


Maintenance of local adaptation despite gene flow in a coastal songbird

Jonathan D. Clark,^{1,2,3}  Phred M. Benham,⁴ Jesus E. Maldonado,^{1,5} David A. Luther,⁶
 and Haw Chuan Lim^{5,6,7}

¹Department of Environmental Science and Policy, George Mason University, Fairfax, Virginia 22030

²Current Address: Department of Natural Resources and the Environment, University of New Hampshire, Durham, New Hampshire 03824

³E-mail: jonathan.clark@unh.edu

⁴Museum of Vertebrate Zoology, University of California, Berkeley, Berkeley, California 94720

⁵Center for Conservation Genomics, Smithsonian Conservation Biology Institute, Washington, D.C. 20013

⁶Department of Biology, George Mason University, Fairfax, Virginia 22030

⁷E-mail: hlim22@gmu.edu

Received March 15, 2021

Accepted March 19, 2022

Adaptation to local environments is common in widespread species and the basis of ecological speciation. The song sparrow (*Melospiza melodia*) is a widespread, polytypic passerine that occurs in shrubland habitats throughout North America. We examined the population structure of two parapatric subspecies that inhabit different environments: the Atlantic song sparrow (*M. m. atlantica*), a coastal specialist, and the eastern song sparrow (*M. m. melodia*), a shrubland generalist. These populations lacked clear mitochondrial population structure, yet coastal birds formed a distinct nuclear genetic cluster. We found weak overall genomic differentiation between these subspecies, suggesting either recent divergence, extensive gene flow, or a combination thereof. There was a steep genetic cline at the transition to coastal habitats, consistent with isolation by environment, not isolation by distance. A phenotype under divergent selection, bill size, varied with the amount of coastal ancestry in transitional areas, but larger bill size was maintained in coastal habitats regardless of ancestry, further supporting a role for selection in the maintenance of these subspecies. Demographic modeling suggested a divergence history of limited gene flow followed by secondary contact, which has emerged as a common theme in adaptive divergence across taxa.

KEY WORDS: Bill divergence, contact zone, divergent selection, ecological speciation, saltmarsh syndrome.

Widespread species often use considerably different habitats across their geographic range (e.g., Aldrich 1984; Aldrich and James 1991; Song et al. 2020) exposing subpopulations to an array of varied selection pressures from the environment (Darwin 1859; Mayr 1963; Coyne and Orr 2004; Price 2008; Nosil 2012). As a result, regional populations can develop local adaptations, that is, heritable dissimilarities in morphology, physiology, or behavior suited to their local habitats. When there is strong divergent selection between locally adapted populations in contact, local adaptations can be maintained despite the homogenizing effects of gene flow (Yeaman and Otto 2011;

Butlin et al. 2014; Marques et al. 2016). Hence, studying the genetic and phenotypic divergence of interbreeding populations that maintain local adaptations can shed light on the early stages of ecological speciation with gene flow (Rosenblum and Harmon 2011; Sendell-Price et al. 2020), as local adaptations maintained by divergent selection can eventually lead to speciation when reproductive barriers form directly as a by-product of adaptation or indirectly as a result of drift due to isolation by environment (IBE; Pinho and Hey 2010; Servedio et al. 2011; Feder et al. 2012; Wang and Bradburd 2014). However, detecting population genetic structure between natural populations with extensive

gene flow requires high-resolution genetic methods that, until recently, were often cost prohibitive or unfeasible, because populations that maintain contrasting local adaptations and interbreed in a contact zone due to a lack of intrinsic reproductive barriers often exhibit weak genomic differentiation.

More natural examples of local adaptation despite gene flow are needed to understand how the complex interactions between assortative mating, genomic architecture, the strength of selection, genetic drift, gene flow, and changes in the distribution of habitats through time result in the evolution and maintenance of local adaptation in widespread species (Dobzhansky and Pavlovsky 1957; Schluter and Conte 2009; Via 2009; Yoder et al. 2010; Nosil 2012; Song et al. 2020). One vertebrate species that provides an excellent model to examine the development and maintenance of locally adapted populations is the song sparrow (*Melospiza melodia*). The song sparrow, a polytypic North American passerine with 25 recognized subspecies (Patten and Pruett 2009), has long been of interest to taxonomists and evolutionary biologists, especially the concentrated diversity of subspecies in western North America (Swarth 1923; Browning 1978; Aldrich 1984; Pruett et al. 2008; Pruett and Winker 2010; Mikles et al. 2020). The remarkable phenotypic diversity of the song sparrow is likely a result of its ubiquity in varied shrubland habitats across the continent, including those of high altitudes, hot continental deserts, and subarctic islands (Patten et al. 2004; Patten and Pruett 2009).

In eastern North America, there are only two recognized subspecies of song sparrow: the eastern song sparrow (*M. m. melodia*) and the Atlantic song sparrow (*M. m. atlantica*). The eastern song sparrow is a common, widespread generalist that occurs east of the Rocky Mountains from the Appalachian Mountains north to Newfoundland and west to the Northwest Territories of Canada (Fig. 1a), where it primarily inhabits mesic successional shrublands and thus is often associated with human environments (Aldrich and Coffin 1980; Patten and Pruett 2009). The Atlantic song sparrow is a coastal specialist that is restricted to the saltmarsh edges and sand dunes of the Atlantic coast of the United States from the Outer Banks of North Carolina (NC) north to Long Island, New York (NY; Todd 1924; Patten and Pruett 2009). The selection regime of coastal habitats presents serious evolutionary challenges to terrestrial vertebrates (Greenberg et al. 2006), and principal among them are the limited access to freshwater and high ambient temperatures, which necessitate efficient osmoregulatory and water conservation capabilities. Yet several passerellid sparrows—including the Atlantic song sparrow—have colonized coastal habitats and converged on a suite of characteristic traits known as the “saltmarsh syndrome”: larger and elongated bills, darker and grayer plumage, and increased salt excretion abilities (Greenberg et al. 2006; Deane-Coe et al. 2018a; Walsh et al. 2019; Benham and Cheviron 2020).

The Atlantic song sparrow is distinguished from the eastern song sparrow by its larger bill, grayer plumage, and increased black streaking (Fig. 2). A larger bill has been well studied for its functional advantages in hot, freshwater-limited coastal habitats (Greenberg et al. 2012; Danner and Greenberg 2015; Danner et al. 2016; Luther and Danner 2016). On average, the bill of the Atlantic song sparrow has >16% more surface area than the eastern song sparrow, enabling the Atlantic song sparrow to dissipate approximately 30% more body heat through the bill and potentially increase water savings by reducing reliance on evaporative cooling (Greenberg et al. 2012). Additionally, the Atlantic song sparrow tends to have nasal respiratory turbinates with proportionally more surface area than the eastern song sparrow (Danner et al. 2016), which could further reduce evaporative water loss by condensing and retaining exhaled moisture during normal respiration (Geist 2000).

Both the abilities to more efficiently use the bill to radiate dry heat and retain water during respiration are hypothesized to impart fitness benefits to organisms in hot, freshwater-limited coastal habitats. For example, male Atlantic song sparrows with the largest bills were found to sing at twice the rate of the males with the smallest bills in the wild, suggesting that the thermoregulatory advantage of a larger bill enables increased physical activity and potentially imparts increased reproductive success (Luther and Danner 2016). However, a smaller bill is better able to reduce heat loss during the winter, because blood flow to the bill cannot be completely restricted to reduce heat loss through the bill at cold temperatures (Greenberg et al. 2012). Therefore, selection favors smaller bills away from the coast in the habitats of the eastern song sparrow (Danner and Greenberg 2015). The habitat-specific trade-off for large-billed individuals suggests that divergent selection may be the primary force maintaining the distinctiveness of these two subspecies despite suspected gene flow in their contact zone.

Exploring patterns of bill size divergence in conjunction with genomic data will provide insight into the mechanisms that produced and maintain these two parapatric subspecies. (1) The Atlantic song sparrow and eastern song sparrow may have diverged long ago during a period of allopatry and currently be in contact with minimal barriers to gene flow. In this case, there may be a clear signal of neutral population genetic structure between the most distant populations and a broad contact zone with shallow clinal genetic variation, the width of which would be dependent on the time since secondary contact. Further, the maintenance of adaptive trait divergence would likely be minimal, also exhibiting a shallow, uniform clinal pattern. (2) Similarly, if the subspecies experienced recent divergence and there are minimal barriers to gene flow, then clinal morphological and genetic variation with a uniform, shallow slope could be expected, such as that produced by IBD. (3) If the subspecies are deeply diverged

