How Veeries vary: whole-genome sequencing reveals genetic differentiation between boreal and southern montane populations in a long-distance migratory bird

Abigail A. Kimmitt[1], Teresa M. Pegan[1], Andrew W. Jones[2,3], Kevin Winker[4], & Benjamin M. Winger[1*]

[1] Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109 USA

[2] Department of Ornithology, Cleveland Museum of Natural History, Cleveland, OH 44106 USA

[3] Present Address: Spring Island Trust, Okatie, SC 29909 USA

[4] University of Alaska Museum, Fairbanks AK 99775 USA

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), New York State Museum (Jeremy Kirchman), University of Washington Burke Museum (Sharon Birks), Royal Alberta Museum (Jocelyn Hudon), University of Kansas Biodiversity Institute (Town Peterson, Robert Moyle, Mark Robbins), Cleveland Museum of Natural History (Courtney Brennan), University of Michigan Museum of Zoology (Brett Benz) and the University of Alaska Museum. For assistance in the field, we thank Courtney Brennan, Brett Benz, Susanna Campbell, Shane DuBay, Laura Gooch, Eric Gulson-Castillo, Ethan Gyllenhaal, Heather Skeen, Vera Ting, and Brian Weeks. We also thank Kristen Ruegg and Teia Schweizer for training and support in their whole-genome library preparation methods. We thank Nicole Adams, Gideon Bradburd, Zach Hancock, and Leonard Jones for discussion and feedback. 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.

Funding statement

This material is based upon work supported by the National Science Foundation under Grant No. 2146950 to BMW. Funding for AWJ was from the William A. and Nancy R. Klamm Chair of Ornithology and the Richard and Jean Hoffman Ornithological Endowment at the Cleveland Museum of Natural History.

Ethics statement

For permits to collect specimens in the field, we thank the United States Fish and Wildlife Service, United States Forest Service, Michigan Department of Natural Resources, Minnesota Department of Natural Resources, Colorado Department of Natural Resources, Idaho Department of Fish and Game, North Carolina Wildlife Resources Commission, Ohio Division of Wildlife, Oregon Department of Fish and Wildlife, Pennsylvania Game Commission, Vermont Fish and Wildlife Department, Vermont Agency of Natural Resources, West Virginia Department of Natural Resources, Canadian Wildlife Service of Environment and Climate Change Canada, and Manitoba Fish and Wildlife. Field sampling was approved by the University of Michigan Animal Care and Use Committee (# PRO00010608).

Conflict of Interest Statement

The authors declare no conflicts of interest.

Author Contributions

- AWJ, AAK, TMP, and BMW conceived the idea of the study.
- AAK, TMP, and BMW developed methodology.
- AAK and BMW analyzed the data.
- AAK and BMW wrote the original draft of the manuscript.
- TMP, AWJ and KW reviewed and edited the manuscript.
- AWJ, KW, and BMW contributed substantial materials or funding for the study.

Data accessibility

Raw genetic data generated from this study is available on the Sequence Read Archive (Accession #s pending). Code used for analysis is available on Dryad Digital Repository.

Abstract

In high-latitude species with high dispersal ability, such as long-distance migratory birds, populations are often assumed to exhibit little genetic structure due to high gene flow or recent postglacial expansion. We sequenced over 120 low-coverage whole genomes from across the breeding range of a long-distance migratory bird, the Veery (*Catharus fuscescens*), revealing strong evidence for isolation by distance. Additionally, we found distinct genetic structure between boreal, western montane U.S., and southern Appalachian sampling regions. We suggest that population genetic structure in this highly migratory species is detectable with the high resolution afforded by whole-genomic data because, similar to many migratory birds, the Veery exhibits high breeding site fidelity, which likely limits gene flow. Resolution of isolation by distance across the breeding range was sufficient to assign likely breeding origins of individuals sampled in this species' poorly understood South American nonbreeding range, demonstrating the potential to assess migratory connectivity in this species using genomic data. As the Veery's breeding range extends across both historically glaciated and unglaciated regions in North America, we also evaluated whether contemporary patterns of structure and genetic diversity are consistent with historical population isolation in glacial refugia. We found that patterns of genetic diversity did not support southern montane regions (southern Appalachians or western U.S. mountains) as glacial refugia. Overall, our findings suggest that isolation by distance yields subtle associations between genetic structure and geography across the breeding range of this highly vagile species even in the absence of obvious historical vicariance or contemporary barriers to dispersal.

Keywords: population structure, isolation by distance, site fidelity, genetic diversity, phylogeography, migratory birds

Lay summary

- Describing how populations in a species differ genetically is important for understanding that species' evolutionary history.
- Migratory birds have high dispersal abilities, potentially reducing genetic structure. However, many migratory birds return to the same breeding site year after year, which could reduce gene flow between populations.
- We sequenced >120 genomes to detect population genetic differentiation in a common songbird, the Veery (*Catharus fuscescens*).
- We found that genetic similarity between samples decreased with increasing geographic distance (i.e., isolation by distance) and that populations in southern Appalachia were distinct from samples in the rest of the breeding range.
- We did not detect differences in genetic diversity patterns between populations, contrary to predictions about putative glacial refugia.
- We assigned likely geographic breeding region to birds sampled in the poorly understood wintering range.
- Despite the Veery's long-distance migration and high dispersal ability, isolation by distance produces subtle but detectable population structure across its breeding range.

Introduction

Resolving genetic population structure in wild populations is important for understanding a species' spatial and demographic evolutionary history as well as identifying microevolutionary processes underlying adaptation and population differentiation (Manel et al. 2003; Edwards et al. 2015; Lou et al. 2021). For species with high dispersal ability, however, resolving spatial genetic structure can be particularly challenging, as greater dispersal capabilities are associated with higher gene flow and minimal genetic structure (Slatkin 1987; Bohonak 1999; Claramunt et al. 2012; Medina et al. 2018). Seasonally migratory species, which often travel long distances between breeding and wintering grounds each year, are typically considered to have high dispersal, as their vagility should reduce the impact of geographic barriers on dispersal-related movements (Paradis et al. 1998; Medina et al. 2018; Everson et al. 2019; Claramunt 2021). Yet, in many bird species, long-distance seasonal migration is associated with limited dispersal between breeding sites, as adult migratory birds frequently exhibit high interannual fidelity to their breeding sites (Winger et al. 2019). Natal dispersal patterns, however, remain poorly understood in small-bodied migratory birds. Breeding site fidelity and natal philopatry have the potential to limit gene flow across the breeding range, such that long-distance migrants could still exhibit genetic structure or isolation by distance despite their long seasonal journeys and high dispersal potential. Here, we combine thorough range-wide geographic sampling with whole-genome sequencing to investigate whether genetic structure can be resolved in the Veery (*Catharus fuscescens*), a Nearctic-Neotropical long-distance migratory songbird.

The Veery is an ideal species to test for the presence of spatial genetic structure in migratory birds, given its long-distance migrations and high adult breeding site fidelity (Outlaw et al. 2003; Heckscher et al. 2011; Hobson and Kardynal 2015). The Veery breeds across wet

129 forested habitats of the boreal and the temperate-boreal transition ('hemiboreal') belt, coastal
130 forests of the northeastern U.S. and Canada, the Appalachian Mountains, and riparian canyons in
131 the mountains of western North America (Fig. 1; Heckscher et al. 2020). Previous work has
132 delineated five phenotypic subspecies based on subtle geographic breeding population
133 differences in plumage coloration (Phillips 1991; Pyle 1997), but concordance between genetic
134 data and subspecific designation has not been evaluated.

135 Contemporary breeding ranges of long-distance migratory birds, such as the Veery, are
136 typically found at mid or high latitudes, such that Pleistocene glacial cycles presumably forced
137 populations into fragmented habitat when ice sheets advanced (Hewitt 2004; Svenning et al.
138 2015). Isolation of populations in putative glacial refugia is thought to have generated discrete
139 population structure that is detectable in contemporary populations through measures of genetic
140 diversity and heterozygosity (Bohonak 1999; Weir and Schluter 2004). Indeed, this pattern has
141 been observed in several North American migratory bird species with molecularly distinct
142 populations (e.g., Ruegg and Smith 2002; Barrowclough et al. 2004; Milá et al. 2007; Spellman
143 and Klicka 2007; Manthey et al. 2011; van Els et al. 2012; Winker et al. 2023). Molecular
144 signatures in multiple North American bird species have supported several glacial refugia—
145 which might also have been occupied by the Veery—including south of the glaciers to the east
146 (e.g., southern Appalachian Mountains) and west (e.g., southern Rocky Mountains), and offshore
147 of Newfoundland (e.g., Grand Banks) (Hewitt 2004; Soltis et al. 2006).

