

UCLA

UCLA Previously Published Works

Title

Evolutionary divergence and adaptive capacity in morphologically distinct song sparrow subspecies

Permalink

<https://escholarship.org/uc/item/97r6898d>

Journal

Conservation Genetics, 24(6)

ISSN

1566-0621

Authors

Mikles, Chloe S

Arcese, Peter

Lovette, Irby J

et al.

Publication Date

2023-12-01

DOI

10.1007/s10592-023-01547-w

Supplemental Material

<https://escholarship.org/uc/item/97r6898d#supplemental>

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

1 **EVOLUTIONARY DIVERGENCE AND ADAPTIVE CAPACITY IN**
2 **MORPHOLOGICALLY DISTINCT SONG SPARROW SUBSPECIES**

3 Chloe S. Mikles^{1,2}, Peter Arcese³, Irby J. Lovette^{1,4}, Stephanie M. Aguilon^{1,4,5}, Yvonne L. Chan⁶,
4 Phred M. Benham⁷, Katherine Carbeck³, and Jennifer Walsh^{1,4*}

5 ¹Fuller Evolutionary Biology Program, Cornell Laboratory of Ornithology, Cornell University,
6 Ithaca, NY 14850

7 ²Current address: Hopkins Marine Station of Stanford University, Pacific Grove, CA 93950

8 ³Department of Forest and Conservation Sciences, University of British Columbia, 2424 Main
9 Mall, Vancouver, BC T6T1Z4

10 ⁴Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA

11 ⁵Current address: Department of Biology, Stanford University, Stanford, CA 94305

12 ⁶Iolani School, 563 Kamoku Street, Honolulu, HI 96826

13 ⁷The Museum of Vertebrate Zoology, University of California, Berkeley, 3101 UC Berkeley Rd,
14 Berkeley, CA 94720

15 *Corresponding author: jennifer.walsh.emond@gmail.com

16

17 **Keywords:** *Melospiza melodia*, Adaptive Capacity, Conservation Units, Subspecies Delineation,
18 Local Adaptation

19

20 **ABSTRACT**

21 Spatial variation in the environment can affect population fitness and individual phenotype by
22 facilitating natural selection and local adaptation, and thereby enhance the diversity and adaptive
23 capacity and persistence of species at regional to continental scales. The song sparrow subspecies
24 complex endemic to the San Francisco Bay region, which has received over a century of close
25 study, presents an opportunity to evaluate the adaptive potential of distinct subspecies faced with
26 habitat loss, population decline, and threats of future environmental change. We used whole-
27 genome sequences from 39 individuals representing five morphologically distinct song sparrow
28 subspecies to evaluate the role of neutral and adaptive evolutionary processes in driving
29 divergence within physiologically challenging habitats across multiple environmental clines. We
30 found that natural selection for traits explained by ecological variables, including temperature

31 and salinity, are drivers of adaptive genetic variation in these song sparrows. Differentiation was
32 highest for candidate loci under selection (compared to neutral markers), as predicted if local
33 ecological processes are at least partially responsible for the rapid radiation of these subspecies.
34 Our findings inform management aimed at conserving and prioritizing population-level diversity
35 in species displaying local adaptation and inhabiting a diverse range of environments.

36

37 **INTRODUCTION**

38 Climate change is occurring at rates likely to exceed the capacity of many populations to
39 adapt via contemporary evolution, casting doubt on their persistence (Anderson et al. 2012).
40 Predicting ‘winners and losers’ in this race remains challenging due to uncertainties about how to
41 characterize the adaptive potential of populations, the effects of historic and contemporary gene
42 flow and natural selection on the pace of local evolution, and the appropriate metrics for
43 prioritizing populations for conservation (Bay et al. 2018, Coates et al 2018). A growing body of
44 literature aims to define populations’ adaptive capacity, or the ability to cope with environmental
45 change through phenotypic plasticity and evolution (Bay et al. 2017) and to use this information
46 to help inform conservation initiatives. Historically, and in the absence of genetic data, biologists
47 often rely on phenotypic traits to delineate populations for conservation, especially those traits
48 thought to be under genetic control, influential to fitness, or diagnostic of population identity
49 (Haig & Winker 2010). The increasing application of modern genomic methods for systematics,
50 however, provides the opportunity to evaluate adaptive and neutral evolutionary processes,
51 offering additional support for conserving population-level variation (e.g., Oh et al. 2019).

52 Although these approaches are promising, from a taxonomic perspective, there are
53 concerns that the increased resolution from genomic data may lead to over-splitting if all
54 genetically distinctive populations are classified as full species (Coates et al. 2018, Winker
55 2021). For birds, most phenotypically distinctive populations have historically been described
56 and formally named as taxonomic subspecies, and in turn these subspecies have often become
57 populations targeted for increased conservation efforts and legal protection (Barrowclough et al.
58 2016). Some taxonomists, however, have criticized the prioritization of subspecies in
59 conservation, particularly when the subspecies' classification is based on delineations supported
60 by small numbers of neutral genetic markers, or on morphological traits that are clinal across

61 space (Zink 2004, 2010, Zink & Barrowclough 2008). Such debates arise via the worthy desire to
62 discretize biological diversity, often as a requirement of legal proceedings and legislative
63 policies, versus the contrasting view of speciation and differentiation as an ongoing continuum, a
64 perspective increasingly supported by genomic data (Henderson et al. 2020).

65 This debate is particularly well developed with respect to a San Francisco Bay complex
66 of five resident song sparrow subspecies: *Melospiza melodia heermanni*, *M. m. maxillaris*, *M. m.*
67 *samuelis*, *M. m. pusillula*, and *M. m. gouldii* (Figure 1a,c). Song sparrows are widely distributed
68 across North America and display phenotypic variation among their 25 named and 52 described
69 subspecies. The San Francisco Bay region contains the highest concentration of individual
70 subspecies, with five recognized subspecies occupying distinct habitats within a 100x70km
71 region that spans a selective gradient of salt and freshwater habitats. *Maxillaris*, *samuelis*, and
72 *pusillula* are particularly notable for their year-round residency restricted to separate but
73 geographically proximal salt marshes; they occupy a challenging niche that only 25 other
74 vertebrate species have been able to successfully colonize, of which nearly all are of
75 conservation concern due to habitat loss (Greenberg et al. 2006). To this end, subspecies
76 delineations in this system are supported by evidence suggesting rapid adaptation to saline
77 environments (Basham & Mewaldt 1987), differences in coloration (Miller 1956), and genetic
78 isolation (Ferrell 1966), all amounting to substantial evidence of parapatric differentiation (Mayr
79 1942). In contrast, neither microsatellite (Chan & Arcese 2002) nor mitochondrial DNA (Fry &
80 Zink 1998) data are coincident with these phenotypes (Patten & Pruett 2009). For example, Zink
81 (2004, 2010) concluded that neither subspecies delineations nor neutral genetic markers offer
82 evidence of evolutionary significance or conservation value for these localized populations of
83 sparrows. Most recently, reduced representation genome sequencing provided increased
84 resolution, with subtle separation of the subspecies at microgeographic scales (Mikles et al.
85 2020). Though the underlying mechanisms of such genomic divergence remain unclear,
86 differentiation is not explained by isolation by distance (Chan & Arcese 2002, Mikles et al.
87 2020). Here, we use the San Francisco Bay song sparrows as a case study to ask if their
88 subspecies delineations reflect evolutionary processes responsible for the rapid diversification of
89 these populations over the last 10,000 years (Chan & Arcese 2002). If so, then these subspecies
90 classifications would have utility in indicating their evolutionary distinctiveness, and this would

91 in turn imply that the subspecies level of classification has potential merit for recognizing
92 conservation units in other avian taxa.