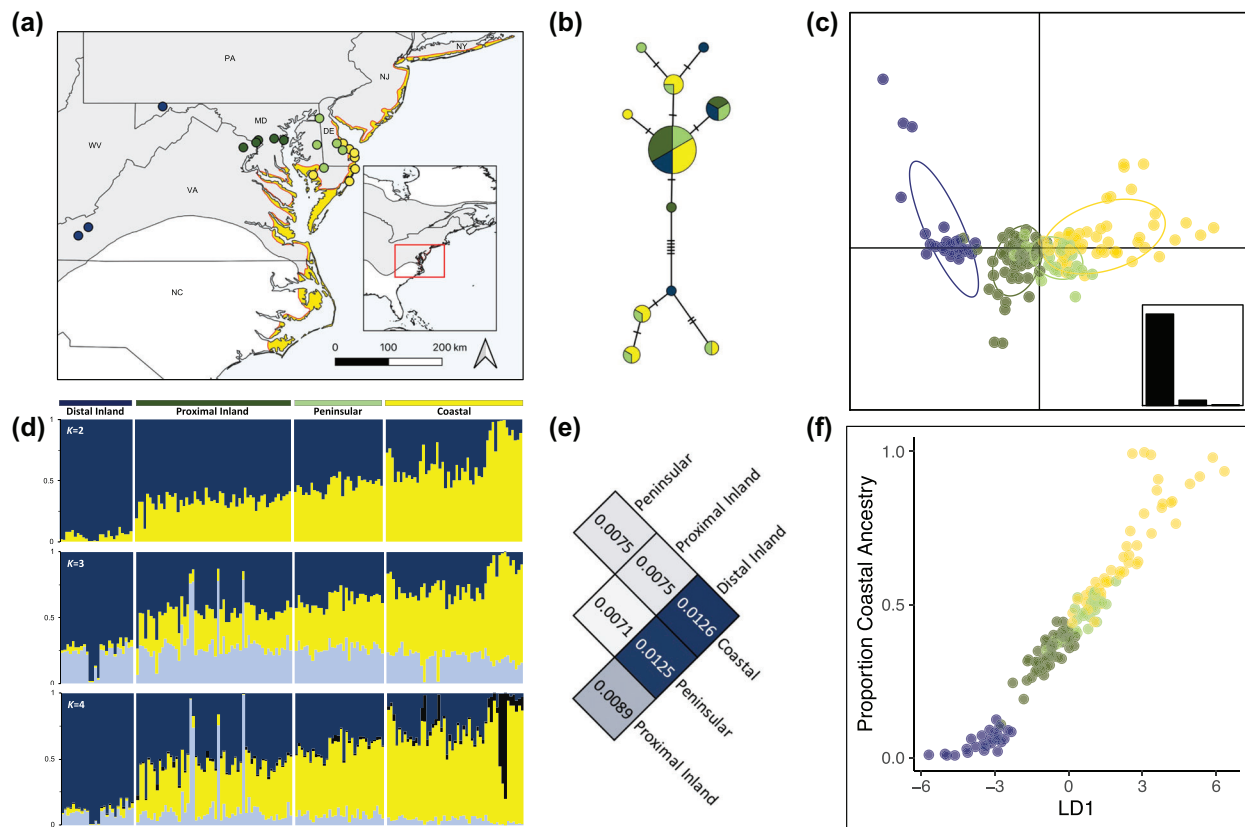


Figure 1. Population structure throughout the southern contact zone of the eastern song sparrow and Atlantic song sparrow in Delaware (DE), Maryland (MD), and Virginia (VA). Navy, dark green, light green, and yellow correspond to distal inland, proximal inland, peninsular, and coastal sites, respectively. (a) Map of sampling sites. The yellow area represents the approximate breeding range of the Atlantic song sparrow in the region, whereas the eastern song sparrow breeds throughout the gray area shown (sensu Patten and Pruett 2009). Inset map of the east coast of the United States shows the study area outlined in red. (b) Minimum spanning network of ND2 haplotypes for the subset of 47 individuals representing all four sampling regions. (c) Discriminant analysis of principal components (DAPC) plot of LD1 (horizontal axis) and LD2 (vertical axis) with seven principal components retained (6.97% cumulative variance) for the 11,695 unlinked biallelic SNPs generated by RADseq for 163 individuals along the coastal-inland gradient. Inset is a Scree plot showing the relative eigenvalues for LD1, LD2, and LD3. (d) Bar plots of genetic ancestry assignment for each individual by sampling region from *STRUCTURE*. The best supported model was $K = 2$. Yellow corresponds to the coast-associated genetic cluster, whereas navy corresponds to the inland-associated genetic cluster. (e) Comparison matrix of overall F_{ST} between each region. (f) Relationship of LD1 with assignment to the coast-associated genetic cluster when $K = 2$ for each individual.

due to extended historical allopatry and there are currently strong barriers restricting most gene flow, then they should exhibit clear, punctuated genetic structure and discrete differences in adaptive traits. (4) But, if the subspecies are recently diverged and gene flow is partly limited by IBE, then there should be a cline in both genetics and morphology that has a steep, abrupt slope coinciding with a habitat transition. Additionally, in this last case, if divergent selection is the primary force maintaining local adaptation in coastal populations, then larger bill size in coastal habitats should be maintained in the face of some gene flow from inland populations, as selection would inhibit introgression of the maladaptive alleles that result in a smaller bill.

To distinguish among these different predictions, we characterized the population genetic structure and its relationship to differences in bill size between the parapatric Atlantic song sparrow and eastern song sparrow to assess how these subspecies are maintained despite the opportunity for gene flow in their contact zone. We evaluated population genetic structure across the southern portion of these subspecies' contact zone by sequencing a mitochondrial marker and thousands of genome-wide single-nucleotide polymorphisms (SNPs). We then assessed genetic differentiation, tested for isolation by distance (IBD), and related the level of genetic admixture to variation in bill size. We also tested different models of demographic history to assess whether these

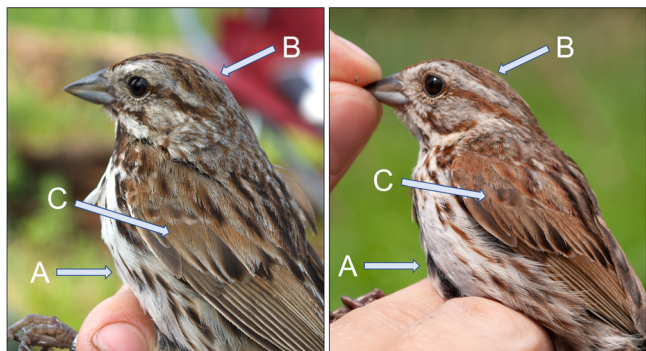


Figure 2. Plumage differences between the Atlantic song sparrow (left) and eastern song sparrow (right). In addition to having a longer and thicker bill than the eastern song sparrow, the Atlantic song sparrow differs in several subtle plumage traits. (a) More prominent ventral streaking with a higher dominance of black pigmentation. (b) More prominent black streaking on the head. (c) A decrease in rusty feather edging and grayer plumage overall.

populations likely diverged with gene flow or are in secondary contact following divergence in allopatry.

Methods

DNA SAMPLES

Blood samples were collected from 220 birds in 2011 representing the Atlantic song sparrow, eastern song sparrow, and potentially admixed individuals along a coastal-inland transect on the southern portion of their contact zone (Tables S1 and S2). The easterly portion of the sampling transect started at the Atlantic shore of Delaware (DE) and Maryland (MD) on the Delmarva Peninsula and moved inland, terminating at sites in western Maryland and southwestern Virginia (VA; Fig. 1a). Additionally, several tissue samples were acquired from the Smithsonian National Museum of Natural History (USNM) from Atlantic song sparrows collected in saltmarshes of the Chesapeake Bay.

We divided the sampling sites across the transect into four regional categories based on distance from coastal habitats (Fig. 1a). All sites in coastal sand dunes or saltmarshes, that is, the habitats associated with the Atlantic song sparrow, were grouped (as “coastal”). Inland sites where song sparrows were collected on the Delmarva Peninsula—putatively the eastern subspecies—were grouped (as “peninsular”), because they may have been in direct contact with Atlantic song sparrows, allowing gene flow. Sites on the mainland where eastern song sparrows were collected were split into two groups: those <100 km inland (as “proximal inland”) and those >100 km inland (as “distal inland”) from coastal habitats.

MITOCHONDRIAL SEQUENCING AND ANALYSIS

A subset of 48 individuals representing both subspecies from across the sampling transect was selected for mitochondrial DNA sequencing (Table S1). A partial sequence of the second subunit of the mitochondrial nicotinamide adenine dinucleotide dehydrogenase gene (ND2), a marker commonly used in population genetic analyses of vertebrate species, was amplified with the primers L5215 and H6113 using a protocol adapted from prior studies on other avian species (Zou et al. 2007; Gawin et al. 2014). In short, 30 cycles of polymerase chain reaction (PCR) were performed with a denaturing temperature of 94°C, an annealing temperature of 57°C, and an extension temperature of 72°C. Amplification products were purified using Sera-Mag SpeedBeads magnetic particles solution (Millipore Sigma) before being thermal cycle sequenced with a BigDye Terminator version 3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and purified with Sephadex G-50 Superfine (GE Healthcare). Sequencing was performed on an Applied Biosystems PRISM 3130xl Genetic Analyzer (Thermo Fisher Scientific) at the George Mason University Microbiome Analysis Center. Chromatograms for each individual were assembled into a contig with the software *Geneious Prime* (version 11.1). The contigs were then manually corrected, if necessary, and trimmed to a length of 818 bp. The generated consensus sequences were aligned in *Geneious Prime* using the *Align/Assemble* function on default settings to a partial mitochondrial reference sequence from a specimen of *M. m. montana* collected in Nevada County, California (GenBank Accession FJ236290; MVZ 173560). Following sequence assembly, one individual was excluded from analyses due to poor sequence quality, bringing the analyzed sample size to 47 individuals.