148 Here, we employ range-wide genomic sampling to test patterns of genetic differentiation
149 and diversity across the species' range and evaluate the phylogeographic history of the species.
150 The only previous phylogeographic work on this species evaluated mitochondrial differentiation
151 between the eastern and western extremes of the breeding range (Newfoundland versus

Washington), identifying subtle but distinct genetic differentiation (Topp et al. 2013). In this study, we use range-wide sampling and low-coverage whole genome-sequencing (lcWGS) to investigate subtle patterns of spatial genetic differentiation and evaluate the concordance of phenotypic subspecies descriptions with patterns of genetic differentiation. The development of cost-effective lcWGS allows inference based on orders of magnitude more loci than reduced representation genome sequencing (Lou et al. 2021), which might facilitate detection of subtle genetic patterns not otherwise evident (Novembre et al. 2008).

We also assessed contemporary genetic structure across the breeding range in light of historic processes associated with geographic isolation in different refugia versus population expansion from a single glacial refugium (Le Corre and Kremer 1998; Mimura and Aitken 2007; Meirmans 2012; Westram et al. 2013; Wahlsteen et al. 2023). If southern Appalachia and the western regions were historic glacial refugia for the Veery, we predict higher nucleotide diversity and heterozygosity in these populations given their likely long-term population stability as source populations for an expansion into post-glacial higher latitude habitat.

Given that we produced the first detailed phylogeographic study of this species, an additional goal of our study was to use samples from the winter range to identify nonbreeding birds' breeding population of origin. The nonbreeding distribution of this species, which occurs entirely within South America, is poorly understood (Remsen Jr 2001; Heckscher et al. 2020). Veeries are known to exhibit intra-tropical movements during the overwintering period, as observed from geolocation data from populations breeding in Delaware (Heckscher et al. 2011) and British Columbia (Hobson and Kardynal 2015). Individuals spend the early portion of the northern winter in the Amazon basin south of the Amazon River in November–December, before moving northwest to a second wintering site likely in response to seasonal flooding patterns

(Heckscher et al. 2011; Heckscher et al. 2015; Hobson and Kardynal 2015). With limited tracking data available, assessing migratory connectivity between breeding and nonbreeding ranges remains a challenge but is critical for identifying the ecological and conservation links between stages in the annual cycle (Webster et al. 2002; Ambrosini et al. 2019). Therefore, we used our data to determine whether breeding origin can be identified for wintering samples given genetic differentiation across the breeding range. Through this analysis, we tested the utility of our breeding grounds dataset for use in future research centered around migratory connectivity in this species and its poorly understood nonbreeding distributions.

Methods

(a) *Study system and sampling*

We used 121 frozen or ethanol-preserved *C. fuscescens* tissue samples from our institutions' museum collections or provided by other museum collections (Fig. 1; Table S1). We also included 3 blood samples from Newfoundland provided by the New York State Museum (Fig. 1; Table S1). Fieldwork by the authors was approved by our Institutional Animal Care and Use Committees and all relevant permitting authorities (see Acknowledgments). All samples were collected during the breeding season, except for 4 individuals that were collected on their wintering grounds in South America in October–November (Bolivia: $n = 3$, Paraguay: $n = 1$; hereafter, 'nonbreeding birds') that we included to assess migratory connectivity. Our sample size for nonbreeding birds is small but includes most nonbreeding tissue samples available in North American museum collections. Specifically, these samples represent four out of only six available tissue samples from the overwintering period that are published in a compendium of museum collections (www.vertnet.org). Given the sampling dates and locations, the four

198 nonbreeding birds were likely collected on their first wintering site (Heckscher et al. 2011;
199 Heckscher et al. 2015; Hobson and Kardynal 2015).
200 We extracted DNA using DNeasy Blood and Tissue Kits (Qiagen Sciences, Germantown,
201 MD, USA) and prepared libraries for low-coverage whole genome sequencing using a modified
202 Illumina Nextera protocol (Therkildsen and Palumbi 2017; Schweizer et al. 2021). All libraries
203 were sequenced on NovaSeq (200 samples per lane) using services provided by the University of
204 Michigan Advanced Genomics Core.
205
206 *(b) Data processing*
207 We trimmed remaining adaptors and low-quality bases from demultiplexed data with
208 AdapterRemoval v2.3.1 using the –trimns and –trimqualities options (Schubert et al. 2016). We
209 also removed low-quality read ends using fastp v0.23.2 (Chen et al. 2018b) with the --cut_right
210 option to mitigate the potential for batch effects arising from differences between sequencing
211 runs (Lou and Therkildsen 2022). Following trimming steps, samples had a mean of 4.8x
212 coverage of the genome (range= 2.59–28.38 billion bases; 2.3x–25.1x coverage).
213 All samples were confirmed to be tissues from *C. fuscescens* using BLAST in Geneious
214 (v. 2021.2.2) on at least one mitochondrial gene from each individual as described in a previous
215 study (Kimmitt et al. 2023). As a chromosome-assembled genome of *C. fuscescens* was not
216 available, we aligned all samples to a reference genome of a close relative, *C. ustulatus*
217 (GenBank assembly accession number GCA_009819885.2bCatUst1.pri.v2, coverage = 60.58x)
218 using bwa mem (Li and Durbin 2010) and Samtools (Li et al. 2009). We removed overlapping
219 reads using clipOverlap in bamUtil (Jun et al. 2015), marked duplicate reads with
220 MarkDuplicates, and assigned all reads to a new read group with AddOrReplaceReadGroups

using picard (http://broadinstitute.github.io/picard/). All bam files were then indexed using Samtools (Li et al. 2009). The mean mapping rate across all samples used in analyses was 97.43% (range 93.98–98.43%). We then used GATK v3.7 (Van der Auwera et al. 2013) to re-align samples around indels by applying RealignerTargetCreator to the entire dataset and using IndelRealigner for each sample.

Genotype likelihoods from low-coverage sequencing data were calculated using the GATK model in ANGSD v0.9.40 (Korneliussen et al. 2014). Given the genotype uncertainty associated with low-coverage sequencing, all results were analyzed in a genotype likelihood framework, as this method uses probability-based inference to account for sequencing error (Korneliussen et al. 2014; Lou et al. 2021). Parameters used for each ANGSD analysis are described further below or detailed in Table S2.

(c) Population structure

We calculated genotype likelihoods for all sites with a SNP p-value < 0.05 across the entire genome using ANGSD (Table S1). We then filtered mis-mapped or paralogous SNPs out of the dataset using ngsParalog v1.3.2 (https://github.com/tplinderoth/ngsParalog; Linderoth 2018). ngsParalog is designed for low-coverage sequencing data and implements a likelihood method to find mapping problems.

We used PCAngsd v1.10 (Meisner and Albrechtsen 2018) to conduct Principal Component Analyses (PCA) to visually assess spatial genetic structure. As PCA can be sensitive to genomic inversions that could obscure geographic structure (Novembre et al. 2008; Tian et al. 2008; Novembre and Peter 2016), we first ran PCAngsd separately for each chromosome using all 124 samples to identify possible inversions. At least six chromosomes exhibited evidence of

clustering associated with putative inversions, so we analyzed each chromosome further for inversions using lostruct (Li and Ralph 2019) as implemented using PCAngsd with scripts available from https://github.com/alxsimon/local_pcangsd. All microchromosomes with evidence of inversions (n = 8) as well as all sex chromosomes were removed from the dataset. For the remaining chromosomes, we then ran PCAngsd with the --admix option to estimate admixture proportions using a non-negative matrix factorization algorithm so that we could produce genome-wide PCAs and admixture plots. Two individuals sampled from Nova Scotia had an aberrantly high PCA covariance (> 0.2) such that they were visual outliers on the PCA (see Fig. S1); therefore, we excluded one of these individuals from the final PCAs to better facilitate visual assessment of range-wide structure patterns.

We implemented the *find.clusters* function from the R package *adegenet* (Jombart et al. 2010) using the covariance matrix produced by PCAngsd; *find.clusters* runs successive *K*-means with an increasing number of clusters (*K*) and then performs a goodness of fit analysis (BIC) to identify the optimal *K*. We also used a Mantel test in the *ade4* package v. 1.7–19 (Thioulouse and Dray 2007), with 1000 permutations, to determine whether genetic distance (using the proxy 1 – PCA covariance; Novembre et al. 2008) varied significantly with geographic distance between samples. As a continuous population genetic analysis, Mantel tests do not rely on pre-assigned population clusters, such that we could investigate both discrete and continuous population structure.

