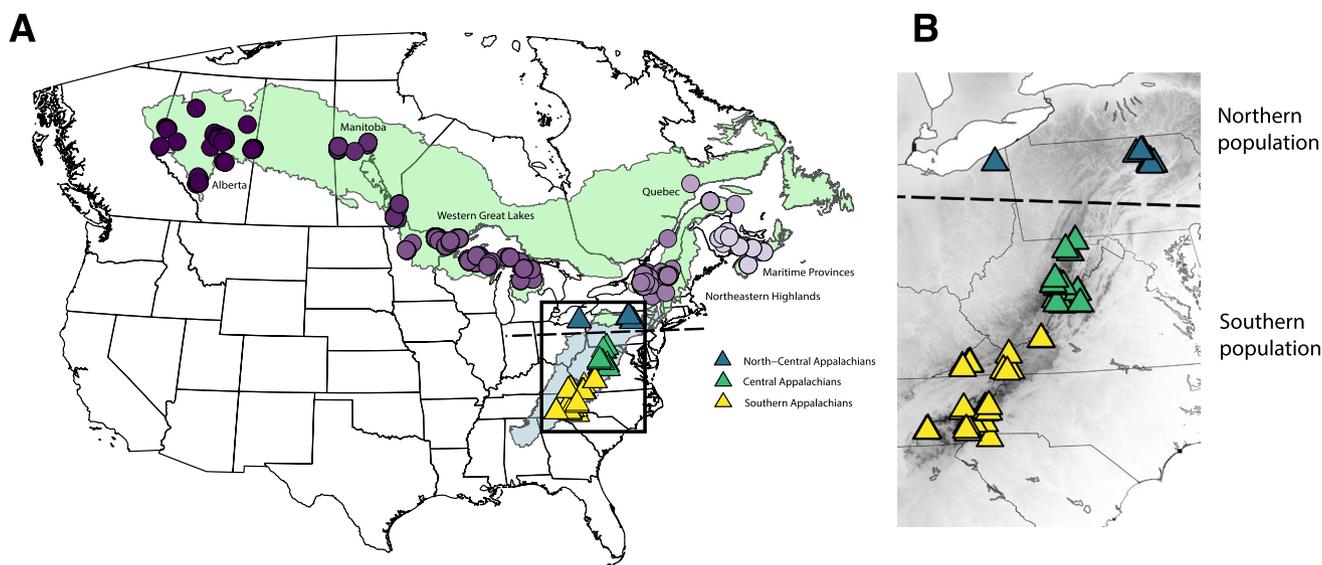


FIGURE 1 | (A) Map of specimen sampling locations for all 12 species. We sampled an average of 75.6 individuals (range = 35–94) per species across the full Boreal-Appalachian breeding range. Each point represents an individual, but in some cases, multiple individuals were sampled from the same location, such that points are overlapping. The boreal forest (green) and the Appalachian forests (light blue) were designated following “level 1” and “level 2” ecoregions, respectively defined by Omernik and Griffith (2014). Samples from the boreal region are shown in purple, with shading corresponding to longitude. (B) The inset map focuses on the three sampling regions associated with the Appalachian Mountains south of the boreal forest belt: (1) North-Central Appalachians (teal), (2) Central Appalachians (green), (3) Southern Appalachians (yellow). Elevation is represented in grayscale, with darker colors showing higher elevation. In both maps, the dotted line demarcates the geographic boundary between “Northern” and “Southern” populations that exhibited genetic differentiation in 11 of 12 study species (Figure 2).

Appalachian populations. Knowledge of Appalachian populations' genetic diversity and connectivity with larger boreal populations is relevant to their conservation management, as genetic diversity across a species' range is critical for long-term population persistence and adaptive potential (Exposito-Alonso et al. 2022; Laikre et al. 2020). Therefore, in addition to testing for genetic structure, we evaluated intraspecific variation in levels of population genetic diversity in boreal versus southern Appalachian populations in each of the study species.

Specifically, we tested whether the geography of genetic diversity in each species supports alternative hypotheses that derive from the effects of historic versus contemporary processes on population demographic trajectories. Much of the contemporary boreal forests was covered by ice during the Last Glacial Maximum, whereas the southern and central Appalachians were south of the Laurentide Ice Sheet (Lyman and Edwards 2022). This pattern has led to the suggestion that the Appalachian Mountains served as a glacial refugium from which populations expanded into the boreal forests after the ice sheet retreated (Lyman and Edwards 2022; Soltis et al. 2006), a hypothesis that predicts higher levels of genetic diversity in contemporary southern (Appalachian) populations than northern (boreal forest) populations (Excoffier 2004; Hewitt 1999; Provan and Bennett 2008; Ralston et al. 2021). However, an alternative hypothesis deriving from modern differences between ecoregions in geographic range and census population sizes (Lyman and Edwards 2022; Soltis et al. 2006) predicts the opposite patterns in genetic diversity, wherein Appalachian populations may exhibit lower genetic diversity than the larger boreal population. Assessing relative levels of genetic diversity in boreal versus Appalachian populations will elucidate the relevance of population genetics to conservation of Appalachian diversity.

Whole genome phylogeographic sampling creates the opportunity not only to assess spatial population genetic patterns, but also to evaluate the genomic context of these patterns. For example, genomic divergence can reflect effects of few loci with very strong structure (e.g., Sodeland et al. 2016; Toews et al. 2016; Harringmeyer and Hoekstra 2022), or the subtle effects of many loci (e.g., Novembre et al. 2008; Michel et al. 2010; Sendell-Price et al. 2020). Therefore, after determining patterns of spatial genetic structure in each species, we examined whether genetic differentiation tended to be diffuse or localized in the genome. Positive selection in a population, or local adaptation, is expected to drive large differences in allele frequencies in localized positions on the genome, whereas neutral evolutionary processes are more likely to result in differentiation that is more widely distributed throughout the genome (Lewontin and Krakauer 1973; Hahn 2019). If genetic structure in species is primarily driven by neutral phylogeographic processes, we expect that differentiation should be distributed diffusely across the genome. Alternatively, if local adaptation to environmental differences between divergent populations drives genetic structure within species, we expect more localized peaks of genomic differentiation around loci responsible for the selected traits. Finally, where we observed concordant population structure in closely related species, we tested for parallel patterns of genomic differentiation, that is,

evidence that the same genomic regions drive genetic differentiation across multiple species. Although scans for molecular parallelism are more typically used to determine the molecular basis of known phenotypic convergence (Manceau et al. 2010; Lee and Coop 2019; Sackton and Clark 2019), we apply this approach to investigate whether parallel genomic adaptation (e.g., shared candidate genes between co-distributed species) might underlie congruent phylogeographic patterns of genetic variation.

2 | Methods

2.1 | Study System and Sampling

We examined population structure in 12 co-distributed bird species for which comprehensive tissue sampling was available across the breeding range, spanning the boreal forest belt and the Appalachian Mountains (Table S1). Our study system includes one woodpecker species (Piciformes), while the remaining species are from seven families of songbirds (Passeriformes) (Table 1). Species differ in migratory distance, but otherwise have similar life histories (e.g., mating system, age to first breeding season) and general habitat associations (mixed coniferous-deciduous forest) (Winger and Pegan 2021). Sampling across the boreal forest belt extended from Alberta through the western Great Lakes and northeastern highlands of the United States to the Maritime provinces (Figure 1, Figure S1). In the northeast, our boreal belt sampling region includes parts of the northern Appalachian Mountains as geologically defined (e.g., Green Mountains of Vermont), but our reference to Appalachian sampling refers specifically to the peninsula of montane temperate forests extending south of the boreal belt (Figure 1). We refer to these Appalachian sampling regions as the north-central Appalachians (Pennsylvania and Ohio; teal triangles, Figure 1), the central Appalachians (West Virginia, Maryland, and northern Virginia; green triangles, Figure 1), and the southern Appalachians (North Carolina and southern Virginia; yellow triangles, Figure 1).

We included in our study 656 previously sequenced low-coverage genomes from across the boreal belt from a study focused on genetic structure within the boreal belt region (Pegan et al. 2025b) and 47 samples from the Appalachian range of one species, *Catharus fuscescens*, from a range-wide phylogeographic study (Kimmitt et al. 2024a). Following an identical protocol as these studies, we sequenced an additional 216 samples from the three Appalachian sampling regions of the remaining 11 species to achieve sampling in both the boreal and Appalachian ranges of these species. All DNA sequenced for this manuscript was extracted using DNeasy Blood and Tissue Kits (Qiagen Sciences, Germantown, MD, USA) from frozen or ethanol-preserved tissue samples from existing natural history collections or blood samples that were collected during the breeding season (Table S1, see Acknowledgments). We prepared libraries for low-coverage sequencing using a modified Illumina Nextera protocol (Schweizer and DeSaix 2023; Therkildsen and Palumbi 2017) and then sequenced using paired-end sequencing of 150 bp reads. Libraries were sequenced on an Illumina HiSeq, NovaSeq 6000, or NovaSeq X platform.

