Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird

Honors Thesis
Presented to the College of Agriculture and Life Sciences,
Cornell University
in Partial Fulfillment of the Requirements for the
Biological Sciences Honors Program

by

Chloe Sofia Mikles
May 2020

Advised by:

Irby J. Lovette

Director, Fuller Evolutionary Biology Program, Cornell Lab of Ornithology

Jennifer Walsh

Postdoctoral Fellow, Fuller Evolutionary Biology Program, Cornell Lab of Ornithology

ABSTRACT

Elucidating forces capable of driving species diversification in the face of gene flow remains a key goal in evolutionary biology. Song sparrows, *Melospiza melodia,* occur as 25 subspecies in diverse habitats across North America, are among the continent's most widespread vertebrate species, and are exemplary of many highly variable species for which the conservation of locally adapted populations may be critical to their range-wide persistence. I focus here on six morphologically distinct subspecies resident in the San Francisco Bay region, including three salt-marsh endemics and three residents in upland and riparian habitats adjacent to the Bay. I used reduced-representation sequencing to generate 2,773 SNPs to explore genetic differentiation, spatial population structure, and demographic history. Clustering separated individuals from each of the six subspecies, indicating subtle differentiation at microgeographic scales. Evidence of limited gene flow and low nucleotide diversity across all six subspecies further supports a hypothesis of isolation among locally adapted populations. This research suggests that natural selection for genotypes adapted to salt marsh environments and changes in demography over the past century have acted in concert to drive the patterns of diversification reported here. My results offer evidence of microgeographic specialization in a highly polytypic bird species long discussed as a model of sympatric speciation and rapid adaptation, and they support the hypothesis that conserving locally adapted populations may be critical to the range-wide persistence of similarly highly variable species.

INTRODUCTION

While geographic isolation is viewed as a predominant driver of divergence among populations, a substantial and growing literature suggests that evolutionary diversification can also arise via ecological selection across spatially heterogeneous environments at fine spatial scales (e.g., Kavanagh et al. 2010; Richardson & Urban, 2013; Krueger-Hadfield et al., 2013; Langin et al., 2015). Theory suggests that adaptive divergence at microgeographic scales (i.e., within the potential dispersal radius of an organism) can occur when natural selection overpowers the homogenizing effects of gene flow (Wright, 1969; Hendry et al., 2001; Richardson et al., 2014). Despite supporting theory (Bolnick, 2006; Richardson et al., 2014), empirical studies of adaptation at microgeographic scales remain scarce, perhaps because sampling often occurs at scales exceeding the dispersal capabilities of the organisms under study (Richardson et al., 2014). Nevertheless, a number of classic examples of fine-scaled structuring in bird, plant, insect, and fish populations have appeared (Postma & van Noordwijk, 2005; Antonovics, 2006; Nosil & Crespi, 2004; Hendry, Taylor & McPhail, 2002). As a consequence, an emerging consensus suggests that local adaptation at microgeographic scales may be more common than once thought. This conclusion challenges the idea that population divergence requires periods of geographic isolation and supports the hypothesis that the conservation of locally adapted populations may be crucial to ensuring the persistence of species across their range (e.g., Aitken and Whitlock, 2013; Ho and Zhang, 2018, Reid et al. 2018).

Whereas it may be expected that adaptive divergence be greater when gene flow and genetic drift between populations are low (Charmantier et al., 2016), there is a growing appreciation for the complexity that arises when gene flow and selection are acting in concert. Gene flow can result in increased variation and genetic novelty, thus promoting divergence and

adaptive potential (Garnet et al., 2007) or it can have an opposite, homogenizing effect that prevents divergence (Slatkin, 1985). Elucidating such processes thus requires further empirical research on species displaying evident diversification at fine spatial scales. Given that some of the best examples of microgeographic adaptation involve mainly sedentary organisms such as plants (Steiner & Berrang, 1990; Antonovics, 2006; Hendrick et al., 2016), my research aimed to investigate such processes in a more mobile organism, wherein the signal of microgeographic adaptation may be less evident and underappreciated as a factor affecting the rate of diversification under gene flow.

Here I leverage a song sparrow "subspecies complex" resident in the San Francisco Bay region of California, USA (referred to as the Bay or bay region throughout this manuscript), to evaluate genomic differentiation at fine spatial scales (70 x 100 km) and spanning multiple fresh-to-salt water habitat gradients. At the broadest scale, the song sparrow (*Melospiza melodia*) is one of North America's most abundant, adaptable, and wide-spread bird species. However, at finer scales, song sparrows display differences in song, size, and plumage coloration that imply a history of local adaptation, particularly where birds reside on or near territories year-round (Arcese et al., 2002; Pruett & Winker, 2005). Despite being found in close geographic proximity, five recognized and phenotypically distinct subspecies of song sparrow reside year-round in the San Francisco Bay region (Figure 1a; Grinnell & Miller, 1944). *M. m. samuelis* (Samuels Song Sparrow in the marshes of the San Pablo Bay), *M. m. maxillaris* (Suisun Song Sparrow in the marshes of Suisun Bay), and *M. m. pusillula* (Alameda Song Sparrow in the marshes of South San Francisco Bay) are endemic subspecies found only in the saline marshes of their three respective sub-bays. *M. m. gouldii* (Marin Song Sparrow) occupies the surrounding upland regions along the coast and surrounding the Bay, and *M. m. heermanni* (Heermann's Song

Sparrow) occupies riparian zones east of the Bay (Chan & Arcese, 2002). A sixth putative subspecies occupying upland habitat around the Bay, *M. m. santaecrucis,* was described by Grinnell (1901) and is thought to be the result of hybridization between *M. m. gouldii* and *M. m. heermanni* (Patten & Pruett, 2009). In 2008, four of these six subspecies were classified by the State of California as "Species of Special Concern:" *pusillula*, *maxillaris, samuelis, and heermanni* (formerly *M. m. mailliardi;* Northern Central Valley population). The further loss of marsh habitat represents a severe threat to the persistence of these endemic subspecies (Sauer et a. 2014) by raising the likelihood of extinction in a potentially unique group of song sparrows adapted to saline conditions.

The diversification of song sparrows in this region has received over a century of investigation into their morphological, genetic, and behavioral divergence (Grinnell, 1913; Huxley, 1942; Miller, 1947, 1956; Marshall, 1948a,b; Johnston, 1956a,b; Mayr, 1963, Chan & Arcese, 2002, 2003). Results to date have revealed geographic structure in plumage and morphology, and blood type, but less evidence in behavioral or life history traits (Johnston, 1956a,b; Mulligan, 1963; Ferrell, 1966; Chan & Arcese, 2002, 2003). Several mechanisms have been proposed to explain the diversification of song sparrows in San Francisco Bay, including habitat selection (Grinnell, 1913), habitat selection coupled with isolation (Marshall, 1948b), drift among populations of small demographic size (Miller, 1947; Ferrell 1966), geographic isolation in different arms of the San Francisco Bay (Mayr, 1942), strong selection favoring phenotypic variation despite ongoing gene flow (Aldrich, 1984; Zink & Dittmann, 1993), and phenotypic plasticity in the absence of genetic variation (Zink & Dittmann, 1993; Smith, 1998). In contrast, prior studies of genetic variation in microsatellite (Chan & Arcese, 2002) and

mitochondrial DNA (Zink & Dittmann, 1993; Fry & Zink, 1998) suggest that divergence among subspecies in the bay region is low, despite representing diagnosable phenotypes.

The observed morphological variation among San Francisco Bay song sparrows is consistent with well-documented patterns of phenotypic convergence observed in tidal marsh sparrows in traits with putative adaptive functions to tidal environments (e.g., Basham & Mewaldt 1987, Grenier & Greenberg, 2006; Greenberg & Olsen, 2010; Luttrell et al., 2015; Tattersall et al., 2017; Walsh et al., 2019). For example, *maxillaris* has a 40% greater bill depth–a hypothesized mechanism for facilitating evaporative heat loss in freshwater-limited and exposed environment (Greenberg et al. 2012; Tattersall et al., 2017)–than its upland counterpart *gouldii,* and experimental studies suggest significant salt tolerance in *pusillula*, whereas individual *gouldii* were unable to maintain bodyweight when consuming saline solutions (Basham & Mewaldt, 1987). Taken together, these patterns are consistent with a hypothesis of phenotypic differentiation driven by strong selection on functional loci that influence traits intimately linked to individual performance in a local environment, and which reflect the outcome of adaptative evolution at microgeographic scales in the absence of differentiation at neutral loci.

To test this hypothesis, I further elucidated the microgeographic patterns of genetic structure in San Francisco Bay song sparrows by replicating the sampling design of Chan & Arcese (2002) to assess microsatellite diversification in the bay region using updated genomic methods and high-resolution sequencing. Specifically, my objectives were to: 1) characterize patterns of song sparrow diversification in the San Francisco Bay, 2) employ genomic markers to delineate population boundaries at a fine spatial scale, and 3) combine population genetic and demographic analyses with outlier scans for loci under divergent selection to estimate the

contributions of previously proposed mechanisms underlying observed patterns of divergence in the region. To do so, I am re-visiting existing hypotheses on the divergence of song sparrows in this region to evaluate the potential roles of genetic drift, geographic isolation, and local adaptation in divergence, and to speculate on the roles that local adaptation to tidal marsh environments may play in structuring populations. To conclude, I will discuss my results in the context of evolutionary theory and provide information to aid in the management of endemic song sparrows in the bay region based on degree of evolutionary divergence.

MATERIALS AND METHODS

Study System and Sampling

All birds were sampled during the breeding season (March to May) by Chan & Arcese (2002). Individuals were sampled in 1999 from both tidal salt marshes and the upland riparian areas in the bay region (Table 1; Figure 1a). In total, I obtained samples from 160 individuals from 12 different populations representing all six of the putative subspecies in the region.

Molecular Methods

Genomic DNA was extracted from blood samples using Qiagen DNeasy blood and tissue extraction kits (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. I used double-digest restriction site-associated DNA sequencing (ddRAD) according to the protocol of Peterson et al. (2012) with modifications following Thrasher et al. (2017) to construct libraries and generate genomic data for single nucleotide polymorphism (SNP) discovery. In brief, ~200-500ng of DNA was used for each individual (concentrations ranged from 8 to 57 ng/μL); concentrations were determined using the Qubit fluorometer and dsDNA broad range assay kit (ThermoFisher Scientific, Q32853, Life Technologies, Carlsbad, CA, USA). DNA was digested with *Sbf*I and *Msp*I restriction enzymes (New England Biolabs, MA, USA) and ligated to one of

20 P1 adaptors and a P2 adaptor using T4 DNA ligase (New England Biolabs). Samples with similar DNA concentrations were then separately pooled into a total of eight index groups. Samples were purified with MagNA beads prepared according to the protocol of Rohland and Reich (2012) to remove the enzymes and small DNA fragments. Fragments were size selected between 400 and 700 bp using the Blue Pippin (Sage Science, MA, USA) to ensure the same loci were recovered in all index groups. I performed low-cycle PCR with Phusion High-Fidelity DNA Polymerase (New England Biolabs) for each index group, purified with SPRI beads to eliminate small fragments, and then visualized the product on 1% agarose gel and fragment Bioanalyzer (Agilent Technologies, CA, USA). The index groups were combined and sequenced on one Illumina HiSeq 2500 lane (read length 100 bp, single end) using the rapid run mode at the Cornell University Biotechnology Resource Center.

Data Processing and SNP Calling

Sequence quality was assessed using FASTQC version 0.11.8. (www.bioinformatics.babraham.ac.uk/projects/fastqc). I removed three samples due to poor quality. The remaining 157 individuals were filtered for quality using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit), removing sequences with Phred quality scores below 10 and sequences with more than 5% of bases with Phred quality scores below 20. I demultiplexed sequences using the command "process_radtags" in STACKS version 1.48 (Catchen et al., 2011) and additionally filtered samples to only retain reads that passed the Illumina chastity filter, contained an intact *Sbf*I RAD site, contained a unique sample barcode, and did not contain Illumina indexing adapters. To account for differences in length, the remaining filtered and demultiplexed reads were trimmed to 94 base pairs at the 3' end using FASTX TRIMMER (FASTX-Toolkit).

Sequences were aligned to a song sparrow (*Melospiza melodia*) reference genome (Feng et al., in review) using BOWTIE2 version 2.3. Mapped reads were then analyzed using the ref_map.pl pipeline in STACKS. I allowed five mismatches between sample loci and required a minimum of ten identical raw reads to make a stack. I ran the Populations module in STACKS for one population (-p) and required that a locus be present in a minimum of 80% of individuals to be processed (-r). In addition to obtaining all SNPs per locus (10,270), I created a subset of SNPs that included only the first SNP per stack (2,773). To avoid bias associated with physical linkage (O'Leary et al. 2018), I used the reduced, unlinked dataset with 2,773 SNPs for all analyses, unless otherwise stated.

I used the "relatedness" function within VCFTOOLS to obtain pairwise relatedness statistics between all individuals (Yang et al., 2011) prior to downstream analyses. Individuals that are very closely related have the potential to skew patterns of population structuring. The expectation for this function is that comparisons of theoretically unrelated individuals within populations are assigned a value of 0 and individuals compared against themselves have a value of 1.