We calculated pairwise F_{ST} between three distinct populations that were revealed by the PCA-based clustering methods mentioned above (see also Results and Figs. 1-2): (1) “western” (i.e., western United States including Washington, Oregon, Idaho, and Colorado; n = 22), (2) “southern Appalachian” (i.e., West Virginia and North Carolina, n = 29), and (3) “boreal” (i.e.,

Canada from Alberta to Newfoundland, Western Great Lake states and Northeastern United States; $n = 69$) (Fig. 1). To make our F_{ST} calculations computationally tractable, we randomly downsampled the reference genome, including SNPs and invariant sites, to create a set of loci that consisted of stretches of 2 kb loci at least 10 kb apart (yielding approximately 12% of the whole genome). We used scripts modified from https://github.com/markravinet/genome_sampler and excluded loci from regions flagged by the inversion filters. Sites flagged by ngsParalog were also removed from the subsampled dataset and stored the loci in a BED file. We generated a site allele frequency (SAF) file in ANGSD with the -doSaf parameter and -sites filter to include only subsampled loci. We used winsfs (Rasmussen et al. 2022) to create 2-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 (i.e., Bhatia estimator) in ANGSD to calculate pairwise F_{ST} between unbalanced sample sizes.

Finally, to assess the direction of gene flow between the three populations, we calculated a directionality index (ψ) from the 2D SFS with a custom script from (Adams et al. 2023) using equation 1b from (Peter and Slatkin 2013). Balanced sample sizes are necessary to calculate ψ (Peter and Slatkin 2013), such that we randomly selected 22 individuals three times from both the southern Appalachian and the boreal populations to created new SAF and 2D SFS files between each population pair for a total of 15 SFS files.

(d) Genetic diversity and heterozygosity

Genetic diversity measure, pairwise $\theta_{\pi,}$ requires pre-assigned populations for analysis. Therefore, for each of the three populations identified by the clustering analysis above (western, boreal, and southern Appalachian), we estimated population-level summary statistics for genetic

diversity from the subsampled loci using ANGSD and winsfs. Since the sample size of the more geographically expansive boreal population was much larger than the other two populations, we randomly selected 30 individuals from the boreal population for population-level genetic diversity analyses. We generated a SAF using only subsampled loci in ANGSD that excluded flagged sites by ngsParalog and microchromosomes with detected putative inversions. We used winsfs to produce and fold a population-level 1-dimensional (1D) SFS. Pairwise θ_π was calculated for each chromosome separately using the saf2theta and thetaStat functions in ANGSD. We compared θ_π by chromosome among populations using a one-way analysis of variance (ANOVA). For each population we also calculated the total pairwise π / total number of sites across all chromosomes.

Individual-level heterozygosity was also estimated by creating individual-level SAF files with the same subsampled loci used in genetic diversity estimates. We then used these SAF files with winsfs to generate individual 1D SFS. Individual heterozygosity was calculated as the number of polymorphic sites divided by the total sites in each individual's 1D SFS (Kersten et al. 2021). Exploratory analyses suggested that samples with very low (< 4x) genomic coverage exhibited low individual-level heterozygosity relative to samples above 4x coverage. Therefore, we filtered all samples with less than 4x coverage (n = 35) out of the dataset for this analysis (retaining samples sizes of n = 15 for western U.S., n = 56 for boreal, and n = 14 for Southern Appalachian regions). We then compared population differences in individual heterozygosity using ANOVA and Tukey multiple pairwise comparisons. Since the three populations span large geographic ranges, we also tested for the presence of gradients of heterozygosity across latitude or longitude using linear models: 1) a western gradient across latitude including samples from Alberta and the western population, 2) a boreal gradient across longitude, and 3) an eastern

gradient across latitude including samples from the southern Appalachian population as well as PA, OH, VT, Nova Scotia, and Newfoundland (Fig. 1).

Results

(a) Population structure

The PCA pattern indicated isolation by distance, as the shape of the PCA reflects a sinusoidal curve typical of continuous structure (Fig. 2A; Novembre and Stephens 2008). The Mantel test confirmed isolation by distance, as genetic distance (1 – PCA covariance) was positively associated with geographic distance (Pearson's correlation coefficient, $r = 0.38$, $p = 0.001$) using samples across the full breeding range (Fig. 3A). We also visually inferred two distinct population clusters in the range-wide dataset of breeding individuals ($n = 120$) in the PCA analysis, such that individuals from the southern Appalachian sampling regions (North Carolina and West Virginia) clustered separately from all other individuals, demonstrating a genetic break between the southern Appalachian samples and the northern Appalachian samples (i.e., Pennsylvania, Ohio, and Vermont; Fig. 2A). We confirmed that two clusters were the best fit for the data using the *find.clusters* tool. The admixture plot ($K = 2$) showed a gradual shift in population ancestry across the geographic range.

We next removed southern Appalachian samples from the dataset to determine if we could detect finer-scale genetic structure in the more genetically similar samples from western and boreal sampling regions. In the boreal and western populations dataset, the relationship between genetic distance and geographic distance was stronger than in the range-wide dataset ($r = 0.78$, $p = 0.001$; Fig. 3B). Without the southern Appalachian samples included, we also found that the western population was distinct from a boreal population, which also includes the

336 northern Appalachian samples (Fig. 2B) and confirmed that two clusters was the best fit for this
337 subset of the data using *find.clusters*. Finally, we ran a PCA on the boreal genetic population,
338 which includes northern Appalachian samples, to determine if subpopulations would be
339 detectable on a further reduced geographic scale. We visually noted that 3 out of 4 of the samples
340 from Newfoundland sorted separately on the PCA, suggesting that this isolated population could
341 be distinct from other boreal populations. However, *find.clusters* did not assign distinct clusters
342 associated with geography within the boreal population samples, consistent with the observed
343 overlap among sampling regions within the PCA (Fig. 2). Finally, the relationship between
344 genetic distance and geographic distance was weakest in the boreal population samples only (r =
345 0.22, p = 0.001).
346 Based on the PCA and clustering results, we conducted analyses of population
347 differentiation and genetic diversity (next section) using three identified populations across the
348 sampling regions (Fig. 1): southern Appalachian (i.e., WV, NC), western US (i.e., WA, OR, ID,
349 and CO; hereafter the 'western' population), and the boreal belt (Alberta to Newfoundland)
350 including the northern Appalachians (i.e., PA, OH, VT), hereafter the 'boreal' population.
351 Pairwise F_{ST} values were < 0.02 between all three populations, indicating low levels of
352 population differentiation. Weighted pairwise population-level F_{ST} was 0.008 between the boreal
353 and southern Appalachian population and 0.006 between the boreal and western populations. F_{ST}
354 was highest between the southern Appalachian and western populations (0.014).
355 Finally, the directionality index was low (ψ < 0.05) for all pairwise population
356 comparisons (Table S3). Since ψ is positive, this might indicate that the boreal population has
357 been a source population for expansion; however, these values were not significantly different
358 from zero, supporting an isolation by distance model or a population expansion model in which

populations are equidistant from the origin of expansion and are exhibiting comparable levels of gene flow between populations (Peter and Slatkin 2013; Adams et al. 2023).

(b) Genetic diversity and heterozygosity

Neither nucleotide diversity (pairwise θ_π) estimated per chromosome nor individual-level heterozygosity differed significantly between populations (Table 1; Fig. S2A: $F_{2,87} = 0.02$, $p = 0.982$; Fig. S2B: $F_{2,81} = 1.09$, $p = 0.341$). Across a latitudinal gradient in the montane west, heterozygosity scaled positively with latitude (Fig. 4A). Across a longitudinal gradient in the boreal belt, western samples had significantly higher heterozygosity than eastern samples (Fig. 4B). Finally, there were no significant latitudinal differences across the eastern latitudinal gradient from North Carolina to Newfoundland (Table 2; Fig. 4C).

(c) Breeding population assignment for nonbreeding samples

We leveraged our thorough sampling of the breeding range to assess the likely breeding populations for the 4 nonbreeding samples from South America based on their location in a PCA of all individuals. One nonbreeding bird (collected in Bolivia in November) clustered with the Appalachian breeding samples from West Virginia, and the remaining three individuals (collected in Bolivia in November or Paraguay in October) clustered with boreal breeding samples (Fig. 2A). A PCA containing only the boreal individuals (Fig. 2C) suggested that one nonbreeding sample likely originated from either Manitoba or the Western Great Lakes, whereas the other two samples associate with the Western Great Lakes or the northern Appalachians (i.e., VT, PA, or OH). However, without distinct clusters in the boreal-only analysis, we refrain from

confidently assigning these nonbreeding samples to breeding populations more specific than the broader boreal population.