A minimum spanning network was generated with the program *PopART* at the default tolerance value ($\epsilon = 0$) to provide a visualization of the diversity and relative abundance of ND2 haplotypes and the pattern of mitochondrial population division between the subspecies (Bandelt et al. 1999; Leigh and Bryant 2015). Divergence of ND2 haplotypes between populations was assessed using a hierarchical analysis of molecular variance (AMOVA) executed in *PopART* with 1000 permutations to test if the genetic distances between the haplotypes of individuals were related to their designated population grouping (Excoffier et al. 1992). Population categories were nested into two broader groups for analysis: mainland (distal inland and proximal inland) and coast-associated (peninsular and coastal). Despite that they are taxonomically classified as different subspecies, birds from peninsular and coastal regions were nested together because their proximity might have allowed for extensive gene flow.

RESTRICTION-SITE ASSOCIATED DNA SEQUENCING

To generate nuclear sequence data for all 220 individuals, we used a dual-indexed, three-enzyme restriction site-associated

DNA sequencing (RADseq) approach (3RAD; see Bayona-Vázquez et al. 2019 for details). The three restriction enzymes used were: *NheI*, *EcoRI*, and *XbaI* (New England Biosystems). For each sample, approximately 100 ng of genomic DNA suspended in a 10 μ L volume was used for library preparation. During library preparation, after restriction digestion and ligation of the inline indexed adapters, samples were pooled into subgroups of eight that shared an iTru7 index. A random iTru5 index was used to allow for decloning of PCR duplicates.

Following library preparation, size selection of 450–600 bp adapter-ligated, amplified restriction fragments was done using a 2% e-Gel precast agarose electrophoresis system (Thermo Fisher Scientific). After size selection, the concentration of the final pooled library of all samples was quantified using a Qubit 4 Fluorometer (Invitrogen). The concentration of the size-selected final library was too low for sequencing, so it was amplified a final time with five cycles of PCR and then purified and quantified. The final pooled library of all samples was sequenced as 150 bp paired-end reads on one lane of an Illumina HiSeq 4000 at the Duke Center for Genomic and Computational Biology.

BIOINFORMATIC WORKFLOW

The raw sequencing reads were received demultiplexed by the iTru7 index shared by each sub-pool of eight samples. The program *STACKS* (version 2.3) was then used to demultiplex each index barcode group of eight into individual samples by the two inline barcodes annealed in library preparation (Catchen et al. 2013; Rochette et al. 2019; see Supporting Information for details on genotyping). After demultiplexing, three individuals lacking data on location of collection and four individuals with <1000 decloned read pairs were removed from the sample set.

No public reference genome existed for the song sparrow at the time of this study; therefore, SNPs were called via the de novo map pipeline (*denovo_map.pl*) built into *STACKS*. Optimal parameters for genotyping in *STACKS* ($M = 4$, $n = 4$, and $m = 3$) were selected following methods outlined previously in Paris et al. (2017) and Rochette and Catchen (2017). The 32 individuals with <10,000 genotyped loci or a mean coverage <8 \times were excluded at this point.

Population genetic analyses were conducted on SNPs with a minimum allele frequency of 5% and maximum heterozygosity of 70% from catalog loci found in at least 80% of individuals, as recommended by Rochette and Catchen (2017), which were output for analyses using the *populations* module. The mean length of loci in the final dataset was 241.72 bp, and 60% of loci had read overlap, meaning that the sequenced libraries were shorter than intended. The mean length of overlapping loci was 224.55 bp with a mean overlap of 28.41 bp. After filtering, the genotype dataset contained 15,590 RAD-loci (11,695 of which were polymorphic) totaling 3.77 Mbp of sequence and 34,712

biallelic SNPs. At this point, 18 individuals identified as missing data for more than 20% of the filtered loci were removed. After all quality control measures, the total number of genotyped individuals available for analysis was brought from the initial 220 down to 163 (distal inland = 26, proximal inland = 56, peninsular = 32, coastal = 49) with a mean decloned sequence depth of 41 \times . For population structure analyses, only the first SNP of each locus was used to minimize the effects of linkage disequilibrium.

NUCLEAR POPULATION STRUCTURE

We conducted a discriminant analysis of principal components (DAPC) to explore the population genetic structure of the subspecies (Jombart et al. 2010). The DAPC was performed in *R* with the package *adegenet* on a *genind* object output using the *populations* module of *STACKS* (Jombart 2008). Sampling region was used as the a priori grouping. The function *xvalDapc* was used to run cross-validation to identify the optimal number of principal components (PCs) to retain. The DAPC was then executed using the function *dapc*.

Genotype data from 3RAD were output by *STACKS* in *STRUCTURE* file format using the *populations* module and analyzed using the program *STRUCTURE* (version 2.3.4) to evaluate genetic ancestry with a Bayesian clustering method based on Markov Chain Monte Carlo (MCMC) estimation (Pritchard et al. 2000; Porras-Hurtado et al. 2013). *STRUCTURE* was used to generate a matrix of membership coefficients (the individual *Q*-matrix), which are scores indicating the portion of the genome per individual that belongs to each inferred ancestral genetic cluster (hereafter ancestry assignment). The program was run using the admixture model with the assumed number of parental populations (*K*) set to values of 1–4 with a burn-in period of 10,000 MCMC steps and 50,000 post-burn-in MCMC steps.

For each setting of *K*, five independent runs of *STRUCTURE* were conducted, then the results were evaluated using the online program *STRUCTURE HARVESTER* to identify the best supported value of *K* via the Evanno method (Evanno et al. 2005; Earl and vonHoldt 2012). The program *CLUMPP* was then used to summarize all runs for each value of *K* into one consensus ancestry assignment for each individual (Jakobsson and Rosenberg 2007).

Pairwise fixation index (F_{ST}) between sampling regions was used to assess extent of differentiation between the defined regional groups by using the *fstats* option of the *populations* module in *STACKS* to produce a matrix of overall F_{ST} values between each pair of regional groups. Moreover, the genome-wide F_{ST} distribution between divergent taxa can provide qualitative information about the extent of gene flow, with few highly divergent loci and a concentration of near-zero values suggesting breeding between demes (Seehausen et al. 2014; Peñalba et al. 2019;

Sendell-Price et al. 2020). Therefore, we also examined the distribution of per-SNP F_{ST} for all SNPs between the distal inland and coastal populations.

To determine if the genetic patterns discerned could be explained by IBD across the sampling area, a Mantel test was performed in *R* using functions from *adegenet* (Mantel 1967). First, the *genind* object was converted to *GenePop* data format with the function *genind2genpop*. The data were then used to generate a matrix of Edwards' genetic distances (Cavalli-Sforza and Edwards 1967) between site pairs with the *dist.genepop* function. Next, a geographic distance matrix was generated for each pair of sites using the *rdist.earth* function from the *fields* package. The Mantel test was executed on these distance matrices using the function *mantel.randtest* with 10,000 permutations. The analysis was then repeated, once excluding coastal sites to assess IBD within the eastern song sparrow subspecies, and once with only coastal sites to assess IBD within the Atlantic subspecies.

ADMIXTURE AND PHENOTYPE CLINES

Ancestry assignments from *STRUCTURE* analysis ($K = 2$) were used to assess how geographic and morphological variation related to estimates of admixture. Average assignments to the identified coast-associated genetic cluster for each region were plotted against distance inland from the coast to summarize shifts in ancestry across the coastal-inland transect.

To evaluate the correlation between genomic and bill size (a phenotype hypothesized to be under divergent selection) changes across the coastal-inland transect, we used measurements of bill width (W) and height (H) at the anterior edge of the nares and length from the anterior edge of the nares to the bill tip (N) to estimate bill surface area for 118 adult male sparrows by calculating the lateral surface area of a cone with a base diameter equal to the average of bill width and depth:

$$\text{Bill surface area} = \pi N \left(\frac{W + H}{4} \right).$$

Ancestry assignment was compared to bill surface area using a simple linear regression in *R* with the *lm* function. An analysis of variance (ANOVA) and a Tukey's honestly significant difference (Tukey's HSD) test were conducted to test for a difference in bill size between each of the regions. Both tests were performed in *R* using the functions *aov* and *TukeyHSD*, respectively. Prior to conducting the ANOVA and Tukey's HSD test, the distribution of bill surface area of each group was tested to confirm conformation to a normal distribution by using the function *shapiro.test* to perform a Shapiro-Wilk test of normality.

INFERENCE OF DEMOGRAPHIC HISTORY

To estimate the demographic history of our song sparrow populations, the fit of a series of demographic models was

compared to the observed three-dimensional site frequency spectrum (SFS) using the program *Moments* (Jouganous et al. 2017). The SFS was generated from a VCF file of the 15,150 unlinked SNPs (filtered without a minimum allele frequency threshold) output using the *populations* module in *STACKS*. To simplify the tested demographic models, proximal inland individuals were excluded. The VCF file was converted to a *Moments* input format using a custom python script (*vcf2dadi.py*) written by Isaac Overcast (<https://github.com/isaacovercast/WCS/>). The VCF file was imported and downprojected within *Moments* to maximize the number of SNPs with complete coverage across all individuals. Following downprojection of the data within *Moments*, the input dataset comprised 11,709 SNPs segregating across 107 individuals: 26 eastern song sparrows from the distal inland population, 32 song sparrows from the peninsular population, and 49 Atlantic song sparrows from the coastal population.