Characterizing Patterns of Genomic Divergence in the San Francisco Bay

I employed several programs to identify patterns of genetic substructure while investigating patterns of diversification among sparrow populations around the bay region. I used Populations in STACKS to measure variation across the 12 populations representing the six subspecies by calculating pairwise F_{ST} and the observed and expected heterozygosity. Using the R package "POPPR" (Kamvar et al., 2015), I analyzed hierarchical genetic structuring using analysis of molecular variance (AMOVA). I conducted 999 permutations to test for significance in F_{IT}, F_{IS}, and F_{ST}.

To visualize genetic clustering among subspecies, I conducted principle component analysis (PCA) using the R package "SNPRELATE" (Zheng et al., 2012). I analyzed patterns of genetic structure in a Bayesian framework using the program STRUCTURE version 2.3.4 (Pritchard et al., 2000). I conducted 10 runs for each value of K=1-7 for our six subspecies; each run consisted of a 200,000 iteration burn-in followed by 300,000 sampling iterations. I implemented the admixture model and used an allele frequency prior that was estimated from an initial run of K=1 (λ=0.29). I determined the most likely number of clusters (K) using the method described by Evanno et al. (2005) as implemented in STRUCTUREHARVESTER version 0.6.94 (Earl & vonHoldt, 2012). I averaged results across all of our runs using the greedy algorithm in CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007) and visualized results using DISTRUCT version 1.1 (Rosenberg, 2003).

Spatial Patterns and the Influence of Geographic & Environmental Isolation

I evaluated the role of spatial patterns, specifically geographic isolation, in influencing genetic structure among populations in the bay region. I first used the R package "CONSTRUCT" (Bradburd et al., 2018) to visualize continuous and discrete population structure using both spatial and nonspatial models. CONSTRUCT differs from traditional model-based clustering methods designed to detect discrete population structure (i.e., STRUCTURE) in that it jointly models the effects of both discrete population structure and continuous isolation by distance on sample relationships (Bradburd et al., 2018). I ran CONSTRUCT for K=1-7 both with and without spatial information, with 10,000 iterations for three independent chains, and subsequently compared these models using cross-validation analysis. If the predictive accuracy of the spatial model is higher than the non-spatial model, it is assumed that spatial patterns (i.e., isolation by distance) are contributing to the observed genetic structure. I also tested for isolation by distance

(IBD) using the R package "ADEGENET" (Jombart and Ahmed 2008), using locations from all 12 sampling sites.

I additionally visualized spatial population structure and genetic dissimilarity between populations using estimated effective migration surfaces (EEMS; Petkova et al., 2016). EEMS models the genetic and geographic relationships using matrices of average pairwise genetic distance and geographic distances between individuals. The program models effective migration based off of the stepping stone model and Markov Chain Monte Carlo sampling from the posterior probability distribution of the genetic dissimilarities. I calculated pairwise genetic distance using the "bed2diffs" function in EEMS and geographically referenced all individuals using respective sampling site locations. I ran EEMS using three independent chains each with 10,000,000 MCMC iterations with a 20,000,000 burn-in and a population grid density of 700 demes, adjusting all other parameters until achieving a proposal acceptance rate between 20 and 30%. I used the R package "rEEMSplots" to plot model fits (to confirm chain convergence for each individual run), and to generate contour maps of effective migration and nucleotide diversity.

I evaluated whether the magnitude of genetic differentiation (F_{ST}) between the salt marsh populations and the upland populations could be attributed exclusively to genetic drift. To do this, I used linear regression analyses to correlate pairwise estimates of F_{ST} between each salt marsh population vs. *gouldii* (chosen as a representative upland population) as a response variable and various measures of genetic diversity as the predictor variables (Funk et al., 2016). For the predictor variables, I used estimates of observed and expected heterozygosity (estimated in STACKS using only variant sites) and nucleotide diversity (estimated in STACKS using all positions, variant and fixed). Under the assumption that genetic drift is responsible for driving

high levels of genetic differentiation, I expect to see a negative correlation between low within population diversity and pairwise F_{ST} between marsh and upland populations.

Lastly, I evaluated whether the magnitude of genetic differentiation (F_{ST}) could be attributed to local differences in environment. I performed partial-mantel tests using the R package "VEGAN" (Oksanen et al., 2019) to test for evidence of adaptive differentiation between subspecies considering environmental factors, in addition to pairwise F_{ST} and geographic distance. I obtained environmental data on water salinity and air temperatures of different marshes and upland habitats from USGS and NOAA repositories (Schraga and Cloern 2017, NOAA, 1999).

Identifying Putative Outliers

To assess whether selection is contributing to the fine scale diversification in the bay region, I scanned for outliers using two different approaches. For both of these approaches, I used the full dataset (10,270 SNPs). First, I identified outlier SNPs using BAYESCAN version 2.1 (Foll & Gaggiotti, 2008). I scanned all SNPs in pairwise comparisons across all of our subspecies and allowed for a false discovery rate of 5%. Second, I used per locus SNP F_{ST} estimates (calculated in VCFTOOLS; Danecek et al., 2011) to identify putative outliers in my dataset. For each pairwise comparison, I characterized a SNP as having elevated levels of differentiation if it had an F_{ST} estimate greater than 5 standard deviations above the mean. To assess the contribution of neutral versus potentially adaptive SNPs in delineating population boundaries, I repeated the PCA using only the SNPs identified as putatively under selection using either method. For the F_{ST} approach, a SNP had to be classified as elevated in more than one pairwise comparison. For all outlier SNPs, I assessed the homology and the distribution of

loci across the Zebra finch (*Taeniopygia guttata*) genome using BLASTn (Altschul et al., 1990). To assign matches, I required an E value of $<1.0x10^{-5}$ and >70% identity score.

Demographic Models

**This section was contributed by co-author of the corresponding manuscript, Phred Benham, and is included here for the purpose of allowing me to evaluate all potential drivers of diversification in this system. For this section, I retain the pronoun "we," as I did not personally conduct the analysis.*

To estimate the demographic history of song sparrows in the bay region, we fit a series of demographic models to the site frequency spectrum using the program MOMENTS (Jouganous et al., 2017). To maximize the number of SNPs with coverage across individuals we excluded individuals with more than 5% missing data and created a site frequency spectrum with a VCF file that had zero missing data. This VCF comprised 4,366 SNPs across 82 individuals: 13 *M. m. gouldii,* 17 *M. m. heermanni,* 15 *M. m. maxillaris*, 20 *M. m. pusillula*, and 17 *M. m. samuelis* (because *santaecrucis* is not a recognized subspecies in the bay region, we have excluded these individuals from demographic analyses). The VCF was converted to a MOMENTS input format using a custom python script from Isaac Overcast (https://github.com/isaacovercast/WCS/blob/master/vcf2dadi/vcf2dadi.py).

We inferred divergence times, population sizes, and gene flow following an initial split between *M. m. pusillula* and all other subspecies. This split reflects patterns of population structure indicated in other analyses and inference based on a two-dimensional site frequency spectrum allows us to more easily explore the possibility of bottleneck or other population histories. For all demographic models we compared a history of continuous gene flow to a history without continuous gene flow following the population split. We also compared a history

of constant population size, exponential growth, or a bottleneck in either *M. m. pusillula*, the other subspecies, or in both populations (see Figure S1 for illustration of all models compared). We additionally estimated divergence time and migration history among three populations (*M. m. pusillula, M. m. gouldii,* and all others) by comparing all combinations of three-population tree in preliminary analyses, and founding a splitting history of (*pusillula*, (*gouldii*, all others)) as best supported. Although five population demographic analyses are possible in MOMENTS these analyses are computationally demanding and marked population structure was not detected in all subspecies.

For all two-population models, 15 optimizations were run from different starting parameters using the perturb function in MOMENTS with a maximum number of iterations set to 30. To ensure that a global optimum for a given model had been reached we ran one final optimization with 75 iterations using the parameter values estimated from the shorter run with the highest likelihood. Given the larger parameter numbers and decreased computational speed for three and five population models, we ran 10 optimizations with maximum iterations of 10, and a final run with iterations of 50. We calculated demographic parameter values from the estimated value of theta ($4N_e\mu L$; L is sequence length) based on the assumption of a 1 year generation time for song sparrows (Arcese et al. 2002), the average substitution rate for Passeriformes (3.3×10^{-9} substitutions/site/year; Zhang *et al.* 2014), and total sequence length equal to 260,662 bp (2,773 loci x 94 bp /loci). To calculate uncertainty around each parameter estimate, we generated 100 bootstrapped site frequency spectra using the bootstrap function within MOMENTS. Bootstrapped spectra were then used to calculate parameter uncertainties using the Godambe information matrix (GIM; Coffman *et al.* 2016). Large parameter uncertainties based on the GIM method can indicate that the distribution is non-normal. To account for this we

also estimated parameter uncertainties using a non-parametric bootstrap approach where we fit the same demographic model to each of the 100 bootstrapped site frequency spectra to generate a distribution of parameter estimates. To infer the best fit model, we performed a likelihood ratio test with a GIM-based adjustment to account for the use of composite likelihoods (Coffmann *et al.* 2016).

RESULTS

I ran the STACKS pipeline on a total of 157 individuals (totaling 300,342 reads). My samples had an average alignment rate to the reference genome of 95.2% and an average depth of coverage of 35.7X. I removed four samples with more than 50% missing data and one sample due to 83% relatedness to another individual, resulting in a final sample size of 152 sparrows.

Genomic Divergence in the San Francisco Bay

PCA revealed a modest signal of genetic variation among song sparrow populations, particularly the isolation of *M. m. pusillula* from other subspecies (Figure 1b). PC1 and PC2 explained 1.47% and 1.25% of the variation, respectively. While subsequent axes explained similarly low amounts of variation, PC2 and PC3 suggest a pattern of differentiation between the

Subspecies	Locality	County, State	# of Individuals	H_O	π
M. m. gouldii	Mark's Marsh, Tomales Bay, Audubon Canyon Ranch (MM)	Marin County, California	24	0.1394	0.00135
	Tennessee Valley (TV)	Marin County, California	1	0.2157	0.002
M. m. heermanni	The Nature Conservancy Cosumnes River Preserve (CO)	Sacramento County, California	25	0.1081	0.001
M. m. santaecrucis	Los Gatos Creek County Park (LG)	Santa Clara County, California	18	0.1084	0.0011
	San Pedro Valley County Park (SP)	San Mateo County, California	7	0.1201	0.0013
M. m. maxillaris	Goodyear Slough Unit, Grizzly Island Wildlife Area (GS)	Solano County, California	14	0.1097	0.0011
	Rush Ranch Open Space, Solano County Farmlands and Open Space Foundation (RR)	Solano County, California	15	0.1102	0.0012
M. m. pusillula	Dumbarton Marsh, Don Edwards San Francisco Bay National Wildlife Refuge (DM)	Alameda County, California	14	0.1030	0.0011
	Palo Alto Baylands Nature Preserve (PB)	San Mateo County, California	15	0.1030	0.0011
M. m. samuelis	Petaluma River Mouth, California Department Fish and Game (PM)	Sonoma County, California	6	0.1123	0.0012
	Sonoma Creek, San Pablo Bay National Wildlife Refuge (SC)	Solano County, California	19	0.1071	0.0011
	China Camp State Park (CC)	Marin Co, California	2	0.1443	0.0015

Table 1. Sampling information for six subspecies of song sparrow (*Melospiza melodia)*. For each subspecies, I report sampling location, sample size, observed heterozygosity (H_O), and nucleotide diversity (π).