Discussion

We found evidence of isolation by distance across the breeding range of the Veery, a long-distance migratory songbird, as well as population clustering of the western, boreal, and southern Appalachian sampling regions. PCA revealed geographically nested patterns of genetic clustering (Fig. 2) and a pattern of genetic covariance between individual samples that decayed with geographic distance (Fig. 3). Our results suggest that breeding site fidelity, which acts to temper natal and breeding dispersal distances, appears to be sufficiently strong to yield spatial genetic structure in the absence of extrinsic barriers to dispersal.

Previous phenotypic assessments of the Veery described five subspecies based on plumage color variation associated with the following regions: (1) Newfoundland and central Quebec (*C. f. fuliginosus*), (2) the eastern United States and Canada (including all Appalachian populations; *C. f. fuscescens*), (3) Great Plains of Canada and western Great Lakes (*C. f. levyi*), (4) British Columbia and the Rocky Mountains (*C. f. salicicolus*), and the (5) western United States east of the Cascade Mountains (*C.f. subpallidus*) (Heckscher et al. 2020). Identification of phenotypic subspecies, however, has been disputed, as geographic variation in plumage is obscured by individual variation (Pyle 1997). Alternate phenotypic subspecific delineation includes a southern Appalachian subspecies from Georgia to West Virginia (*C.f. pulichorum*) and excludes *C.f. subpallidus* (Pyle 1997). Our genetic results are not fully consistent with the boundaries of these phenotypically described subspecies, as we found only three differentiated populations across the range of the Veery, with the southern Appalachian population the most

404 distinct. The boreal and northern Appalachian PCA (Fig. 2c) revealed that 3 of our 4 samples
405 from Newfoundland clustered together separately from the other boreal samples, suggesting a
406 subtle genetic difference in that sampling region. Although our quantitative analysis of
407 population clustering did not support Newfoundland samples as a discrete population, it is
408 possible that increased sampling from the Maritime Provinces of Canada would bolster detection
409 of a distinct genetic cluster associated with the subspecies described from this region.
410 Additionally, we note that low sampling density in western Canada could potentially produce the
411 pattern of discrete phylogeographic clustering between the boreal and western U.S. populations,
412 rather than a continuous pattern of genetic structure between western and boreal populations, if
413 admixed individuals occur between our sampling in central Alberta and the northwest U.S.
414 Nevertheless, we conclude that the genetic structure detected in our study does not align with the
415 phenotypically described subspecies, such that phenotypic differences are unlikely driven by
416 historical population isolation and differentiation (Zamudio et al. 2016). Instead, subtle plumage
417 differences across the range could reflect local selection on a small number of plumage genes
418 without genome-wide divergence (e.g., McCormack et al. 2012; Toews et al. 2016) or
419 phenotypic plasticity in response to environmental conditions (e.g., Mason and Taylor 2015;
420 López-Rull et al. 2023).

421 Our data also allowed us to determine the general breeding origins of the very few
422 wintering site genetic samples available. Understanding migratory connectivity—the geographic
423 links between wintering, stopover, and breeding sites—is critical (Webster and Marra 2005;
424 Marra et al. 2006; Somveille et al. 2021), as conditions on the wintering grounds can have carry-
425 over effects on breeding season fitness (Norris and Taylor 2006; Harrison et al. 2011; Ambrosini
426 et al. 2019). Individual tracking can reveal movement patterns across the annual cycle

(Stutchbury et al. 2009; Fraser et al. 2012; Batbayar et al. 2021; Rushing et al. 2021), but is both time intensive and accompanied by several challenges associated with sample size and data recovery (Ruegg et al. 2017). The Veery's complex movements between two wintering regions in the tropics (Heckscher et al. 2011; Heckscher et al. 2015; Hobson and Kardynal 2015) add another challenge to using tracking information to identify the breeding population of an individual. Genetic data from whole-genome sequencing has been used previously to identify an individual's population of origin (e.g., Manel et al. 2002; Nielsen et al. 2009; Hess et al. 2011; Ruegg et al. 2014) and might be a robust alternative method to tracking methods, as it is cost effective at a large scale and can be used to detect subtle breakpoints in continuous population structure (Turbek et al. 2023). Using PCA to identify putative population of origin of nonbreeding samples, we identified one individual from the southern Appalachian population and three individuals from the boreal population (Fig. 2). This clustering of nonbreeding samples with breeding samples allowed putative regional breeding assignment despite absence of distinct genetic clusters within the boreal population. Population assignment is typically conducted using a panel of genetic markers or loci that consistently differ between distinct populations (Veale et al. 2012; Chen et al. 2018a; Sylvester et al. 2018); however, these techniques are ineffectual across wide ranges without pronounced population structure, such as the boreal forest belt for the Veery. By combining lcWGS with range-wide sampling, PCAs can detect finer structure, such that regional breeding area assignment might be possible in regions with high gene flow.

We also evaluated geographic patterns of genetic diversity to test whether contemporary genetic patterns reflect historic isolation of populations in glacial refugia. Phylogeographic hypotheses have suggested that populations geographically closer to putative refugia (i.e., source

450 populations) should harbour higher levels of genetic diversity due to founder effects and greater
451 geographic isolation (i.e., 'southern richness and northern purity' hypothesis) (Hewitt 1999;
452 Excoffier 2004; Eckert et al. 2008; Provan and Bennett 2008; Excoffier et al. 2009). We
453 therefore hypothesized that if the southern Rockies and southern Appalachia were glacial refugia
454 for the Veery, we would detect lower genetic diversity in the boreal population in comparison.
455 Using cluster-based analyses, we first found that the western, boreal, and Appalachian
456 populations did not differ in any measures of genetic diversity (Table 1, Fig. S2), inconsistent
457 with this hypothesis (Hewitt 2004; Provan and Bennett 2008; Ralston et al. 2021). Geographic
458 gradients in genetic diversity, however, might further show signatures of range expansion
459 dynamics (Provan and Bennett 2008; Peter and Slatkin 2015; Adams et al. 2023). We found that
460 individual heterozygosity was positively correlated with latitude across the western montane
461 region and negatively with longitude across the boreal forest belt (Fig. 4**)**. These patterns also do
462 not align with expected patterns of higher genetic diversity in the south associated with
463 northward postglacial expansion (Miller et al. 2020; Adams et al. 2023). However, comparable
464 or higher genetic diversity has also been observed at the leading expansion front (Vandepitte et
465 al. 2017; Wang et al. 2017; Bors et al. 2019) likely due to continued high gene flow with the
466 source population (Miller et al. 2020; Adams et al. 2023). Expansions that occur at a rapid pace
467 are also likely to retain higher heterozygosity at the expansion front (Goodsman et al. 2014).
468 Therefore, our heterozygosity results may alternatively provide weak support for rapid
469 expansions out of western and northeastern refugia. The directionality index, however, was close
470 to zero between all pairwise comparison ($\psi < 0.05$), suggesting that the data might better fit an
471 isolation by distance rather than expansion model (Peter and Slatkin 2013; Adams et al. 2023).
472 Ultimately, our results do not provide compelling evidence for a glacial refugium in

Newfoundland or the southern Rockies, because the subtle patterns found are also consistent with continuous processes of gene flow between populations across the range.

In conclusion, we were able to resolve detailed spatial genetic structure in the Veery despite the high dispersal potential in this species, and we observed evidence for both continuous and discontinuous structure across the range. Given the resolution that we achieved through low-coverage, whole-genome sequencing and range-wide sampling, we were also able to assign regions of origin to individuals collected on their wintering grounds, which has important implications for assessing migratory connectivity at a larger scale than enabled by traditional tracking methods. Finally, based on the patterns of population differentiation and genetic diversity in this species, we conclude that gene flow, isolation by distance, and site fidelity likely play a more important role in shaping current population genetic structure and diversity in this species than historic isolation.

Literature Cited

Adams, N. E., R. R. Bandivadekar, C. Battey, M. W. Clark, K. Epperly, K. Ruegg, L. A. Tell, and R. A. Bay (2023). Widespread gene flow following range expansion in Anna's Hummingbird. *Molecular Ecology* 32:3089–3101.

Ambrosini, R., A. Romano, and N. Saino (2019). Changes in migration, carry-over effects, and migratory connectivity. *Effects of climate change on birds*:93–107.