TABLE 1 | Whole genome analyses include population differentiation (F_{ST}) and population-level genetic diversity (mean \pm standard deviations of pairwise θ_π calculated across the chromosomes) for each species.

Family	Species	Whole genome					Mitochondrial					
		N		Pairwise θ_π		P	N		Nei's		Nucleotide diversity (Southern/Northern)	Tajima's D
		$N_{\text{Southern}}/N_{\text{Northern}}$	F_{ST}	Southern (mean \pm SD)	Northern (mean \pm SD)		$N_{\text{Southern}}/N_{\text{Northern}}$	F_{ST}	D	D		
Picidae	<i>Sphyrapicus varius</i>	3/63	0.006	—	—	—	3/63	—	—	—	—/0.003	-2.42
Virionidae	<i>Vireo solitarius</i>	28/56	0.006	0.0102 \pm 0.0013	0.0069 \pm 0.0013	<0.0001	27/53	0.014	0.004	0.004	0.004/0.003	-2.02
Regulidae	<i>Regulus satrapa</i>	28/50	0.004	0.0075 \pm 0.0019	0.0068 \pm 0.0019	0.0002	28/50	0.000	0.001	0.001	0.001/0.001	-2.66
Certhiidae	<i>Certhia americana</i>	10/48	0.004	0.0052 \pm 0.0011	0.0050 \pm 0.0009	0.3806	10/48	0.076	0.012	0.012	0.002/0.002	-2.07
Troglodytidae	<i>Troglodytes hiemalis</i>	9/26	0.005	0.0062 \pm 0.0018	0.0060 \pm 0.0016	0.0818	1/24	—	—	—	—/0.029	-2.31
Turdidae	<i>Catharus fuscus</i>	29/65	0.008	0.0157 \pm 0.0035	0.0159 \pm 0.0038	0.6903	29/64	0.001	0.002	0.002	0.002/0.003	-2.20
Turdidae	<i>Catharus guttatus</i>	5/63	0.005	—	—	—	5/63	—	—	—	—/0.001	-2.42
Passerellidae	<i>Junco hyemalis</i>	12/58	0.014	0.0063 \pm 0.0007	0.0063 \pm 0.0006	0.4994	6/64	0.086	0.017	0.017	0.001/0.001	-1.68
Parulidae	<i>Setophaga coronata</i>	9/71	0.004	0.0093 \pm 0.0009	0.0094 \pm 0.0008	0.3184	9/68	0.005	0.005	0.005	0.002/0.002	-2.06
Parulidae	<i>Setophaga fusca</i>	20/51	0.001	0.0122 \pm 0.0009	0.0113 \pm 0.0007	<0.0001	20/51	0.012	0.002	0.002	0.002/0.001	-2.65
Parulidae	<i>Setophaga magnolia</i>	22/68	0.006	0.0121 \pm 0.0008	0.0122 \pm 0.0008	0.5364	22/60	0.055	0.005	0.005	0.002/0.001	-2.34
Parulidae	<i>Setophaga virens</i>	25/68	0.003	0.0056 \pm 0.0005	0.0077 \pm 0.0004	<0.0001	25/65	0.029	0.006	0.006	0.005/0.005	-1.75

Note: Number of individuals used for calculating whole genome and mitochondrial statistics are reported as $N_{\text{Southern}}/N_{\text{Northern}}$. We included p-values from a Wilcoxon-rank sum test used to compare mean pairwise θ_π between Southern and Northern populations. For comparison to traditional mitochondrial analyses, ND2 statistics include population differentiation (F_{ST} , Nei's D), population-level genetic diversity (nucleotide diversity), and species-level Tajima's D. We excluded *S. varius* and *C. guttatus* from pairwise θ_π analyses given the very small sampling size of the Southern population ($n \leq 5$ individuals). We also excluded *S. varius*, *C. guttatus*, and *T. hiemalis* from Southern genetic diversity measures and the two measures of population differentiation, due to very low numbers of sequenced mitochondrial genomes ($n \leq 5$ individuals).

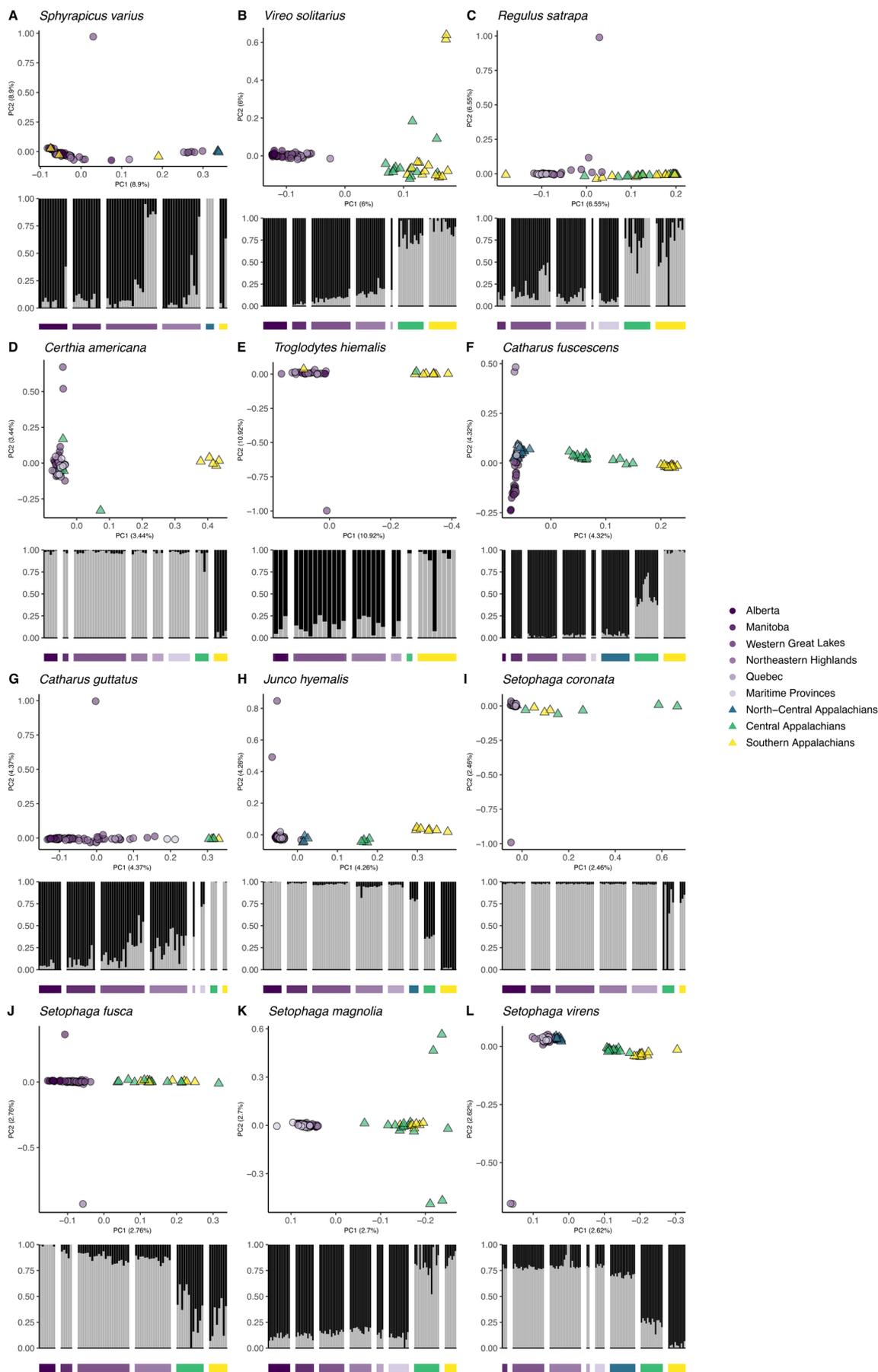


FIGURE 2 | Legend on next page.

FIGURE 2 | Principal Component Analysis (PCA) (top) and admixture plots ($K=2$) (bottom) for each species. Scientific name is at the top of each panel. Colors represent sampling location across the Boreal-Appalachian range. We denote Appalachian sampling regions as triangles, and boreal belt sampling regions as circles with purple shading corresponding to longitude in Figure 1. All species exhibit spatial patterns of genetic differentiation associated with a phylogeographic break in the central or southern Appalachians, except for *Sphyrapicus varius* (panel A).

2.2 | Data Processing

Following the methods used by Pegan et al. (2025b) and Kimmitt et al. (2024a), we trimmed adapters and low-quality bases from demultiplexed data using the `-trimns` and `-trimqualities` options in AdapterRemoval v2.3.1 (Schubert et al. 2016). To reduce potential sequencing batch effects, we removed low-quality read ends using fastp v0.23.2 (Chen et al. 2018) with the `--cut_right` option (Lou and Therkildsen 2022).

