Title: Spatial variation in population genomic responses to over a century of anthropogenic change within a tidal marsh songbird

Authors: Phred M. Benham[1], Jennifer Walsh[2], Rauri C. K. Bowie[1]

Author affiliations:

(1) Museum of Vertebrate Zoology & Department of Integrative Biology, University of California, Berkeley, Berkeley, CA

(2) Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, Ithaca, NY.

Corresponding author: Phred M. Benham; 3101 Valley Life Sciences Building, Berkeley, CA 94720-3160; 603-443-2744; phbenham@gmail.com

KEYWORDS: conservation genomics; tidal marshes; Passerellidae; anthropogenic change; museum collections, exome capture

ABSTRACT: Combating the current biodiversity crisis requires the accurate documentation of population responses to human-induced ecological change. To this end, museum collections preserve a record of population responses to anthropogenic change that can provide critical baseline data on patterns of genetic diversity, connectivity, and population structure. We leveraged spatially-replicated time series of specimens to document population genomic responses to the destruction of nearly 90% of coastal habitats occupied by the Savannah sparrow (*Passerculus sandwichensis*) in California. Spatial-temporal analyses of genetic diversity from 219 sparrows collected between 1889-2017 showed that the amount of habitat lost was not predictive of genetic diversity loss. Despite experiencing the greatest levels of habitat loss, we found that genetic diversity in the San Francisco Bay Area remained relatively high. Over the past century, immigration into the Bay Area from interior populations has also increased. This may have minimized genetic diversity declines, but likely led to the erosion of divergence at loci associated with tidal marsh adaptation. Tracing the genomic trajectories of multiple populations over time provided unique insights into how shifting patterns of gene flow through time in response to human-induced habitat loss may contribute to negative fitness consequences.

INTRODUCTION: Habitat loss is a primary driver of biodiversity loss [1]. At the genetic level, habitat-induced diversity loss can exacerbate extinction risk through increased mutational load and a reduced capacity to adapt to future environmental change [2-3]. Consequently, the preservation of genetic diversity is now recognized as a major priority for the conservation of global diversity [4]. A predictable relationship between habitat and genetic diversity loss exists that can inform how much of a species' range must be protected to maintain target levels of genetic diversity [5]. However, across a species' distribution local adaptation, population size, and gene flow will

influence the rate and kinds (e.g. neutral vs. locally adapted variants) of genetic diversity being lost in response to habitat loss and degradation. For instance, habitat deterioration can turn populations into sinks with increased immigration from a source population needed to sustain the population. Changes in rate of gene flow could lead to the evolutionary rescue of a small population [6] or could lead to outbreeding depression if immigrants from source populations are maladapted [7]. Despite mixed theoretical and empirical support for these different outcomes, a detailed understanding of how interactions among different ecological and demographic processes shape patterns of genetic variation in species threatened by habitat loss is lacking. Addressing this gap will be critical for guiding habitat protection and restoration efforts to conserve target levels of genetic diversity.

In this case, looking to past population genetic responses to habitat loss will provide critical insights into the impacts of habitat degradation on genetic diversity. Natural history collections typically preserve specimens collected from the past 150-200 years, a period marked by dramatic human-induced landscape transformation, that can be used to reconstruct population genomic responses to past habitat loss [8-11]. In particular, with the increased ease of generating genomic-scale data from historic DNA sources, analyzing replicated time-series of specimens from across a species distribution will provide important case-studies documenting the role of habitat deterioration in disrupting migration, local adaptation, and drift among populations. To assess how human-induced changes in these meta-population dynamics may impact temporal patterns of genetic diversity, we leverage a spatially replicated time-series of Savannah sparrow (*Passerculus sandwichensis*) specimens that were densely sampled from throughout the state of California over the past 128 years.

The Savannah sparrow is a widespread North American songbird, with 17 subspecies breeding across a range of open habitats from Alaska to Guatemala [12]. Within California, four subspecies span a landscape that has experienced dramatic, but spatially variable, transformations due to human activity (Fig. 1a). The coastal subspecies *P. s. alaudinus* and *P. s. beldingi* primarily occupy tidal marsh habitats in northern and southern California, respectively. These coastal populations exhibit local adaptation to high salinity and flooding in tidal marshes with prior work documenting physiological, behavioral, and genomic divergence from other inland, freshwater-associated California populations of the species [13-17]. Coastal populations have experienced nearly 90% human-induced habitat loss in certain estuaries [18-19] and are now listed as either state endangered species (*P. s. beldingi;* [20]) or bird species of special concern (*P. s. alaudinus;* [21]). In contrast, *P. s. nevadensis* and *P. s. brooksi* of northern and eastern California are of least conservation concern with habitats remaining relatively intact.