Barrowclough, G. F., J. G. Groth, L. A. Mertz, and R. Gutiérrez (2004). Phylogeographic structure, gene flow and species status in Blue Grouse (*Dendragapus obscurus*). *Molecular Ecology* 13:1911–1922.

Batbayar, N., K. Yi, J. Zhang, T. Natsagdorj, I. Damba, L. Cao, and A. D. Fox (2021). Combining tracking and remote sensing to identify critical year-round site, habitat use and migratory connectivity of a threatened waterbird species. *Remote Sensing* 13:4049.

Bohonak, A. J. (1999). Dispersal, gene flow, and population structure. *The Quarterly review of biology* 74:21–45.

Bors, E. K., S. Herrera, J. A. Morris Jr, and T. M. Shank (2019). Population genomics of rapidly invading lionfish in the Caribbean reveals signals of range expansion in the absence of spatial population structure. *Ecology and Evolution* 9:3306–3320.

Chen, K. Y., E. A. Marschall, M. G. Sovic, A. C. Fries, H. L. Gibbs, and S. A. Ludsin (2018a). 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.

Chen, S., Y. Zhou, Y. Chen, and J. G. Gu (2018b). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34:i884–i890.

Claramunt, S. (2021). Flight efficiency explains differences in natal dispersal distances in birds. *Ecology* 102:e03442.

Claramunt, S., E. P. Derryberry, J. 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* 279:1567–1574.

Eckert, C. G., K. E. Samis, and S. C. Lougheed (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* 17:1170–1188.

518 Edwards, S. V., A. J. Shultz, and S. C. Campbell-Staton (2015). Next-generation sequencing and
519 the expanding domain of phylogeography. *Folia Zoologica* 64:187–206.
520 Everson, K. M., J. F. McLaughlin, I. A. Cato, M. M. Evans, A. R. Gastaldi, K. K. Mills, K. G.
521 Shink, S. M. Wilbur, and K. Winker (2019). Speciation, gene flow, and seasonal
522 migration in *Catharus* thrushes (Aves: Turdidae). *Molecular Phylogenetics and Evolution*
523 139:106564.
524 Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range
525 expansion: lessons from the infinite-island model. *Molecular Ecology* 13:853–864.
526 Excoffier, L., M. Foll, and R. J. Petit (2009). Genetic consequences of range expansions. *Annual*
527 *Review of Ecology, Evolution, and Systematics* 40:481–501.
528 Fraser, K. C., B. J. Stutchbury, C. Silverio, P. M. Kramer, J. Barrow, D. Newstead, N. Mickle, B.
529 F. Cousens, J. C. Lee, and D. M. Morrison (2012). Continent-wide tracking to determine
530 migratory connectivity and tropical habitat associations of a declining aerial insectivore.
531 *Proceedings of the Royal Society B: Biological Sciences* 279:4901–4906.
532 Goodsman, D. W., B. Cooke, D. W. Coltman, and M. A. Lewis (2014). The genetic signature of
533 rapid range expansions: How dispersal, growth and invasion speed impact heterozygosity
534 and allele surfing. *Theoretical population biology* 98:1–10.
535 Harrison, X. A., J. D. Blount, R. Inger, D. R. Norris, and S. Bearhop (2011). Carry-over effects
536 as drivers of fitness differences in animals. *Journal of Animal Ecology* 80:4–18.
537 Heckscher, C. M., M. R. Halley, and P. M. Stampul (2015). Intratropical migration of a Nearctic-
538 Neotropical migratory songbird (*Catharus fuscescens*) in South America with
539 implications for migration theory. *Journal of Tropical Ecology* 31:285–289.

Heckscher, C. M., S. M. Taylor, J. W. Fox, and V. Afanasyev (2011). Veery (*Catharus fuscescens*) wintering locations, migratory connectivity, and a revision of its winter range using geolocator technology. *The Auk* 128:531–542.

Heckscher, C. M., L. R. Bevier, A. F. Poole, W. Moskoff, P. Pyle, and M. A. Patten. 2020. Veery (*Catharus fuscescens*), version 1.0. In Birds of the World. (P. G. Rodewald, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

Hess, J., A. Matala, and S. Narum (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular ecology resources* 11:137–149.

Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68:87–112.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359:183–195.

Hobson, K. A., and K. J. Kardynal (2015). Western Veeries use an eastern shortest-distance pathway: New insights to migration routes and phenology using light-level geolocators. *The Auk: Ornithological Advances* 132:540–550.

Jombart, T., S. Devillard, and F. Balloux (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics* 11:1–15.

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.

563 Kersten, O., B. Star, D. M. Leigh, T. Anker-Nilssen, H. Strøm, J. Danielsen, S. Descamps, K. E.
564 Erikstad, M. G. Fitzsimmons, and J. Fort (2021). Complex population structure of the
565 Atlantic Puffin revealed by whole genome analyses. *Communications Biology* 4:922.
566 Kimmitt, A. A., T. M. Pegan, A. W. Jones, K. S. Wacker, C. L. Brennan, J. Hudon, J. J.
567 Kirchman, K. Ruegg, B. W. Benz, and R. Herman (2023). Genetic evidence for
568 widespread population size expansion in North American boreal birds prior to the Last
569 Glacial Maximum. *Proceedings of the Royal Society B: Biological Sciences*
570 290:20221334.
571 Korneliussen, T. S., A. Albrechtsen, and R. Nielsen (2014). ANGSD: analysis of next generation
572 sequencing data. *BMC bioinformatics* 15:1–13.
573 Le Corre, V., and A. Kremer (1998). Cumulative effects of founding events during colonisation
574 on genetic diversity and differentiation in an island and stepping-stone model. *Journal of*
575 *evolutionary biology* 11:495–512.
576 Li, H., and R. Durbin (2010). Fast and accurate long-read alignment with Burrows–Wheeler
577 transform. *Bioinformatics* 26:589–595.
578 Li, H., and P. Ralph (2019). Local PCA shows how the effect of population structure differs
579 along the genome. *Genetics* 211:289–304.
580 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
581 Durbin, and G. P. D. P. Subgroup (2009). The sequence alignment/map format and
582 SAMtools. *Bioinformatics* 25:2078–2079.
583 Linderoth, T. 2018. Identifying population histories, adaptive genes, and genetic duplication
584 from population-scale next generation sequencing. University of California, Berkeley.

López-Rull, I., C. Salaberría, and J. A. Fargallo (2023). Plastic plumage colouration in response to experimental humidity supports Gloger's rule. *Scientific reports* 13:858.

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.

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.

Manel, S., P. Berthier, and G. Luikart (2002). Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. *Conservation Biology* 16:650–659.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet (2003). Landscape genetics: combining landscape ecology and population genetics. *Trends in ecology & evolution* 18:189–197.

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

Marra, P. P., D. Norris, S. Haig, M. Webster, J. Royle, K. Crooks, and M. Sanjayan (2006). Migratory connectivity. In Conservation Biology Series, vol. 14, Cambridge, UK:157.

Mason, N. A., and S. A. Taylor (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a H olarctic songbird. *Molecular Ecology* 24:3009–3025.

McCormack, J. E., J. M. Maley, S. M. Hird, E. P. Derryberry, G. R. Graves, and R. T. Brumfield (2012). Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular Phylogenetics and Evolution* 62:397–406.

Medina, I., G. M. Cooke, and T. J. Ord (2018). Walk, swim or fly? Locomotor mode predicts genetic differentiation in vertebrates. *Ecology Letters* 21:638–645.

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology* 21:2839–2846.

Meisner, J., and A. Albrechtsen (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 210:719–731.

Milá, B., J. E. McCormack, G. Castañeda, R. K. Wayne, and T. B. Smith (2007). Recent postglacial range expansion drives the rapid diversification of a songbird lineage in the genus *Junco*. *Proceedings of the Royal Society B: Biological Sciences* 274:2653–2660.

Miller, T. E., A. L. Angert, C. D. Brown, J. A. Lee-Yaw, M. Lewis, F. Lutscher, N. G. Marculis, B. A. Melbourne, A. K. Shaw, and M. Szűcs (2020). Eco-evolutionary dynamics of range expansion. *Ecology* 101:e03139.

Mimura, M., and S. Aitken (2007). Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity* 99:224–232.

Nielsen, E. E., J. Hemmer-Hansen, P. F. Larsen, and D. Bekkevold (2009). Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* 18:3128–3150.

Norris, D. R., and C. M. Taylor (2006). Predicting the consequences of carry-over effects for migratory populations. *Biology Letters* 2:148–151.