We assembled mitochondrial genomes for each sample using NOVOplasty (v. 4.3.1) (Dierckxsens et al. 2017). Using at least one mitochondrial gene from each sample, we used BLAST in Geneious (v. 2021.2.2) to confirm species identity of each sample and check for evidence of cross-contamination and hybridization during extractions and/or library preparation (as described in Kimmitt et al. 2023, 2024a). We removed 12 samples that failed to sequence, showed evidence of cross-contamination, or showed evidence of species misidentification.

We aligned samples from each species in our dataset to a closely related species with an available chromosome-level reference genome (Table S2; Toews et al. 2016; Manthey et al. 2021; Friis et al. 2022) using `bwa mem` (Li and Durbin 2010) and Samtools (Li et al. 2009). We removed overlapping reads using `clipOverlap` in `bamUtil` (Jun et al. 2015), marked duplicate reads with `MarkDuplicates`, and assigned all reads to a new read group with `AddOrReplaceReadGroups` using `picard` (<http://broadinstitute.github.io/picard/>). We then indexed all `bam` files using Samtools (Li et al. 2009). We discarded one sample (*Troglodytes hiemalis*) given low mapping quality ($< 50\%$). We aligned samples around indels using GATK (v. 3.7) (Van der Auwera et al. 2013) by applying `RealignerTargetCreator` to the entire dataset and using `IndelRealigner` for each sample.

After filtering, we included a total of 906 samples in the final analyses. Genotype likelihoods from low-coverage sequencing data were calculated using the GATK model in ANGSD v0.9.40 (Korneliussen et al. 2014). We analyzed results using a genotype likelihood framework given the genotype uncertainty associated with low-coverage sequencing (Korneliussen et al. 2014; Lou et al. 2021). The options used for ANGSD analyses are detailed in Table S3.

2.3 | Population Structure

We used ANGSD to calculate genotype likelihoods for all sites across the genome that were determined to be polymorphic (i.e., SNPs) at an alpha level of 0.05 (Korneliussen et al. 2014). We filtered paralogous or mis-mapped SNPs out of the dataset using `ngsParalog` v1.3.2 (<https://github.com/tplinderroth/ngsParalog>; Linderroth 2018), a program that implements a likelihood

method to detect mapping problems specifically in low-coverage sequencing data. To visually assess the spatial population genetic structure of each species, we conducted Principal Component Analyses (PCA) with PCAngsd v1.10 (Meisner and Albrechtsen 2018). Because genomic inversions can obscure geographic structure in PCAs (Novembre et al. 2008; Tian et al. 2008; Novembre and Peter 2016), we used visualizations from single-chromosome PCAs and `lostruct` (Li and Ralph 2019) to detect putative inversions across all chromosomes and all 12 species as implemented in PCAngsd (https://github.com/alxsi/mon/local_pcsangsd). We removed blocks of macrochromosomes and full microchromosomes with evidence of inversions from the dataset (Table S4). All sex chromosomes were also removed from the dataset because the quality and the sex of the reference genomes varied, which could introduce bias into our analyses. After filtering out all inversions, we produced genome-wide PCAs and admixture plots using the `--admix` option in PCAngsd to both determine the best fit for the number of ancestry populations (i.e., K) as well as to estimate admixture proportions for that K using a non-negative matrix factorization algorithm. Differentiation on PC2 in *Junco hyemalis* in the genome-wide PCA was associated with sex. Since we could not account for the cause of this pattern, we removed 19 females from the final PCA as well as all other whole genome analyses. In some species, PCA revealed outliers along the non-geographic PC axis (PC 2), for which the source of variation is unknown. Removal of these outliers produced subsequent outliers without changing the nature of the genetic structure associated with geography. Therefore, we have retained all points in the PCA figures that had not been removed for previously described reasons (Figure 2).

We calculated genome-wide pairwise F_{ST} between two distinct genetic populations that emerged from the admixture analyses and visually in our PCAs in most species (see Section 3 and Figures 1 and 2): (1) a “Northern” population that included the boreal belt and the north-central Appalachians and (2) a “Southern” population that included the southern and central Appalachians (Figure 1). Generation of site frequency spectra with genotype likelihood data is computationally intensive, so we created a subset of loci for each reference genome, including both variant and invariant sites, by randomly subsampling stretches of 2 kb loci at least 10 kb apart (yielding about 10% of the whole genome), using scripts modified from https://github.com/markravinet/genome_sampler. We removed all sites flagged by `ngsParalog` from the subsampled dataset and stored the loci in a BED file. Using the `-doSaf` command and the `-sites` filter with our BED file, we generated a site allele frequency (SAF) file in ANGSD for these subsampled loci. We used `winsfs` (Rasmussen et al. 2022) to create folded two-dimensional (2D) site frequency spectra (SFS) between each population pair. We then used the F_{ST} index and `stats` function with the option `-whichFst 1` (the Bhatia estimator, which is recommended for uneven sample sizes; Bhatia et al. 2013) in ANGSD to calculate global pairwise F_{ST} .

2.4 | Genetic Diversity

For population-level genetic diversity measures (pairwise θ_π), we used the same genetic populations (Northern and Southern) described above and calculated genetic diversity measures using site allele frequencies and site frequency spectra. We used the same subsampled and filtered set of loci including invariant sites, as previously described. For all 12 species, the sample size across the Northern population was much larger than the sample size for the Southern population, reflecting the differences in geographic size of the landscape covered by each population. Therefore, for each species, we created five randomly subsampled groups of individuals from the Northern population, each matching the sample size of that species' Southern population. For each species, we then generated population-level site allele frequency (SAF) files in ANGSD with the `-doSaf` parameter and excluded sites flagged by `ngsParalog` and putative inversions. We generated population-level 1-dimensional (1D) site frequency spectra (SFS) using `winsfs` (Rasmussen et al. 2022). We calculated pairwise θ_π for each chromosome using the `saf2theta` and `thetaStat` functions in ANGSD. For the Northern populations, we repeated these analyses individually on the five random subgroups for each species and then averaged per-chromosome θ_π among these subgroups. We compared each species' average Northern population pairwise θ_π to the Southern population's pairwise θ_π using a nonparametric Wilcoxon rank-sum test. We excluded *Sphyrapicus varius* and *Catharus guttatus* from pairwise θ_π analyses given the very small sampling size of the Southern population ($n \leq 5$ individuals).

2.5 | Mitochondrial Analyses

To put our whole genome results in the context of a long history of previous avian comparative phylogeographic analyses from other regions that relied on mitochondrial genes (e.g., Zink 1996; Klicka and Zink 1997; Smith et al. 2014; Winger and Bates 2015), we also measured key population statistics using the mitochondrial gene ND2 across all 12 species, which was a byproduct of the whole genome sequencing (described above). We calculated the average pairwise fixation index as well as Nei's distance between the Southern and Northern populations using R packages `adegenet` v 2.1.5 (Jombart 2008) and `hierfstat` v 0.5-10 (Goudet 2005). For population comparisons in our mitochondrial analyses (F_{ST} and Nei's D), we excluded *C. guttatus*, *S. varius*, and *T. hiemalis* due to low sample sizes or mitochondrial genomes failing to assemble from the Southern population (which yielded $n \leq 5$ individuals). We tested for signatures of population expansion in each species by calculating Tajima's D (Tajima 1989) and we calculated nucleotide diversity for both populations using the R package `pegas` v 1.1 (Paradis 2010). We also created haplotype networks for ND2 using `pegas`.

2.6 | Evaluating the Genomic Extent of Differentiation and Shared Patterns of Divergence

For each species, we calculated genome-wide pairwise F_{ST} and created Manhattan plots to evaluate patterns of differentiation across the genome between assigned Northern and Southern

populations. We generated population-level SAF files in ANGSD using the variant sites file generated for population genetics methods. Using `winsfs`, we then generated a folded 2D SFS for each species. We estimated pairwise F_{ST} across 50kb sliding windows with a window step of 10kb using the `stats2` function in ANGSD. We used a conservative approach commonly used to identify F_{ST} window outliers that were 5 standard deviations above the global F_{ST} mean (e.g., Walsh et al. 2019; Moreira and Smith 2023). We then visualized genome-wide patterns of divergence by creating Manhattan plots for each species using the function "manhattan" in the R package `qqman` v.0.1.9 (Turner 2018).

We also tested for shared signatures of elevated genetic differentiation associated with population differences in allele frequencies by identifying shared F_{ST} outlier windows. Owing to the need to use the same reference genome for this analysis (Fraser and Whiting 2020), we restricted this test of molecular parallelism to groups of species that are all mapped to the same reference genome (i.e., four warbler species mapped to the *Setophaga coronata* reference genome (Toews et al. 2016), two thrush species mapped to the *Catharus ustulatus* reference genome (Louder et al. 2024), and three species from unique families but all mapped to the *Certhia americana* reference genome (Manthey et al. 2021); see Table S2). We estimated pairwise F_{ST} across 10 kb non-overlapping windows using the `stats2` function in ANGSD and then identified outliers that were 5 standard deviations above the global F_{ST} mean. We identified the positions of overlapping outlier windows between any two species that were mapped to the same reference genome. For species mapped to the *S. coronata* and *C. americana* reference genomes (Baiz et al. 2021; Manthey et al. 2021), annotations were constructed using MAKER (Cantarel et al. 2008); we therefore obtained the NCBI ID associated with the proteins with the highest confidence in each annotation. We then used the NCBI ID to identify the gene or protein name, description, and function using the NCBI database (<https://www.ncbi.nlm.nih.gov/>). For species mapped to the *C. ustulatus* reference genome (Louder et al. 2024), the annotation was constructed using the NCBI Eukaryotic Genome Annotation Pipeline, such that the annotation had gene ID information provided. We used the NCBI database to collect any additional information about gene function.