A series of three-population demographic models were fit to the joint SFS to infer divergence times, effective population sizes (N_e), and gene flow (Fig. 3). Models of two different branching histories were compared. First, (scenario A) an initial split between non-coastal and coastal song sparrow populations followed by divergence between distal inland and peninsular populations was modeled. Alternatively, (scenario B) a model was produced for a history where distal inland populations first diverged from the peninsular and coastal populations, followed by divergence between the latter two populations. For each of the two branching histories, four different histories of gene flow among the three populations were tested: (model 1) no gene flow among populations; (model 2) only recent gene flow among populations after divergence of the peninsular population, but with no gene flow between coastal and distal song sparrows; (model 3) only recent gene flow among all populations after divergence of the peninsular population; (model 4) gene flow among all populations throughout the history of divergence. Models allowed for individual migration rates to encompass zero; therefore, model 4 was effectively inclusive of the possibility of divergence with gene flow followed by contemporary gene flow between adjacent populations. Support for model 2 or model 3 would suggest a history of isolation followed by secondary contact, while support for model 4 would be more consistent with a history of primary divergence along the coastal-inland gradient.

For each model tested, 15 optimizations were run from different starting parameters using the *perturb* function in *Moments* with a maximum number of iterations set to 25. To ensure that a global optimum for a given model had been reached, we ran one final optimization with 50 iterations using the parameter values estimated from the shorter optimization run with the highest likelihood. We derived demographic parameter values from the estimated value of Θ , where $\Theta = 4N_e\mu L$ based on a generation time of 2.5 years for the song sparrow (Bird et al. 2020),

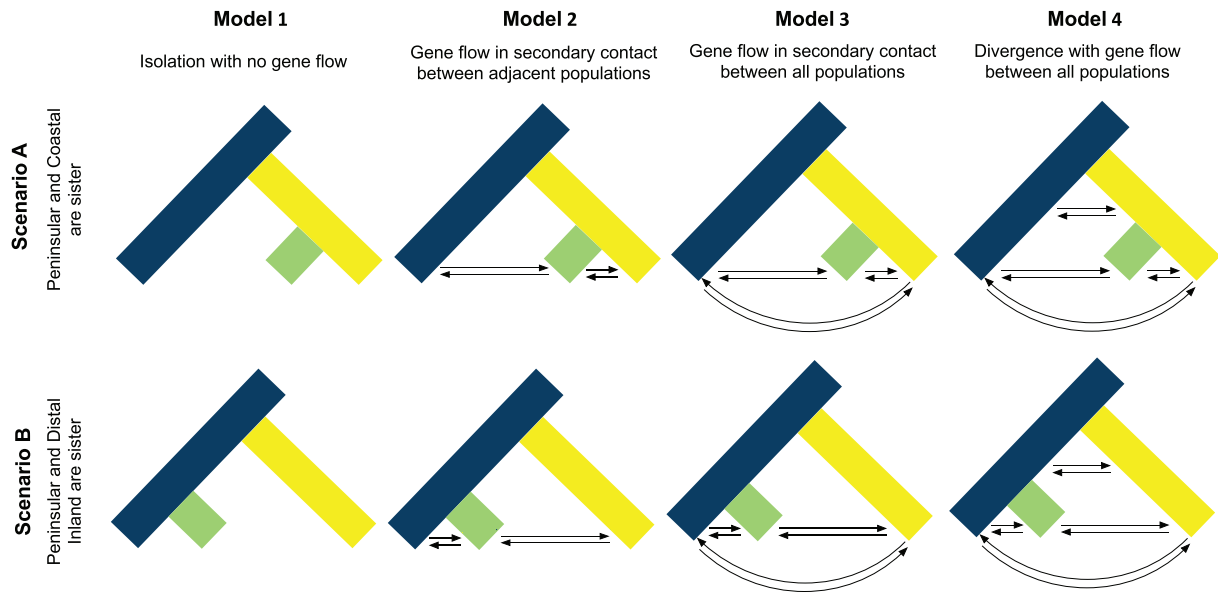


Figure 3. Population divergence models tested. Yellow, light green, and navy represent coastal, peninsular, and distal inland populations, respectively.

the mean genomic substitution rate (μ) for Passeriformes of 3.3×10^{-9} substitutions/site/year (Zhang et al. 2014), and the total length of sequence length (L) from which SNPs were called, which was 2,901,679 bp—(15,590 total loci \times 241.72 bp mean length) \times (11,709 loci in SFS / 15,150 total RAD-loci). To calculate uncertainty around each parameter estimate, we generated 100 bootstrapped SFS using the bootstrap function within *Moments*. Bootstrapped spectra were then used to calculate parameter uncertainties using the Godambe information matrix (GIM) with log-transformed parameter values (Coffman et al. 2016). We compared differences in model fit to the observed SFS using an AIC approach. For discussion, migration rates from the best fit model—that is, proportion of individuals in the population whose parents were from the donor population—were converted to effective migrants per generation (mpg) by multiplying by the estimated N_e of the receiving population.

IDENTITY OF DIVERGENT LOCI

To determine the putative identity of highly divergent markers in the Atlantic song sparrow genome, we examined loci with the highest sequence divergence (ϕ_{ST}) based on the *fstats* outputs of *STACKS* between two Atlantic song sparrow populations, Assateague Island ($n = 13$) and Ocean City, MD ($n = 14$), and two independent distal inland eastern song sparrow populations in Blacksburg, VA ($n = 9$) and Frostburg, MD ($n = 11$). Any locus that was greater than four standard deviations from the mean differentiation (i.e., $Z\phi_{ST} > 4$) for all four pairwise coastal-inland comparisons was identified as an outlier locus.

Outlier loci were then matched to a publicly available annotated reference genome of the white-throated sparrow (*Zonotrichia albicollis*; Taxid: 44394; *Zonotrichia albicollis*_1.0.1 reference Annotation Release 102) using *megaBLAST* (Zhang et al. 2000). Divergent loci were assessed as potential candidates of selection based on their proximity to genes in the white-throated sparrow genome. Candidate loci were then aligned to the zebra finch genome (*Taeniopygia guttata*; Taxid: 59729; bTaeGut2.pat.W.v2 reference Annotation Release 105) to determine putative chromosomal location. Finally, the extent of divergence regionally (i.e., between all distal inland and all coastal birds) was assessed for the identified outlier loci.

Results

MITOCHONDRIAL SEQUENCING AND ANALYSIS

From the 47 analyzed individuals, there were 14 variable nucleotide positions across the 11 unique ND2 haplotypes identified. The hierarchical AMOVA confirmed that there was no significant difference in ND2 sequences between the sampling regions ($\phi_{ST} = 0.064$, $P = 0.156$; Table S3). However, there was some mitochondrial differentiation between the grouped sampling regions ($\phi_{CT} = 0.142$, $P < 0.001$).

Of the 11 haplotypes identified, one was present in more than half of the sequenced individuals, common in both subspecies, and found across the sampling transect. Although the minimum spanning network had two distinct haplotype groups, it did not strictly exhibit clear population genetic structure between subspecies (Fig. 1b). Of the 47 individuals analyzed, 81% were

within ≤ 2 mutations of the most abundant haplotype, whereas the remaining 19% were ≥ 6 mutations from it. Notably, the mitochondrial reference sequence (not included in the analysis) from an individual from California (i.e., $>3,400$ km away) was an exact match to the most abundant haplotype.

NUCLEAR POPULATION STRUCTURE

Cross-validation determined that the optimum number of PCs to retain for the DAPC was 7 (6.97% cumulative variance), and three discriminant axes were retained. A clinal relationship was apparent along the main linear discriminant axis of the DAPC (LD1), which corresponded to regional groupings and distance from the coast (Fig. 1c). Peninsular individuals overlapped with both coastal and proximal inland individuals. Distal inland individuals from sites in western Maryland and southwestern Virginia clustered together despite their substantial geographic distance from each other—roughly equal to their distance from the coast.

STRUCTURE HARVESTER indicated that the pattern of ancestry was best represented by $K = 2$ (Figs. 1d, S1). When $K = 4$, individuals dominated by the additional clusters were few proximal inland individuals from one site and few coastal individuals from another site, thus additional clusters after $K = 2$ were not indicative of any large-scale population genetic structure. There was an obvious correlation of LD1 with an individual's ancestry assignment by *STRUCTURE* at $K = 2$, and both indicated clinal genetic differentiation along the coastal-inland gradient (Fig. 1f).

The two clusters that resulted from $K = 2$ clearly separated distal inland individuals from coastal individuals. All individuals from distal inland sites had an assignment to the coast-associated cluster of $<12.4\%$. Nonetheless, most individuals across the transect were substantially admixed; only 31 of 163 birds had an ancestry assignment $>90\%$ to either cluster. Mean assignment of individuals to the coast-associated cluster for proximal inland and peninsular sites was 34% and 47%, respectively. Individuals from the coastal population had highly variable levels of admixture with a mean assignment to the coast-associated cluster of 67% (max = 99.7%, min = 44.0%, SD = 16.7%). However, individuals from Assateague Island ($n = 13$), a narrow barrier island off the coast of Maryland, collectively had the highest assignment to the coast-associated cluster with a mean of 90.2% (max = 99.7%, min = 74.0%, SD = 7.8%).