The contrasting histories of habitat loss experienced by different sparrow populations provide a natural experiment for exploring the temporal dynamics of genetic diversity change in response to human activities. We sequenced genomic data from 219 individuals sampled from 1889-2017, representing nine spatially replicated time-series and all four subspecies. With this dataset, we asked: (1) how has habitat loss impacted patterns of population structure and gene flow among populations? (2) Does habitat loss predict genetic diversity loss? And (3) how have temporal changes in genetic diversity and gene flow among populations impacted loci associated with local adaptation to different habitats? Addressing these questions in concert with time-series data provides a novel framework for evaluating how human-induced ecological change can disrupt historic patterns of local adaptation, gene flow, and population size.

87 **RESULTS:**

88 *Spatial and temporal inference of population structure*: We sequenced 145 historical (sampled
89 pre-1970) and 74 modern (post-2000) Savannah sparrow specimens using an exome capture
90 approach (see supplemental results, Fig. S1-S3, and Table S1 for capture performance). Although
91 whole genome sequencing of degraded DNA is increasingly feasible [22-23], target capture
92 approaches have proven to be a highly targeted and cost-effective tool for analyzing temporal
93 genetic change [24-25]. Principial components analysis of all 219 individuals (60,765 SNPs)
94 confirmed that modern and historical samples collected from the same locality clustered together
95 (Fig. 1b). The first principal component explained 2.8% of the variance in the data and separated
96 out three clusters: (1) *Passerculus sandwichensis beldingi* subspecies from south coastal
97 California; (2) birds from Morro Bay in San Luis Obispo County (*P. s. alaudinus*); and (3) birds
98 from the rest of California (*P. s. alaudinus, brooksi,* and *nevadensis*). The second principal
99 component explained 1.1% of the variance in the data and divided birds from the *P. s. nevadensis*
100 subspecies in eastern California, from birds in northwest California (*P. s. alaudinus* and *P. s.*
101 *brooksi*), and the remaining birds in the *P. s. alaudinus* subspecies from the Bay Area and central
102 California coast. The third principal component (0.9% of variance) separated modern from
103 historical samples and the fourth component (0.8% of variance) split northwest California birds
104 from other populations. PCA of other datasets (downsampled coverage, CpG sites removed, C->T
105 and G->A sites removed) revealed similar patterns of population clustering (supplemental Figure
106 S4).

107 To further test for temporal changes in population structure and admixture among
108 populations, we analyzed a dataset of 65,583 unlinked SNPs and 239 individuals (an additional 20
109 from outside California) within the program DyStruct [26]. DyStruct resembles other model-based

clustering programs, but explicitly accounts for the temporal dynamics of allele frequency change due to genetic drift in time-series data. The highest likelihoods were for K=3 to 5 (log-likelihood= -1383.63 to -1383.76; Fig. 1c). All three values of K identified birds from south coastal California and Mexico as a distinct cluster, sparrows from the northern California coast as a second genetic cluster, and birds of eastern California, Wyoming, Alaska, and Washington state forming a third cluster. The main difference between K=3 and 4 was that Morro Bay birds were recognized as a distinct cluster, and at K=5 modern birds from the northern California coast were identified as a distinct cluster. This analysis also revealed a key finding: temporal variation in population structure, with modern birds from the central and northern California coast exhibiting a 17% (Morro Bay) to 64% (Humboldt Bay) increase in the proportion of ancestry shared with inland eastern California birds. These temporal changes point to a recent increase in the level of gene flow from inland eastern California populations into coastal populations.

Spatial and temporal patterns of genetic diversity: We next explored how variation in the scale of human-induced habitat loss influenced patterns of genetic diversity across sparrow populations. First, effective diversity surfaces estimated in EEMS [27] show similar patterns of genetic diversity among populations in both the historical (<1940) and modern (>2000) datasets (Fig. 2a & 2b) with southern California birds consistently exhibiting lower effective diversity rates across time relative to northern California birds. Second, the southern California birds exhibited significantly lower levels of nucleotide diversity (π) and Watterson's theta (θ) compared to northern birds (Fig. 2c; historic one-way ANOVA analysis [π: F=58.2, df=7, p<<0.0001; θ: F=252.9, df=7, p<<0.0001]; modern one-way ANOVA analysis [π: F=101.9, df=7, p<<0.0001; θ: F=301.7, df=7, p<<0.0001]).