93 **METHODS**

94 ***Whole genome re-sequencing and variant discovery***

95 We sequenced the genomes of 43 song sparrows representing five subspecies from the
96 San Francisco Bay area (Figure 1a; Table S1). All birds were sampled during the breeding
97 season (March to May, 1999) by Y. Chan and P. Arcese; adults were captured in mist nets,
98 measured and blood sampled, and released (see Chan & Arcese 2002). We extracted genomic
99 DNA using the DNeasy blood and tissue kit (Qiagen, CA, USA) and quantified DNA
100 concentrations using the Qubit dsDNA High Sensitivity Assay Kit (Life Technologies). Using
101 200 ng of DNA from each sample, we prepared individually barcoded libraries with a 550 bp
102 insert size following the protocol for the TruSeq Nano DNA Library Prep kit (Illumina,
103 California, USA). We then sequenced libraries for *heermanni*, *maxillaris*, and *samuelis* were
104 sequenced on a single Illumina NextSeq lane at the Cornell Institute for Biotechnology core
105 facility. We obtained raw sequences for *gouldii* and *pusillula* from Walsh et al. (2019a).

106 We assessed library quality using FASTQC version 0.11.8. We used ADAPTERREMOVAL
107 version 2.1.1 for sequence trimming, adapter removal, and quality filtering, requiring a minimum
108 Phred quality score of 20 and merged overlapping paired-end reads. We aligned filtered reads to
109 the Song Sparrow reference genome (Feng et al. 2020) using the default settings in BWA 0.7.4
110 (Li & Durbin 2009) and obtained alignment statistics from QUALIMAP version 2.2.1
111 (Okonechnikov et al. 2016). We removed three samples for >50% missing data, and one for 35%
112 relatedness to another individual (selecting to keep the individual with higher quality mapping
113 statistics). We used SAMTOOLS version 1.9 (Li et al. 2009) to convert all resulting BAM files to
114 SAM files and to sort and index files, and PICARD TOOLS version 2.19.2 to add index groups and
115 mark duplicates. We used the *Haplotype Caller* module in GATK version 3.8.1 (McKenna et al.
116 2010) for single nucleotide polymorphism (SNP) discovery and genotyping and used the
117 following filtering parameters to remove variants: QD < 2, FS > 60.0, MQ < 30.0, and
118 ReadPosRankSum < -8.0. We additionally filtered out variants that were not biallelic, had minor
119 allele frequencies less than 5%, mean coverage less than 2X or more than 50X, and more than
120 20% missing data. This resulted in a total of 1,630,425 SNPs across all five subspecies.

121 *Characterizing Subspecific Niche Variation*

122 We used existing range maps, bioclimatic variables (WorldClim V2; Fick & Hijmans
123 2017) and Random Forest (RF; Breiman 2001) to characterize environmental niches for each
124 subspecies to assess possible patterns of ecological divergence in this system (Elith &
125 Leathwick, 2009). Briefly, we first georeferenced range maps in Chan & Arcese (2002), Patten et
126 al. (2004) and Patten & Pruett (2009) using *georeferencing* and *editor* tools in ArcGIS (version
127 10.7.1; ESRI, 2019) to obtain a contiguous map of the range of each subspecies (Figure 1a). To
128 characterize variation in their niches we used *a priori* knowledge on the effect of climatic
129 conditions and events on song sparrow demography to identify 11 candidate variables: Annual
130 Mean Temperature (bio1), Temperature Seasonality (bio4), Mean Temperature of Driest Quarter
131 (bio9), Mean Temperature of Warmest Quarter (bio10), Mean Temperature of the Coldest
132 Quarter (bio11), Annual Precipitation (bio12), Precipitation Seasonality (bio15), Precipitation of
133 Wettest Quarter (bio16), Precipitation of Driest Quarter (bio17), Precipitation of Warmest
134 Quarter (bio18), Precipitation of Coldest Quarter (bio19) (Table S2) extracted at 30 arc-second
135 spatial resolution from WorldClim, spanning the period of 1970-2000. As an additional
136 candidate variable, we estimated salinity as a proxy using Euclidean distance to the edge of the
137 nearest saline wetland using the *spatial analyst* tool in ArcGIS and the Areas of Conservation
138 Emphasis Saline Wetlands geospatial data layer, which represents wetland area and location at
139 the HUC12 watershed level (California Department of Fish and Wildlife, 2019). This salinity
140 proxy was determined assuming there is a plateau, and thus was calculated as the Euclidean
141 distance squared, resulting in the further distances being much larger and therefore less affected
142 by salt.

143 We used RF to classify and delineate subspecies ranges by their environmental
144 conditions. Within the RF algorithm, trees are produced using a bootstrapped sample of ‘bagged’
145 data that comprise ~64% of all observations, and they are tested against the remaining ‘out-of-
146 bag’ (OOB) data to estimate prediction error (OOB error) as the percentage of incorrectly
147 classified observations. We carried out the tuning and fit of RF in R version 4.0.2 (R Core Team,
148 2020), using the “CARET” and “RANDOMFOREST” packages (Liaw & Wiener 2002; Kuhn, 2020).
149 Our model was trained using five repetitions of a 10-fold cross validation scheme, wherein we
150 randomly split the data into 50% training (n = 30,000) and 50% evaluation data (n = 30,000) to

151 avoid overfitting (Kuhn & Johnson 2013). These data were generated by randomly selecting
152 60,000 pixels across our study area and associated with subspecies ranges (Fig 1a). This process
153 generated a final model trained with 11 predictor variables with the best hyperparameter values
154 ($mtry = 2$; $ntree = 500$) according to the RMSE parameter (Kuhn & Johnson 2013). To reduce
155 prediction bias, we fit a balanced RF model by growing each tree from random samples of the
156 data with an equal number from each class. We adopted the Gini index to evaluate variable
157 importance, wherein a higher Gini importance indicates a variables' importance in maintaining
158 predictive power in the RF model. We then evaluated model performance using unseen
159 evaluation data for accuracy, sensitivity, specificity, and Kappa (Kuhn & Johnson 2013). To
160 assess whether the observed subspecies niches were statistically different among the song
161 sparrows, we used an analysis of variance (ANOVA) on the top five most important variables
162 from the final RF model. Variables in this analysis were normalized to bring values to range
163 from 0-1 to assess effect size.

164 ***Neutral Genomic Population Structure & Subspecies Delineation***

165 To characterize patterns of genetic structure among subspecies, we ran ADMIXTURE
166 version 1.2.3 (Alexander et al. 2009) using a filtered data set (4,961 SNPs) that contained no
167 missing data, was pruned to avoid linkage using the script *ldPruning.sh*
168 (<https://github.com/speciationgenomics/scripts/blob/master/ldPruning.sh>), and had putatively
169 adaptive SNPs removed (see below). For this analysis, we investigated five population clusters,
170 using the default settings. For all other analyses of population structure, we assessed patterns and
171 genomic diversity based on the full data set (1,630,425 SNPs) and a putatively neutral SNP data
172 set. To target putatively neutral SNPs, we excluded SNPs that were mapped to exons or intervals
173 within 25 kb of exons (see Kawakami et al. 2014), resulting in 308,973 SNPs. We performed a
174 principal component analysis (PCA) on the full and neutral data sets using the “SNPRELATE”
175 package in R (Zheng et al. 2012). For both data sets, we additionally characterized genome-wide
176 patterns of divergence between subspecies by calculating pairwise F_{ST} values for each
177 comparison using VCFTOOLS (Danecek et al. 2011). We calculated F_{ST} for 25 kb windows across
178 our scaffolds and for individual SNPs, dropping windows with fewer than 10 SNPs. Using
179 pairwise F_{ST} estimates for the full data set, we tested for isolation by distance using a Mantel test
180 in R. We quantified genetic diversity by estimating individual heterozygosity and nucleotide

181 diversity in 25 kb windows using VCFTOOLS; for these calculations, we removed all missing
182 data from both datasets, as estimates can be biased by missing data (Schmidt et al., 2021). For
183 the full and neutral data sets, we counted the number of private alleles within each subspecies
184 using BCFTOOLS (Li et al. 2009).