Despite detectible clustering, overall differentiation between the distal inland and coastal populations was low ($F_{ST} = 0.0126$). Other pairwise F_{ST} values were lower, ranging from 0.0071 to 0.0089 (Fig. 1e). The genome-wide per-SNP F_{ST} distribution for distal inland and coastal birds was consistent with patterns expected for regularly interbreeding and/or recently diverged populations, with a large proportion of SNPs with a zero or near-zero F_{ST} (Fig. S2).

The Mantel test confirmed that the effect of IBD across the sampling transect was not significant ($r = 0.152$, $P = 0.205$; Fig. S3). Geographic distances between sites ranged from 4 to 521 km, yet seven of the 10 most genetically distant site pairs were <200 km apart. IBD was still not significant when coastal populations were excluded ($r = 0.132$, $P = 0.257$) nor when only coastal populations were included ($r = -0.119$, $P = 0.583$).

ADMIXTURE AND PHENOTYPE CLINES

When assignment to the coast-associated cluster for each region was plotted against the approximate distance inland from coastal habitats, a geographic cline in ancestry was apparent (Fig. 4a). This cline was steepest between coastal and peninsular sites. In adult male eastern song sparrows ($n = 66$), bill surface area increased with assignment to the coast-associated cluster ($R^2 = 0.302$, $P < 0.001$). However, bill surface area in the Atlantic song sparrow ($n = 36$) was unrelated to admixture proportion ($R^2 = 0.003$, $P = 0.739$; Fig. 4b).

The ANOVA confirmed that there was a significant difference in bill size between at least one pair of groups ($F(2, 99) = 26.64$, $P < 0.001$). The Tukey's HSD test determined that bill size of the proximal inland population (mean = 85.6 mm², SD = 4.87 mm²) was significantly smaller than both the coastal population (mean = 94.3 mm², SD = 6.39 mm², $P < 0.001$) and peninsular population (mean = 92.3 mm², SD = 4.42 mm², $P < 0.001$; Fig. S4). However, bill size did not differ significantly between the coastal population and peninsular population ($P = 0.320$).

INFERENCE OF DEMOGRAPHIC HISTORY

The top demographic model (Fig. 5; Table S4) supported a history of divergence between distal inland birds and a peninsular-coastal clade (scenario A) approximately 268 kya (95% confidence interval [CI]: [216, 376]) followed by divergence between coastal and peninsular birds around 64 kya (95% CI: [51, 91]). This best model (model 3) indicated a lack of gene flow between the distal inland population and the ancestral population of the peninsular and coastal populations, then gene flow resumed between all pairs of populations after the peninsular and coastal populations diverged (log-likelihood = -3477.65 ; AIC = 6979.3; Δ AIC = 0). Notably, despite the classification of peninsular and distal inland populations as the same subspecies, a model that placed them as sister (scenario B, model 3) was less likely (log-likelihood = -3481.45 ; AIC = 6986.90; Δ AIC = 7.61).

Based on the best model, after divergence between peninsular and the coastal populations, migration rates were greatest from the coastal population to peninsular sites ($m_4 = 2.51 \times 10^{-3}$, 158.1 mpj; see Fig. 5 for CIs). The peninsular population exhibited modest, but non-zero, emigration to both the inland

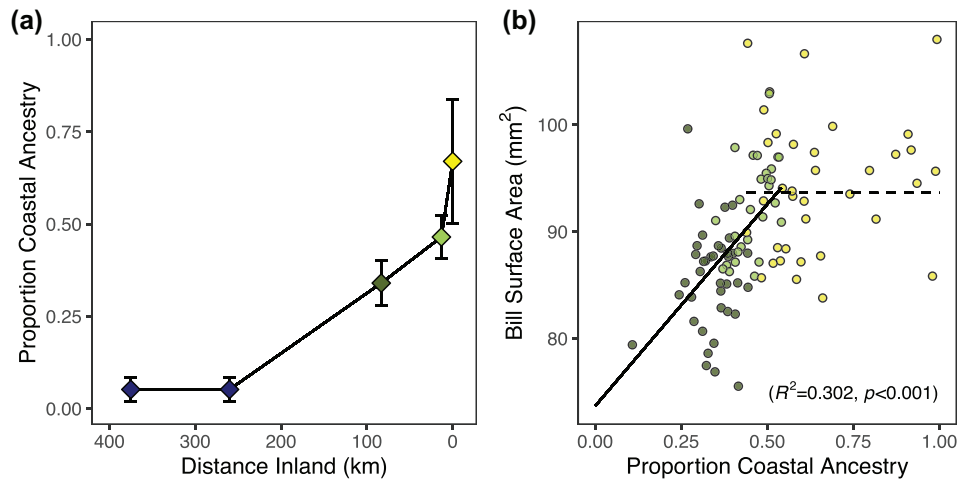
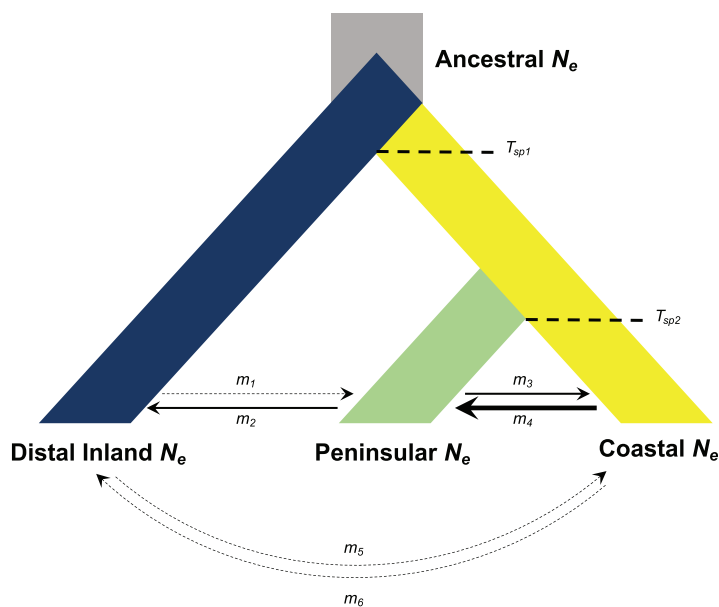


Figure 4. Clines in ancestry assignment in relation to geographic and morphological variation. Colors correspond to region: navy is distal inland, dark green is proximal inland, light green is peninsular, and yellow is coastal. (a) Regional assignment to the coast-associated genetic cluster from *STRUCTURE* analysis when $K = 2$ plotted against distance inland from the Atlantic song sparrow breeding range. Diamonds are the mean value for the sampling region with bars showing standard deviation. (b) Relationship between ancestry assignment and an ecologically relevant morphological trait, bill surface area, for a subset of adult males for which morphological metadata were available. There is a significant, positive linear relationship between bill surface area and the proportion of coastal ancestry for individuals of the non-coastal populations ($R^2 = 0.302, P < 0.001$; solid line), suggesting that gene flow influences this trait. For the coastal population, however, there is no relationship between bill surface area and proportion of coastal ancestry ($R^2 = 0.003, P = 0.739$). The dashed line shows the mean bill surface area for the coastal population.



Parameter	Point Estimate	95% CI
Log-Likelihood	-3477.65	
Θ	78.50	50.76 – 106.24
Ancestral N_e	2,049	1,325 – 2,774
Historical Peninsular/Coastal N_e	5,593	501 – 16,250
Modern Distal Inland N_e	30,996	24,179 – 45,263
Modern Peninsular N_e	62,881	36,405 – 118,298
Modern Coastal N_e	17,968	14,122 – 26,020
T_{sp1}	267,796	216,209 – 375,774
T_{sp2}	63,786	50,825 – 90,916
m_1 (Inland → Peninsular)	7.48E-05	0 – 5.86E-04
m_2 (Peninsular → Inland)	2.73E-04	1.65E-04 – 3.26E-04
m_3 (Peninsular → Coastal)	3.20E-04	1.89E-04 – 3.82E-04
m_4 (Coastal → Peninsular)	2.51E-03	4.05E-04 – 3.52E-03
m_5 (Inland → Coastal)	3.31E-04	0 – 9.42E-04
m_6 (Coastal → Inland)	5.60E-05	0 – 3.09E-04

Figure 5. Best fit demographic model from *Moments* analysis. Colors correspond to region: navy is distal inland, dark green is proximal inland, light green is peninsular, and yellow is coastal. The most likely model supported divergence in allopatry followed by gene flow in secondary contact. Migration rates shown represent proportion of individuals in the population whose parents were from the donor population. Divergence times (T_{sp1} and T_{sp2}) are in years. Since divergence, gene flow was greatest from coastal (*M. m. atlantica*) populations inland to peninsular populations (bold arrow). Confidence intervals for some migration rates encompassed zero (dashed arrows).