3 | Results

3.1 | Genome Sequencing

Following filtering, samples across all 12 species had a mean of 5.0 \times coverage across the genome (range = approximately 0.91 \times to 23.5 \times coverage). The mean mapping rate across all samples used in analyses was 89.3% (range = 53.0%–98.4%) (Table S1).

3.2 | Population Genetics

We observed concordant patterns of spatial genetic differentiation between the boreal ("Northern") and Appalachian ("Southern") populations on genome-wide PCA plots in 11 out

of 12 of the species (all species except the woodpecker, *S. varius*, which did not exhibit genetic differentiation between boreal and Appalachian populations; Figure 2). Our dataset, however, contained very few southern Appalachian samples ($n=3$) and no central Appalachian samples from *S. varius*, due to the paucity of available genetic samples, which could explain the lack of spatial genetic pattern detected. Admixture analyses further supported $K=2$ as the best fit for the number of ancestry groups and indicated distinct differences between the Southern and Northern populations. The geographic break between these Southern and Northern clusters was most typically observed between the northern Appalachian sampling region (Pennsylvania/Ohio) and the central Appalachian sampling region (West Virginia/northern Virginia; Figures 1 and 2). We note, however, that nuclear genetic differentiation between Southern and Northern populations is very low across all species ($F_{ST} \leq 0.01$; Table 1), indicating shallow divergence.

3.3 | Genetic Diversity

Pairwise θ_π did not significantly differ between Northern and Southern populations in 6 out of 10 species included in the analyses (Table 1). In three species (*Setophaga fusca*, *Regulus satrapa*, and *Vireo solitarius*), the Southern population exhibited significantly higher genetic diversity (pairwise θ_π) than the Northern population, whereas in one species (*Setophaga virens*),

the Northern population exhibited significantly higher genetic diversity than the Southern population.

3.4 | Mitochondrial Analyses

In the mitochondrial gene (ND2), we detected little to no population differentiation ($F_{ST} \leq 0.09$), and low levels of Nei's genetic distance (Nei's $D \leq 0.017$) between Southern and Northern populations. There was no observable population clustering in the haplotype networks (Figure S2). We found little nucleotide diversity across species ($\pi \leq 0.005$, except for *T. hiemalis*), and differences in diversity between populations were not detected in mitochondrial analyses. We also observed negative Tajima's D for all 12 species, which is consistent with the expectations for populations with recent demographic expansion (Table 1).

3.5 | The Genomic Extent of Differentiation and Shared Molecular Patterns of Divergence

By characterizing regions of elevated differentiation across the genome, we detected outlier windows ranging from 203 windows in *S. varius* to 519 in *Setophaga magnolia*. Outlier windows accounted for an average of 0.4% of the sampled windows of each species (Figure 3). Outlier windows also

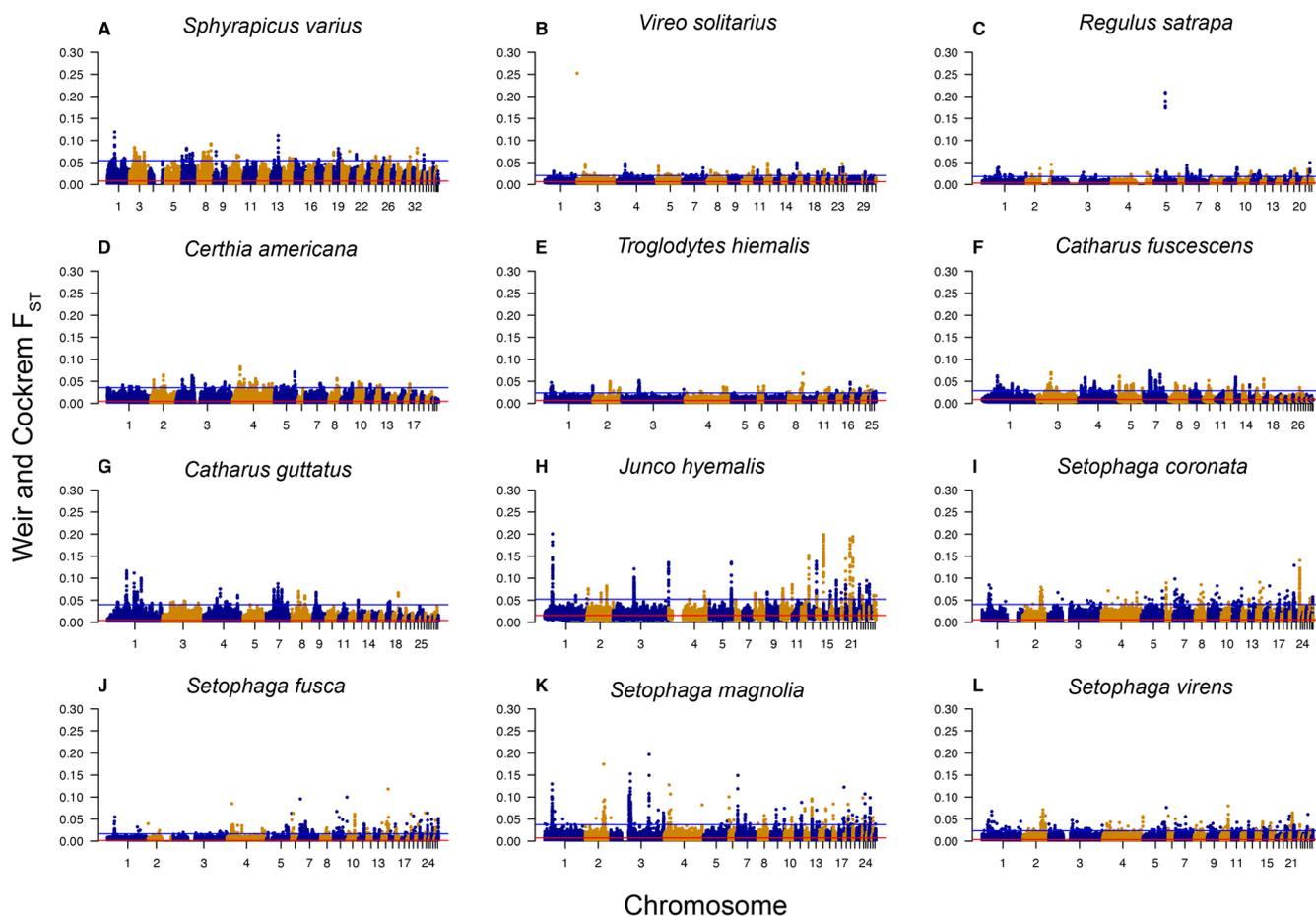


FIGURE 3 | Manhattan plots of pairwise F_{ST} between the Southern and Northern population across 50 kb sliding windows with a window step of 10 kb calculated in ANGSD. Alternating navy and gold colors differentiate chromosomes. The orange line demarcates the global mean F_{ST} , and the blue line demarcates five standard deviations above the global mean.

exhibited low levels of differentiation (F_{ST} average = 0.04; range = 0.01–0.25) and were not concentrated in any specific genomic regions.

We did not identify any shared windows of elevated differences across all species in either set that included greater than 2 species (i.e., four parulid species, and three species mapped to *C. americana*). Considering only shared outlier windows containing annotated proteins, we identified a total of 20 candidate genes for selection between the Southern and Northern populations shared by two or three parulid species. Only one candidate gene was shared by three species (*S. fusca*, *S. magnolia*, and *S. virens*), ATP6V0B (ATPase H⁺ transporting V0 subunit b), which is a gene involved in proton transmembrane transport (Table S5). The candidate genes shared between two species are also primarily involved in cellular functioning, including development, maintenance, and inflammatory response (Table S5). We identified one candidate gene between Southern and Northern populations in the *R. satrapa* and *T. hiemalis* that plays a role in cell adhesion and migration (the ITGB2 gene, or integrin subunit beta 2). In the pair of *Catharus* species, we also identified only one candidate gene shared by both species, IGFBP7 (insulin like growth factor binding protein 7), which plays a role in regulating cell growth.