628 Novembre, J., and M. Stephens (2008). Interpreting principal component analyses of spatial
629 population genetic variation. *Nature genetics* 40:646–649.
630 Novembre, J., and B. M. Peter (2016). Recent advances in the study of fine-scale population
631 structure in humans. *Current opinion in genetics & development* 41:98–105.
632 Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S.
633 Bergmann, and M. R. Nelson (2008). Genes mirror geography within Europe. *Nature*
634 456:98–101.
635 Outlaw, D. C., G. Voelker, B. Mila, and D. J. Girman (2003). Evolution of long-distance
636 migration in and historical biogeography of *Catharus* thrushes: a molecular phylogenetic
637 approach. *The Auk* 120:299–310.
638 Paradis, E., S. R. Baillie, W. J. Sutherland, and R. D. Gregory (1998). Patterns of natal and
639 breeding dispersal in birds. *Journal of Animal Ecology* 67:518–536.
640 Peter, B. M., and M. Slatkin (2013). Detecting range expansions from genetic data. *Evolution*
641 67:3274–3289.
642 Peter, B. M., and M. Slatkin (2015). The effective founder effect in a spatially expanding
643 population. *Evolution* 69:721–734.
644 Phillips, A. (1991). The known birds of North and Middle America, Part II: Bombycillidae;
645 Sylviidae to Sturnidae; Vireonidae. *Published by the author, Denver, Colorado*.
646 Provan, J., and K. D. Bennett (2008). Phylogeographic insights into cryptic glacial refugia.
647 *Trends in ecology & evolution* 23:564–571.
648 Pyle, P. 1997. Identification guide to North American birds: a compendium of information on
649 identifying, ageing, and sexing" near-passerines" and passerines in the hand. Slate Creek
650 Press, Bolinas, CA.

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:ukab018.

Rasmussen, M. S., G. Garcia-Erill, T. S. Korneliussen, 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.

Remsen Jr, J. (2001). True winter range of the Veery (*Catharus fuscescens*): lessons for determining winter ranges of species that winter in the tropics. *The Auk* 118:838–848.

Ruegg, K. C., and T. B. Smith (2002). Not as the crow flies: a historical explanation for circuitous migration in Swainson's Thrush (*Catharus ustulatus*). *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269:1375–1381.

Ruegg, K. C., E. C. Anderson, R. J. Harrigan, K. L. Paxton, J. F. Kelly, F. Moore, and T. B. Smith (2017). Genetic assignment with isotopes and habitat suitability (GAIAH), a migratory bird case study. *Methods in Ecology and Evolution* 8:1241–1252.

Ruegg, K. C., E. C. Anderson, K. L. Paxton, V. Apkenas, S. Lao, R. B. Siegel, D. F. DeSante, F. Moore, and T. B. Smith (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology* 23:5726–5739.

Rushing, C. S., A. M. Van Tatenhove, A. Sharp, V. Ruiz-Gutierrez, M. C. Freeman, P. W. Sykes Jr, A. M. Given, and T. S. Sillett (2021). Integrating tracking and resight data enables unbiased inferences about migratory connectivity and winter range survival from archival tags. *The Condor* 123:duab010.

Schubert, M., S. Lindgreen, and L. Orlando (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* 9:1–7.

674 Schweizer, T. M., M. G. DeSaix, and K. C. Ruegg (2021). LI-Seq: A Cost-Effective, Low Input
675 DNA method for Whole Genome Library Preparation. *bioRxiv*:2021.2007.2006.451326.
676 Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*
677 236:787–792.
678 Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis (2006). Comparative
679 phylogeography of unglaciated eastern North America. *Molecular Ecology* 15:4261–
680 4293.
681 Somveille, M., R. A. Bay, T. B. Smith, P. P. Marra, and K. C. Ruegg (2021). A general theory of
682 avian migratory connectivity. *Ecology Letters* 24:1848–1858.
683 Spellman, G. M., and J. Klicka (2007). Phylogeography of the white-breasted nuthatch (*Sitta*
684 *carolinensis*): diversification in North American pine and oak woodlands. *Molecular*
685 *Ecology* 16:1729–1740.
686 Stutchbury, B. J., S. A. Tarof, T. Done, E. Gow, P. M. Kramer, J. Tautin, J. W. Fox, and V.
687 Afanasyev (2009). Tracking long-distance songbird migration by using geolocators.
688 *Science* 323:896–896.
689 Svenning, J.-C., W. L. Eiserhardt, S. Normand, A. Ordonez, and B. Sandel (2015). The influence
690 of paleoclimate on present-day patterns in biodiversity and ecosystems. *Annual Review of*
691 *Ecology, Evolution, and Systematics* 46:551–572.
692 Sylvester, E. V., P. Bentzen, I. R. Bradbury, M. Clément, J. Pearce, J. Horne, and R. G. Beiko
693 (2018). Applications of random forest feature selection for fine-scale genetic population
694 assignment. *Evolutionary applications* 11:153–165.

695 Therkildsen, N. O., and S. R. Palumbi (2017). Practical low-coverage genomewide sequencing
696 of hundreds of individually barcoded samples for population and evolutionary genomics
697 in nonmodel species. *Molecular ecology resources* 17:194–208.
698 Thioulouse, J., and S. Dray (2007). Interactive multivariate data analysis in R with the ade4 and
699 ade4TkGUI packages. *Journal of Statistical Software* 22:1–14.
700 Tian, C., R. M. Plenge, M. Ransom, A. Lee, P. Villoslada, C. Selmi, L. Klareskog, A. E. Pulver,
701 L. Qi, and P. K. Gregersen (2008). Analysis and application of European genetic
702 substructure using 300 K SNP information. *PLoS genetics* 4:e4.
703 Toews, D. P., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, and I. J.
704 Lovette (2016). Plumage genes and little else distinguish the genomes of hybridizing
705 warblers. *Current Biology* 26:2313–2318.
706 Topp, C. M., C. L. Pruett, K. G. McCracken, and K. Winker (2013). How migratory thrushes
707 conquered northern North America: a comparative phylogeography approach. *PeerJ*
708 1:e206.
709 Turbek, S. P., W. C. Funk, and K. C. Ruegg (2023). Where to draw the line? Expanding the
710 delineation of conservation units to highly mobile taxa. *Journal of Heredity* 114:300–
711 311.
712 Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine,
713 T. Jordan, K. Shakir, D. Roazen, and J. Thibault (2013). From FastQ data to high-
714 confidence variant calls: the genome analysis toolkit best practices pipeline. *Current*
715 *protocols in bioinformatics* 43:11.10.11–11.10.33.

van Els, P., C. Cicero, and J. Klicka (2012). High latitudes and high genetic diversity: phylogeography of a widespread boreal bird, the Gray Jay (*Perisoreus canadensis*). *Molecular Phylogenetics and Evolution* 63:456–465.

Vandepitte, K., K. Helsen, K. Van Acker, J. Mergeay, and O. Honnay (2017). Retention of gene diversity during the spread of a non-native plant species. *Molecular Ecology* 26:3141–3150.

Veale, A., M. Clout, and D. Gleeson (2012). Genetic population assignment reveals a long-distance incursion to an island by a stoat (*Mustela erminea*). *Biological Invasions* 14:735–742.

Wahlsteen, E., E. V. Avramidou, G. Bozic, R. M. Mediouni, B. Schuldt, and H. Sobolewska (2023). Continental-wide population genetics and post-Pleistocene range expansion in field maple (*Acer campestre* L.), a subdominant temperate broadleaved tree species. *Tree Genetics & Genomes* 19:15.

Wang, E., R. E. Van Wijk, M. S. Braun, and M. Wink (2017). Gene flow and genetic drift contribute to high genetic diversity with low phylogeographical structure in European Hoopoes (*Upupa epops*). *Molecular Phylogenetics and Evolution* 113:113–125.

Webster, M. S., and P. P. Marra (2005). The importance of understanding migratory connectivity and seasonal interactions. *Birds of Two Worlds: The Ecology and Evolution of Temperate-Tropical Migration.*

Webster, M. S., P. P. Marra, S. M. Haig, S. Bensch, and R. T. Holmes (2002). Links between worlds: unraveling migratory connectivity. In Trends in ecology & evolution, vol. 17. Johns Hopkins University Press, Baltimore, MD 2:76–83.