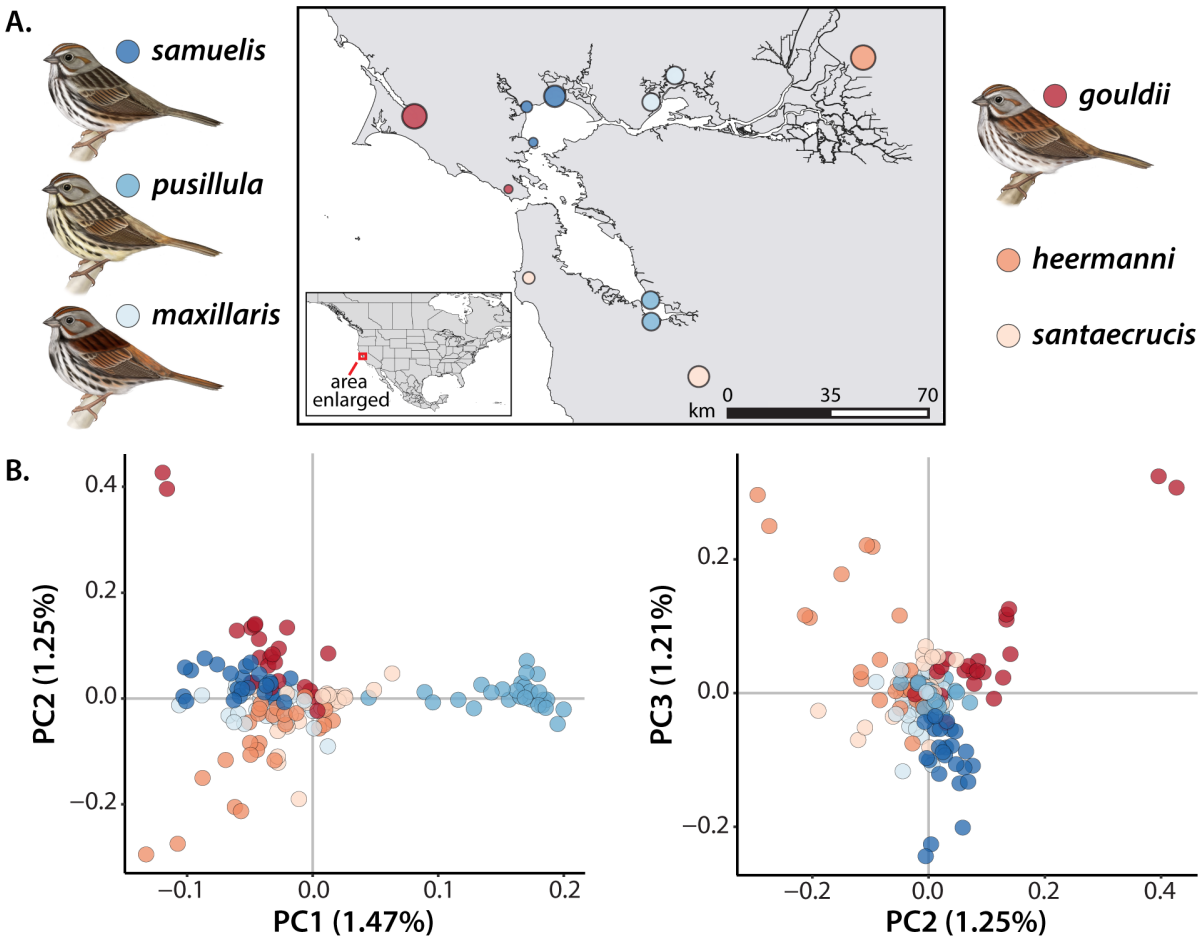


Figure 1. Map of the study site with illustrations of song sparrow subspecies to demonstrate morphological variation (A) and principle component analyses (PCA; B). The map of the San Francisco Bay shows the 12 sampling sites and the respective subspecies sampled (size of points indicates the number of individuals sampled at each site). *M. m. pusillula, maxillaris,* and *samuelis* are the three subspecies endemic to the salt marsh, and *gouldii, heermanni,* and *santaecrucis* are the upland subspecies. PC1 demonstrates strong clustering of *pusillula,* PC2 explains more of the divergence in the upland subspecies, *gouldii* and *heermanni,* and PC3 explains variation between saltmarsh and upland populations.

	gouldii	*heermanni*	*santaecrucis*	*maxillaris*	*pusillula*
gouldii					
heermanni	0.0178				
santaecrucis	0.0168	0.0156			
maxillaris	0.0157	0.0136	0.0137		
pusillula	0.0235	0.0224	0.0197	0.0190	
samuelis	0.0181	0.0169	0.0166	0.0138	0.0232

Table 2. Pairwise F_{ST} estimates for the six subspecies of song sparrow in San Francisco Bay. The average F_{ST} across all subspecies is 0.018. *M. m. pusillula* accounts for the most variation, and *M. m. santaecrucis* the least.

upland subspecies, *gouldii* and *heermanni* from those restricted to the Bay (Figure 1b; Figure S2). AMOVA results revealed that variation among individuals (F_{IS}; P=0.082, 1.877%) and among subspecies (F_{ST}; P=0.001, 1.984%) was lower than expected, but that variation within individuals (F_{IT}; P=0.003, 96.139%) was greater than expected. Only p-values from F_{ST} and F_{IT} are statistically significant. Average observed average heterozygosity was similar across all of the subspecies with the exception of *pusillula*, which was slightly lower (Table 1). Pairwise F_{ST} estimates supported the results of the PCA (Table 2). The global F_{ST} is 0.018 ($0.0136<\mu<0.0235$), which represents a moderate amount of differentiation given the small spatial scale. The highest differentiation occurs between *pusillula* and all other subspecies ($0.0196<\mu<0.0235$) and the lowest differentiation is between *santaecrucis* and all other subspecies ($0.0136<\mu<0.0196$). Pairwise F_{ST} estimates further suggest increased differentiation among the three marsh subspecies (*maxillaris, pusillula, samuelis*) than within their upland counterparts. No pairwise comparisons were statistically significant.

Results from STRUCTURE were consistent with my PCA, with *pusillula* separated from the other populations most strongly. With all six subspecies included, the ΔK method in STRUCTURE identified K=2 to be the optimal number of clusters, reflecting a clear split of *pusillula* from other subspecies (Figure 2). I subsequently performed a STRUCTURE run without *pusillula* to see if there was additional structuring being obscured. I estimated a new allele frequency prior ($\lambda=0.31$) for the five subspecies included in this latter run, and the ΔK method supported K=3 as the optimal number of clusters (Figure 2). Based on Q values (an individual is assigned to a cluster based on majority percent cluster assignment), hierarchical STRUCTURE results identified one defined cluster comprised of *samuelis* individuals (80% of *samuelis* individuals were assigned to this cluster). STRUCTURE also identified a second, moderately less

defined cluster, that corresponded to *gouldii* individuals (54% of *gouldii* individuals were assigned to this cluster). The third cluster was comprised of individuals from the three remaining subspecies (*maxillaris, heermanni,* and *santaecrucis*; Figure 2), which appear to be undifferentiated from each other.

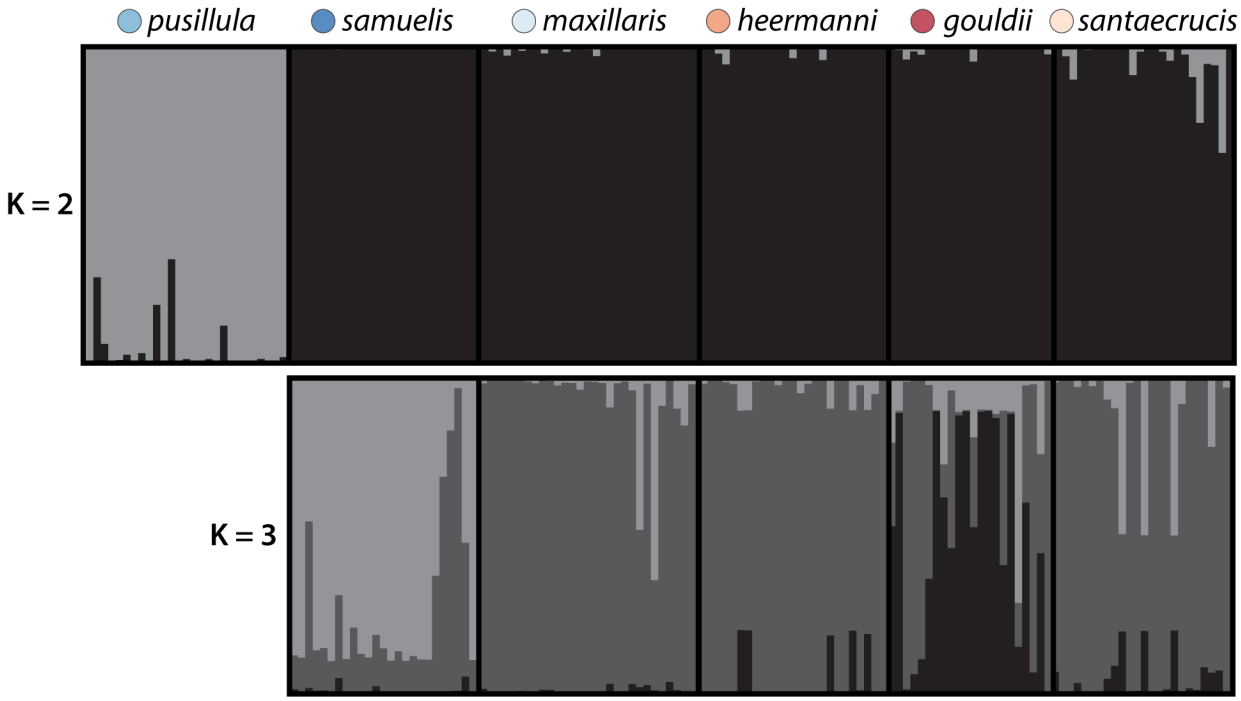


Figure 2. Results from the STRUCTURE analysis. For all subspecies, K=2 is most optimal, separating *pusillula* from the rest of the subspecies. A hierarchical run excluding *pusillula* suggests K=3 as the best, separating the marsh endemic *samuelis* and the upland *gouldii.*

Spatial Patterns and the Influence of Geographic & Environmental Isolation

Cross-validation analysis of the CONSTRUCT results revealed no significant difference between the spatial and non-spatial models, indicating that isolation by distance is not responsible for the differentiation being observed (Figure S3). I found no significant correlation between genetic and geographic distance among the song sparrow populations in the bay region (P=0.995, Figure S4).

Effective migration rates, estimated using EEMS, are low into *maxillaris* and *pusillula* populations, and nucleotide diversity rates are notably low within *samuelis* and *pusillula*. The effective migration rates are visualized and plotted on a $\log_{10}$ scale as a contour map (Figure 3) with colors representing estimated (a) average migration rates and (b) average nucleotide diversity surfaces ranging from low (dark red) to high (dark blue; Figure 3). There are apparently no strict barriers to migration, but notably lower migration rates and genetic diversity for each of the three populations restricted to single arms of the Bay.

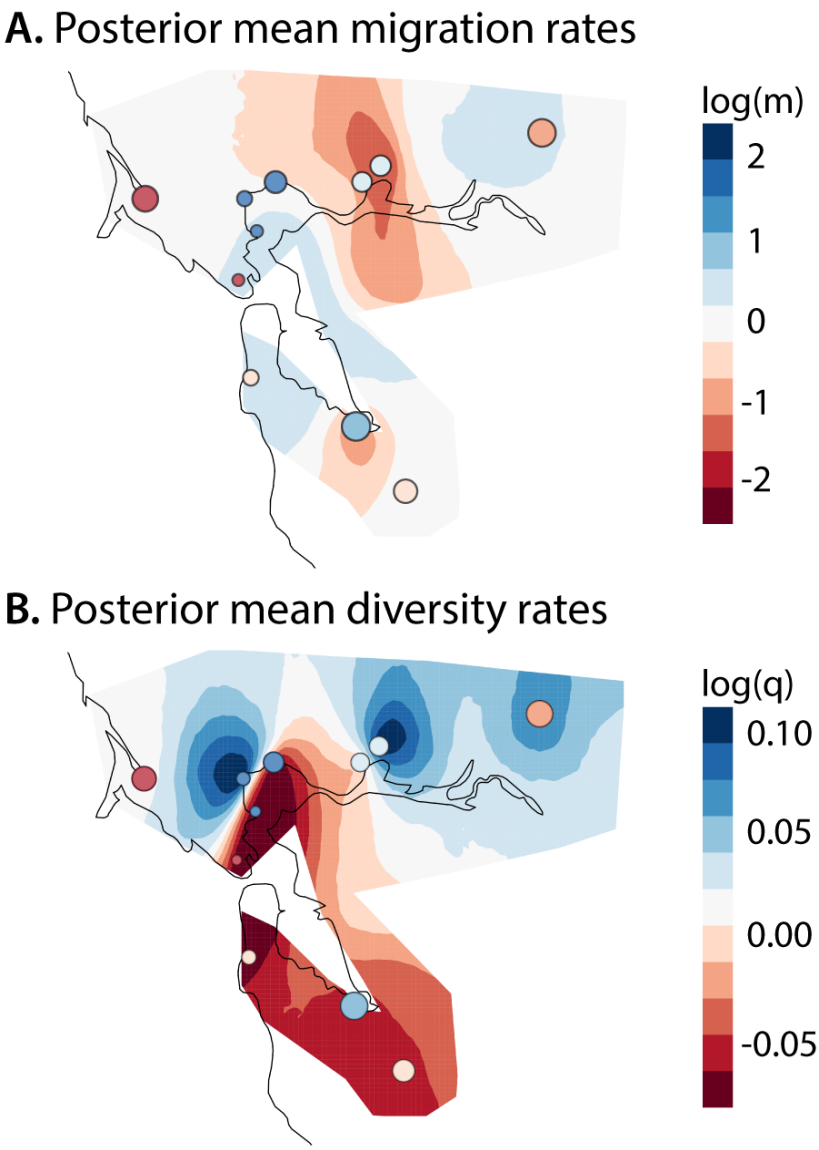


Figure 3. Estimates of effective migration and mean nucleotide diversity in San Francisco Bay song sparrow populations. Colored circles represent sampling locations with colors indicating subspecies as shown in Figure 1. Estimates of both migration rates (A) and average nucleotide diversity surfaces (B) are on a log scale, with warmer colors representing low values and cooler colors representing high values. Effective migration rates are low in marsh subspecies *maxillaris* and *pusillula*, and average nucleotide diversity is notably low in the marsh subspecies *samuelis* and *pusillula*.

I did not observe clear evidence of genetic drift through comparisons of pairwise F_{ST} between each saltmarsh population and *gouldii* (as a representative upland group) and measures of genetic diversity (Figure S5; H_O: R^2=-0.118, P=0.572; H_E: R^2=0.266, P=0.135; π: R^2=-0.145, P=0.645). However, I did observe a notable positive trend in a partial-mantel test correlating pairwise F_{ST} and salinity (while controlling for distance; P=0.07; Figure S6). There was no significant effect of temperature on pairwise F_{ST} (P=0.5).