4 | Discussion

We detected spatially concordant genetic divergence between the Southern (Appalachian) and Northern (boreal) populations of 11 out of the 12 sympatrically breeding birds we analyzed (Figure 2). Identification of genetically distinct populations using whole genome sequencing can be used to assign intraspecific conservation units, or designated groups used to inform management decisions to conserve species-level genetic diversity (DeSaix et al. 2023; Paetkau 1999). By considering shared patterns of genetic population differentiation across several species, conservation efforts can prevent genetic diversity loss while taking a more inclusive community rather than species-level approach. In our study of a co-distributed species assemblage, our results consistently point to small populations persisting in the southern Appalachian Mountains as harboring genetic diversity that is distinct from much larger, widespread boreal populations of the same species. The degree of genetic structure, however, was subtle, as genetic differentiation between the boreal and Appalachian populations was low across all species at the level of the nuclear genome ($F_{ST} \leq 0.01$) and a mitochondrial gene, ND2 ($F_{ST} \leq 0.09$).

Our results detecting genetic structure between northern and southern populations are similar to those in previous studies of boreal-Appalachian migratory bird species, including *Cardellina canadensis* (Miller et al. 2024) and *C. fuscescens* (Kimmitt et al. 2024a; a species also included in our current study). By contrast, a previous study of another warbler species, *Setophaga caerulescens*, indicated low genetic differentiation between boreal and Appalachian populations based on microsatellite and mitochondrial DNA (Davis et al. 2006). Mitochondrial DNA in our study was likewise unable to recover genetic divergence across this region, showing low divergence and no population-level structure in the haplotype networks in all 12 species. Therefore, the subtle

spatial structure we observe between boreal and Appalachian populations may require whole genome data to detect.

The concordant phylogeographic breaks we detected are unlikely explained by a specific geographic barrier. Although the population abundance of our study species may be lower in the high elevation habitat of southern Pennsylvania's Valley and Ridge geologic province (Fink et al. 2024), near where the phylogeographic break occurs in most species, these species are nevertheless continuously distributed throughout the Appalachians. And yet, concordant patterns of discrete structure between southern and central Appalachian populations suggest restricted gene flow between these regions. We argue that this reduced gene flow may be driven by philopatry and breeding site fidelity, whereby migratory birds, despite flying long distances to their nonbreeding locations, tend to return to the same geographic region and habitat type as where they were fledged or bred in a previous year (Conklin et al. 2024; Pegan et al. 2025b; Winger et al. 2019). That is, despite their mobility, migratory species may have lower than expected dispersal between these ecogeographic regions, leading to genetically distinct populations in the Appalachians, a region of great conservation concern.

We also tested whether genetic diversity patterns reflected either (1) historic range expansions out of southern refugia that resulted in higher diversity in southern populations or (2) contemporary differences in geographic area and census population size that resulted in lower diversity in southern populations. We did not detect consistent population differences in genetic diversity across the species, as 6 out of 10 species tested did not differ in genetic diversity measures between the Southern and Northern populations. In three species (*S. fusca*, *R. satrapa*, and *V. solitarius*), genetic diversity in the Southern population was higher than diversity in the Northern population, suggesting that southern Appalachian populations could have acted as source populations for postglacial northern expansion in these species. By contrast, in one species (*S. virens*), the Northern population exhibited higher genetic diversity than the Southern population, which is more consistent with the prediction that populations isolated in small geographic regions should have lower genetic diversity (Berthier et al. 2002). However, the fact that these opposing patterns of genetic diversity were not broadly observed across our study species casts doubt on their explanatory power.

Our comparative system using whole genome techniques also provided an opportunity to evaluate the genomic context of differentiation underlying spatial structure. For all species, peaks of differentiation across the genome were relatively low ($F_{ST} \leq 0.25$) with outlier windows spread across the genome, suggesting that concordant patterns of spatial structure emerge from species-specific processes that likely reflect neutral genetic differentiation (Hahn 2019; Lewontin and Krakauer 1973). If selection via local adaptation is driving differentiation patterns in similar ways among co-distributed species, then concordant phylogeographic breaks may be associated with shared patterns of molecular evolution, including the genes underlying differentiation between populations (McGaughan et al. 2022). However, we found no shared windows of elevated differentiation across the four *Setophaga* species and only detected a total of 20 candidate genes between species pairs and triplets (Table S4). Similarly, we only identified one unique potential candidate gene shared

between the two *Catharus* species, as well as one unique potential candidate gene in only two out of the three species mapped to the *C. americana* reference genome (Table S4).

The lack of evidence for shared patterns of genetic differentiation in our study, along with the results of several recent comparative studies, suggests that molecular parallelism can be difficult to detect or often does not explain concordant phylogeographic patterns of population differentiation. A previous study of shared genomic differentiation in parulid warblers in the western portions of their ranges in the Canadian Rocky Mountains (which included two of our species, *S. virens* and *S. coronata*) also did not find shared F_{ST} peaks between pairs of hybridizing subspecies or species (Irwin et al. 2018). The authors concluded that the pairs were experiencing differing selective forces or differing genomic responses to selection. Additionally, two recent comparative studies on passerines did not find strong support for molecular parallelism, as no candidate genes were shared between all four focal species in each study, and only 2–33 candidate genes were shared between two or three species (Recuerda et al. 2024; Walsh et al. 2019). By contrast, two woodpecker species, *Dryobates pubescens* and *Dryobates villosus*, which exhibit convergent geographic patterns of plumage coloration, shared about 12% of the outlier windows corresponding to 139 shared candidate genes for selection (Moreira and Smith 2023), providing stronger evidence for molecular parallelism compared to our study and the other studies in passerines.

The comparative nature of our study suggests caution when interpreting outlier loci as candidates for selection, given the absence of molecular parallelism detected in closely related species with similar environmental pressures. Many single-species empirical studies report a high prevalence of outliers and conclude evidence for candidate genes under selection that might explain population differentiation (Ahrens et al. 2018; McGaughan et al. 2022). However, outlier detection methods are known to be prone to false positives (Ahrens et al. 2018), and heterogeneity in relative divergence metrics like F_{ST} across the genome can also be generated by intrinsic genomic architecture (Cruickshank and Hahn 2014; Thom et al. 2024). Our findings further emphasize the complexity of the mechanistic underpinnings of spatial population structure by suggesting that idiosyncratic molecular pathways among species can still produce concordant patterns of spatial population differentiation.

Collectively, our results reveal strong concordance of subtle but distinct genetic structure between the boreal and southern Appalachian populations across most of the co-distributed migratory songbirds tested. The presence of genetically distinct populations across several species associated with adjacent but unique ecoregions can inform conservation and management efforts. The field of conservation has long debated whether conservation policy should prioritize preserving ecosystem, species, or genetic diversity (Bowen 1999). With the advances and accessibility of genomic sequencing and resources, a recent emphasis has been placed on preserving genetic diversity, as it is critical for the adaptive potential of the species (Kahilainen et al. 2014) and plays a role in community and ecosystem resilience (Stange et al. 2021). In addition to conservation genetic approaches focusing on vulnerable species for species-level conservation (e.g., Dufresnes et al. 2023; Miller et al. 2024; Senior et al. 2022),

some work has also focused on community genetics among interacting species (Schweitzer et al. 2008; Stange et al. 2021) or keystone species (Chacón-Vargas et al. 2020; de Góes Maciel et al. 2024; Thompson et al. 2019) to evaluate impact on ecosystem resilience. Our study highlights how comparative phylogeography, a classic discipline for studying the geographic context of genetic variation (Edwards et al. 2022), could be used as a broader link between preserving genetic diversity and ecosystem management, by using evidence of concordant patterns of intraspecific conservation units among co-occurring species to inform habitat conservation policy.

Acknowledgments

For providing tissue and blood samples, we thank the American Museum of Natural History (Brian Smith, Joel Cracraft, Paul Sweet, Peter Capainolo, Tom Trombone), Cleveland Museum of Natural History (Courtney Brennan), Cornell University Museum of Vertebrates (Irby Lovette, Vanya Rohwer, Mary Margaret Ferraro, Charles Dardia), Smithsonian Institution National Museum of Natural History (Chris Milensky, Brian Schmidt, Gary Graves), New York State Museum (Jeremy Kirchman), North Carolina Museum of Natural Sciences (Brian O'Shea, John Gerwin), Royal Alberta Museum (Jocelyn Hudon), Royal Ontario Museum (Mark Peck), University of Alaska Museum (Kevin Winker), University of Minnesota Museum of Natural History (Keith Barker), University of California, Berkeley Museum of Vertebrate Zoology (Rauri Bowie and Carla Cicero), University of Michigan Museum of Zoology (Brett Benz), University of Washington Burke Museum (Sharon Birks), Theresa Burg, and Ellen Ketterson. We thank Nicole Adams, Gideon Bradburd, Zach Hancock, and Leonard Jones for discussion and code assistance. Next-generation sequencing for this project was carried out in the Advanced Genomics Core at the University of Michigan. This research was also supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor.