185 ***Genotype-Environment Associations***

186 To assess whether the observed genomic differentiation among song sparrows in the San
187 Francisco Bay area was a result of ecological divergence, we scanned for SNPs associated with
188 five environmental variables identified by RF as most important in delineating ecological niches
189 of the subspecies. These variables included temperature seasonality, mean temperature of the
190 coldest quarter, annual mean temperature, salinity, and annual precipitation. We tested for
191 Genotype-Environment Associations (GEA) through a combination of multivariate and
192 univariate approaches: we ran a redundancy analysis (RDA; multivariate) using the *rda* function
193 in the R package VEGAN 2.4-5 (Oksanen et al., 2017), and a latent factor mixed model (LFMM;
194 univariate) using the *lfmm* function of the LFMM package in R. Both methods are robust to a
195 range of underlying demographic processes and sampling designs (Rellstab et al., 2015; Forester
196 et al., 2018), while providing a balance between error rates and detection power (Carvalho et al.,
197 2020). For both methods, we imputed missing genotypes by using the most common genotype at
198 each SNP across all individuals. Because temperature variables were correlated (pairwise
199 Pearson correlation coefficients > 0.7), we ran a PCA on all temperature variables using the
200 *prcomp* function in R and used the first and second principal components as predictors in the
201 GEA analyses (Frichot et al., 2013). We used results from ADMIXTURE to define the number
202 of latent factors used as $K=3$ (Supplemental Material, Figure S1).

203 RDAs can offer a robust approach to detecting correlations between genotype and
204 environmental data, particularly compared to other differentiation-based outlier scans (Rellstab et
205 al., 2015; Forester et al., 2018). We used an RDA to test for multilocus signatures of selection for
206 multiple environmental variables, and evaluated the significance of the RDA using an analysis of
207 variance (ANOVA) with 999 permutations. Loci with a loading $\pm 4 SD$ from the mean loading on
208 the first three constrained ordination axes were considered candidates under environmental
209 selection following Forester et al. (2018). We used a Pearson's correlation (r) to identify
210 environmental variables exhibiting the strongest association with each candidate locus.

211 For the LFMM analysis, we conducted 5 independent runs of 10,000 iterations and a
212 5,000 iteration burn-in. We corrected association p -values based on empirical genomic inflation
213 factors (λ) to control for false discoveries (François et al., 2016). To do this, we inspected the
214 distribution of p -values to ensure that they followed a normal distribution following François et
215 al., (2016). For all variables, we set $\lambda = 0.1$. When generating a list of candidate SNPs, we used
216 the Benjamini-Hochberg algorithm (Benjamini & Hochberg, 1995) with a maximum false
217 discovery rate of 10^{-5} .

218 ***Evolutionary Distinctiveness and Identification of Conservation Units***

219 Based on the GEA methods above, we identified shared candidate SNPs that correlated
220 with environmental variables, which we refer to from here on as a putatively adaptive SNP data
221 set. To provide a measure of adaptive diversity we calculated pairwise F_{ST} and individual
222 heterozygosity with the putatively adaptive SNP data set. To prioritize groups for conservation,
223 we calculated a measure of population distinctiveness for each subspecies. To do this, we
224 calculated the *Shapley metric* (SH; Volkman et al. 2014), which can be calculated from
225 unrooted trees. F_{ST} estimates from pairwise comparisons of all five subspecies were used to build
226 a NeighborNet network using the *neighborNet* function in the R package “PHANGORN” (Schliep
227 2011, Hudson & Bryant 2006). This network was then used to estimate the genetic contribution
228 of individual tips (Volkman et al. 2014), with higher SH values indicating higher priority for
229 management.

230 **RESULTS**

231 Whole genome re-sequencing yielded a mean of 16,043,533 reads per individual with the
232 following sample sizes per subspecies: *gouldii* ($n = 10$), *heermanni* ($n = 8$), *samuelis* ($n = 6$),
233 *pusillula* ($n = 9$), and *maxillaris* ($n = 6$). The mean alignment rate was 97.5%, the mean coverage
234 was 2.60X, and the mean missing data was 12% (Table S1).

235 ***Subspecific Niche Variation***

236 Climate varied predictably across the five subspecies ranges. The top five most influential
237 variables in delineating subspecies niches were temperature seasonality (bio4), mean temperature
238 of the coldest quarter (bio11), annual mean temperature (bio1), annual precipitation (bio12), and
239 distance to saltwater (Figure 1b). The influence of the top five climate predictors varied
240 significantly among subspecies ranges (Figure 1d; $p < 0.001$), except coldest quarter (bio11) and

241 annual mean temperature (bio1) did not differ between the ranges of *gouldii* (normalized mean \pm
242 SD: 0.47 ± 0.17) and *maxillaris* (0.47 ± 0.05 ; $p > 0.99$), and *gouldii* (0.46 ± 0.19) and *samuelis*
243 (0.48 ± 0.05 ; $p > 0.05$), respectively. Distance to saltwater over the ranges of the marsh and
244 upland subspecies varied as predicted, given their known ecotypes. Distance to saltwater also
245 varied between the two upland subspecies (*gouldii*: 0.13 ± 0.18 ; *heermanni*: 0.22 ± 0.25 ; $p <$
246 0.001).

247 The final RF model displayed high accuracy ($97.92\% \pm 0.21$), sensitivity ($91.81\% \pm$
248 0.48), and specificity ($98.49\% \pm 0.19$), and a mean error rate of $2.27\% (\pm 0.08)$ and Kappa of
249 $96.75\% (\pm 0.32)$. Classification error differed among subspecies such that upland subspecies had
250 the lower error rates (*gouldii* = 2.25% ; *heermanni* = 0.96%) compared to marsh subspecies
251 (*maxillaris* = 8.58% ; *pusillula* = 15.14% ; *samuelis* = 16.72%).

252 ***Neutral Genomic Population Structure & Subspecies Delineation***

253 We observed subtle differentiation among the five subspecies. In the full data set, results
254 from ADMIXTURE supported $K=1$ as the best supported cluster, yet we detected observable
255 structuring up to $K=3$. Under this model, clusters corresponded to 1) *pusillula*, 2) *gouldii*, and 3)
256 *heermanni*, *maxillaris*, and *samuelis* (Figure S1;a). Based on ~ 1.6 million SNPs, we observed
257 some clustering by subspecies on a PCA (Figure S1;b), with the three salt marsh subspecies
258 separating along axis one (3.92% of variation explained). The most prominent clustering in this
259 PCA was separation of *pusillula* and *samuelis* from the rest of the subspecies. This pattern of
260 separation along PC axes was still present but less pronounced for the neutral data set (Figure
261 S2). Genome-wide F_{ST} estimates further suggest moderate levels of divergence among San
262 Francisco Bay area song sparrows (Table S3, S4). Mean F_{ST} estimates based on the full data set
263 ranged from 0.011 (*maxillaris* vs *heermanni*) to 0.047 (*pusillula* vs *samuelis*; Table S3). Per SNP
264 F_{ST} estimates ranged from $0-1$ in each of the 10 pairwise comparisons. There was no significant
265 difference between F_{ST} estimates based on the full data versus F_{ST} estimates based on the neutral
266 data set (Paired Sample T-test; $df = 9$, $t = -1.67$, $p = 0.12$). We found no significant correlation
267 between genetic and geographic distance ($p = 0.129$; Figure S5).

268 We observed slight differences in heterozygosity across the five subspecies (Figure S3),
269 with *pusillula* and *gouldii* populations exhibiting the lowest mean heterozygosity (0.33 and 0.35 ,
270 respectively). Mean heterozygosity was highest in *samuelis* (0.46). We observed a significant

271 difference between individual heterozygosity estimates based on the full versus neutral data set
272 (Paired Sample T-test; $df = 38$, $t = -14.7$, $p < 2.2e-16$). Patterns of nucleotide diversity were
273 notably similar among the five subspecies (Figure S4). Based on the full data set, the number of
274 private alleles observed in each subspecies was: 13,247 (0.82%) in *gouldii*, 11,258 (0.70%) in
275 *heermanni*, 4,279 (0.29%) in *maxillaris*, 16,140 (0.98%) in *pusillula*, and 10,156 (0.62%) in
276 *samuelis*. The number of private alleles observed in the neutral data set was proportionally
277 comparable: 0.81% in *gouldii*, 0.88% in *heermanni*, 0.37% in *maxillaris*, 1% in *pusillula*, and
278 0.78% in *samuelis*.