Identifying Putative Outliers

Based on two approaches for outlier detection, I detected several candidate SNPs that are putatively under selection. BAYESCAN identified two outliers from the full dataset (0.019% of all SNPs; Figure S7). The two identified outlier SNPs had F_{ST} values of 0.11193 and 0.11335. Using a F_{ST} cut-off approach, I identified a number of SNPs that exhibited elevated F_{ST} estimates (functionally defined here as 5 standard deviations above the mean), ranging from 19 (*gouldii-pusillula*) to 37 (*maxillaris-pusillula*) putative outliers across the fifteen pairwise comparisons. Of these putative outliers, 103 elevated sites were shared across more than one pairwise comparison. PCA using only these 103 shared SNPs resulted in a similar pattern as the unlinked dataset, with a slightly more defined separation between salt marsh and upland populations (Figure S8). Several of these putative outlier regions aligned to annotated regions of the Zebra Finch genome (Table S1).

Demographic History

**This section was contributed by co-author to the corresponding manuscript, Phred Benham.*

An isolation with continuous migration (IM) model was the best fitting model to the joint site frequency spectrum of *pusillula* and all other subspecies (Figure. 4; log-likelihood: -608.19). An IM model with exponential growth in the other subspecies population (LL: -607.6) and an IM

model with a bottleneck within the *pusillula* population (LL: -607.93) both showed slightly greater likelihoods relative to the IM constant size model. However, likelihood ratio tests comparing the IM constant to the growth model (adjusted D: 0.0139; P=0.453) and IM constant versus IM bottleneck (adjusted D: 0.535; P=0.232) both show that increasing model complexity does not result in a significantly better fit to the site frequency spectrum. Under the IM constant population size model we inferred that *pusillula* diverged from the other subspecies ~61 kya (95% CI: 1-321,258 ya). Following the split *pusillula* maintained an effective population size of ~26,573 (95% CI: 6,888-54,117) and the other four subspecies an N_e of 256,940 (95% CI: 172,250-420,972). Migration was found to be slightly greater from other subspecies into *pusillula* (1.37×10^{-4}) than in the opposite direction (2.72×10^{-5}). Uncertainties surrounding divergence time between *pusillula* and the other subspecies were large with standard deviations inferred using the GIM method exceeding the parameter estimate. Estimating uncertainties using the nonparametric bootstrap approach still indicated broad 95% confidence intervals from 1,806 - 358,553 ya (Table S2). This wide distribution reflects a bimodal distribution of divergence times with most divergence time estimates less than 100 kya, but a second cluster showing divergence times >250 kya (Figure S9). The three-population model (log-likelihood: -1659.93; Figure S10) showed a much deeper divergence between *pusillula* and the other subspecies at ~527,190 generations ago (95% CI: 423,881-731,576) followed by divergence between *gouldii* and all other subspecies more recently, ~2,156 generations ago (95% CI: 1-95,196). Effective population size estimates were larger for *pusillula* (50,984; 95% CI: 33,456-70,809), but there was overlap in confidence intervals between two-populations and three-population analyses. *Gouldii* exhibited the largest effective population size (233,035; 95% CI: 186,496-284,377), while the remaining three subspecies combined had an estimated N_e comparable to *pusillula*

(53,102; 95% CI: 9,610-103,264). Migration rate was also found to be highest between *gouldii* and *pusillula* (3.01×10^{-5}; 95% CI: $0\text{-}2.0\times10^{-4}$), though confidence intervals did encompass zero, suggesting the possibility of no gene flow between these subspecies.

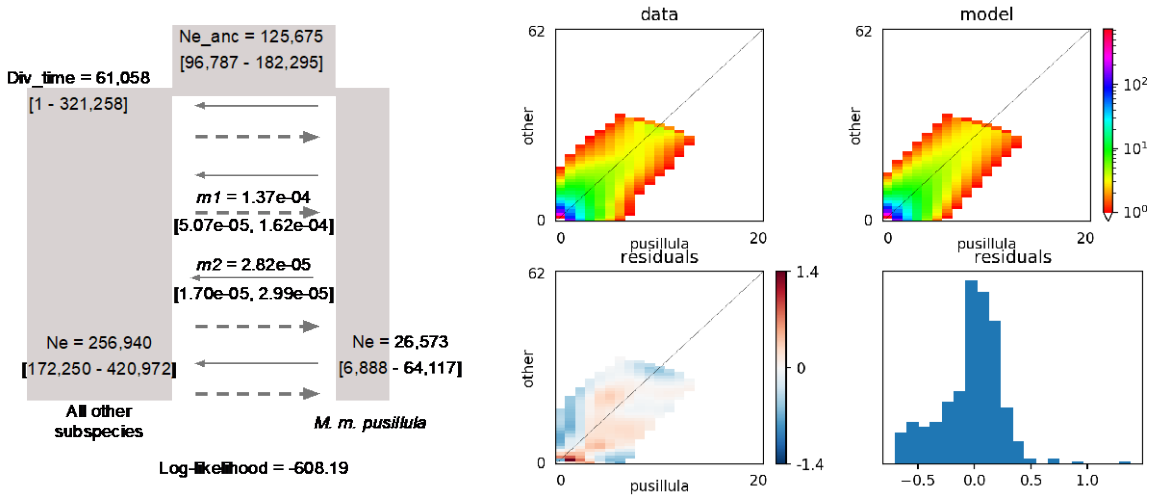


Figure 4. *This figure and associated analyses were contributed by Phred Benham. (Left-hand panel) Schematic of best-supported model, where *M. m. pusillula* splits from all other subspecies ~61 kya. Following divergence continuous gene flow continued between both populations with slightly greater migration into *M. m. pusillula.* The four panels on the right side of the plot show how well the best-fit model fits the observed site frequency spectrum. The top-left panel shows the observed SFS. The top-right panel shows the simulated SFS based on the best-fit model and parameters. The bottom panels show the residuals from fitting observed data to model data.

DISCUSSION

My study describes patterns of genetic structure in San Francisco Bay song sparrows to elucidate forces capable of driving species diversification in the face of ongoing gene flow. Within a 70 x 100 km region, I identified signals of differentiation among subspecies, specifically, observable differentiation of *pusillula* and *samuelis* from upland *gouldii* populations. The remaining subspecies in the Bay (*heermanni, santaecrucis, maxillaris*) appear largely undifferentiated from each other. Despite hypotheses that *santaecrucis* is a hybrid

between *gouldii* and *heermanni,* structure results suggest possible admixture from *samuelis* as well–although future studies that combine increased sampling efforts along a transect would be more appropriate to address this in full. While the observed differentiation in all cases was subtle, the patterns observed here are notable given the fine spatial scale over which this study was conducted. Moreover, signals of genomic differentiation between tidal marsh and upland clades coupled with evidence of isolation by environment support earlier suggestions that natural selection has played at least a partial role in driving this differentiation (e.g., Miller, 1956; Johnston, 1956b; Basham & Mewaldt, 1987; Chan & Arcese, 2003). I estimated that the present-day diversification of song sparrows in the bay region occurred recently (within the last 500,000 generations). While the split between *pusillula* and other subspecies pre-dates marsh formation, estimates for the other groups are more consistent with the recent formation of salt marsh habitats within the bay region. However, the divergence times estimated here were not well-resolved (see below). Overall, my results offer new insight into the roles of natural selection and demography in shaping divergence of song sparrow populations in this region.

Several mechanisms are proposed to explain the diversification of song sparrows in the bay region. Aldrich (1984) suggested that ecological gradients including salinity, climate, seasonality, and background coloration have all contributed to microevolutionary change in song sparrow across their range, but found little support for historical effects of geographic isolation or genetic drift on phenotype. Patten and Pruett (2009) described song sparrows as a ring species due to the pattern of subspecific morphological variation over their geographic range. My results provide further insights into the origins of rapid differentiation with the San Francisco Bay region, specifically with respect to the roles of three potential drivers of divergence: isolation, drift, and natural selection. First, my analyses using CONSTRUCT failed to explain observed

patterns better than non-spatial models, offering no evidence of isolation by distance among these sparrow populations. Although geographic isolation has been proposed as a mechanism shaping diversification in song sparrows within and outside the bay region (e.g., Miller 1947, Ferrell 1966, Pruett et al. 2009), the spatial scale over which isolation by distance occurs should be proportional to the dispersal distance of an organism and the extent of gene flow between populations. Thus, this scenario seems less likely given that I expect the dispersal capabilities of these fully flight-capable sparrows to far exceed the spatial scale over which I found genetic divergence.

Evidence for a role of drift in promoting population divergence among subspecies would include low genetic diversity, small effective population sizes, and/or a history of population bottlenecks. EEMS analyses show lower nucleotide diversity within tidal marsh populations of song sparrow (Figure 3). While analysis of the site frequency spectrum did not support a history of bottlenecks in these populations, previous analyses using microsatellite loci from many of the same individuals (Chan & Arcese, 2002) and whole genome sequencing (WGS) data from *M. m. pusillula* and *M. m. gouldii* (Walsh et al., 2019) reveal contrasting results. Specifically, WGS analysis found evidence for recent bottlenecks (<250 years ago) in both *M. m. pusillula* and *M. m. gouldii,* and much smaller current N_e: 324 and 179, respectively. It is likely that there is more power in the larger WGS dataset to detect fluctuations in N_e (e.g. Terhorst & Song, 2015), as well as to detect more recent fluctuations in population size compared to the RAD data. Additionally, the difference in population sampling between our studies could contribute to different results. Gene flow with unsampled populations can lead to erroneous inferences of population bottlenecks (e.g. Nielsen & Beaumont, 2009) as well as bias estimation of effective population size (e.g. Lynch & Sethuraman 2019). While further analyses will be required to

resolve these differences in demographic history among different datasets, the inferred N_e and patterns of genetic diversity across all datasets exceed those observed in other species where drift appears to have played a significant role in population differentiation. For example, Funk et al. (2016) found that in island fox (*Urocyon littoralis*) populations, N_e was inferred to be as low as 2.1 (range: 2.1-89.7) and heterozygosity as low as 0.016 (range: 0.016-0.231). Despite the potentially strong influence of drift for those fox populations, the authors still detected signatures of selection in comparison of their genomes across island populations. In sum, my results and prior work on song sparrows from the San Francisco bay region provide mixed support for drift driving population differentiation. In concert, while it is probable that drift contributed to the observed patterns to an extent, it is highly unlikely that drift alone shaped patterns and levels of population differentiation among these subspecies.

I found support for selection in driving diversification of sparrow populations around the Bay, possibly related to adaptation to saltwater versus freshwater habitats. The hypothesized role of selection in shaping patterns of differentiation is not unreasonable given the steep ecological gradients and adaptive challenges associated with freshwater-saline ecotones (Greenberg & Maldonado, 2006; Walsh et al., 2019). These patterns are notably different, however, from comparisons of tidal marsh Savannah Sparrows in San Francisco Bay, which exhibited low levels of genetic differentiation, even across ecological gradients (Benham & Cheviron, 2019). Pairwise comparisons resulted in the identification of 103 shared SNPs that exhibited elevated signals of differentiation (F_{ST} greater than 5 SD above the mean). Some of these differentiated regions (Table S1) corresponded to genes with putative functions in osmoregulatory processes, including genes associated with ion homeostasis (Di Ciano-Oliveira et al., 2006), regulation of MAPK cascades (Vom Dahl et al., 2001; Kultz et al., 2001), and microtubule cytoskeleton

organization (Di Ciano-Oliveira et al., 2006). Of particular interest are several genes with putative links to cytoskeleton organization (MYH13, MYO18, MYO3, TTBK2, PAK3; Table S3) that are associated with signals of elevated divergence between upland and salt marsh populations of San Francisco Bay song sparrows. Previous work has shown that the cytoskeleton, which is responsible for maintaining cell morphology, is a likely candidate for regulating cell volume response in the face of osmotic changes (Di Ciano-Oliveira et al., 2006; Bober et al., 2015). Cytoskeletal reorganization may allow cells to resist volume changes through the reinforcement of cell structure (Di Ciano-Oliveira et al., 2006), a potentially important adaptation to saltwater environments. Putative selection for genes linked to cytoskeletal organization have been identified in other freshwater-saline comparisons, including in savannah sparrows (Walsh et al., 2019) and saltmarsh sparrows (Walsh et al., 2018). Savannah sparrows (Goldstein et al. 1990; Benham & Cheviron, 2020) and Chilean seaside cinclodes (Sabat et al., 2004) both show elevated plasma osmolality in high salinity environments, suggesting that genes related to cytoskeletal reorganization may be important for increased salinity tolerance in these populations. These findings paired with whole-genome comparisons of *M. m. gouldii* and *M. m. pusillula*, which identified similar candidate genes with putative links to tidal marsh adaptations (Walsh et al., 2019) support a role for microgeographic adaptation in shaping patterns of divergence in this region.

Independent of their mechanism of origin, these newly characterized fine scale patterns of diversification among song sparrow populations in the San Francisco Bay region have tangible conservation and management implications. The loss of marsh habitat threatens the breeding and nesting grounds for all three Bay-endemic subspecies (*maxillaris, pusillula,* and *samuelis*), which are the only extant populations of song sparrow adapted to life in salt marshes.