Ethics Statement

Samples sequenced newly for this study were provided from existing natural history tissue collections. For providing permits that authorized the collection of samples contributed by the authors, we thank the United States Fish and Wildlife Service, United States Forest Service, North Carolina Wildlife Resources Commission, Ohio Division of Wildlife, Pennsylvania Game Commission, Vermont Fish and Wildlife Department, Vermont Agency of Natural Resources, and West Virginia Department of Natural Resources. Sampling protocols were approved by the University of Michigan Animal Care and Use Committee (# PRO00010608).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The raw sequence data used in this study has been uploaded to the NCBI Sequence Read Archive under BioProjects PRJNA1043688 (Kimmit et al. 2024b) and PRJNA1130443 (Pegan et al. 2025a).

References

- Ahrens, C. W., P. D. Rymer, A. Stow, et al. 2018. "The Search for Loci Under Selection: Trends, Biases and Progress." *Molecular Ecology* 27, no. 6: 1342–1356.
- Avise, J. C., J. Arnold, R. M. Ball, et al. 1987. "Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population

- Genetics and Systematics." *Annual Review of Ecology and Systematics* 18, no. 1: 489–522.
- Baiz, M. D., A. W. Wood, A. Brelsford, I. J. Lovette, and D. P. Toews. 2021. "Pigmentation Genes Show Evidence of Repeated Divergence and Multiple Bouts of Introgression in *Setophaga* Warblers." *Current Biology* 31, no. 3: 643–649.
- Berthier, P., M. A. Beaumont, J.-M. Cornuet, and G. Luikart. 2002. "Likelihood-Based Estimation of the Effective Population Size Using Temporal Changes in Allele Frequencies: A Genealogical Approach." *Genetics* 160, no. 2: 741–751.
- Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price. 2013. "Estimating and Interpreting F_{ST} : The Impact of Rare Variants." *Genome Research* 23, no. 9: 1514–1521.
- Bowen, B. W. 1999. "Preserving Genes, Species, or Ecosystems? Healing the Fractured Foundations of Conservation Policy." *Molecular Ecology* 8: S5–S10.
- Campbell, I., K. Behrens, C. Hesse, and P. Chaon. 2021. *Habitats of the World: A Field Guide for Birders, Naturalists, and Ecologists*. Princeton University Press.
- Cantarel, B. L., I. Korf, S. M. Robb, et al. 2008. "MAKER: An Easy-to-Use Annotation Pipeline Designed for Emerging Model Organism Genomes." *Genome Research* 18, no. 1: 188–196.
- Chacón-Vargas, K., V. H. García-Merchán, and M. J. Sanín. 2020. "From Keystone Species to Conservation: Conservation Genetics of Wax Palm *Ceroxylon quindiuense* in the Largest Wild Populations of Colombia and Selected Neighboring Ex Situ Plant Collections." *Biodiversity and Conservation* 29, no. 1: 283–302.
- Chen, K., E. A. Marschall, M. G. Sovic, A. C. Fries, H. L. Gibbs, and S. A. Ludsin. 2018. "assign POP: An R Package for Population Assignment Using Genetic, Non-Genetic, or Integrated Data in a Machine-Learning Framework." *Methods in Ecology and Evolution* 9: 439–446.
- Claramunt, S., E. P. Derryberry, J. V. Remsen Jr., and R. T. Brumfield. 2012. "High Dispersal Ability Inhibits Speciation in a Continental Radiation of Passerine Birds." *Proceedings of the Royal Society B: Biological Sciences* 279, no. 1733: 1567–1574.
- Cogbill, C. V., and P. S. White. 1991. "The Latitude-Elevation Relationship for Spruce-Fir Forest and Treeline Along the Appalachian Mountain Chain." *Vegetatio* 94: 153–175.
- Conklin, J. R., Y. I. Verkuil, M. J. Lefebvre, et al. 2024. "High Dispersal Ability Versus Migratory Traditions: Fine-Scale Population Structure and Post-Glacial Colonisation in Bar-Tailed Godwits." *Molecular Ecology* 33, no. 15: e17452.
- Cruickshank, T. E., and M. W. Hahn. 2014. "Reanalysis Suggests That Genomic Islands of Speciation Are due to Reduced Diversity, Not Reduced Gene Flow." *Molecular Ecology* 23, no. 13: 3133–3157.
- Davis, L. A., E. H. Roalson, K. L. Cornell, K. D. McClanahan, and M. S. Webster. 2006. "Genetic Divergence and Migration Patterns in a North American Passerine Bird: Implications for Evolution and Conservation." *Molecular Ecology* 15, no. 8: 2141–2152.
- de Góes Maciel, F., S. O'Rourke, M. Jones, et al. 2024. "Loss of Genetic Diversity and Isolation by Distance and by Environment in Populations of a Keystone Ungulate Species." *Conservation Genetics* 25, no. 4: 939–953.
- DeSaix, M. G., E. C. Anderson, C. M. Bossu, et al. 2023. "Low-Coverage Whole Genome Sequencing for Highly Accurate Population Assignment: Mapping Migratory Connectivity in the American Redstart (*Setophaga ruticilla*)." *Molecular Ecology* 32, no. 20: 5528–5540.
- Dierckxsens, N., P. Mardulyn, and G. Smits. 2017. "NOVOPlasty: De Novo Assembly of Organelle Genomes From Whole Genome Data." *Nucleic Acids Research* 45: e18–e18.
- Dufresnes, C., L. Dutoit, A. Brelsford, et al. 2023. "Inferring Genetic Structure When There Is Little: Population Genetics Versus Genomics of the Threatened Bat *Miniopterus schreibersii* Across Europe." *Scientific Reports* 13, no. 1: 1523.
- Edwards, S. V., V. V. Robin, N. Ferrand, and C. Moritz. 2022. "The Evolution of Comparative Phylogeography: Putting the Geography (And More) Into Comparative Population Genomics." *Genome Biology and Evolution* 14, no. 1: eval176.
- Excoffier, L. 2004. "Patterns of DNA Sequence Diversity and Genetic Structure After a Range Expansion: Lessons From the Infinite-Island Model." *Molecular Ecology* 13, no. 4: 853–864.
- Exposito-Alonso, M., T. R. Booker, L. Czech, et al. 2022. "Genetic Diversity Loss in the Anthropocene." *Science* 377, no. 6613: 1431–1435.
- Fink, D., T. Auer, A. Johnston, et al. 2024. *eBird Status and Trends, Data Version: 2023; Released: 2025*. Cornell Lab of Ornithology. <https://doi.org/10.2173/WZTW8903>.
- Fraser, B. A., and J. R. Whiting. 2020. "What can be Learned by Scanning the Genome for Molecular Convergence in Wild Populations?" *Annals of the New York Academy of Sciences* 1476: 23–42.
- Freeman, B. G., M. N. Scholer, V. Ruiz-Gutierrez, and J. W. Fitzpatrick. 2018. "Climate Change Causes Upslope Shifts and Mountaintop Extirpations in a Tropical Bird Community." *Proceedings of the National Academy of Sciences* 115, no. 47: 11982–11987.
- Freeman, B. G., Y. Song, K. J. Feeley, and K. Zhu. 2021. "Montane Species Track Rising Temperatures Better in the Tropics Than in the Temperate Zone." *Ecology Letters* 24, no. 8: 1697–1708.
- Friis, G., J. Vizueta, E. D. Ketterson, and B. Milá. 2022. "A High-Quality Genome Assembly and Annotation of the Dark-Eyed Junco *Junco hyemalis*, a Recently Diversified Songbird." *G3* 12: jkac083.
- Goudet, J. 2005. "Hierfstat, A Package for R to Compute and Test Hierarchical F-Statistics." *Molecular Ecology Notes* 5: 184–186.
- Grueber, C. E. 2015. "Comparative Genomics for Biodiversity Conservation." *Computational and Structural Biotechnology Journal* 13: 370–375.
- Hahn, M. W. 2019. *Molecular Population Genetics*. Oxford University Press.
- Harringmeyer, O. S., and H. E. Hoekstra. 2022. "Chromosomal Inversion Polymorphisms Shape the Genomic Landscape of Deer Mice." *Nature Ecology & Evolution* 6: 1965–1979.
- Hewitt, G. M. 1999. "Post-Glacial Re-Colonization of European Biota." *Biological Journal of the Linnean Society* 68, no. 1–2: 87–112.
- Irwin, D. E., B. Milá, D. P. Toews, et al. 2018. "A Comparison of Genomic Islands of Differentiation Across Three Young Avian Species Pairs." *Molecular Ecology* 27: 4839–4855. Wiley Online Library.
- Jombart, T. 2008. "adegenet: A R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24: 1403–1405.
- Jun, G., M. K. Wing, G. R. Abecasis, and H. M. Kang. 2015. "An Efficient and Scalable Analysis Framework for Variant Extraction and Refinement From Population-Scale DNA Sequence Data." *Genome Research* 25: 918–925.
- Kahilainen, A., M. Puurtinen, and J. S. Kotiaho. 2014. "Conservation Implications of Species–Genetic Diversity Correlations." *Global Ecology and Conservation* 2: 315–323.
- Kimmit, A. A., T. M. Pegan, A. W. Jones, et al. 2023. "Genetic Evidence for Widespread Population Size Expansion in North American Boreal Birds Prior to the Last Glacial Maximum." *Proceedings of the Royal Society B: Biological Sciences* 290, no. 1991: 20221334.
- Kimmit, A. A., T. M. Pegan, A. W. Jones, K. Winker, and B. M. Winger. 2024a. "How Veeries Vary: Whole Genome Sequencing Resolves