279 ***Genotype-Environment Associations***

280 For the RDA, the first three components explained 36.15%, 27%, and 20.4% of the
281 variation, respectively and the full model was significant ($p = 0.027$). Temperature PC1 showed
282 significant variation with song sparrow genotypes ($p = 0.014$) and captured approximately 90%
283 of the variation driven by all three temperature variables. Annual precipitation (bio12; $p = 0.14$),
284 temperature PC2 ($p = 0.84$), and salinity ($p = 0.1$) did not show significant variation with
285 genotype. The first two axes of the RDA largely separated *pusillula* and *gouldii* from the other
286 three subspecies (Figure 2a). RDA1 appeared to associate more with salinity and RDA2
287 associated with the remaining environmental variables. We saw distinct clustering of individuals
288 by subspecies along axes two and three of the RDA, with axis three separating salt marsh
289 populations from upland populations (Figure 2b). RDA3 appeared to be associated with all the
290 variables, with salinity and annual precipitation being negatively correlated with temperature.
291 Based on our cutoff of ± 4 SD, we identified 171 candidate SNPs that correlated with
292 environmental variables. These included 143 SNPs associated with annual precipitation, 25
293 SNPs associated with temperature PC1, 1 SNP associated with temperature PC2, and 2 SNPs
294 associated with salinity. Correlations between these candidate SNPs and their most strongly
295 associated environmental variable were moderate, averaging 0.63 (r range = 0.13 – 0.74).

296 LFMM identified substantially more candidate loci than the RDA. LFMM identified
297 allele frequencies of 11,451 SNPs as significantly associated with environmental variables (282
298 SNPs for salinity; 8,172 SNPs for annual precipitation; 39 SNPs for temperature PC1; 2,958
299 SNPs for temperature PC2). Given the discrepancy in the number of outliers identified by
300 LFMM compared to our other methods, we only retained those outliers that were identified by

301 both the RDA and LFMM. For the downstream identification of genes linked to putative regions
302 under selection, we retained 144 SNPs that were identified as outliers by both RDA and LFMM
303 (Table S5).

304 *Evolutionary Distinctiveness and Identification of Conservation Units*

305 Mean F_{ST} estimates based on the 144 candidate SNPs ranged from 0 (*samuelis* vs
306 *heermanni* and *gouldii* vs *heermanni*) to 0.058 (*pusillula* vs *samuelis*; Table S6). Pairwise F_{ST}
307 estimates did not differ significantly between putatively adaptive SNPs compared to estimates
308 based on the full data set (Paired Sample T-test; $df = 9$, $t = -2.56$, $p = 0.8$). The overall pattern of
309 heterozygosity among subspecies based on outlier SNPs was the same as that based on the full
310 data set, with *gouldii* and *pusillula* having the lowest per-individual heterozygosity estimates.
311 We observed a significant difference between individual heterozygosity estimates based on the
312 full versus adaptive data set (Paired Sample T-test; $df = 38$, $t = 2.66$, $p = 0.01$). The NeighborNet
313 network for subspecies in the San Francisco Bay is non-tree like, and the close placement of
314 groups to each other on the network is as expected given their recent divergence times (Figure
315 2c). We do see the placement of *pusillula* at a more isolated tip in the network, which is
316 consistent with our other metrics of population structure. The *pusillula* subspecies had the
317 highest SH rank, as expected based on the network (0.029), followed by *maxillaris* (0.01). The
318 other subspecies largely clustered together in the network and had lower, and comparable, SH
319 rankings: 0.008 for *samuelis*, 0.006 for *heermanni*, and 0.007 for *gouldii*.

320 **Discussion**

321 An extraordinary amount of attention has been dedicated to the study of the San
322 Francisco Bay song sparrows (Grinnell 1909, Huxley 1942, Grinnell & Miller 1944, 1956,
323 Marshall 1948a,b, Johnston 1956a,b, Mayr 1963, Chan & Arcese 2002, 2003, Mikles et al.
324 2020). Notable for their high concentration of morphologically distinct subspecies within a small
325 geographic area, these song sparrows present a tractable system for investigating replicated
326 colonization of marsh environments across a habitat gradient variable in salinity, temperature,
327 and precipitation, and one now drastically altered by anthropogenic influence. By combining
328 genomic data with random forest niche modeling, we identified evidence of local selection and
329 putative ecological divergence over a fine spatial scale in the song sparrow subspecies of the San
330 Francisco Bay. We posit that ecological variables linked to microgeographic habitat variation in

331 the Bay are primarily responsible for the rapid radiation of these subspecies. Our work further
332 suggests that selection for adaptive phenotypes, rather than neutral processes linked to drift or
333 divergence time alone, is the primary driver of diversification in this system, however we discuss
334 these conclusions within the context of alternative mechanisms in greater detail below. We
335 conclude that these song sparrows offer lessons for the application of how genomic data can be
336 applied to the characterization and conservation of local genetic diversity.

337 *Evolution and Conservation in the San Francisco Bay ecoregion*

338 Saltwater marshes are inherently challenging environments that require specialized
339 adaptations in the vertebrate species that have colonized them (Greenberg et al. 2006). The
340 fragmented and patchy distribution of salt marsh habitats creates further challenges when
341 prioritizing conservation efforts among small populations subject to rapid environmental change.
342 Despite being the largest estuary on the west coast of North America, 90% of the San Francisco
343 Bay marsh habitats have been converted to human use (San Francisco Bay Estuary Project 1991,
344 Takekawa et al. 2006) and those remaining are at risk due to sea level rise in the next century
345 (Thorne et al. 2018). Given the associated conservation challenges, active management of tidal
346 marsh endemics is warranted. By identifying environmental drivers of locally adapted
347 populations in these ecosystems, we can better understand how to preserve the full range of
348 endemic phenotypes represented in the region.

349 Fine-scale mapping of habitats which appear to be homogenous, such as salt marshes,
350 supported our hypothesis that the ranges of song sparrow subspecies in the San Francisco Bay
351 area varied in microclimate. While we expected variation to be high between salt marsh and
352 upland subspecies ranges, we also observed variation within both salt marsh and upland habitats,
353 consistent with the hypothesis that environmental heterogeneity can facilitate local adaptation at
354 fine spatial scales (Miller 1956, Mayr 1963, Ferrell 1966). Temperatures diverged most between
355 the niches of the two upland subspecies, with *gouldii* experiencing the coolest conditions on the
356 coast, and *heermanni* experiencing the warmest in inland habitats. Additionally, salinity varies
357 greatly between upland and salt marsh subspecies but also within the estuary, with the South San
358 Francisco Bay roughly 33 times saltier than the Suisun Bay, which receives large freshwater
359 inputs from the Sacramento and San Joaquin Rivers (Schrage and Cloern, 2017). Given ample
360 evidence of rapid adaptation to saline environments in other species of new world sparrows

361 (Walsh et al. 2019a, Walsh et al. 2019b, Benham et al. 2020), our discovery of 144 regions that
362 may be associated with an adaptive response to environmental variation is not surprising.
363 However, this is the first use of whole genome data to explore local adaptation among subspecies
364 of song sparrows over a microgeographic scale. Our detailed characterization of covariation in
365 the spatial distributions of habitat and genotype at fine geographic scales offers strong support for
366 the hypothesis that selection has contributed to the rapid diversification of locally adapted types
367 in this system. We acknowledge that despite advances in the approaches to identifying GxE
368 associations, there are several challenges with these methods (Hoban et al. 2016). Significant
369 GEAs can alternatively arise from both neutral population genetic and demographic processes
370 (Hoban et al. 2016). Moreover, linked selection via background selection or hitchhiking can
371 result in increased genomic divergence between populations, with loci correlating strongly with
372 environmental variables by chance (Cruickshank & Hahn 2014). To disentangle these processes,
373 our findings warrant future work with a broader spatial and environmental sampling scheme that
374 can help to develop a robust assessment of local adaptation in song sparrows. However, the lack
375 of strong neutral genetic structure among the subspecies studied here, coupled with the absence
376 of IBD based on multiple marker types (Chan & Arcese 2002, Mikles et al. 2020, this study)
377 offers support for our hypothesis of ecological divergence. Moreover, the identification of
378 outliers associated with candidate genes that have previously been linked to tidal marsh
379 adaptations including HSP90B1 (Wan et al. 2017) and PHF20 (Walsh et al. 2019a) provides
380 compelling candidates for future validation. Despite the above caveats, we feel that our work
381 identifies new and important signals of genetic diversity among these populations, which in turn
382 reflects evolutionary distinctiveness of populations that is potentially beneficial to the persistence
383 of these populations both locally and regionally.