It has been estimated that the current marsh habitat occupies only 15% of its pre-industrial acreage (Marshall & Dedrick, 1994), and projected sea level rise poses serious threats to the persistence of these intertidal ecosystems. Based on current population declines, updated information on local adaptive capacity and potential management units within the Bay has significant conservation relevance. My research provides evidence for genetic structuring on a fine spatial scale, with a putative role of local adaptation in driving population diversification. A major challenge facing conservation and evolutionary biologists is the identification and preservation of biological diversity and ecological viability in response to anthropogenic change. Characterizing and maximizing genetic variation should be a key goal in any management endeavor, as increased genetic diversity can enhance a population's resilience in the face of changing environmental conditions (Reed & Frankham, 2003). Based on our findings, I recommend that local management initiatives consider the potential for locally adapted populations within the bay region, with a particular focus on the *pusillula* subspecies given the clear differentiation of this group from the remainder of the populations in the region. More generally, these factors may be particularly important if considering any future translocation or captive breeding efforts.

Conclusions

My results offer evidence of local adaptation at a microgeographic scale in subspecies of song sparrows residing in the marshes and uplands surrounding San Francisco Bay, including signals of differentiation in two of the three salt marsh-dwelling forms. Although correlative with regard to mechanism, my results offer support for the hypothesis that context-dependent natural selection on individual performance has played a role in the diversification of song sparrows in this region; and specifically, that local adaptation to salt marsh environments and genetic drift

have each contributed to the variation in population structure we describe. Although differentiation across subspecies was modest overall, a putative role for natural selection in shaping these patterns was suggested by whole-genome analyses of two of the subspecies studied here (Walsh et al., 2019). I propose that whole-genome comparisons of song sparrows targeting loci or genomic pathways under selection, and thus putatively linked to underlying local adaptation, are now warranted. Incorporating such information into scenario planning and management decisions in the San Francisco Bay region offers further opportunities to maximize the persistence of this and similarly variable species by developing plans most likely to conserve the adaptive scope of organisms at their appropriate geographic scales.

ACKNOWLEDGEMENTS

I first thank my mentors, Jennifer Walsh and Irby Lovette, for their guidance and support throughout my time at Cornell. In addition to Walsh and Lovette, I thank my other co-authors on the corresponding manuscript to this thesis: Stepfanie Aguillon, Yvonne Chan, Peter Arcese, and Phred Benham. I also thank the members of the Fuller Evolutionary Biology Lab, specifically D. Hooper, N. Hofmeister, J. Berv, and B. Butcher, for feedback and comments on an earlier draft of this manuscript. J. Ditner for illustrating the song sparrows based on specimens provided by the Cornell University Museum of Vertebrates and the California Academy of Sciences, and S. Billerman for feedback on the phenotypic variation portrayed by these illustrations. G. Camenisch and L. Keller provided an updated, annotated reference genome that was used for RAD sequence alignment. Funding for this work was provided by the Rawlings Cornell Presidential Research Scholarship Program and the Cornell Lab of Ornithology.

REFERENCES

Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, G. Coop, and N. Chen. 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLoS Genet. 13:e1006911.

Aitken, S. and M. Whitlock. 2013. Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annual Review of Ecology, Evolution, and Systematics 44:367-388.

Aldrich, J. W. 1984. Geographical variation in size and proportions of Song Sparrows (*Melospiza melodia*). Ornithological Monographs 35:1-134.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Antonovics, J. (2006). Evolution in closely adjacent plant populations X: long-term persistence of pre-reproductive isolation at a mine boundary. *Heredity*, *97*(1), 33–37.

Arcese, P., Sogge, M.K., Marr, A.B., Patten, M.A. 2002. Song Sparrow (*Melospiza melodia*), The Birds of North America Online (A. Poole, Ed.). Ithaca: Cornell Lab of Ornithology; Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/704.

Atwater, B.F. 1979. Ancient processes at the site of Southern San Francisco Bay: movement of the crust and changes in sea level. In: *San Francisco Bay: The* Urbanized Estuary (T. J. Conomos, ed.), pp. 31–45. Pacific Division/American Association for the Advancement of Science, San Francisco.

Basham, M. P., Mewaldt, L. R., & Marin, B. T. (1987). Salt water tolerance and the distribution of South San Francisco Bay song sparrows. *The Condor*, *89*(4), 697–709.

Benham, P. M., & Cheviron, Z. A. (2019). Divergent mitochondrial lineages arose within a large, panmictic population of the Savannah sparrow (Passerculus sandwichensis). *Molecular ecology*, *28*(7), 1765-1783.

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*, *74*(1), 57–72. https://doi.org/10.1111/evo.13886

Bober, B. G., Love, J. M., Horton, S. M., Sitnova, M., Shahamatdar, S., Kannan, A., & Shah, S. B. (2015). Actin–myosin network influences morphological response of neuronal cells to altered osmolarity. *Cytoskeleton*, *72*(4), 193-206.

Bradburd, G.S. (2019). conStruct: Models Spatially Continuous and Discrete Population Genetic Structure. R package version 1.0.3. https://CRAN.R-project.org/package=conStruct

Bradburd, G. S., Coop, G. M., & Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space. *Genetics*, *210*(1), 33-52.

Camacho, C., Canal, D., & Potti, J. (2016). Natal habitat imprinting counteracts the diversifying effects of phenotype-dependent dispersal in a spatially structured population. *BMC evolutionary biology*, *16*(1), 158.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, *1*(3), 171-182.

Chan, Y., and P. Arcese. 2002. Subspecific differentiation and conservation of song sparrows (*Melospiza melodia*) in the San Francisco Bay region inferred by microsatellite loci analysis. The Auk 119:641-657.

Chan, Y., and P. Arcese. 2003. Morphological and microsatellite differentiation in *Melospiza melodia* (Aves) at a microgeographic scale. Journal of Evolutionary Biology 16:939-947.

Charmantier, A., Doutrelant, C., Dubuc-Messier, G., Fargevieille, A., & Szulkin, M. (2016). Mediterranean blue tits as a case study of local adaptation. *Evolutionary Applications*, *9*(1), 135-152.

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2015). Computationally efficient composite likelihood statistics for demographic inference. *Molecular biology and evolution*, *33*(2), 591-593.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Ciano-Oliveira, C. D., Sirokmány, G., Szászi, K., Arthur, W. T., Masszi, A., Peterson, M., ... & Kapus, A. (2003). Hyperosmotic stress activates Rho: differential involvement in Rho kinase-dependent MLC phosphorylation and NKCC activation. *American Journal of Physiology-Cell Physiology*, *285*(3), C555-C566.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Edelaar, P., A. M. Siepielski, and J. Clobert. 2008. Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. *Evolution***62**:2462–2472.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, *14*(8), 2611-2620.

Ferrell, G.T. 1966. Variation in blood group frequencies in populations of song sparrows of the San Francisco Bay Region. *Evolution* **20**: 369–382.

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977-993.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., ... & Polato, N. R. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (Urocyon littoralis). *Molecular Ecology*, *25*(10), 2176-2194.

Fry, A.J. & Zink, R.M. 1998. Geographic analysis of nucleotide diversity and song sparrow (Aves: Emberizidae) population history. *Mol. Ecol.* **7**: 1303–1313.

Garnett, S. T., Sayer, J., & Du Toit, J. (2007). Improving the effectiveness of interventions to balance conservation and development: a conceptual framework. *Ecology and society*, *12*(1).

Goldstein, D. L. (2006). Osmoregulatory biology of saltmarsh passerines. *Studies in Avian biology*, *32*, 110.

Greenberg, R. S., & Maldonado, J. E. (2006). Diversity and endemism in tidal marsh vertebrates. *Vertebrates of Tidal Marshes*.

Greenberg, R., & Olsen, B. (2010). Bill size and dimorphism in tidal-marsh sparrows: island-like processes in a continental habitat. *Ecology*, *91*(8), 2428-2436.

Grenier, J. L., & Greenberg, R. U. S. S. E. L. L. (2006). Trophic adaptations in sparrows and other vertebrates of tidal marshes. *Studies in Avian Biology*, *32*, 130.

Grinnell, J. (1901). The Santa Cruz Song Sparrow, with Notes on the Salt Marsh Song Sparrow. *The Condor*, *3*(4), 92–93. https://doi.org/10.2307/1361163

Grinnell, J. 1913. The species of the mammalian genus Sorex of west-central California. Univ. Calif. Publ. Zool. 10:179–195.
Grinnell, J., and A. H. Miller. 1944. The distribution of the birds of California. Artemesia Press, Lee Vining, California.
Hendry, A. P., Day, T., & Taylor, E. B. (2001). Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution*, *55*(3), 459-466.
Hendry, A. P., Taylor, E. B., & McPhail, J. D. (2002). Adaptive divergence and the balance between selection and gene flow: Lake and stream stickleback in the Misty system. *Evolution*, *56*(6), 1199–1216.
Ho, W., and J. Zhang. 2018. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. Nature Communications 9:1–11.
Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**:579–588.
Huxley, J. 1942. *Evolution, the Modern Synthesis*. Harper and Brothers, New York.
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*(14), 1801-1806.
Johnston, R.F. 1956a. Population structure in salt marsh song sparrows. Part I. Environment and annual cycle. *Condor* **58**: 24–44.
Johnston, R.F. 1956b. Population structure in salt marsh Song Sparrows. Part II: Density, age structure, and maintenance. *Condor* **58**: 254–271.
Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. doi:10.1093/bioinformatics/btr521
Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*, *206*(3), 1549-1567.
Kamvar Z.N., Brooks J.C. and Grünwald N.J. 2015. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front. Genet. 6:208.
Kavanagh, K. D., T. O. Haugen, F. Gregersen, J. Jernvall, & L. A. Vøllestad. 2010. Contemporary temperature-driven divergence in a Nordic freshwater fish under conditions commonly thought to hinder adaptation. *BMC Evol. Biol.*, **10**:350.
Krueger-Hadfield, S. A., Roze, D., Mauger, S., & Valero, M. (2013). Intergametophytic selfing and microgeographic genetic structure shape populations of the intertidal red seaweed C hondrus crispus. *Molecular ecology*, *22*(12), 3242-3260.
Kültz, D., & Avila, K. (2001). Mitogen-activated protein kinases are in vivo transducers of osmosensory signals in fish gill cells. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *129*(4), 821-829.
Langin, K. M., Sillett, T. S., Funk, W. C., Morrison, S. A., Desrosiers, M. A., & Ghalambor, C. K. (2015). Islands within an island: repeated adaptive divergence in a single population. *Evolution*, *69*(3), 653-665.
Luttrell, S. A., Gonzalez, S. T., Lohr, B., & Greenberg, R. (2014). Digital photography quantifies plumage variation and salt marsh melanism among Song Sparrow (Melospiza melodia) subspecies of the San Francisco Bay. *The Auk: Ornithological Advances*, *132*(1), 277-287.

Lynch, M. D., & Sethuraman, A. (2019). The Effects of Gene Flow from Unsampled" Ghost" Populations on the Estimation of Evolutionary History under the Isolation with Migration Model. *bioRxiv*, 733600.

Marshall, J.T.J. 1948a. Ecologic races of song sparrows in the San Francisco Bay region. Part I. Habitat and abundance. *Condor* **50**: 193–215.

Marshall, J.T.J. 1948b. Ecologic races of song sparrows in the San Francisco Bay region. Part II. Geographic variation. *Condor* **50**: 233–256.

Marshall, J. T., and K. G. Dedrick. 1994. Endemic song sparrows and yellowthroats of San Francisco Bay. Studies in Avian Biology 15:316-327.

Mayr, E. 1942. *Systematics and the Origin of Species*. Harper and Brothers, New York.

Mayr, E. 1963. *Animal Species and Evolution*. Harvard University Press, Cambridge.

Mikles, C.S, Aguillon, S.M., Chan, Y.L, Arcese, P., Benham, P.M., Lovette, I.J., Walsh, J. 2019. Song Sparrow SNP data, Dryad.

Miller, A.H. 1947. Panmixia and population size with reference to birds. *Evolution* **1**: 186–190.

Miller, A.H. 1956. Ecologic factors that accelerate formation of races and species of terrestrial vertebrates. *Evolution* **10**: 262–277.

Mulligan, J.A. 1963. A description of song sparrow song based on instrumental analysis. *Proceedings XIII International Ornithological Conference*, pp. 272–284. American Ornithologists Union and Museum of Zoology, Louisiana State University, Baton Rouge, LA.

Nielsen, R., & Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Molecular ecology*, *18*(6), 1034-1047.

NOAA, National Climate Data Center. Integrated Mapping Tool Monthly Summaries for California [Online]. https://gis.ncdc.noaa.gov/maps/clim/summaries/monthly. (Accessed 24 January 2020).

Nosil, P., & Crespi, B. J. (2004). Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in TImena cristinae walking-sticks. *Evolution*, *58*(1), 102–112.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., & McGlinn, D. (2019). vegan: Community ecology package. R package version 2.5-4. R Foundation for Statistical Computing.

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular ecology*, *27*(16), 3193-3206.

Patten, M. A., and C. L. Pruett. 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(1):33-62.

Peterson, B. K., J. N Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLOSONE 7:e37135.

Petkova, D., J. Novembre, and M. Stephens. 2015. Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics 48:94-100.

Postma, E., and A. van Noordwijk. 2005. Gene flow maintains a large clutch size at a small spatial scale. Nature 433(7021):65-68.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Pruett, C. L., and K. Winker. 2005. Northwestern song sparrow populations show genetic effects of sequential colonization. Molecular Ecology 14:1421-1434.

Reed, D. H., & Frankham, R. (2003). Correlation between fitness and genetic diversity. *Conservation biology*, *17*(1), 230-237.