($m_2 = 2.73 \times 10^{-4}$, 8.5 mpg) and coastal ($m_3 = 3.20 \times 10^{-4}$, 5.7 mpg) populations. Other migration rates in the best fit model were small with CIs encompassing zero (Fig. 5). Although it is worth considering the possibility that the limited gene flow between the distal inland population and others may be in part an artifact of a sampling gap created by excluding proximal inland populations from the analysis, limited gene flow in the distal inland population is consistent with the results of genetic clustering.

IDENTITY OF DIVERGENT LOCI

Two RAD-loci were identified as ϕ_{ST} outliers for all four coastal-inland comparisons of the subset. The first (CLocus-6013, length = 199 bp, number of SNPs = 1) was 607 bp downstream of the gene *SENP5* in the white-throated sparrow genome (MegaBLAST e -value = 9×10^{-89}). The second (CLocus-29705, length = 290 bp, number of SNPs = 1) was in an intergenic region of the white-throated sparrow genome (e -value = 4×10^{-64}). The nearest protein coding sequences to CLocus-29705 were *CD109* (175 kbp upstream) and *COL12A1* (283 kbp downstream). Notably, CLocus-29705 was also 567 kbp upstream of *SENP6*, a gene closely related to *SENP5*. In the zebra finch genome, CLocus-6013 aligned to chromosome 9 and CLocus-29705 aligned to chromosome 3.

When birds from all coastal and all distal inland sites were compared as regional populations, CLocus-6013 was the locus with the greatest sequence divergence between the distal inland and coastal regions ($\phi_{ST} = 0.276$), and CLocus-29705 had the third greatest sequence divergence ($\phi_{ST} = 0.237$). The two SNPs in these outlier loci also had F_{ST} values well above average for comparison of the distal inland and coastal regions (mean SNP $F_{ST} = 0.013$, CLocus-6013 SNP $F_{ST} = 0.161$, CLocus-29705 SNP $F_{ST} = 0.131$; Fig. S1). The frequency of the coast-associated allele in the distal inland, proximal inland, peninsular, and coastal regions, respectively, was 0.08, 0.36, 0.34, and 0.56 for CLocus-6013 and 0.31, 0.61, 0.53, and 0.68 for CLocus-29705.

Discussion

The song sparrow provides many valuable opportunities to investigate the maintenance of phenotypic differentiation between parapatric populations that experience gene flow. Here, by examining how genome-wide patterns of population structure relate to variation in a divergent phenotype in conjunction with assessing models of demographic history, we provide an empirical contribution to the broader understanding of the process by which local adaptation evolves and is maintained. We demonstrate that in the Atlantic song sparrow, there is likely a major role for natural selection in the maintenance of a local adaptation in the face of maladaptive gene flow.

MAINTENANCE OF LOCAL ADAPTATION

Natural selection can play an important role in the maintenance of local adaptations in the face of gene flow (Tigano and Friesen 2016; e.g., Chavarria-Pizarro et al. 2019). In the Atlantic song sparrow, larger bill size was maintained in coastal birds despite significant genetic admixture. Thus, due to pressures from the coastal selection regime, populations of the Atlantic song sparrow appear to have maintained the phenotypically determinant alleles for a larger bill despite gene flow. However, in transitional habitats near the coast—where coastal selection pressures are relaxed—bill size varies with the extent of genetic admixture in backcrossed individuals. Therefore, selection pressures against large bills in near-coastal habitats are likely weaker than selection pressures for larger bills in coastal habitats.

A regularly occurring theme emerging from recent studies of local adaptation between populations with gene flow is the formation of local adaptations during an initial period of allopatry (e.g., Le Moan et al. 2016, 2019, Luzuriaga-Aveiga et al. 2021). Our findings suggest that these subspecies likely experienced a period of little-to-no gene flow during the beginning of their divergence, followed by extensive—but asymmetrical—gene flow as a result of secondary contact. Allopatry offers an opportunity for selection to alter allele frequencies without the diluting effects of gene flow from another population, promoting adaptive differentiation. Then, for taxa in secondary contact to maintain phenotypic differences despite interbreeding, local selection (or sexual selection) must sufficiently penalize introgression of maladaptive alleles at phenotypically determinant loci to overcome gene flow. Overall, our findings were notably similar to patterns characterized between an interbreeding pair of parapatric, altitudinally differentiated South American tanager sister species (Luzuriaga-Aveiga et al. 2021).

Contemporary admixture between the Atlantic song sparrow and eastern song sparrow subspecies could be a result of postglacial range shifts resulting in gene flow between populations that occupied formerly isolated refugia (Fry and Zink 1998). Moreover, recent anthropogenic changes to the landscape may have also played a role in allowing increased gene flow between these subspecies, as human development along the Mid-Atlantic coast has brought the eastern song sparrow closer to the Atlantic song sparrow throughout its range (Aldrich and Coffin 1980) or allowed emigrant Atlantic song sparrows to colonize these peripheral habitats where coastal selection is relaxed. The small N_e and restricted range of the Atlantic song sparrow make it vulnerable to genetic swamping by immigrants from a maladapted gene pool (Todesco et al. 2016); thus, if gene flow into coastal habitats increases due to coastal development, the phenotypic distinctions between these subspecies may eventually erode.

POTENTIAL FOR ECOLOGICAL SPECIATION

The possibility for ecological speciation with gene flow is predicated on the ability for selection to overcome migration in interbreeding taxa, and for pre- and/or postmating isolation mechanisms to develop in the process (Rundle and Nosil 2005; Pinho and Hey 2010). In the absence of allopatry or assortative mating, alleles under divergent selection generally require strong selection pressure to become fixed in their associated environment, as high migration rates regularly move locally advantageous alleles to a different environment and introduce maladaptive alleles to the local gene pool (Pinho and Hey 2010; Yeaman and Otto 2011). Thus, behaviors that facilitate reproductive isolation are one means of aiding in the maintenance of local adaptation, as they reduce gene flow between populations in contact and thereby lower the required degree of selection to prevent fusion (Servedio and Bürger 2020). Moreover, a trait that is both under divergent selection and associated with mating signals can behave as a “magic” trait, altering both sides of the migration-selection balance equation in favor of divergence (Servedio et al. 2011).

Introducing sexual selection on an adaptive trait further increases its selection coefficient and decreases gene flow; therefore, magic traits are an efficient means for non-allopatric populations to move toward ecological speciation. The bill is often implicated as a magic trait in avian evolution (Slabbekoorn and Smith 2002; Servedio et al. 2011), due to its ecological functions and its role in producing mating signals (Ballentine 2006; Huber and Podos 2006; Podos 2010; Ballentine et al. 2013a,b; Grant and Grant 2018). Therefore, divergent selection on the bill provides the opportunity for the development of assortative mating between the Atlantic song sparrow and eastern song sparrow, meriting future studies to characterize any existing differences in song or mate preference between these subspecies.

DETECTING GENES UNDER SELECTION

One of the identified outlier loci was in the genomic vicinity of a collagen-associated gene, *COL12A1*. Its distance from the gene suggests that this locus is unlikely to be a functional part of the gene itself (e.g., a *cis*-regulatory element), but instead that the locus may be associated with the gene through strong linkage disequilibrium (Nielsen 2005). Signatures of selection in collagen genes (*COL11A* and *COL4A5*) have been related to bill size in passerine species, and these same genes are associated with craniofacial development in humans (Bosse et al. 2017; Ravinet et al. 2018). *COL12A1* is associated with osteoblast differentiation and long bone length in mice (Izu et al. 2011). Notably, two collagen genes (*COL4A5* and *COL4A3BP*) were also divergent between a saltmarsh-adapted subspecies and an inland subspecies of song sparrow in western North America (Walsh et al. 2019).

Another gene identified as potentially under selection in the Atlantic song sparrow was a gene involved in the small ubiquitin-

related modifier (SUMO) system, *SENP5*. The SUMO system is a mechanism for the reversible posttranslational modification of protein function in eukaryotes (Nayak and Müller 2014) and can affect gene expression by regulating the function of transcription factors (Rosonina et al. 2010). *SENP5* plays a role in ribosomal development and mitochondrial fragmentation during mitosis by facilitating the maturation of SUMO2 and SUMO3 proteins and deconjugating SUMO2/3 from their substrates (Nayak and Müller 2014). Yet, *SENP5* did not show signals of adaptation in any of the four passerellid species that exhibit a parallel phenotypic response to the saltmarsh selection regime for which there are published whole-genome coastal-inland comparisons (Walsh et al. 2019). However, it is becoming well established that the parallel evolution of complex phenotypes is rarely the result of selection on the same genetic pathways (e.g., Barghi et al. 2019; Therkildsen et al. 2019; Walden et al. 2020; Miranda et al. 2021). In fact, 89% of the divergent genes were species specific between the four coastal-inland passerellid subspecies comparisons published by Walsh et al. (2019).