384 *Evolutionary Distinctiveness and Conservation*

385 Our comparisons of putatively adaptive versus neutral loci suggest a pattern of local
386 adaptation and diversity in the song sparrows that highlights the utility of genomic data sets in
387 resolving population-level patterns of diversification. Our findings show that population
388 differentiation is the highest in San Francisco Bay area song sparrows at putatively adaptive loci.
389 Because all five subspecies were well-delineated over three RDA axes, our findings offer
390 empirical evidence of the value of identifying adaptive variation among populations (Crandall et

391 al. 2000, Fraser & Bernatchez 2001). However, despite broad recognition that assessing genetic
392 distinctiveness is a desirable first step in identifying the appropriate units for conservation (e.g.,
393 Funk et al. 2012, Volkmann et al. 2014), the appropriate metrics for assessing genome-wide
394 differentiation remains uncertain (Fernandez-Fournier 2021). Although relatively few empirical
395 studies have focused on the conservation of genomic diversity to date, a growing literature
396 describes the potentially complementary roles of adaptive and neutral processes in shaping
397 genomic variation within species and its incorporation in conservation planning (Bonin et al.
398 2007, Funk et al. 2012). Our results generally support these suggestions by elucidating a system
399 in which the inclusion of genomic data on traits putatively under selection revealed aspects of
400 diversity that could be overlooked given nuanced variation in the phenotypes of the subspecies
401 studied here.

402 More work is needed to determine how patterns of diversity and distinctiveness may
403 affect the adaptive capacity or evolutionary potential of song sparrow populations at micro-
404 spatial to regional scales. However, given operational definitions of *adaptive capacity* as the
405 ability a species to cope with environmental change (IPCC 2014), and *evolutionary potential* as
406 an attribute determining a species' ability to maintain positive long-term growth rates in novel
407 environmental conditions (Thurman et al. 2020), we suggest each will be maximized by
408 conserving units exhibiting novel variation at loci linked to traits underlying additive genetic
409 variance in individual fitness (Hendry et al. 2018). Specifically, our RDA identified several loci
410 putatively linked to salinity and climate, factors driving selection and local adaptation in a
411 variety of taxa (Kingsolver et al. 2012), consistent with the notion that microgeographic variation
412 in selection can increase divergence at local scales whilst reducing it within populations (Hendry
413 et al. 2018, Funk et al. 2019). In the case of song sparrows, isolation and small effective
414 population size (Mikles et al. 2020) might be expected to increase genetic drift among
415 populations adapted to their contemporary environment but compromise their capacity to
416 accommodate change in the future (Funk et al. 2019). However, because our prior results
417 indicate substantial evidence of contemporary gene flow between the five subspecies studied here
418 (Mikles et al. 2020), it is possible that sufficient admixture currently exists among populations to
419 maintain a capacity to respond to variation in natural selection and environment in future. Given
420 an imperfect understanding of these factors at present, we suggest managers prioritize the

421 conservation of evolutionarily significant units (ESUs) defined broadly, based on morphological,
422 genetic, and/or ecological boundaries likely to reflect underlying adaptive process and maintain
423 genetic variation in fitness (reviewed in Funk et al. 2012). Because phenotypic differentiation at
424 fine spatial scales and in response to ecological gradients known to affect individual fitness is a
425 defining trait of song sparrows in North America (e.g., Aldrich 1984, Arcese et al. 2002, Patten
426 & Pruett 2011), we suggest that conserving morphologically and genetically distinct subspecies
427 across the range is a first step towards conserving resilience and persistence in the species
428 overall. Genomic data, interpreted conservatively, can help reveal adaptive and neutral genetic
429 differentiation that can facilitate the prioritization of cryptic species or populations (Coates et al.,
430 2018). When integrated with diagnosable differences in phenotype, as is the case for San
431 Francisco Bay area song sparrows, whole genome data can help delineate populations with high
432 precision.

433 ***Revisiting the Song Sparrow Subspecies of the San Francisco Bay***

434 The five subspecies of song sparrow studied here meet the recommended criteria for
435 consideration as distinct populations (McCormack & Maley 2015) given that they were (1)
436 classified *a priori* by phenotype, and (2) shown to be differentiated at dozens of loci linked to
437 environmental heterogeneity. We now (3) know that they occupy different environmental niches
438 with respect to climate and salinity. Currently, four of these five subspecies (all but *gouldii*) are
439 listed as ‘species of special concern’ in California, but all song sparrows in California appear to
440 be declining (Sauer 2020). Our use of whole-genome surveys suggests that focusing on adaptive
441 variation can advance management planning in many widespread species with cryptic underlying
442 differences in genetic traits affecting fitness. Quantifying local evolutionary distinctiveness could
443 facilitate predictions on how the influence of climate change, genetic variation, and natural
444 selection may affect potential rates of local evolution (Garant 2020). Moreover, characterizing
445 local adaptive variation may play a role in informing assisted gene flow among these
446 populations, defined as the managed movement of individuals between populations to mitigate
447 local maladaptation (Kelly et al. 2021).

448 **References**

449 Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in
450 unrelated individuals. *Genome Research* 19:1655-1664.