Reid, J.M., J. M. J. Travis, F. Daunt, S.J. Burthe, S. Wanless, C. Dytham. 2018. Population and evolutionary dynamics in spatially structured seasonally varying environments. Biological Reviews 93: 1578-1603

Richardson, J. L., & Urban, M. C. (2013). Strong selection barriers explain microgeographic adaptation in wild salamander populations. *Evolution*, *67*(6), 1729-1740.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in ecology & evolution*, *29*(3), 165-176.

Rohland, N., and D. Reich. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22(5):939-46.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular ecology notes*, *4*(1), 137-138.

Sabat, P., Maldonado, K., Rivera-Hutinel, A., & Farfan, G. (2004). Coping with salt without salt glands: osmoregulatory plasticity in three species of coastal songbirds (ovenbirds) of the genus Cinclodes (Passeriformes: Furnariidae). *Journal of Comparative Physiology B*, *174*(5), 415-420.

Sauer, J. R., J. E. Hines, J. E. Fallon, K. L. Pardieck, D. J. Ziolkowski, Jr., and W. A. Link. 2014. The North American Breeding Bird Survey, Results and Analysis 1966 - 2013. Version 01.30.2015. USGS Patuxent Wildlife Research Center, Laurel, MD.

Schraga, T.S. & Cloern, J.E. Water Quality Measurements in San Francisco Bay by the U.S. Geological Survey, 1969-2015. *Scientific Data* 4:170098. 2017.

Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, *39*(1), 53-65.

Smith, J.I. 1998. Allometric influence on phenotypic variation in the Song Sparrow (*Melospiza melodia*). *Biol. J. Linn. Soc.* **122**: 427–454.

Steiner, K. C., & Berrang, P. C. (1990). Microgeographic adaptation to temperature in pitch pine progenies. *American Midland Naturalist*, 292-300.

Stelkens, R. B., G. Jaffuel, M. Escher, and C. Wedekind. 2012. Genetic and phenotypic population divergence on a microgeographic scale in brown trout. Molecular Ecology 21(12):2896-2915.

Tattersall, G. J., B. Arnaout, and M. R. Symonds. 2017. The evolution of the avian bill as a thermoregulatory organ. *Biol. Rev.***92**:1630–1656.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics*, *49*(2), 303.

Thrasher, D. J., B. G. Butcher, L. Campagna, M. S. Webster, and I. J. Lovette. 2017. Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: a proof of concept in a highly promiscuous bird. Molecular Ecology Resources doi: 10.1111/1755-0998.12771.

Urban, M. C. 2011. The evolution of species interactions across natural landscapes. *Ecol. Lett.* **14**:723–732.

Vom Dahl, S., Dombrowski, F., Schmitt, M., Schliess, F., Pfeifer, U., & Häussinger, D. (2001). Cell hydration controls autophagosome formation in rat liver in a microtubule-dependent way downstream from p38MAPK activation. *Biochemical Journal*, *354*(Pt 1), 31.

Walsh, J., Clucas, G., MacManes, M., Thomas, K., & Kovach, A. (2018). Divergent selection and drift shape the genomes of two avian sister species spanning a saline-freshwater ecotone. *bioRxiv*, 344614.

Walsh, J., Benham, P. M., Deane-Coe, P. E., Arcese, P., Butcher, B. G., Chan, Y. L., ... & Shriver, W. G. (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution letters*, *3*(4), 324-338.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114.

Wright, S. 1969. Evolution and the Genetics of Populations, Vol. **4**. *The Theory of Gene Frequencies*. Univ. Chicago Press, Chicago, IL.

Yang, X., Xu, Y., Shah, T., Li, H., Han, Z., Li, J., & Yan, J. (2011). Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica*, *139*(8), 1045.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., ... & Ödeen, A. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, *346*(6215), 1311-1320.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir. 2012. A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. Bioinformatics.

Zink, R.M. & Dittmann, D.L. 1993. Gene flow, refugia, and evolution of geographic variation in the song sparrow (*Melospiza melodia*). *Evolution* **47**: 717–729.

SUPPLEMENTAL MATERIALS

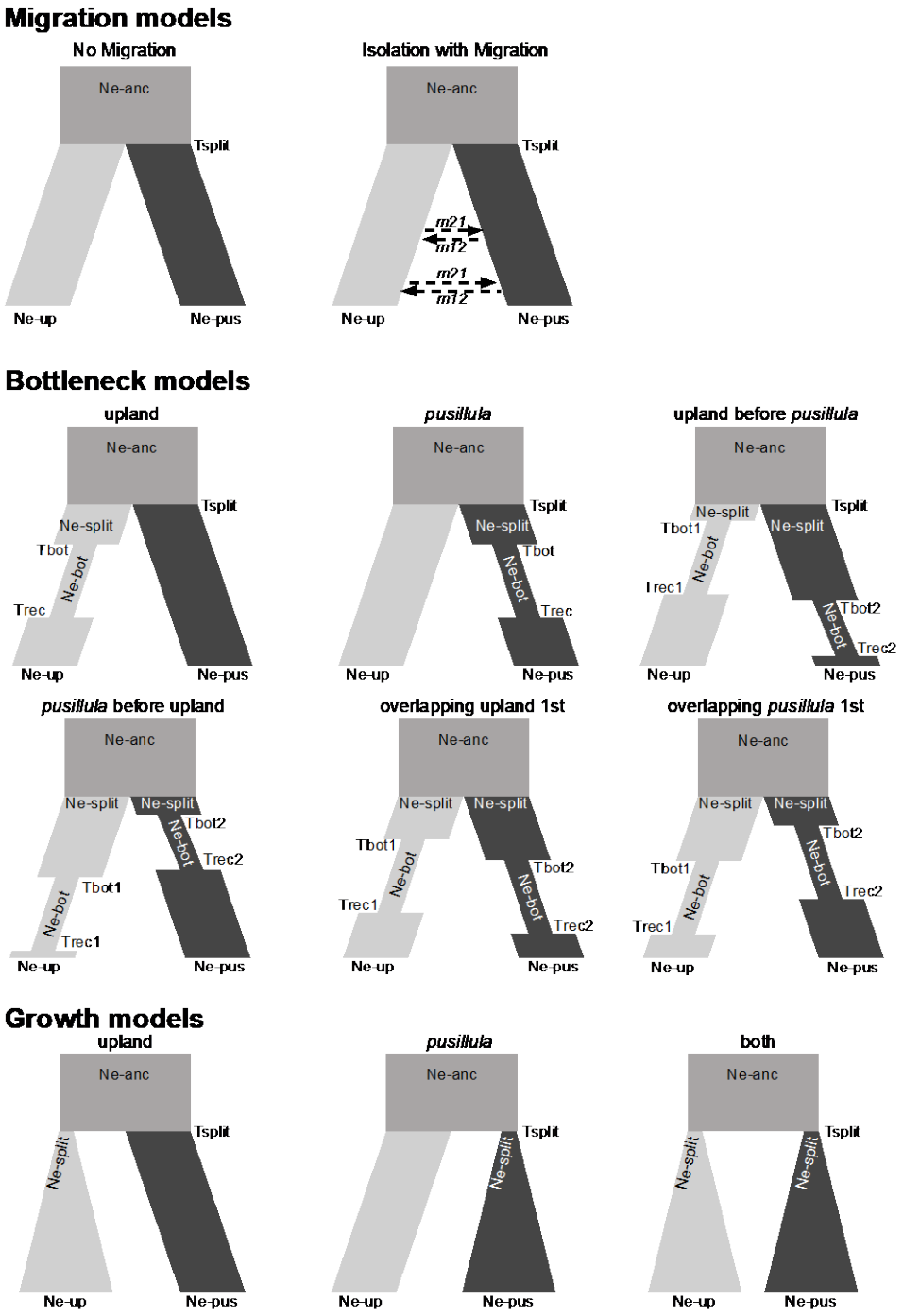


Figure S1. Demographic models fit to the joint site frequency of *M. m. pusillula* (pus) and all other subspecies (upland). We fit all models both with and without migration (*m1, m2*) Parameters estimated are labeled on each model. Population sizes (Ne) during different events are labeled for current Ne (Ne-up, Ne-pus), ancestral Ne (Ne-anc), bottleneck Ne (Ne-bot), end bottleneck Ne (Ne-rec), beginning of growth Ne (Ne-split). Labeling for time parameters: divergence time (Tsplit), bottleneck start time (Tbot), bottleneck end time (Trec).

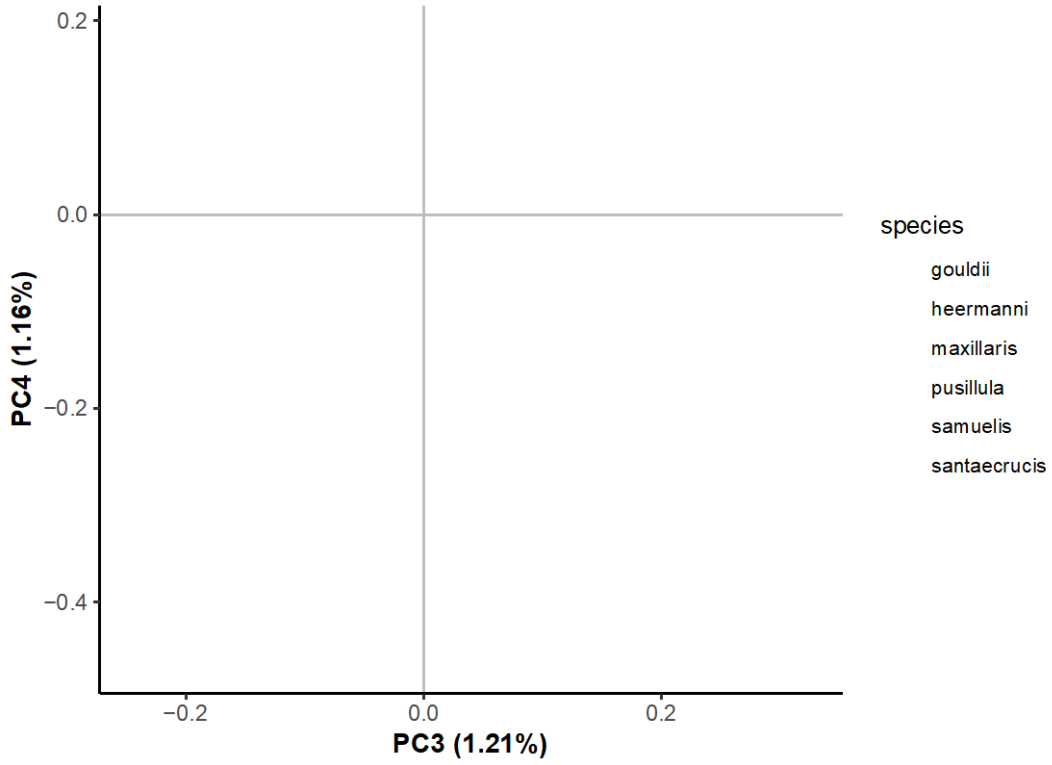


Figure S2. Principle Component Analysis (PCA) for axes PC3 vs. PC4. Six populations of song sparrow are included in the PCA and colors correspond with key below plot. PC3 suggests a pattern of differentiation between the upland subspecies, *santaecrucis* and *heermanni* from those restricted to the Bay.

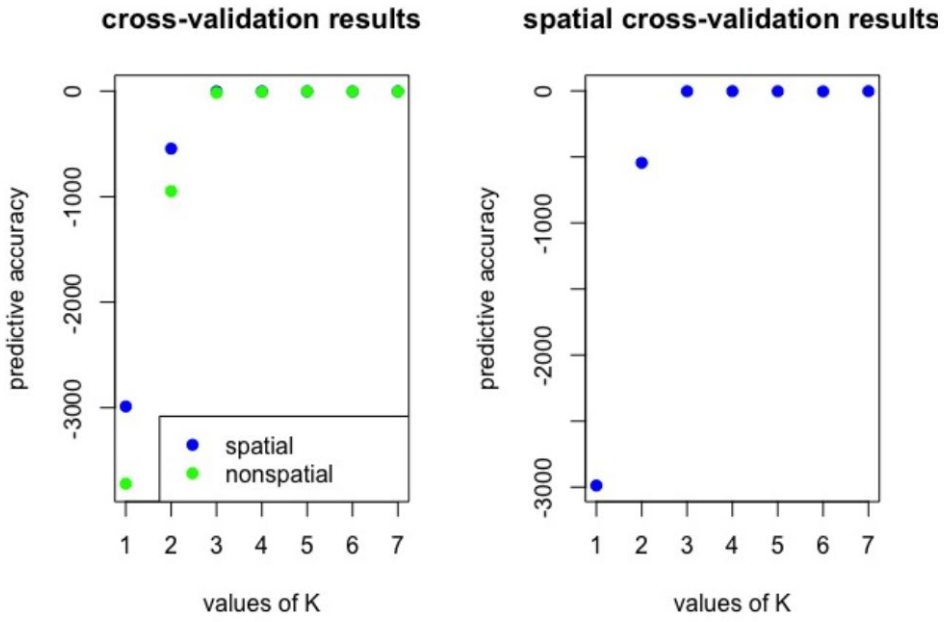


Figure S3. ConStruct cross validation results indicate that there is no significant difference between spatial and nonspatial models, suggesting that isolation by distance is not responsible for shaping patterns of population structure among song sparrow clades.