738 Weir, J. T., and D. Schluter (2004). Ice sheets promote speciation in boreal birds. *Proceedings of*
739 *the Royal Society of London. Series B: Biological Sciences* 271:1881–1887.
740 Westram, A. M., J. Jokela, and I. Keller (2013). Hidden biodiversity in an ecologically important
741 freshwater amphipod: differences in genetic structure between two cryptic species. *PLoS*
742 *One* 8:e69576.
743 Winger, B. M., G. G. Auteri, T. M. Pegan, and B. C. Weeks (2019). A long winter for the Red
744 Queen: rethinking the evolution of seasonal migration. *Biological Reviews* 94:737–752.
745 Winker, K., J. J. Withrow, D. D. Gibson, and C. L. Pruett (2023). Beringia as a high-latitude
746 engine of avian speciation. *Biological Reviews*.
747 Zamudio, K. R., R. C. Bell, and N. A. Mason (2016). Phenotypes in phylogeography: Species'
748 traits, environmental variation, and vertebrate diversification. *Proceedings of the*
749 *National Academy of Sciences* 113:8041–8048.

750
751

Figures

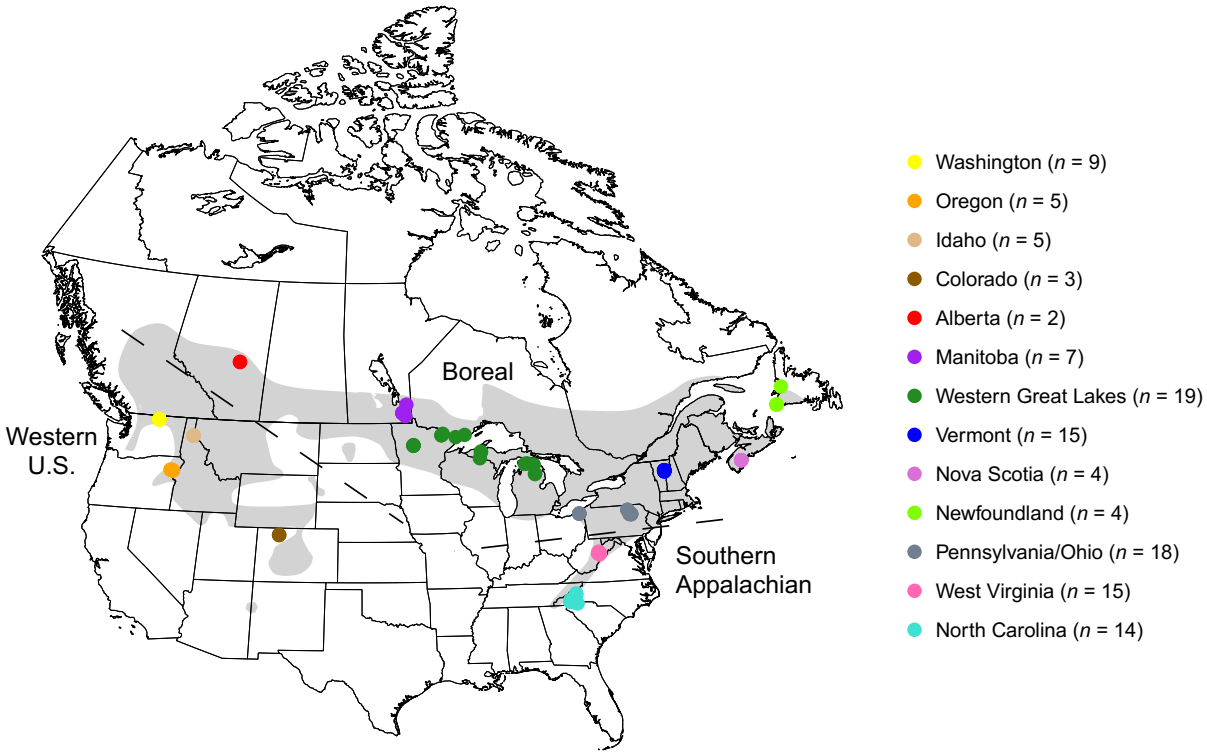


Figure 1. Map of sampling locations for the Veery (*Catharus fuscescens*) in North America. The approximate breeding range is highlighted in light gray (BirdLife International). Each point represents an individual, but in some cases, multiple individuals were collected from the same location, such that points are overlapping. Dotted lines indicate approximate boundaries of the western U.S., boreal, and southern Appalachian populations identified by clustering analyses (Fig. 2). Four individuals sampled on their wintering grounds are not included in the map but were sampled in the Amazon Basin in Bolivia and Paraguay in October–November.

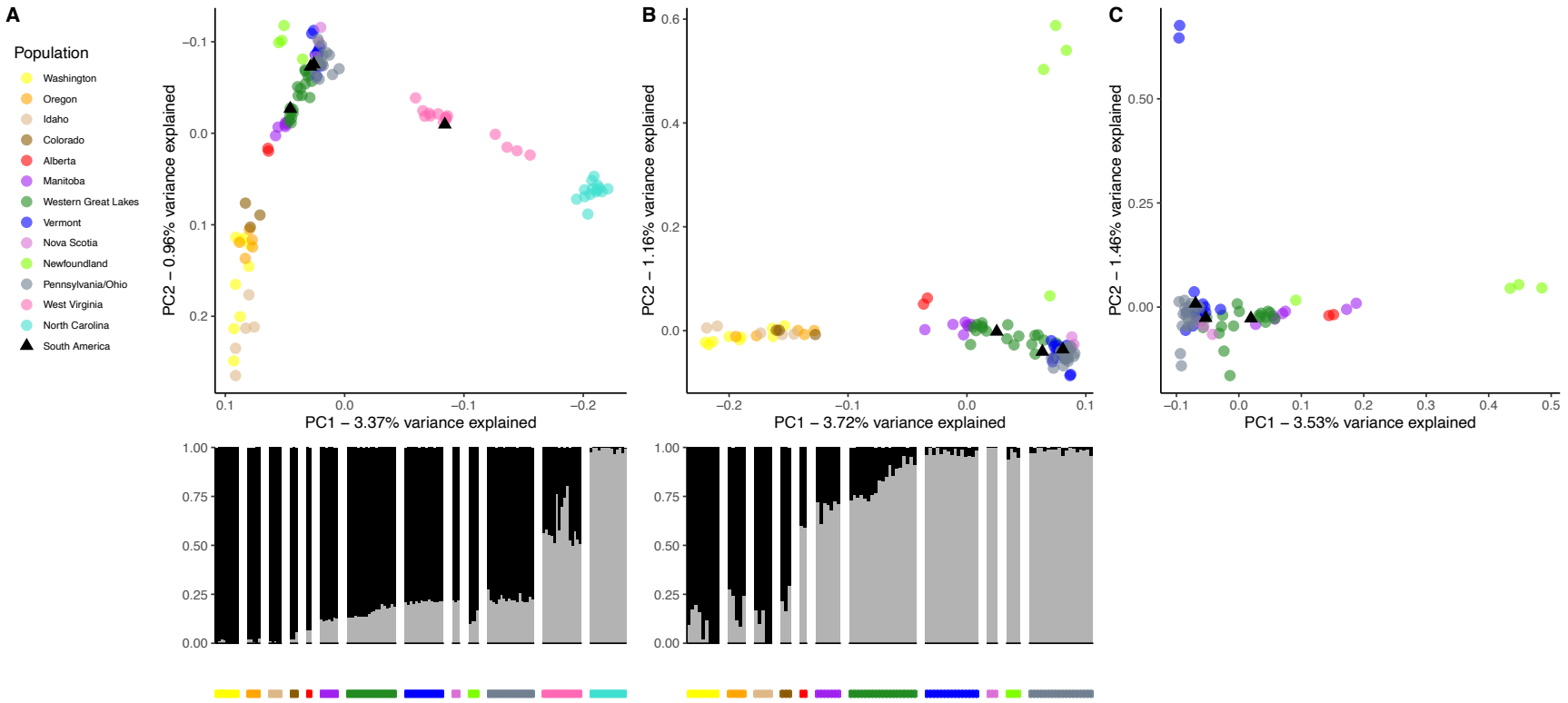
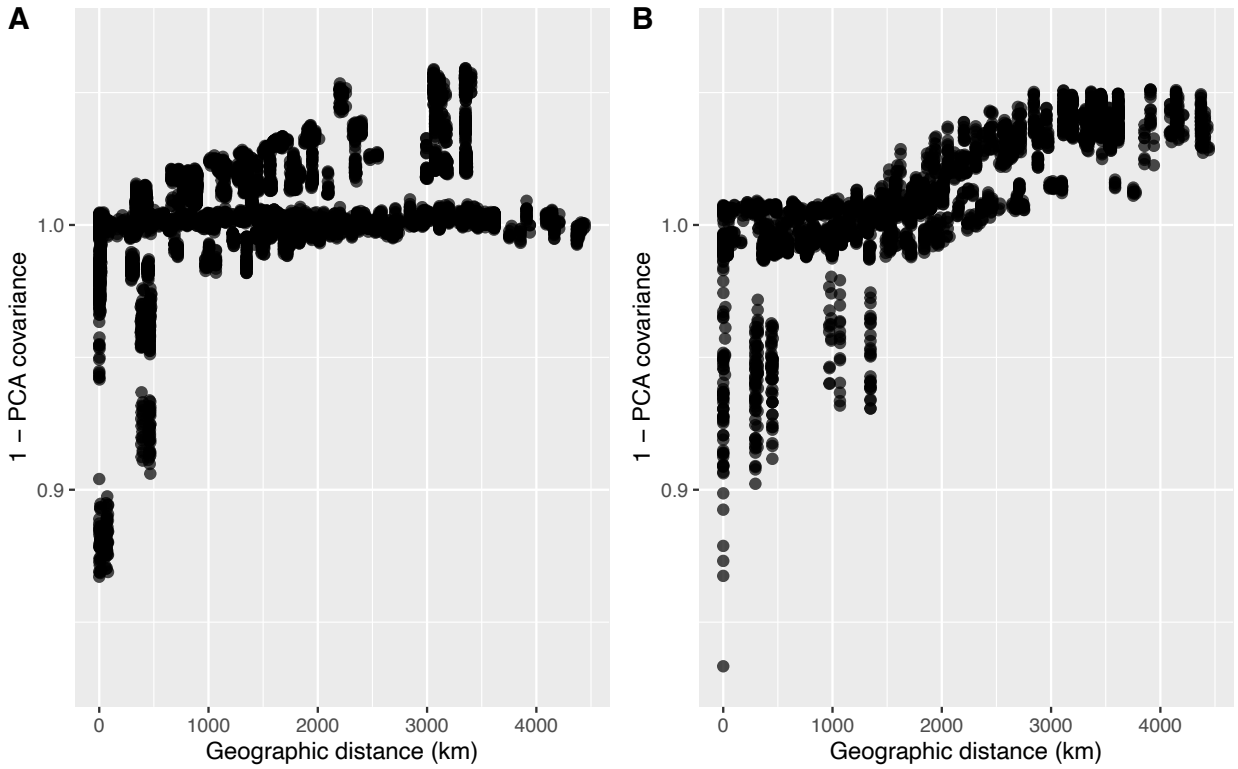