- Genetic Structure in a Long-Distance Migratory Bird." *Ornithology* 141: ukad061. Oxford University Press US.
- Kimmitt, A. A., T. M. Pegan, A. W. Jones, K. Winker, and B. M. Winger. 2024b. *How Veeries Vary: Whole Genome Sequencing Resolves Genetic Structure in a Long-Distance Migratory Bird*. NCBI Sequence Read Archive.
- Klicka, J., and R. M. Zink. 1997. "The Importance of Recent Ice Ages in Speciation: A Failed Paradigm." *Science* 277: 1666–1669.
- Knowles, L. L. 2009. "Statistical Phylogeography." *Annual Review of Ecology, Evolution, and Systematics* 40: 593–612.
- Korneliusson, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15: 1–13.
- Laike, L., S. Hoban, M. W. Bruford, et al. 2020. "Post-2020 Goals Overlook Genetic Diversity." *Science* 367, no. 6482: 1083–1085.
- Lee, K. M., and G. Coop. 2019. "Population Genomics Perspectives on Convergent Adaptation." *Philosophical Transactions of the Royal Society B* 374: 20180236.
- Leppold, A. J., and R. S. Mulvihill. 2011. "The Boreal Landbird Component of Migrant Bird Communities in Eastern North America." In *Boreal Birds of North America*, edited by J. Wells, 73–83. University of California Press.
- Lewis, W. B., R. J. Cooper, R. B. Chandler, et al. 2023. "Climate-Mediated Population Dynamics of a Migratory Songbird Differ Between the Trailing Edge and Range Core." *Ecological Monographs* 93, no. 1: e1559.
- Lewontin, R. C., and J. Krakauer. 1973. "Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms." *Genetics* 74, no. 1: 175–195.
- Li, H., and R. Durbin. 2010. "Fast and Accurate Long-Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 26: 589–595.
- Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25: 2078–2079.
- Li, H., and P. Ralph. 2019. "Local PCA Shows how the Effect of Population Structure Differs Along the Genome." *Genetics* 211: 289–304.
- Linderoth, T. 2018. *Identifying Population Histories, Adaptive Genes, and Genetic Duplication From Population-Scale Next Generation Sequencing*. University of California.
- Lituma, C. M., J. J. Cox, S. F. Spear, et al. 2021. "Terrestrial Wildlife in the Post-Mined Appalachian Landscape: Status and Opportunities." In *Appalachia's Coal-Mined Landscapes: Resources and Communities in a New Energy Era*, 135–166. Springer International Publishing.
- Lou, R. N., A. Jacobs, A. P. Wilder, and N. O. Therkildsen. 2021. "A Beginner's Guide to Low-Coverage Whole Genome Sequencing for Population Genomics." *Molecular Ecology* 30: 5966–5993.
- Lou, R. N., and N. O. Therkildsen. 2022. "Batch Effects in Population Genomic Studies With Low-Coverage Whole Genome Sequencing Data: Causes, Detection and Mitigation." *Molecular Ecology Resources* 22: 1678–1692.
- Louder, M. I., H. Justen, A. A. Kimmitt, et al. 2024. "Gene Regulation and Speciation in a Migratory Divide Between Songbirds." *Nature Communications* 15, no. 1: 98.
- Lyman, R. A., and C. E. Edwards. 2022. "Revisiting the Comparative Phylogeography of Unglaciated Eastern North America: 15 Years of Patterns and Progress." *Ecology and Evolution* 12, no. 4: e8827.
- Manceau, M., V. S. Domingues, C. R. Linnen, E. B. Rosenblum, and H. E. Hoekstra. 2010. "Convergence in Pigmentation at Multiple Levels: Mutations, Genes and Function." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 365: 2439–2450.
- Manthey, J. D., J. Klicka, and G. M. Spellman. 2021. "The Genomic Signature of Allopatric Speciation in a Songbird Is Shaped by Genome Architecture (Aves: *Certhia americana*)." *Genome Biology and Evolution* 13, no. 8: evab120.
- McCaslin, H. M., and J. A. Heath. 2020. "Patterns and Mechanisms of Heterogeneous Breeding Distribution Shifts of North American Migratory Birds." *Journal of Avian Biology* 51, no. 3: 2077–2088.
- McGaughran, A., L. Liggins, K. A. Marske, et al. 2022. "Comparative Phylogeography in the Genomic Age: Opportunities and Challenges." *Journal of Biogeography* 49, no. 12: 2130–2144.
- Medina, I., G. M. Cooke, and T. J. Ord. 2018. "Walk, Swim or Fly? Locomotor Mode Predicts Genetic Differentiation in Vertebrates." *Ecology Letters* 21, no. 5: 638–645.
- Meisner, J., and A. Albrechtsen. 2018. "Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data." *Genetics* 210: 719–731.
- Michel, A. P., S. Sim, T. H. Powell, M. S. Taylor, P. Nosil, and J. L. Feder. 2010. "Widespread Genomic Divergence During Sympatric Speciation." *Proceedings of the National Academy of Sciences of the United States of America* 107: 9724–9729.
- Milanovich, J. R., W. E. Peterman, N. P. Nibbelink, and J. C. Maerz. 2010. "Projected Loss of a Salamander Diversity Hotspot as a Consequence of Projected Global Climate Change." *PLoS One* 5, no. 8: e12189.
- Miller, C. V., C. M. Bossu, J. F. Sarraco, et al. 2024. "Genomics-Informed Conservation Units Reveal Spatial Variation in Climate Vulnerability in a Migratory Bird." *Molecular Ecology* 33, no. 1: e17199. <https://doi.org/10.1111/mec.17199>.
- Mitchell, J. C., T. K. Pauley, D. I. Withers, et al. 1999. "Conservation Status of the Southern Appalachian Herpetofauna." *Virginia Journal of Science* 50, no. 1: 2.
- Mitchell, J. C., S. C. Rinehart, J. F. Pagels, K. A. Buhlmann, and C. A. Pague. 1997. "Factors Influencing Amphibian and Small Mammal Assemblages in Central Appalachian Forests." *Forest Ecology and Management* 96, no. 1–2: 65–76.
- Moreira, L. R., and B. T. Smith. 2023. "Convergent Genomic Signatures of Local Adaptation Across a Continental-Scale Environmental Gradient." *Science Advances* 9, no. 20: eadd0560.
- Novembre, J., T. Johnson, K. Bryc, et al. 2008. "Genes Mirror Geography Within Europe." *Nature* 456: 98–101.
- Novembre, J., and B. M. Peter. 2016. "Recent Advances in the Study of Fine-Scale Population Structure in Humans." *Current Opinion in Genetics & Development* 41: 98–105.
- Omernik, J. M., and G. E. Griffith. 2014. "Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework." *Environmental Management* 54, no. 6: 1249–1266. <https://doi.org/10.1007/s00267-014-0364-1>.
- Paetkau, D. 1999. "Using Genetics to Identify Intraspecific Conservation Units: A Critique of Current Methods." *Conservation Biology* 13, no. 6: 1507–1509.
- Paradis, E. 2010. "pegas: An R Package for Population Genetics With an Integrated–Modular Approach." *Bioinformatics* 26: 419–420.
- Pegan, T., A. Kimmitt, B. Benz, et al. 2025a. *Comparative Phylogeography of North American Migratory Birds [Dataset]*. NCBI Sequence Read Archive. <https://www.ncbi.nlm.nih.gov/sra/PRJNA1130443>.
- Pegan, T., A. Kimmitt, B. Benz, et al. 2025b. "Long-Distance Seasonal Migration to the Tropics Promotes Genetic Diversity but Not Gene Flow in Boreal Birds." *Nature Ecology & Evolution* 9: 957–969.
- Potter, K. M., W. W. Hargrove, and F. H. Koch. 2010. "Predicting climate change extirpation risk for central and southern Appalachian forest tree species." In *Proceedings From the Conference on the Ecology and Management of High-Elevation Forests in the Central and Southern*