133 California's coastal ecosystems experienced more dramatic habitat loss over the past
134 century [18-19; 21] relative to localities in eastern and northern California. Consistent with these
135 differences in landscape change, we found significant increases in π and θ for eastern and northern
136 California populations, but evidence for significant declines in genetic diversity for five out of six
137 tidal marsh populations in coastal California (Fig. 2c). Suisun Bay in the Bay-Delta region of
138 central California was the only coastal population that exhibited significant increases in genetic
139 diversity. Declines in genetic diversity were more evident in θ as opposed to π, which contributed
140 to a significant increase in Tajima's D for all tidal marsh populations (except Suisun Bay). This
141 pattern is indicative of a population undergoing a contraction where rarer variants (as measured by
142 θ) are eliminated before overall declines in heterozygosity (measured by π). Estimates of these
143 parameters using datasets that were less conservatively filtered showed similar patterns
144 (supplemental Fig. S5). For Newport Bay and San Francisco Bay trends in genetic diversity
145 metrics were also evaluated across samples taken from three and six time points, respectively. For
146 Newport Bay birds, significant decreases in π and θ only occurred between the 1960s and 2010s
147 (supplemental Fig. S6). San Francisco Bay on the other hand showed a long-term increase in
148 Tajima's D from the early 1900s to the present, driven by decreasing θ through time (supplemental
149 Fig. S7).

150 We also noted significant differences in the magnitude of genetic diversity decline among
151 tidal marsh populations (one-way ANOVA analysis; π: $F=294.7$, $df=5$, $p<<0.0001$; θ: $F=61.15$,
152 $df=5$, $p<<0.0001$). To determine whether differences in the amount of habitat loss experienced by
153 different tidal marsh populations explained this variation in genetic diversity loss, we compared
154 estimates of percent marsh habitat lost (data from [19]) for each estuary with temporal change in
155 genetic diversity. The amount of tidal marsh loss varied from 2-87%, but this variation did not

156 significantly explain temporal change in π or θ based on simple linear regression models (Fig. 2d).
157 We further tested whether total amount of tidal marsh area (log-transformed) in historic or modern
158 times predicted change in genetic diversity across sites, or whether contemporary extent of tidal
159 marsh explained variation in genetic diversity. All of these relationships were non-significant
160 (Supplemental Table S2). The only significant predictor of modern levels of genetic diversity was
161 the historic area of tidal marsh, which significantly predicted variation in contemporary θ
162 (R^2=0.79; p=0.01) and the relationship with π was nearly significant (R^2=0.53; p=0.06; Fig. 2e).
163 These results indicate that despite widespread habitat loss across all coastal marshes, the
164 populations from the historically largest marshes continue to maintain the highest levels of genetic
165 diversity. This is true even though populations from some of the most expansive marshes (e.g.,
166 San Francisco Bay) experienced the greatest habitat loss and decline in θ, but even in modern
167 times maintain levels of θ comparable to other northern California populations.
168
169 *Temporal change in demography:* To quantify the extent of temporal change in genetic diversity
170 and migration rates among populations we next fit a series of demographic models to the three-
171 dimensional site frequency spectrum of eastern California (*P. s. nevadensis*), Bay Area (*P. s.*
172 *alaudinus*), and Newport Bay (*P. s. beldingi*) birds in GADMA2 [28]. We ran 20 independent runs
173 of GADMA2 for both a historic and modern dataset. For historic birds, the best-fit model (Fig. 3a;
174 Table1; Log-likelihood -1391.83) showed an initial divergence time between Newport Bay and
175 the northern California populations ~254 kya, followed by divergence between the Bay Area and
176 Eastern California populations ~32 kya. Following the initial split, birds from southern California
177 (Newport Bay) grew from an effective population size (Ne) of 0.2 to 16,749 before contracting
178 again to 6,675. In contrast, the northern California population remained large with an Ne of

297,826 before splitting into an eastern California population that maintained a constant Ne of 68,438 and a Bay Area population that declined to an Ne of 16,755. Migration rates were roughly symmetrical from eastern California to the Bay Area (1.6e-04) and vice versa (1.3e-04), which translates to a rate of 8.9 migrants per generation (Nm) from the Bay Area to eastern California and Nm=2.7 in the reverse direction. Overall, the top five of 20 models fit to the historic dataset show a similar demographic history and parameter estimates (Table 1).