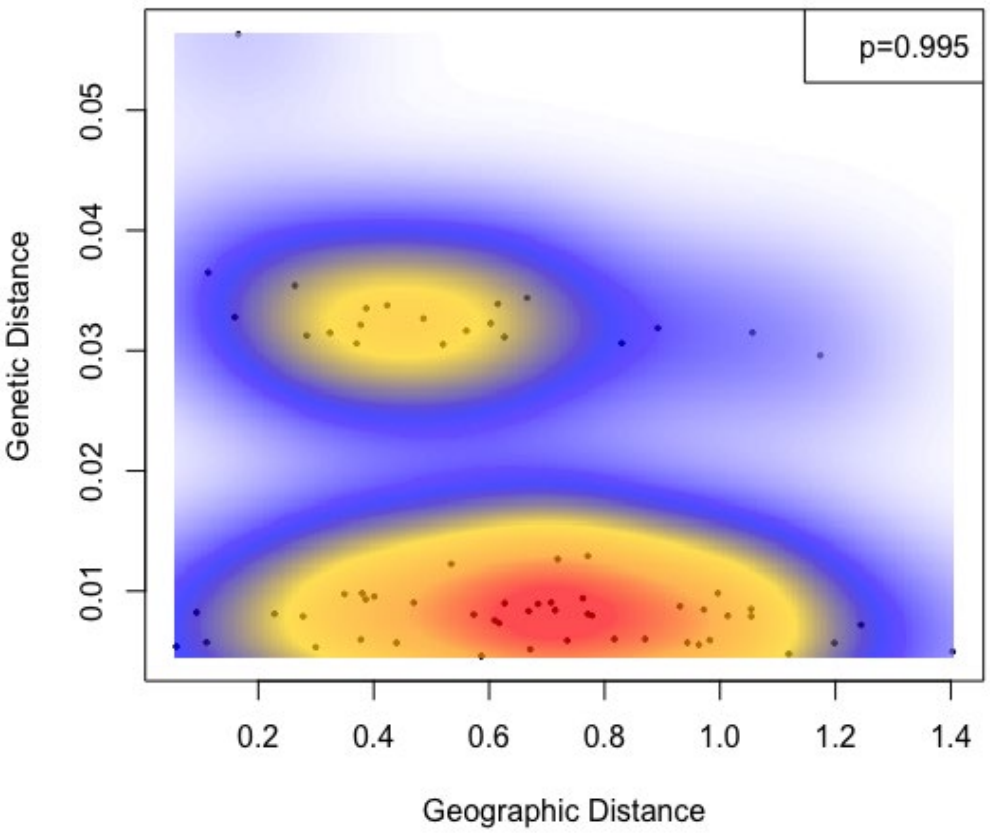


Figure S4. Test for Isolation by Distance. Plot includes a comparison of geographic distances from 12 sampling locations and corresponding genetic distance. There is no significant correlation between genetic and geographic distance among the song sparrow populations in the bay region.

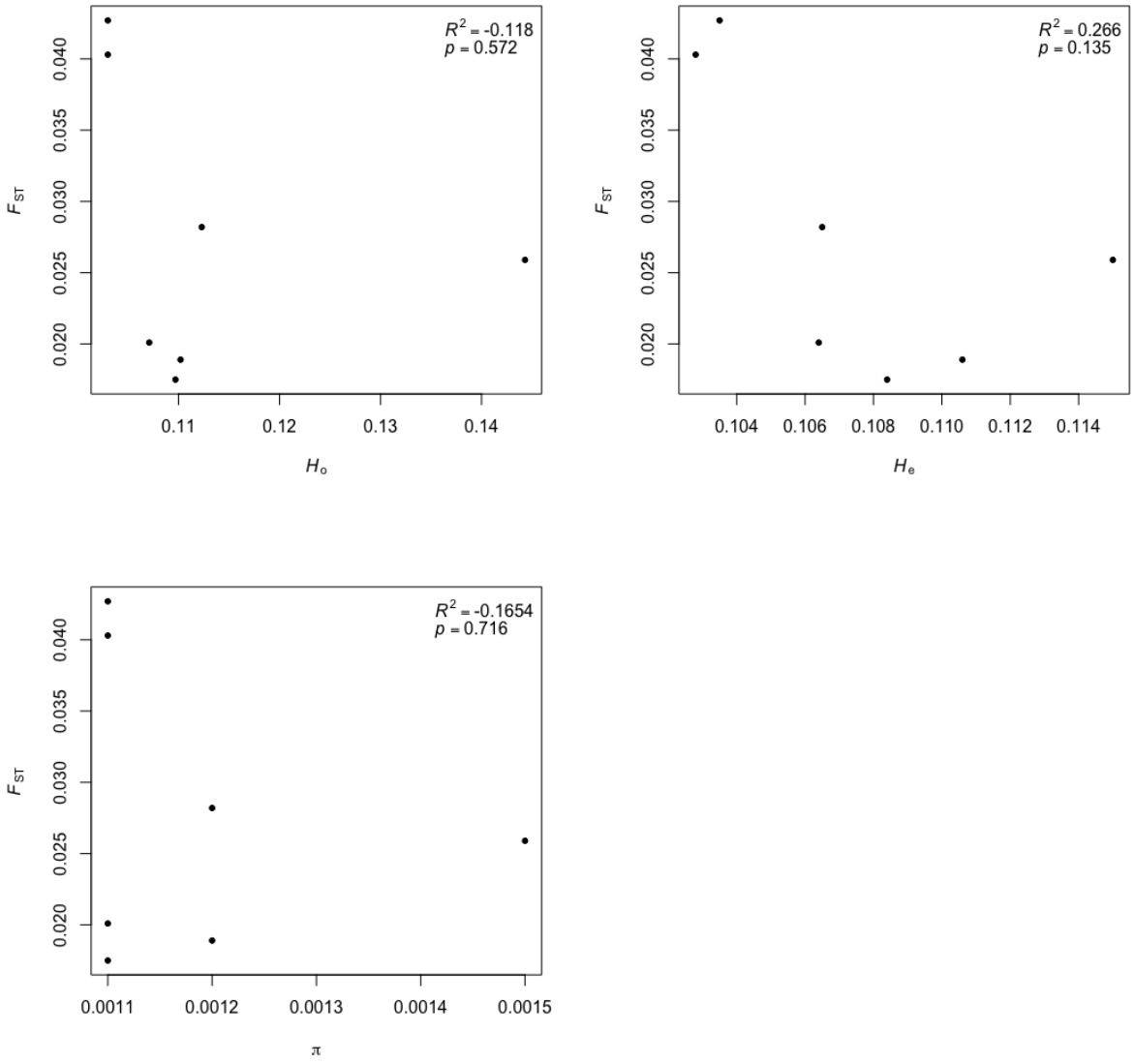


Figure S5. Scatterplots of pairwise F_{ST} values between all salt marsh populations and *gouldii* compared to three measures of genetic diversity (H_O - observed heterozygosity, H_e - expected heterozygosity, and π - nucleotide diversity). R^2 and p values are included and show no significant correlation between genetic differentiation and site-specific genetic diversity (see text).

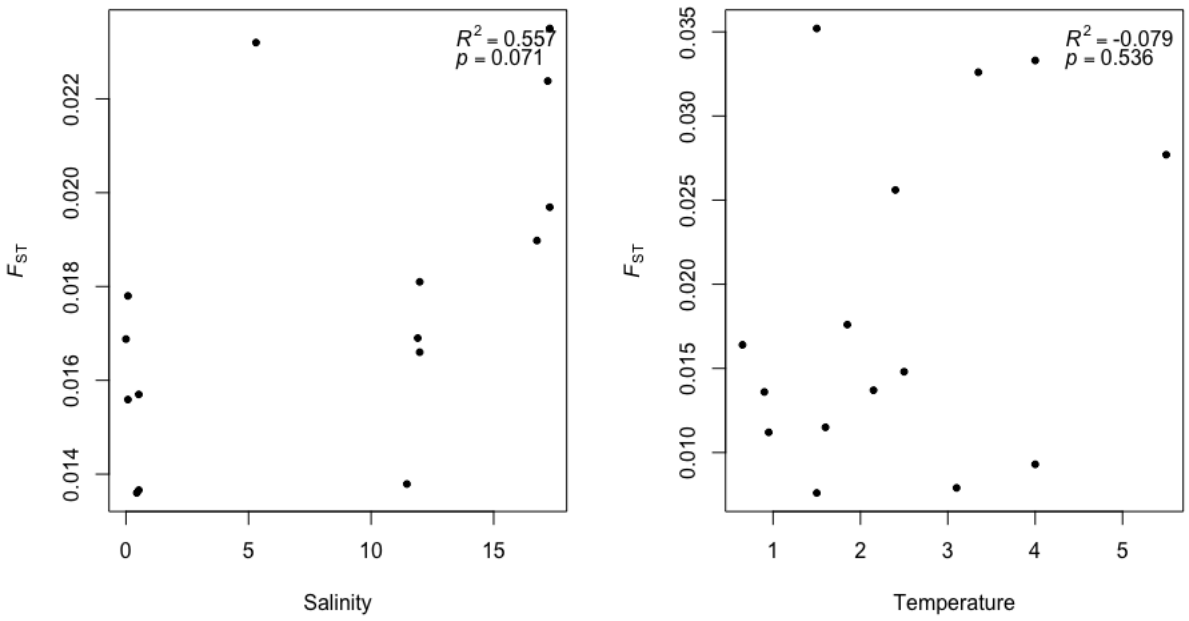


Figure S6. Scatterplots showing the results of partial mantel tests between matrices of pairwise F_{ST} by subspecies and salinity (A) and temperature (B), while controlling for distance. There is a positive trend in the test correlating pairwise F_{ST} and salinity, but no significant effect of temperature on pairwise F_{ST}.

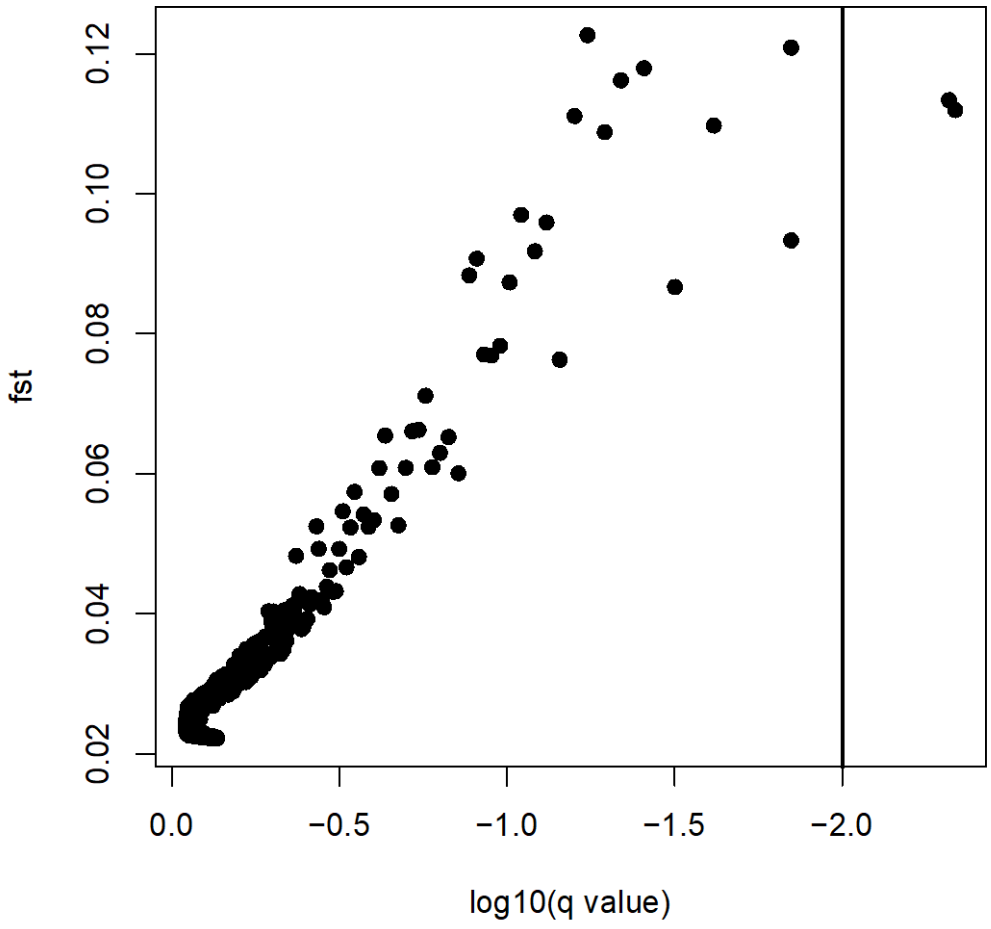


Figure S7. BayeScan analysis using the full SNP dataset (10,270 SNPs) identified two outliers (0.019% of all SNPs) using a q value cutoff of -2.0.