- 451 Aldrich, J. W. 1984. Ecogeographical variation in size and proportions of Song Sparrows
452 (*Melospiza melodia*). Ornithological Monographs 35.
- 453 Anderson, J. T., D. W. Inouye, A. M. McKinney, R. I. Colautti, and T. Mitchell-Olds. 2012.
454 Phenotypic plasticity and adaptive evolution contribute to advancing flowering
455 phenology in response to climate change. *Proceedings of the Royal Society B* 279: 3843-
456 3852.
- 457 Arcese, P., Sogge, M. K., Marr, A. B., & Patten, M. A. (2002). Song sparrow (*Melospiza*
458 *melodia*), The Birds of North America Online. In Poole, A. (Ed.). Washington, DC: The
459 Academy of Natural Sciences, Philadelphia, Pennsylvania, and The American
460 Ornithologists' Union. Retrieved from the Birds of North America Online:
461 <http://bna.birds.cornell.edu/bna/species/704>
- 462 Barrowclough, G. F., J. Cracraft, J. Klicka, and R. M. Zink. 2016. How many kinds of birds are
463 there and why does it matter? *PLoS One* 11(11):e0166307.
- 464 Basham, M. P., and L. R. Mewaldt. 1987. Salt water tolerance and the distribution of South San
465 Francisco Bay Song Sparrows. *The Condor* 89(4):697-709.
- 466 Bay, R. A., N. H. Rose, C. A. Logan, and S. R. Palumbi. 2017. Genomic models predict
467 successful coral adaptation in future ocean warming rates are reduced. *Science Advances*
468 3(11):e1701413
- 469 Bay, R.A., Harrigan, R.J., Underwood, V.L., Gibbs, H.L., Smith, T.B. & Ruegg, K. 2018.
470 Genomic signals of selection predict climate-driven population declines in a
471 migratory bird. *Science*, 359, 83–86.
- 472 Benham, P. M., & Cheviron, Z. A. (2020). Population history and the selective landscape shape
473 patterns of osmoregulatory trait divergence in tidal marsh Savannah sparrows
474 (*Passerculus sandwichensis*). *Evolution*, 74(1), 57–72.
- 475 Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and
476 powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1):289-
477 300.
- 478 Bonin, A., F. Nicole, F. Pompanon, C. Miaud, P. Taberlet. 2007. Population adaptive index: a
479 new method to help measure intraspecific genetic diversity and prioritize populations for
480 conservation. *Conservation Biology* 21(3):697-708.
- 481 Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
482 <https://doi.org/10.1023/A:1010933404324>
- 483 California Department of Fish and Wildlife (CDFW). 2019. Area of Conservation Emphasis
484 (ACE) II Saline Wetlands by Watershed [GIS data]. CDFW Biogeographic Data Branch.
485 (<https://wildlife.ca.gov/Data/Analysis/ACE>). Accessed December 23, 2020.
- 486 Carvalho, C. S., et al. 2020. Combining genotype, phenotype, and environmental data to
487 delineate site-adjusted provenance strategies for ecological restoration. *Molecular Ecology*
488 *Resources* 21:44-58.
- 489 Chan, Y., and P. Arcese. 2002. Subspecific Differentiation and Conservation of Song Sparrows
490 (*Melospiza melodia*) in the San Francisco Bay Region Inferred by Microsatellite Loci
491 Analysis. *The Auk*. <https://doi.org/10.2307/4089964>
- 492 Chan, Y., and P. Arcese. 2003. Morphological and microsatellite differentiation in *Melospiza*
493 *melodia* (Aves) at a microgeographic scale. *Journal of Evolutionary Biology* 16(5):939-47.

494 Clucas, G. V., R. N. Lou, N. O. Therkildsen, and A. I. Kovach. 2019. Novel signals of adaptive
495 genetic variation in northwestern Atlantic cod revealed by whole-genome
496 sequencing. *Evolutionary Applications* 12:1971-1987.

497 Coates, D. J., M. Byrne, and C. Mortiz. 2018. Genetic diversity and conservation units: dealing
498 with the species-population continuum in the age of genomics. *Frontiers in Ecology and*
499 *Evolution* 6:165.

500 Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace, and R. K. Wayne. 2000. Considering
501 evolutionary processes in conservation biology. *TREE* 15(7): 290-295.

502 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of
503 speciation are due to reduced diversity, not reduced gene flow. *Molecular*
504 *ecology*, 23(13), 3133-3157.

505 Danecek, P., A., et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156-
506 2158.

507 Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and
508 prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*,
509 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>

510 ESRI. (2019). ArcMap 10.7.1. In *ESRI*.

511 Feng, S., et al. (2020). Dense sampling of bird diversity increases power of comparative
512 genomics. *Nature*, 587(7833), 252-257.

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

515 Fernandez-Fournier, P., J. M. M. Lewthwaite, and A. Ø. Mooers. 2021. Do we need to identify
516 adaptive genetic variation when prioritizing populations for conservation? *Conservation*
517 *Genetics* 22:205-216.

518 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces
519 for global land areas. *International Journal of Climatology*, 37, 4302–4315.
520 <https://doi.org/10.1002/joc.5086>

521 Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2018. Comparing methods for
522 detecting multilocus adaptation with multivariate genotype-environment associations.
523 *Molecular Ecology* 27(9):2215-2233.

524 François, O., H. Martins, K. Caye, S. D. Schoville. 2016. Controlling false discoveries in
525 genome scans for selection. *Molecular Ecology* 25(2):454-469.

526 Fraser, D. J., and L. Bernatchez. 2001. Adaptive evolutionary conservation: towards a unified
527 concept for defining conservation units. *Molecular Ecology* 10(12):2741-52.

528 Frichot, E., S. D. Schoville, G. Bouchard, and O. François. 2013. Testing for associations
529 between loci and environmental gradients using latent factor mixed models. *Molecular*
530 *Biology and Evolution* 30(7):1687-1699.

531 Fry, A. J., and R. M. Zink. 1998. Geographic analysis of nucleotide diversity and song sparrow
532 (Aves:Emberizidae) population history. *Molecular Ecology* 7(10):1303-13.

533 Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing genomics
534 for delineating conservation units. *Trends in Ecology and Evolution* 27 (9): 489-496.

535 Funk, W. C., B. R. Forester, S. J. Converse, C. Darst, and S. Morey. 2019. Improving
536 conservation policy with genomics: a guide to integrating adaptive potential into U.S.
537 Endangered Species Act decisions for conservation practitioners and geneticists.
538 *Conservation Genetics* 20:115-134.

539 Garant, D. 2020. Natural and human-induced environmental changes and their effects on
540 adaptive potential of wild animal populations. *Evolutionary Applications* 13:1117-1127.

541 Greenberg, R., J. E. Maldonado, S. Droege, and M. V. McDonald. 2006. Tidal marshes: a global
542 perspective on the evolution and conservation of their terrestrial vertebrates. *BioScience*
543 56(8):675-685.

544 Grinnell, J. 1909. Three new Song Sparrows from California. *University of California*
545 *Publications in Zoology* 5, 265–269.

546 Grinnell, J., & Miller, A. H. 1944. *The distribution of the birds of California*. Artemesia Press.

547 Haig, S. M., and K. Winker. 2010. Avian Subspecies: Summary and Prospectus. *Ornithological*
548 *Monographs* 67:172-175.

549 Henderson, E. C., and A. Brelsford. 2020. Genomic differentiation across the speciation
550 continuum in three hummingbird species pairs. *BMC Evolutionary Biology* 20:113.

551 Hendry, A. P., D. J. Schoen, M. E. Wolak, and J. M. Reid. 2018. The contemporary evolution of
552 fitness. *Annual Review of Ecology, Evolution, and Systematics* 49:457-476.

553 Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., ... &
554 Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls,
555 practical solutions, and future directions. *The American Naturalist*, 188(4), 379-
556 397.

557 Hudson, D. H., D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies.
558 *Molecular Biology and Evolution* 23(2):254-267.

559 Huxley, J. (1942). *Evolution, the Modern Synthesis*. Harper and Brothers.

560 Johnston, R. F. (1956a). Population structure in salt marsh song sparrows. Part I. Environment
561 and annual cycle. *Condor*, 58, 24–44.

562 Johnston, R. F. (1956b). Population structure in salt marsh Song Sparrows. Part II: Density, age
563 structure, and maintenance. *Condor*, 58, 254–271.

564 Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A. M., ... Ellegren, H.
565 (2014). Estimation of linkage disequilibrium and interspecific gene flow in
566 *Ficedula* flycatchers by a newly developed 50k single-nucleotide polymorphism
567 array. *Molecular Ecology Resources*, 14(6), 1248–1260.

568 Kelly, E., Kenbi Traditional Owners and Rangers, C. J. Jolly, N. Indigo, A. Smart, J. Webb, and
569 B. Phillips. 2021. No outbreeding depression in a trial of targeted gene flow in an
570 endangered Australian marsupial. *Conservation Genetics* 22:23-33.

571 Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., Carlson, S. M. (2012). Synthetic analyses of
572 phenotypic selection in natural populations: lessons, limitations and future directions.
573 *Evolutionary Ecology* *Evolutionary Ecology* 26(5):1101-1118.

574 Kuhn, M. (2020). Caret: Classification and Regression Training. *R Package Version 6.0-79*,
575 *Version 6.0-86*.

576 Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. In *Applied Predictive Modeling*.
577 <https://doi.org/10.1007/978-1-4614-6849-3>

578 Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler
579 transform. *Bioinformatics* 25(14):1753-60.