Figure 2. Principal Component Analysis (PCA) (*top*) and admixture plots ($K = 2$) (*bottom*) for (A) all samples and (B) all samples excluding the southern Appalachian samples (i.e., excluding West Virginia and North Carolina). (C) PCA of boreal and northern Appalachian samples only. Since we did not identify distinct populations within the boreal group ($K=1$), no admixture plot is included for panel C. Nonbreeding birds collected in the South America are displayed as black triangles to identify potential population of origin. The x-axis and y-axis in panel A is reversed so that points are displayed to reflect the geographic origins. The U-shaped curve shown in panel A, the nested genetic structure in Panel B and C, and the level of admixture across the geographic range suggests

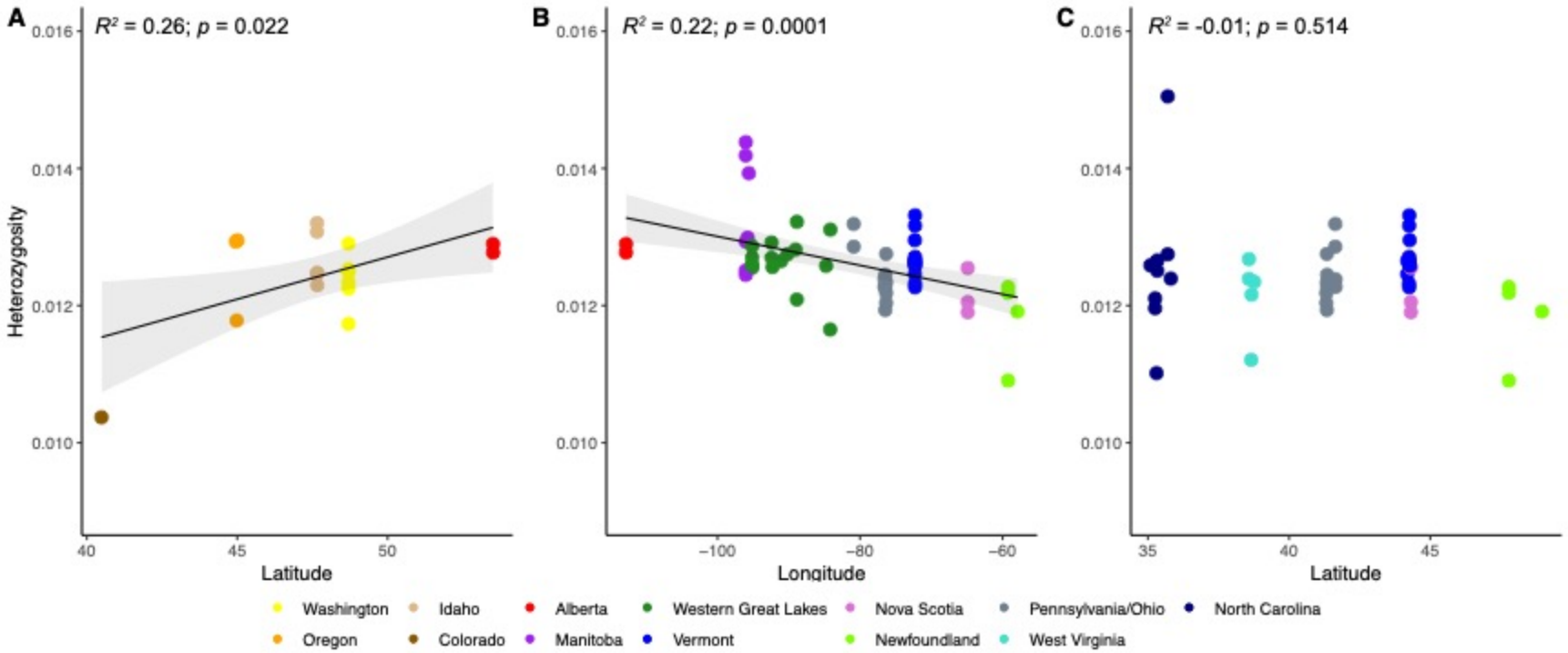
770 isolation by distance across the species' range. The southern Appalachian (West Virginia and North Carolina) were supported as a
771 distinct population from all other samples (A), and the western samples (Washington, Oregon, Idaho and Colorado) were supported as
772 a distinct population when southern Appalachian samples were removed (B). By contrast, all boreal samples were identified as a
773 single population but with evidence of isolation by distance.



774

775 Figure 3. Genetic distance (as measured by 1– PCA covariance) between pairs of individuals is
776 significantly positively correlated with geographic distance (A) using the range-wide dataset
777 (corresponding to panel A in Fig. 2) and (B) the boreal, northern Appalachian, and western US
778 sampling location data (corresponding to panel B in Fig. 2). The correlation recovered from the
779 Mantel test is stronger in the subset of data in panel B (Pearson's correlation coefficient, r =
780 0.78) compared to the range-wide data (Pearson's correlation coefficient, r = 0.38) because of
781 distinct population structure associated with the southern Appalachian population.

782



783

784 Figure 4. Gradients of heterozygosity across (A) the western montane region by latitude, (B) the boreal forest belt by longitude, and
785 (C) the eastern montane region (southern Appalachians northeast to Newfoundland) by latitude. Heterozygosity scales positively with
786 latitude across the western montane region. Heterozygosity is also significantly higher in the west than the east across the boreal forest
787 belt. There is no significant relationship between heterozygosity and latitude across the eastern montane region. Points are colored
788 according to their sampling region.

Tables

Table 1. Nucleotide diversity (pairwise θ_π) estimated per chromosome and individual-level heterozygosity estimated as the number of polymorphic sites divided by the total sites in each individual's 1D site frequency spectrum.

Population	Pairwise θ_π	Individual heterozygosity (*mean* ± *s.e.*)
Western U.S.	0.013	$0.012 \pm 1.82 \times 10^{-4}$
Boreal	0.013	$0.013 \pm 7.48 \times 10^{-5}$
Southern Appalachian	0.013	$0.012 \pm 2.46 \times 10^{-4}$