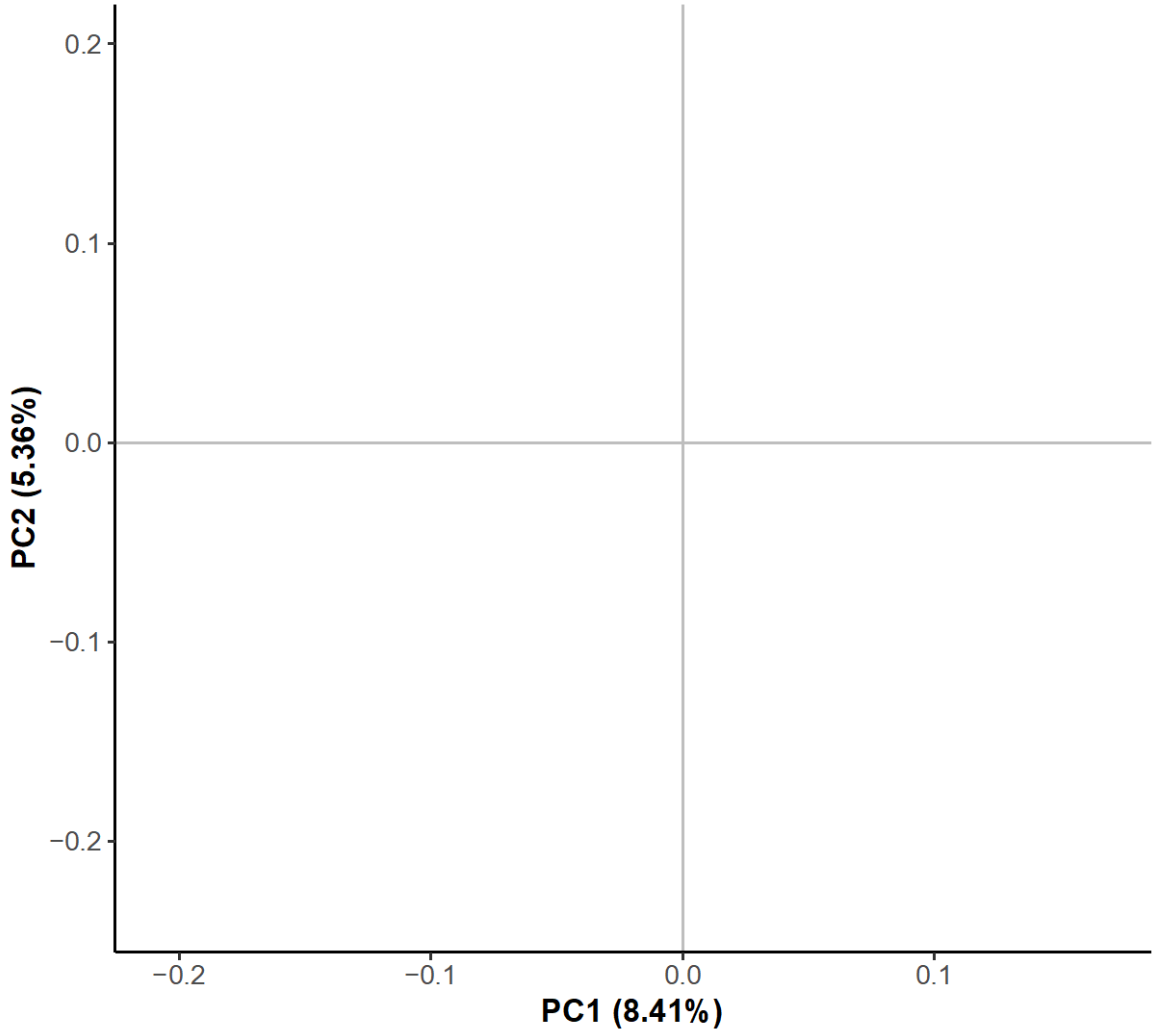


Figure S8. Principle Component Analysis (PCA) using 103 outlier SNPs (F_{ST}>5 standard deviations above the mean). Marsh populations are in blue and upland populations are in red (subspecies coloration as in previous figures).

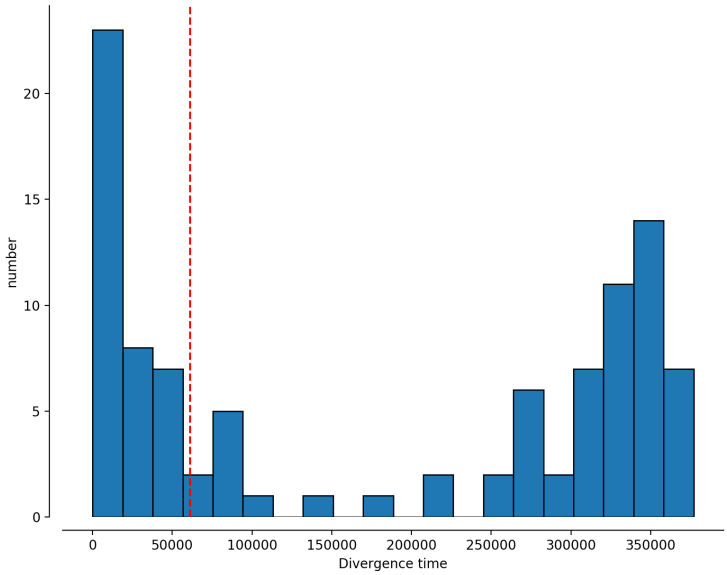
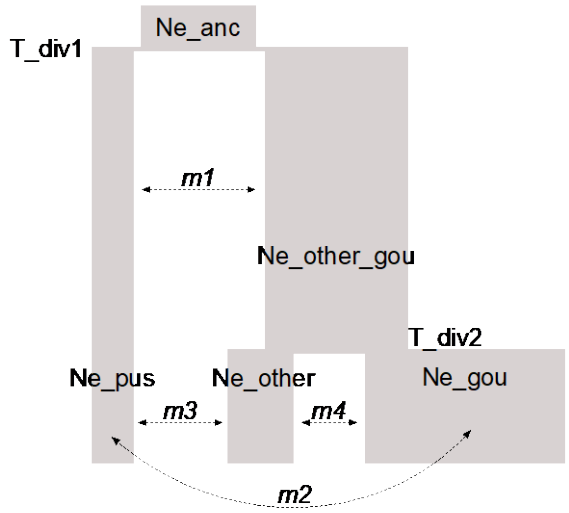


Figure S9. Results of nonparametric bootstrapping approach showing distribution of divergence time estimates from optimizing model fit to 100 resampled bootstrapped site frequency spectrums. Red dashed line shows divergence time estimate from empirical dataset.

a.



b.

Parameter	Point estimate	95% Confidence interval
Likelihood	-1659.93	–
Theta	30.51	28.12 – 32.89
Ne_anc	8,867	8,174 – 9,560
Ne_pus	50,984	33,456 – 70,809
Ne_other_gou	252,871	207,199 – 302,936
Ne_other	53,102	9,610 – 103,264
Ne_gou	233,035	186,496 – 284,377
T_div1	527,190	423,881 – 731,576
T_div2	2,156	1 – 95,196
*m*1	7.78E-05	6.56E-05 – 8.81E-05
m2	3.01E-05	0 – 2.0E-04
m3	6.09E-06	0 – 1.0E-04
*m*4	2.52E-07	0 – 0.0029

Figure S10. Results of fitting an isolation with migration model to a three population site frequency spectrum. (a) schematic of three population demographic model with parameters listed on plot. (b) Table listing corresponding parameter estimates with 95% confidence intervals. Different migration rates denoted as *m*. Ne refers to effective population size. T_div denotes divergence times first between *M. m. pusillula* and all other subspecies (T_div1) followed by divergence between *M. m. gouldii* and other subspecies (T_div2).

Table S1. List of putative outlier regions (Fst estimates greater than 5 standard deviations above the mean). Table includes scaffold ID and position for outlier SNPs, the comparisons in which the regions was elevated, associated Fst estimate, the candidate gene associated with the SNP, E value from the blast results, and putative biological function of the identified candidate gene.

Chrom	Pos	Comparison	FST	Candidate Gene	E value	GO - Biological Process*
Contig315_pilon	4796634	*gouldii-maxillaris*	0.213316	MYH13	4 E -04	cellular response to starvation, muscle contraction
Contig315_pilon	4796644	*gouldii-maxillaris*	0.213316	MYH13	4 E -04	cellular response to starvation, muscle contraction
Contig315_pilon	4796648	*gouldii-maxillaris*	0.231158	MYH13	4 E -04	cellular response to starvation, muscle contraction
Contig1384_pilon	17355	*gouldii-maxillaris, santaecrucis-maxillaris*	0.243503	UDP-glucuronosyltransferase 1-1-like	1 E -31	NA
Contig78_pilon	1245623	*gouldii-pusillula*	0.324166	STAM2	7.00E-22	endosomal transport, intracellular protein transport, membrane organization, multivesicular body assembly
Contig315_pilon	5429527	*gouldii-pusillula, gouldii-santaecrucis*	0.366238	DNAH9	2.00E-35	cell projection organization, microtubule-based movement
Contig1208_pilon	3219420	*gouldii-pusillula, samuelis-pusillula, santaecrucis-pusillula*	0.424732	DOK3	1.00E-11	Ras protein signal transduction
Contig740_pilon	816830	*gouldii-santaecrucis, santaecrucis-maxillaris*	0.194034	FDXACB1	2.00E-35	rRNA base methylation
Contig215_pilon	7916168	*gouldii-heermanni*	0.207494	PYGL	6.00E-04	glucose homeostasis, glycogen metabolic process, response to bacterium
Contig597_pilon	21256414	*gouldii-heermanni*	0.229231	CCDC94	8.00E-38	mRNA processing, RNA splicing, negative regulation of DNA damage response
Contig869_pilon	3006220	*gouldii-heermanni, samuelis-heermanni*	0.239302	MYO18A	7.00E-41	actomyosin structure organization, cell migration, DNA metabolic process, Golgi organization, regulation of macrophage activation
Contig84_pilon	17714673	*gouldii-heermanni*	0.244763	TTBK2	2.00E-42	cerebellar granular layer development, cerebellum development, cillium assembly, microtubule cytoskeleton organization, regulation of cell migration, smoothened signaling pathway
Contig2118_pilon	79720	*heermanni-maxillaris*	0.156108	VDR	4.00E-37	organ morphogenesis, apoptotic signaling pathway, bile acid signaling pathway, calcium ion transport, cellular calcium ion homeostasis, regulation of keratinocyte differentiation
Contig84_pilon	752449	*heermanni-maxillaris, samuelis-heermanni, santaecrucis-maxillaris*	0.22114	TNKS1BP1	1.00E-31	DNA damage response, telomere maintenance
Contig391_pilon	1344696	*heermanni-maxillaris*	0.235947	DENND1A	2.00E-20	endocytic recycling, protein transport, synaptic vesicle endocytosis
Contig84_pilon	1491967	*heermanni-pusillula, heermanni-santaecrucis, samuelis-heermanni, samuelis-maxillaris*	0.369873	PITPNM1	1.00E-05	brain development, lipid metabolic process, phospholipid transport, protein transport
Contig815_pilon	5664552	*heermanni-santaecrucis*	0.190094	PAK3	5.00E-17	activation of protein kinase activity, cell migration, dendrite development, regulation of actin cytoskeleton organization, regulation of MAPK cascade, stress-activated protein kinase signaling cascade
Contig88_pilon	15722076	*pusillula-maxillaris*	0.247609	DNPEP	4.00E-35	peptide metabolic process
Contig215_pilon	7644021	*samuelis-gouldii*	0.280949	TTI1	9.00E-33	invovled in cellular resistance to DNA damage stresses
Contig2077_pilon	58603	*samuelis-maxillaris*	0.15874	TMEM230	1.00E-31	synaptic vesicle transport
Contig384_pilon	391629	*samuelis-maxillaris*	0.190316	ADAR	7.00E-28	cellular response to virus, innate immune response, mRNA processing, regulation of apoptotic process
Contig315_pilon	4971135	*samuelis-santaecrucis, santaecrucis-maxillaris*	0.304479	MYO3	2.00E-35	actin cortical patch localization, response to osmotic stress
Contig78_pilon	1245578	*santaecrucis-maxillaris*	0.165458	STAM2	7.00E-22	intracellular protein transport, membrane organization, multivesicular body assembly
Contig239_pilon	28604156	*santaecrucis-maxillaris*	0.173174	DNAJB12	6.00E-16	cellular response to misfolded protein, ERAD pathway
Contig551_pilon	2061940	*santaecrucis-maxillaris*	0.173239	ZDHHC7	1.00E-24	protein targeting to membrane
Contig384_pilon	391696	*santaecrucis-pusillula*	0.25641	ADAR	7.00E-28	cellular response to virus, innate immune response, mRNA processing, regulation of apoptotic process

*In cases where a gene has many putative biological processes, we do not present the full list here

Table S2. Parameter estimates and uncertainties as 95% confidence interval for the best fit model to the two population site frequency spectrum in *moments*. Asterisk reflects the large uncertainty surrounding divergence time with the lower end of the 95% CI for divergence time being negative and rounded up to a possible divergence time of 1.

		Uncertainties from GIM		Uncertainties from nonparametric bootstrap	
parameter	point estimate	95% CI		95% CI	
Likelihood	-608.19	-	-	-	-
theta	432.41	333.02	627.23	17.39	608.68
Ne_ancestral	125,675	96,787	182,295	5,053	176,902
Ne_pusillula	26,573	6,888	64,117	13,433	234,592
Ne_other subspp	256,940	172,250	420,972	126,065	804,369
Divergence time	61,058	1*	321,258	1,806	358,553
m1 other -> pusillula	1.37E-04	5.07E-05	1.62E-04	6.38E-08	3.15E-04
m2 pusillula -> other	2.82E-05	1.70E-05	2.99E-05	3.02E-08	7.11E-05