580 Li, H., B. Handsaker, A. Wyosoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
581 Durbin, and the 1000 Genome Project Data Processing Subgroup. 2009. The sequence
582 alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-9.

583 Liaw, A., & Wiener, M. (2002). Classification and Regression with Random Forest. *R News*, 2,
584 18–22.

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

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

589 Mayr, E. (1942) Systematics and the Origin of Species (Columbia Univ. Press, New York).

590 Mayr, E. (1963). *Animal species and evolution*. Harvard University Press.

591 McCormack, J. E., and J. M. Maley. 2015. Interpreting negative results with taxonomic and
592 conservation implications: another look at the distinctness of coastal California
593 gnatcatchers. *The Auk* 132(2):380-388.

594 McKenna, A., et al. 2010. *Genome Research* 20(9):1297-1303.

595 Mikles, C. S., S. M. Aguillon, Y. L. Chan, P. Arcese, P. M. Benham, I. J. Lovette, and J. Walsh.
596 2020. Genomic differentiation and local adaptation on a microgeographic scale in a
597 resident songbird. *Molecular Ecology* 29:4295-4307.

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

600 Oh, K. P., Aldridge, C. L., Forbey, J. S., Dadabay, C. Y., & Oyler-McCance, S. J. (2019).
601 Conservation genomics in the Sagebrush Sea: population divergence,
602 demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). *Genome*
603 *biology and evolution*, 11(7), 2023-2034.

604 Oksanen, F. J., et al. 2017. Vegan: community ecology package. R package version 2.4-3.

605 Okonechnikov, K. A. Conesa, F. García-Alcalde. 2016. Qualimap 2: advanced multi-sample
606 quality control for high-throughput sequencing data. *Bioinformatics* 32(2):292-294.

607 Patten, M. A., Rotenberry, J. T., & Zuk, M. (2004). Habitat selection, acoustic adaptation, and
608 the evolution of reproductive isolation. *Evolution*. [https://doi.org/10.1111/j.0014-](https://doi.org/10.1111/j.0014-3820.2004.tb01593.x)
609 [3820.2004.tb01593.x](https://doi.org/10.1111/j.0014-3820.2004.tb01593.x)

610 Patten, M. A., & Pruett, C. L. (2009). The Song Sparrow, *Melospiza melodia*, as a ring species:
611 Patterns of geographic variation, a revision of subspecies, and implications for speciation.
612 *Systematics and Biodiversity*, 7(1), 33–62. <https://doi.org/10.1017/S1477200008002867>

613 R Core Team. (2020). R: A language and environment for statistical computing. In *R: A*
614 *language and environment for statistical computing. R Foundation for Statistical*
615 *Computing, Vienna, Austria*.

616 Sauer, J. R., D. K. Niven, J. E. Hines, D. J. Ziolkowski, Jr, K. L. Pardieck, J. E. Fallon, and W.
617 A. Link. 2017. The North American Breeding Bird Survey, Results and Analysis 1966 -
618 2015. Version 2.07.2017 USGS Patuxent Wildlife Research Center, Laurel, MD.

619 San Francisco Estuary Project (SFEP) (1991) San Francisco Estuary Project status and trends
620 report on wetlands and related habitats in the San Francisco Estuary. ABAG Public report
621 to US-EPA. Oakland, CA: San Francisco Estuary Project

622 Schliep, K. P. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4):592-3.

623 Schmidt, T. L., M. Jasper, A. R. Weeks, and A. A. Hoffmann. 2021. Unbiased population
624 heterozygosity estimates from genome-wide sequence data. *Methods in Ecology and*
625 *Evolution* 12(10):1888-1898.

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

628 Takekawa, J. Y., et al. 2006. Environmental threats to tidal-marsh vertebrates of the San
629 Francisco Bay Estuary. *Studies in Avian Biology* 32:176-197.

630 Thorne, K. G., et al. 2018. U.S. Pacific coastal wetland resilience and vulnerability to sea-level
631 rise. *Science Advances* 4(2): eaao3270.

632 Thurman, L. L., et al. 2020. Persist in place or shift in space? Evaluating the adaptive capacity of
633 species to climate change. *Front Ecol Environ* 18(9): 520-528.

634 Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, R. Holderegger. 2015. A practical guide to
635 environmental association analysis in landscape genomics. *Molecular Ecology* 24(17):4348-
636 4370.

637 Volkmann, L., I. Martyn, V. Moulton, A. Spillner, A. O. Mooers. 2014. Prioritizing populations
638 for conservation using phylogenetic networks. *PLoS ONE* 9(2):e88945.

639 Walsh, J., Kovach, A. I., Olsen, B. J., Shriver, W. G., & Lovette, I. J. (2018). Bidirectional
640 adaptive introgression between two ecologically divergent sparrow
641 species. *Evolution*, 72(10), 2076-2089.

642 Walsh, J., et al. 2019a. Genomics of rapid ecological divergence and parallel adaptation in four
643 tidal marsh sparrows. *Evolution Letters*, 3(4), 324–338.

644 Walsh, J., Clucas, G., MacManes, M., Thomas, K., & Kovach, A. (2019b). Divergent selection
645 and drift shape the genomes of two avian sister species spanning a saline–freshwater
646 ecotone. *Ecology and Evolution*, 9(23), 13477-13494.

647 Wan, Y., Ma, C., Wei, P., Fang, Q., Guo, X., Zhou, B., & Jiang, R. (2017). Dynamic expression
648 of HSP90B1 mRNA in the hypothalamus of two Chinese chicken breeds under
649 heat stress and association analysis with a SNP in Huainan chickens. *Czech Journal of*
650 *Animal Science*, 62(2), 82-87.

651 Winker, K. 2021. An overview of speciation and species limits in birds. *Ornithology*. 138(2),
652 ukab006, <https://doi.org/10.1093/ornithology/ukab006>

653 Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir. A high-performance
654 computing toolset for relatedness and principal component analysis of SNP data.
655 *Bioinformatics* 2012; doi: 10.1093/bioinformatics/bts606

656 Zink, R. M. 2004. The role of subspecies in obscuring avian biological diversity and misleading
657 conservation policy. *Proceedings of the Royal Society B*. 271:561-564.