Because marker-based outlier methods are prone to false positives, we cannot conclusively identify either of these outlier loci as a target of selection based strictly on elevated ϕ_{ST} . Further, the distributions of per-SNP F_{ST} and per-locus ϕ_{ST} suggest that we likely did not detect the most differentiated regions of the genome between these subspecies with RADseq methods. As it did for a congeneric, coastally adapted subspecies, the coastal plain swamp sparrow (*M. georgiana nigrescens*; Deane-Coe et al. 2018a,b), comprehensively identifying regions of the genome under divergent selection in the Atlantic song sparrow will require the use of whole-genome sequencing data. Fortunately, whole-genome studies of the song sparrow have been recently made possible by the creation of an annotated reference genome for the species (Louha et al. 2020). In addition to better characterizing the genomic landscape of divergence between these subspecies, further assessment of adaptive traits and associating them with their underlying causal genetic variants will allow for a more robust understanding of how divergent selection maintains locally adapted populations despite gene flow.

AUTHOR CONTRIBUTIONS

JDC, HCL, DAL, and JEM designed the study. JDC performed laboratory work, drafted the manuscript, and produced figures. JDC and PMB performed data analyses. HCL, PMB, DAL, and JEM provided revisions to the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This project would not have been possible without samples provided by the Smithsonian Institution’s Migratory Bird Center and National Museum of Natural History. In particular, R. Greenberg helped design the collection of blood samples from the field and L. Schreffler collected and stored samples at the Smithsonian Migratory Bird Center. Funding for

sequencing was provided by the Virginia Society of Ornithology. Other funding was provided by George Mason University to HCL. M. Sikaroodi and M. Chua provided training and support for laboratory work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

Mitochondrial sequences are available on GenBank. A FASTA file of the STACKS catalog and VCF file of SNPs used for population structure analyses and demographic modeling are available on Dryad (<https://doi.org/10.5061/dryad.ghx3ffbr5>). Sample and site metadata are included as a table in the Supporting Information.

LITERATURE CITED

- Aldrich, J.W. (1984) Ecogeographical variation in size and proportions of song sparrows (*Melospiza melodia*). *Ornithol. Monogr.*, 35, iii–x, 1–134.
- Aldrich, J.W. & James, F.C. (1991) Ecogeographic variation in the American robin (*Turdus migratorius*). *Auk*, 108, 230–249.
- Aldrich, J.W. & Coffin, R.W. (1980) Breeding bird populations from forest to suburbia after thirty-seven years. *Am. Birds*, 34, 3–7.
- Ballentine, B. (2006) Morphological adaptation influences the evolution of a mating signal. *Evolution; Internation Journal of Organic Evolution*, 60, 1936–1934.
- Ballentine, B., Gkoo, K.W. & Greenberg, R. (2013a) Mechanisms of song divergence between swamp sparrow subspecies. *Behaviour*, 150, 1165–1181.
- Ballentine, B., Horton, B., Brown, E.T. & Greenberg, R. (2013b) Divergent selection on bill morphology contributes to nonrandom mating between swamp sparrow subspecies. *Animal Behaviour*, 86, 467–473.
- Bandelt, H.J., Forster, P. & Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37–48.
- Barghi, N., Tobler, R., Nolte, V., Jakšić, A.M., Mallard, F., Otte, K.A., Dolezal, M., Taus, T., Kofler, R. & Schlötterer, C. (2019) Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biology*, 17, e3000128.
- Bayona-Vásquez, N.J., Glenn, T.C., Kieran, T.J., Pierson, T.W., Hoffberg, S.L., Scott, P.A., Bentley, K.E., Finger, J.W., Louha, S., Troendle, N., et al (2019) Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ*, 7, e7724.
- Benham, P.M. & Cheviron, Z.A. (2020) Population history and the selective landscape shape patterns of osmoregulatory trait divergence in tidal marsh Savannah sparrows (*Passerculus sandwichensis*). *Evolution; Internation Journal of Organic Evolution*, 74, 57–72.
- Bird, J.P., Martin, R., Akçakaya, H.R., Gilroy, J., Burfield, I.J., Garnett, S.T., Symes, A., Taylor, J., Şekercioğlu, Ç.H. & Butchart, S.H.M. (2020) Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, 34, 1252–1261.
- Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G., McMahon, K., Poissant, J., Verhagen, I., et al (2017) Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, 358, 365–368.
- Browning, M.R. (1978) An evaluation of the new species and subspecies proposed in Oberholser's *Bird Life of Texas*. *Proc. Biol. Soc. Wash.*, 91, 85–122.
- Butlin, R.K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., Coyne, J.A., Galindo, J., Grahame, J.W., Hollander, J., et al (2014) Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution; Internation Journal of Organic Evolution*, 68, 935–949.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140.
- Cavalli-Sforza, L.L. & Edwards, A.W.F. (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics*, 19, 233–257.
- Chavarria-Pizarro, T., Gomez, J.P., Ungvari-Martin, J., Bay, R., Miyamoto, M.M. & Kimball, R. (2019) Strong phenotypic divergence in spite of low genetic structure in the endemic Mangrove Warbler subspecies (*Setophaga petechia xanthotera*) of Costa Rica. *Ecol. Evol.*, 9, 13902–13918.
- Coffman, A.J., Hsieh, P.H., Gravel, S. & Gutenkunst, R.N. (2016) Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, 33, 591–593.
- Coyne, J.A. & Orr, H.A. (2004) Speciation. Sinauer Associates, Inc., Sunderland, MA.
- Danner, R.M. & Greenberg, R. (2015) A critical season approach to Allen's rule: bill size declines with winter temperature in a cold temperate environment. *Journal of Biogeography*, 42, 114–120.
- Danner, R.M., Gulson-Castillo, E.R., James, H.F., Dzielski, S.A., Frank, D.C., Sibbald, E.T. & Winkler, D.W. (2016) Habitat-specific divergence of air conditioning structures in bird bills. *Auk*, 134, 65–75.
- Darwin, C. (1859) On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, Lond.
- Deane-Coe, P., Butcher, B.G., Greenberg, R. & Lovette, I.J. (2018a) Whole genome scan reveals the multigenic basis of recent tidal marsh adaptation in a sparrow. *bioRxiv*, <https://doi.org/10.1101/360008>.
- Deane-Coe, P., Greenberg, R., Lovette, I.J. & Harrison, R.G. (2018b) Low genomic divergence and high gene flow between locally adapted populations of the swamp sparrow. *bioRxiv*, <https://doi.org/10.1101/359992>.
- Dobzhansky, T. & Pavlovsky, O. (1957) An experimental study of interaction between genetic drift and natural selection. *Evolution; Internation Journal of Organic Evolution*, 11, 311–319.
- Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.*, 4, 359–361.
- Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, 14, 2611–2620.
- Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Feder, J.L., Egan, S.P. & Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28, 342–350.
- Fry, A.J. & Zink, R.M. (1998) Geographic analysis of nucleotide diversity and song sparrow (Aves: Emberizidae) population history. *Molecular Ecology*, 7, 1303–1313.
- Gawin, D.F., Rahman, M.A., Ramji, M.F.S., Smith, B.T., Lim, H.C., Moyle, R.G. & Sheldon, F.H. (2014) Patterns of avian diversification in Borneo: the case of the endemic mountain black-eye (*Chlorocharis emiliae*). *Auk*, 131, 86–99.
- Geist, N.R. (2000) Nasal respiratory turbinate function in birds. *Physiological and Biochemical Zoology*, 73, 581–589.
- Grant, P.R. & Grant, B.R. (2018) Role of sexual imprinting in assortative mating and premating isolation in Darwin's finches. *Proceedings of the National Academy of Sciences of the USA*, 115, E10879–E10887.