- Appalachian Mountains*. 2009 May 14–15; Slatyfork, WV. *Gen. Tech. Rep. NRS-P-64*, edited by J. S. Rentch and T. M. Schuler, 179–189. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA.
- Provan, J., and K. D. Bennett. 2008. “Phylogeographic Insights Into Cryptic Glacial Refugia.” *Trends in Ecology & Evolution* 23, no. 10: 564–571.
- Ralston, J., and W. V. DeLuca. 2020. “Conservation Lessons From the Study of North American Boreal Birds at Their Southern Periphery.” *Diversity* 12, no. 6: 257.
- Ralston, J., A. M. FitzGerald, T. M. Burg, N. C. Starkloff, I. G. Warkentin, and J. J. Kirchman. 2021. “Comparative Phylogeographic Analysis Suggests a Shared History Among Eastern North American Boreal Forest Birds.” *Ornithology* 138, no. 3: ukab018. <https://doi.org/10.1093/ornithology/ukab018>.
- Rasmussen, M. S., G. Garcia-Erill, T. S. Korneliusen, C. Wiuf, and A. Albrechtsen. 2022. “Estimation of Site Frequency Spectra From Low-Coverage Sequencing Data Using Stochastic EM Reduces Overfitting, Runtime, and Memory Usage.” *Genetics* 222: iyac148.
- Recuerda, M., J. C. H. Montoya, G. Blanco, and B. Milá. 2024. “Repeated Evolution on Oceanic Islands: Comparative Genomics Reveals Species-Specific Processes in Birds.” *BMC Ecology and Evolution* 24, no. 1: 140.
- Rissler, L. J. 2016. “Union of Phylogeography and Landscape Genetics.” *Proceedings of the National Academy of Sciences* 113, no. 29: 8079–8086.
- Rosenberg, K. V., A. M. Dokter, P. J. Blancher, et al. 2019. “Decline of the North American Avifauna.” *Science* 366, no. 6461: 120–124.
- Sackton, T. B., and N. Clark. 2019. “Convergent Evolution in the Genomics Era: New Insights and Directions.” *Philosophical Transactions of the Royal Society B* 374, no. 1777: 20190102.
- Schubert, M., S. Lindgreen, and L. Orlando. 2016. “AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging.” *BMC Research Notes* 9: 1–7.
- Schweitzer, J. A., M. D. Madritch, J. K. Bailey, et al. 2008. “From Genes to Ecosystems: The Genetic Basis of Condensed Tannins and Their Role in Nutrient Regulation in a Populus Model System.” *Ecosystems* 11: 1005–1020.
- Schweizer, T. M., and M. G. DeSaix. 2023. “Cost-Effective Library Preparation for Whole Genome Sequencing With Feather DNA.” *Conservation Genetics Resources* 15, no. 1: 21–28.
- Sendell-Price, A. T., K. C. Ruegg, E. C. Anderson, et al. 2020. “The Genomic Landscape of Divergence Across the Speciation Continuum in Island-Colonising Silvereyes (*Zosterops lateralis*).” *G3: Genes, Genomes, Genetics* 10: 3147–3163.
- Senior, A. F., N. Clemann, M. G. Gardner, K. A. Harrisson, G. M. While, and D. G. Chapple. 2022. “Genetic Structure, Diversity and Distribution of a Threatened Lizard Affected by Widespread Habitat Fragmentation.” *Conservation Genetics* 23, no. 1: 1–15.
- Smith, B. T., J. E. McCormack, A. M. Cuervo, et al. 2014. “The Drivers of Tropical Speciation.” *Nature* 515: 406–409.
- Sodeland, M., P. E. Jorde, S. Lien, et al. 2016. ““Islands of Divergence” in the Atlantic Cod Genome Represent Polymorphic Chromosomal Rearrangements.” *Genome Biology and Evolution* 8: 1012–1022.
- Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis. 2006. “Comparative Phylogeography of Unglaciated Eastern North America.” *Molecular Ecology* 15, no. 14: 4261–4293.
- Stange, M., R. D. Barrett, and A. P. Hendry. 2021. “The Importance of Genomic Variation for Biodiversity, Ecosystems and People.” *Nature Reviews Genetics* 22, no. 2: 89–105.
- Tajima, F. 1989. “Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism.” *Genetics* 123: 585–595.
- Therkildsen, N. O., and S. R. Palumbi. 2017. “Practical Low-Coverage Genomewide Sequencing of Hundreds of Individually Barcoded Samples for Population and Evolutionary Genomics in Nonmodel Species.” *Molecular Ecology Resources* 17, no. 2: 194–208.
- Thom, G., L. R. Moreira, R. Batista, M. Gehara, A. Aleixo, and B. T. Smith. 2024. “Genomic Architecture Predicts Tree Topology, Population Structuring, and Demographic History in Amazonian Birds.” *Genome Biology and Evolution* 16, no. 1: evae002.
- Thompson, T. Q., M. R. Bellinger, S. M. O’Rourke, et al. 2019. “Anthropogenic Habitat Alteration Leads to Rapid Loss of Adaptive Variation and Restoration Potential in Wild Salmon Populations.” *Proceedings of the National Academy of Sciences of the United States of America* 116, no. 1: 177–186.
- Tian, C., R. M. Plenge, M. Ransom, et al. 2008. “Analysis and Application of European Genetic Substructure Using 300 K SNP Information.” *PLoS Genetics* 4: e4.
- Toews, D. P. L., S. A. Taylor, R. Vallender, et al. 2016. “Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers.” *Current Biology* 26: 2313–2318.
- Turner, S. 2018. “qqman: An R Package for Visualizing GWAS Results Using QQ and Manhattan Plots.” *Journal of Open Source Software* 3: 731.
- Van der Auwera, G. A., M. O. Carneiro, C. Hartl, et al. 2013. “From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.” *Current Protocols in Bioinformatics* 43: 11.10.1–11.10.33.
- Walsh, J., P. M. Benham, P. E. Deane-Coe, et al. 2019. “Genomics of Rapid Ecological Divergence and Parallel Adaptation in Four Tidal Marsh Sparrows.” *Evolution Letters* 3, no. 4: 324–338.
- Wiken, E., F. J. Nava, and G. Griffith. 2011. *North American Terrestrial Ecoregions—Level III*, 149. Commission for Environmental Cooperation.
- Winger, B. M., G. G. Auteri, T. M. Pegan, and B. C. Weeks. 2019. “A Long Winter for the Red Queen: Rethinking the Evolution of Seasonal Migration.” *Biological Reviews* 94, no. 3: 737–752.
- Winger, B. M., and J. M. Bates. 2015. “The Tempo of Trait Divergence in Geographic Isolation: Avian Speciation Across the Marañón Valley of Peru.” *Evolution* 69: 772–787.
- Winger, B. M., and T. M. Pegan. 2021. “Migration Distance Is a Fundamental Axis of the Slow-Fast Continuum of Life History in Boreal Birds.” *Auk* 138, no. 4: ukab043.
- Zink, R. M. 1996. “Comparative Phylogeography in North American Birds.” *Evolution* 50: 308–317.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S2:** Reference genomes used to align samples from each species, as well as the number of SNPs analyzed in PCA, and the length of random genome subset used to estimate pairwise θ_{π} . **Table S3:** Parameters used in ANGSD to estimate metrics. Parameter definitions were summarized from the ANGSD website. **Table S4:** A list of chromosomes regions filtered out of the dataset due to evidence of putative inversion polymorphisms. Microchromosomes (indicated by region type) were entirely discarded, such that it is unnecessary to specify filtering range. **Table S5:** Summary of candidate genes identified by outlier windows (50 kb sliding windows with a window step of 10 kb) from genome-wide pairwise F_{ST} analysis shared between two or three species. Outlier windows were identified as 5 standard deviations above the global F_{ST} mean. Using the *Setophaga coronata* annotation, we identified protein matches found within candidate windows. We then identified gene name, description, and function using the NCBI IDs on the NCBI database (<https://www.ncbi.nlm.nih.gov/>). **Figure S1:** (A) Map of specimen sampling locations for each species. Scientific name is at the top of each panel. Each point represents an individual,

but in some cases, multiple individuals were sampled from the same location, such that points are overlapping. The species' range (orange) show the mean abundance of the species during the breeding season in 2023 from the eBird Status and Trends (Fink et al. 2024). Samples from the boreal region are displayed by purple circles, with shading corresponding to longitude. Triangles show one of three potential sampling regions associated with the Appalachian Mountains south of the boreal forest belt, which include: (1) North-Central Appalachians (teal), (2) Central Appalachians (green), (2) Southern Appalachians (yellow). Samples from each species do not always include every sampling region from the dataset. **Figure S2:** Haplotype networks using the ND2 gene generated using the R package *pegas*. The Southern population is in red, and the Northern population is in turquoise. **Table S1:** Summary table of samples used in analyses, which includes institution, species, catalogue number, sex, collection date, locality, and latitude/longitude data for each sample. Genomic coverage (total number of bases) and the percent of filtered reads mapped are also included for every individual. Coverage is equivalent to the total bases/1e9 based on the assumption that avian genomes are approximately 1 GB. Acronyms for institution names are as follows: AMNH = American Museum of Natural History; BURG = blood samples collected by Theresa Burg; CMNH = Cleveland Museum of Natural History; CUMV = Cornell University Museum of Vertebrates; MMNH = Bell Museum of Natural History; MVZ = Museum of Vertebrate Zoology, UC Berkeley; KETT = blood samples collected by Ellen Ketterson; NCSM = North Carolina State Museum; NYSM = New York State Museum; RAM = Royal Alberta Museum; UAM = University of Alaska Museum; UMMZ = University of Michigan Museum of Zoology; USNM = National Museum of Natural History, Smithsonian Institution; UWBM = University of Washington Burke Museum.