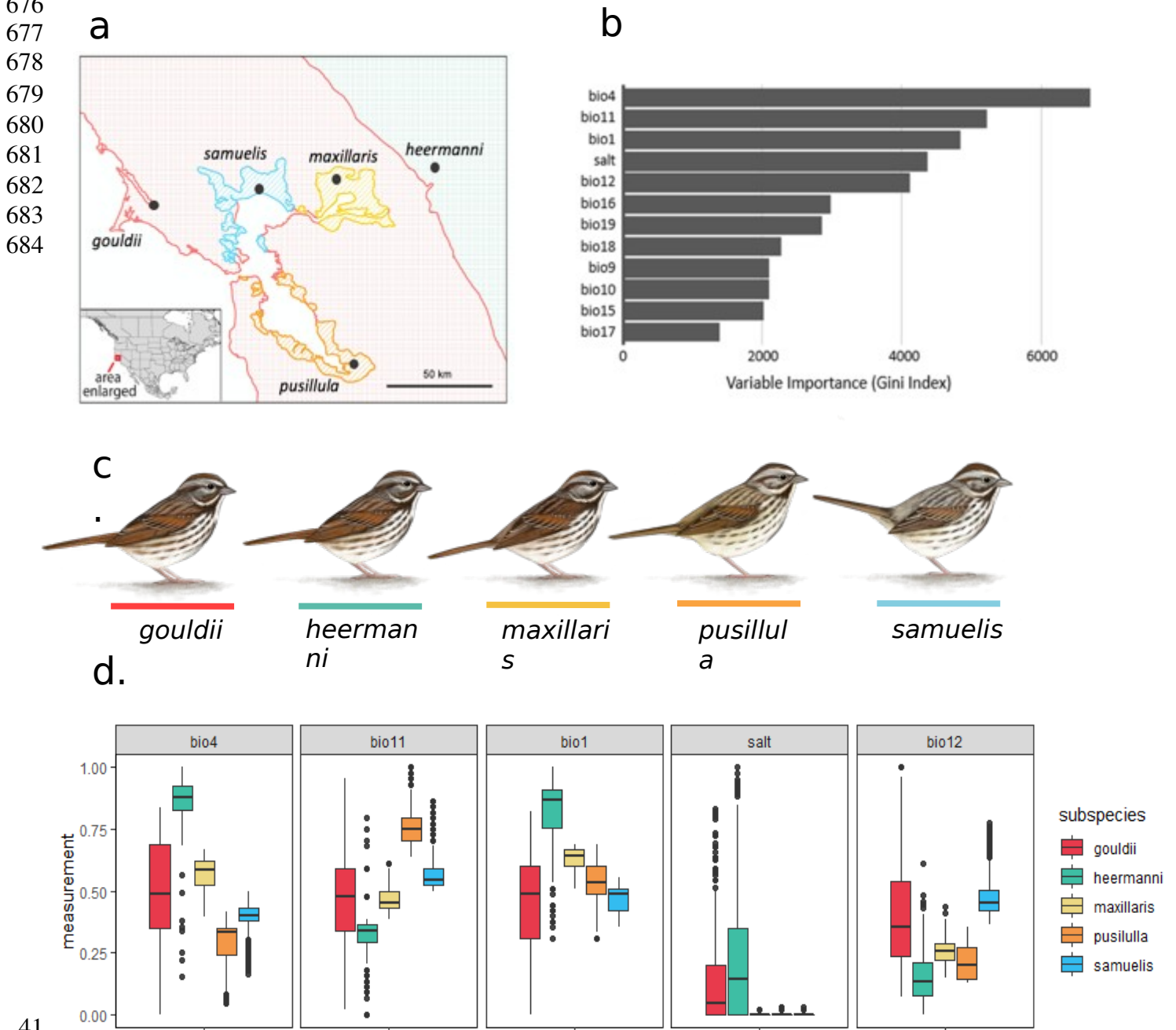
658 Zink R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian
659 phylogeography. *Mol Ecol*. 17:2107–2121.

660 Zink, R. M. 2010. Drawbacks with the use of microsatellites in phylogeography: the song
661 sparrow *Melospiza melodia* as a case study. *Journal of Avian Biology* 1:1-7.

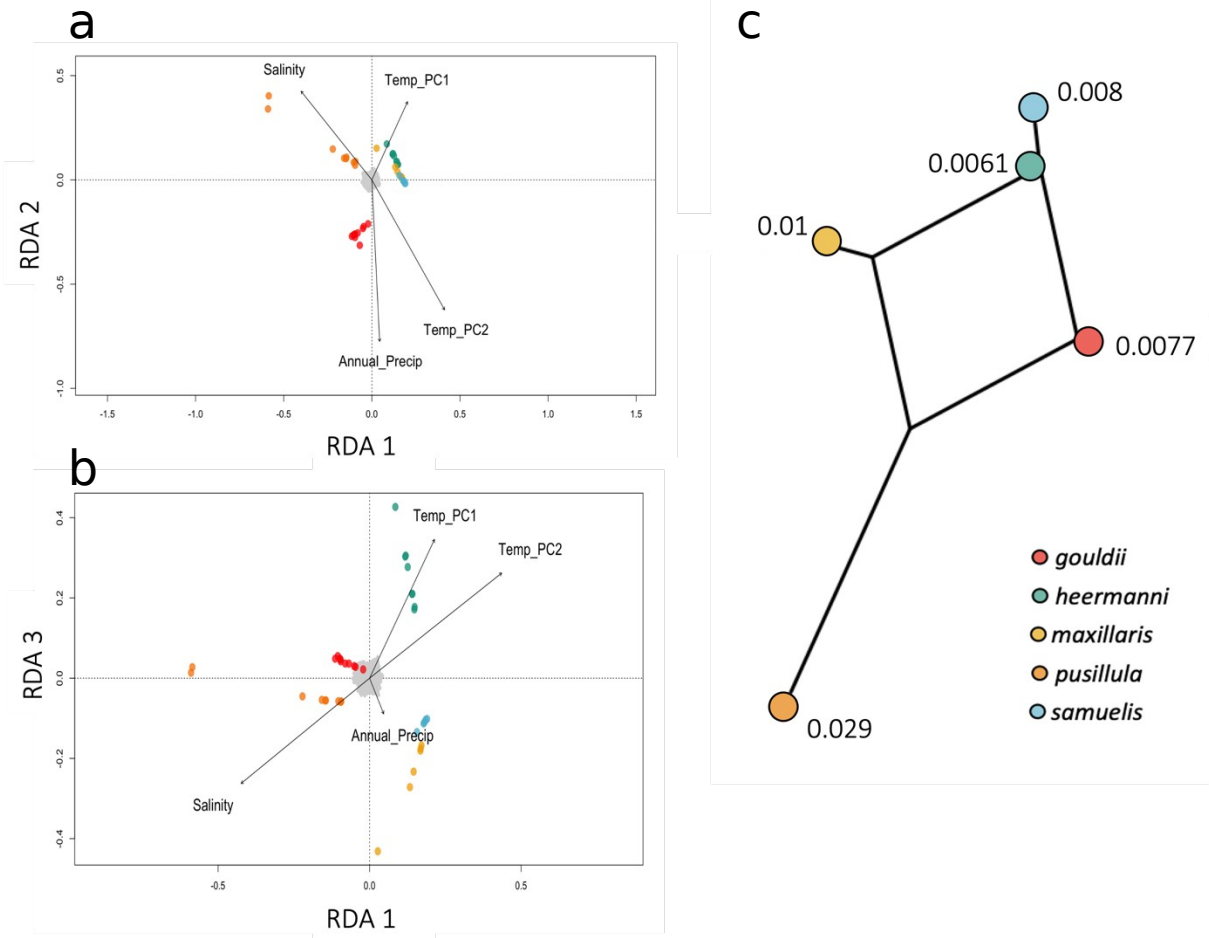
662

663

664 **Figure 1.** (A). San Francisco Bay song sparrow ranges digitized. Black dots indicate the sampling locations from
 665 the five representative populations and the colors represent the ranges for each subspecies in the Bay (*gouldii* in red,
 666 *heermanni* in green, *samuelis* in blue, *maxillaris* in yellow, and *pusillula* in orange. (B). Variable importance output
 667 ordered by their importance estimated by the RF model. Bio4=Temperature Seasonality, Bio11= Mean Temperature
 668 of the Coldest Quarter, Bio1=Annual Mean Temperature, salt= Salinity, bio12= Annual Precipitation, Bio16=
 669 Precipitation of Wettest Quarter, Bio19= Precipitation of Coldest Quarter, Bio18= Precipitation of Warmest Quarter,
 670 Bio9= Mean Temperature of Driest Quarter, Bio10= Mean Temperature of Warmest Quarter, Bio15= Precipitation
 671 Seasonality, Bio17= Precipitation of Driest Quarter. (C). Song sparrow subspecies illustrations by Jillian Ditner
 672 demonstrate subtle morphological differences among the subspecies. (D). Variation in the top 5 most important
 673 variables from RF model by subspecies. Variables were normalized to bring values to range from 0-1. Dark
 674 horizontal lines represent the median, colored boxes show the interquartile range, whiskers indicate the 5th – 95th
 675 percentile, and dots represent outliers.



685 **Figure 2:** Subspecies delineation and evolutionary distinctiveness in San Francisco Bay song sparrows. (A) RDA
 686 axes 1 vs 2 and (B) RDA axes 1 vs 3 support increased divergence among subspecies driven by environmental
 687 factors. (C) NeighborNet depicting the relationship among song sparrows. Circle colors correspond to subspecies.
 688 Values associated with each tip represent SH metrics for each subspecies, indicating prioritization for management.
 689 Legend colors correspond to all panels in this figure.



690
 691