- Greenberg, R., Maldonado, J.E., Droege, S. & McDonald, M.V. (2006) Tidal marshes: a global perspective on the evolution and conservation of their terrestrial vertebrates. *Bioscience*, 56, 675–685.
- Greenberg, R., Cadena, V., Danner, R.M. & Tattersall, G. (2012) Heat loss may explain bill size differences between birds occupying different habitats. *PLOS One*, 7, e40933.
- Greenberg, R., Cammen, K.M., Wilson, A.G., Olsen, B.J., Ballentine, B., McInerney, N.C. & Fleischer, R.C. (2016) Geographic population structure and subspecific boundaries in a tidal marsh sparrow. *Conserv. Genet.*, 17, 603–613.
- Huber, S.K. & Podos, J. (2006) Beak morphology and song features covary in a population of Darwin's finches (*Geospiza fortis*). *Biological Journal of the Linnean Society*, 88, 489–498.
- Izu, Y., Sun, M., Zwolanek, D., Veit, G., Williams, V., Cha, B., Jepsen, K.J., Koch, M. & Birk, D.E. (2011) Type XII collagen regulates osteoblast polarity and communication during bone formation. *Journal of Cell Biology*, 193, 1115–1130.
- Jakobsson, M. & Rosenberg, N.A. (2007) *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.
- Jombart, T. (2008) *ade4*: an R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403–1405.
- Jombart, T., Devillard, S. & Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11, 94.
- Jouganous, J., Long, W., Ragsdale, A.P. & Gravel, S. (2017) Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*, 206, 1549–1567.
- Le Moan, A., Gagnaire, P.-A. & Bonhomme, F. (2016) Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Molecular Ecology*, 25, 3187–3202.
- Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D. & Hemmer-Hansen, J. (2019) Beyond parallel evolution: when several species colonize the same environmental gradient. *bioRxiv*, <https://doi.org/10.1101/662569>.
- Leigh, J.W. & Bryant, D. (2015) *PopART*: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6, 1110–1116.
- Louha, S., Ray, D.A., Winker, K. & Glenn, T.C. (2020) A high-quality genome assembly of the North American song sparrow, *Melospiza melodia*. *G3*, 10, 1159–1166.
- Luther, D. & Danner, R. (2016) Males with larger bills sing at higher rates in a hot and dry environment. *Auk*, 133, 770–778.
- Luzuriaga-Aveiga, V., Ugarte, M. & Weir, J.T. (2021) Distinguishing genomic homogenization from parapatric speciation in an elevationally replacing pair of *Ramphocelus* tanagers. *Molecular Ecology*, 30, 5517–5529.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Marques, D.A., Lucek, K., Meier, J.I., Mwaiko, S., Wagner, C.E., Excoffier, L. & Seehausen, O. (2016) Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLOS Genetics*, 12, e1005887.
- Mayr, E. (1963) *Animal species and evolution*. Belknap Press of Harvard Univ. Press, Cambridge, MA.
- Mikles, C.S., Aguillon, S.M., Chan, Y.L., Arcese, P., Benham, P.M., Lovette, I.J. & Walsh, J. (2020) Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird. *Molecular Ecology*, 29, 4295–4307.
- Miranda, I., Giska, I., Farelo, L., Pimenta, J., Zimova, M., Bryk, J., Dalén, L., Mills, L.S., Zub, K. & Melo-Ferreira, J. (2021) Museomics dissects the genetic basis for adaptive seasonal coloration in the least weasel. *Molecular Biology and Evolution*, 38, 4388–4402.
- Nayak, A. & Müller, S. (2014) SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biology*, 15, 422.
- Nielsen, R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, 39, 197–218.
- Nosil, P. (2012) *Ecological speciation*. Oxford Univ. Press, Oxford, U.K.
- Paris, J.R., Stevens, J.R. & Catchen, J.M. (2017) Lost in parameter space: a road map for Stacks. *Methods in Ecology and Evolution*, 8, 1360–1373.
- Patten, M.A. & Pruett, C.L. (2009) The song sparrow, *Melospiza melodia*, as a ring species: patterns of geographic variation, a revision of subspecies, and implications for speciation. *Systematics and Biodiversity*, 7, 33–62.
- Patten, M.A., Rotenberry, J.T. & Zuk, M. (2004) Habitat selection, acoustic adaptation, and the evolution of reproductive isolation. *Evolution; International Journal of Organic Evolution*, 58, 2144–2155.
- Peñalba, J.V., Joseph, L. & Moritz, C. (2019) Current geography masks dynamic history of gene flow during speciation in northern Australian birds. *Molecular Ecology*, 28, 630–643.
- Pinho, C. & Hey, J. (2010) Divergence with gene flow: models and data. *Annual Review of Ecology, Evolution, and Systematics*, 41, 215–230.
- Podos, J. (2010) Acoustic discrimination of sympatric morphs in Darwin's finches: a behavioural mechanism for assortative mating? *Philos. Trans. R. Soc. B Biol. Sci.*, 365, 1031–1039.
- Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á. & Lareu, M.V. (2013) An overview of *STRUCTURE*: applications, parameter settings, and supporting software. *Front. Genet.*, 4, 98.
- Price, T. (2008) *Speciation in birds*. Roberts and Co., Greenwood Village, CO.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Pruett, C.L. & Winker, K. (2010) Alaska song sparrows (*Melospiza melodia*) demonstrate that genetic marker and method of analysis matter in subspecies assessments. *Ornithol. Monogr.*, 67, 162–171.
- Pruett, C.L., Arcese, P., Chan, Y.L., Wilson, A.G., Patten, M.A., Keller, L.F. & Winker, K. (2008) The effects of contemporary processes in maintaining the genetic structure of western song sparrows (*Melospiza melodia*). *Heredity*, 101, 67–74.
- Ravinet, M., Elgvin, T.O., Trier, C., Aliabadian, M., Gavrilov, A. & Sætre, G.-P. (2018) Signatures of human-commensalism in the house sparrow genome. *Proc. R. Soc. B Biol. Sci.*, 285, 20181246.
- Rochette, N.C. & Catchen, J.M. (2017) Deriving genotypes from RAD-seq short-read data using *Stacks*. *Nature Protocols*, 12, 2640–2659.
- Rochette, N.C., Rivera-Colón, A.G. & Catchen, J.M. (2019) *Stacks 2*: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28, 4737–4754.
- Rosenblum, E.B. & Harmon, L.J. (2011) “Same same but different”: replicated ecological speciation at White Sands. *Evolution; International Journal of Organic Evolution*, 65, 946–960.
- Rosonina, E., Duncan, S.M. & Manley, J.L. (2010) SUMO functions in constitutive transcription and during activation of inducible genes in yeast. *Genes & Development*, 24, 1242–1252.
- Rundle, H.D. & Nosil, P. (2005) Ecological speciation. *Ecol. Lett.*, 8, 336–352.
- Schluter, D. & Conte, G.L. (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences of the USA*, 106, 9955–9962.
- Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Sætre, G.-P., Bank, C., Brännström, Å., et al (2014) Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176–192.

- Sendell-Price, A.T., Ruegg, K.C., Anderson, E.C., Quilodr n, C.S., Van Doren, B.M., Underwood, V.L., Coulson, T. & Clegg, S.M. (2020) The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes (*Zosterops lateralis*). *G3*, 10, 3147–3163.
- Servedio, M.R. & B rger, R. (2020) The effectiveness of pseudomagic traits in promoting divergence and enhancing local adaptation. *Evolution; International Journal of Organic Evolution*, 74, 2438–2450.
- Servedio, M.R., Doorn, G.S.V., Kopp, M., Frame, A.M. & Nosal, P. (2011) Magic traits in speciation: ‘magic’ but not rare? *Trends in Ecology & Evolution*, 26, 389–397.
- Slabbekoorn, H. & Smith, T.B. (2002) Bird song, ecology and speciation. *Philos. Trans. R. Soc. B Biol. Sci.*, 357, 493–503.
- Song, G., Zhang, R., Machado-Stredel, F., Alstr m, P., Johansson, U.S., Irestedt, M., Mays, H.L., McKay, B.D., Nishiumi, I., Cheng, Y., et al (2020) Great journey of great tits (*Parus major* group): origin, diversification and historical demographics of a broadly distributed bird lineage. *Journal of Biogeography*, 47, 1585–1598.
- Swarth, H.S. (1923) The systematic status of some northwestern song sparrows. *Condor*, 25, 214–223.
- Therkildsen, N.O., Wilder, A.P., Conover, D.O., Munch, S.B., Baumann, H. & Palumbi, S.R. (2019) Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science*, 365, 487–490.
- Tigano, A. & Friesen, V.L. (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*, 25, 2144–2164.
- Todd, W.E.C. (1924) A new song sparrow from Virginia. *Auk*, 41, 147–148.
- Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., H bner, S., Heredia, S.M., Hahn, M.A., Caseys, C., Bock, D.G., et al (2016) Hybridization and extinction. *Evol. Appl.*, 9, 892–908.
- Via, S. (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the USA*, 106, 9939–9946.
- Walden, N., Lucek, K. & Willi, Y. (2020) Lineage-specific adaptation to climate involves flowering time in North American *Arabidopsis lyrata*. *Molecular Ecology*, 29, 1436–1451.
- Walsh, J., Benham, P.M., Deane-Coe, P.E., Arcese, P., Butcher, B.G., Chan, Y.L., Cheviron, Z.A., Elphick, C.S., Kovach, A.I., Olsen, B.J., et al (2019) Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evol. Lett.*, 3, 324–338.
- Wang, I.J. & Bradburd, G.S. (2014) Isolation by environment. *Molecular Ecology*, 23, 5649–5662.
- Yeaman, S. & Otto, S.P. (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution; International Journal of Organic Evolution*, 65, 2123–2129.
- Yoder, J.B., Clancey, E., Roches, S.D., Eastman, J.M., Gentry, L., Godsoe, W., Hagey, T.J., Jochimsen, D., Oswald, B.P., Robertson, J., et al (2010) Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, 23, 1581–1596.
- Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., et al (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346, 1311–1320.
- Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7, 203–214.
- Zou, F., Lim, H.C., Marks, B.D., Moyle, R.G. & Sheldon, F.H. (2007) Molecular phylogenetic analysis of the grey-cheeked fulvetta (*Alcippe morrisonia*) of China and Indochina: a case of remarkable genetic divergence in a ‘species’. *Molecular Phylogenetics and Evolution*, 44, 165–174.

Associate Editor: B. Emerson
Handling Editor: Prof. T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. List of samples and associated data.

Table S2. Overview of sampling sites.

Table S3. Results of hierarchical AMOVA of ND2 haplotypes.

Table S4. Best fit model selection criteria for demographic analysis.

Figure S1. Results of *STRUCTURE HARVESTER*.

Figure S2. Per-SNP differentiation (F_{ST}) and per-locus divergence (ϕ_{ST}) distributions between coastal and distal inland song sparrows.

Figure S3. Isolation by distance plot.

Figure S4. Boxplots of bill surface area for adult male song sparrows by sampling region.

Supplemental Methods. Genotyping and parameter selection in *STACKS*.