The best-fit model for the modern dataset (Fig. 3a; Table1; Log-likelihood -1556.50) showed a similar initial divergence time of ~232 kya, but a much earlier divergence time between eastern California and the Bay Area (~125 kya). Newport Bay birds showed a similar history of expansion and decline to a smaller Ne of 4,923. In contrast to the historic dataset, eastern California showed a smaller Ne (31,638) than Bay Area birds (57,309). This difference was consistent across 4 of the 5 top models estimated for the modern dataset. Migration rates were also inferred to be symmetrical between eastern California and the bay area (1.0e-04 and 1.1e-04) with a slightly greater Nm per generation (5.73) from eastern California to the Bay Area than in the opposite direction (Nm=3.48), a reversal from the historic dataset. Unlike the historic dataset, the top 5 models showed extensive variation in a number of parameter estimates, notably related to divergence time between eastern California and the Bay Area (range 3.6 kya to 125 kya), Bay Area Ne (37,157 to 426,253), and Nm from eastern California to the Bay Area (0 to 63.94).

The reasons for the increased variation among models in the modern dataset was unclear, however, greater variation could relate to recent and unaccounted for changes in demographic history due to population declines or shifts in migration rate that were not captured by the demographic model. The latter possibility was previously suggested by DyStruct analysis where a significant increase in eastern California ancestry was observed in modern day Bay Area

202 populations (Fig. 1c). To test this, we performed a series of coalescent simulations in msprime
203 1.2.0 [29] parameterized by the best-fit historic demographic model. Analysis of 26 simulated SNP
204 datasets in DyStruct showed that a shift to a migration rate of 0.15 from eastern California to the
205 Bay Area 5-25 generations ago would be necessary to match observed proportions of eastern
206 California ancestry in Bay Area birds (Fig. 3cd). Simulation results provide strong support for
207 observed DyStruct patterns being the result of significant changes in the rate of gene flow from
208 eastern California to the Bay Area as opposed to other processes such as genetic drift.
209
210 *The influence of demographic change on locally adapted loci:* Finally, we aimed to understand
211 how the inferred changes in gene flow between tidal marsh and freshwater adapted (eastern
212 California) populations of the Savannah sparrow may have impacted divergence in loci putatively
213 associated with local adaptation to tidal marsh habitats. We first compared genome wide mean Fst
214 divergence from eastern California in the four tidal marsh populations (Humboldt Bay, Bay Area,
215 Morro Bay, Newport Bay) for historic and modern datasets. Modern values of Fst divergence
216 showed increases in Morro Bay from 0.054 to 0.062, but a sharp decrease in mean Fst at Humboldt
217 Bay from 0.014 to -0.002 and more modest changes in the other two populations (Table 2). We
218 next estimated Fst and Dxy in sliding windows to identify loci exhibiting evidence for divergent
219 selection between eastern California populations and each of the four tidal marsh populations in
220 the historic dataset. Additionally, we identified SNPs that varied significantly in association with
221 salinity variation across populations using latent factor mixed models (LFMM [30]). Across the
222 four populations, we identified 67 outlier windows in the historic dataset that were in the 95th
223 percentile of Fst, Dxy, and LFMM analyses (Fig. 4a).

To assess whether these outlier loci experienced more or less change in Fst relative to non-outlier loci we compared patterns of ΔFst among tidal marsh populations. ΔFst measures the difference in Fst divergence from eastern California between the modern and historic datasets with negative values signifying decreasing divergence through time. This analysis revealed that all four tidal marsh population have experienced significantly greater declines in Fst divergence within outlier loci relative to non-outlier loci (Fig. 4b). Observed patterns of ΔFst in the Bay Area and Newport Bay significantly exceeded the declines predicted by simulations of a neutral demographic history with no demographic changes over the past 100 years (Supplemental Fig. S8); however, ΔFst for Bay Area outlier loci remained significantly greater than ΔFst estimated from a simulated history of recent migration rate change (Supplemental Fig. S8).

No genes were found to be outliers in all four coastal populations, but two genes were outliers in three populations (Fig. 4a; supplemental Table S2). This includes the estrogen receptor 1 gene (ESR1), which was an outlier in all populations except Humboldt Bay. Moreover, a T->C mutation in the 3' UTR region in the final exon of ESR1 exhibited the greatest Fst divergence in both Newport Bay (Fig. 4c; Fst=0.73) and the Bay Area (Fig. 4d; Fst=0.41). This SNP exhibited large temporal declines in the Bay Area (ΔFst=-0.17), which was likely due to decreasing frequencies of the derived C allele of this SNP in the Bay Area (change in frequency -0.24; Fig. 4e) as opposed to allele frequency change in the ancestral T allele that occurs at higher frequency in eastern California. Again patterns at the ESR1 locus are consistent with temporal changes in gene flow eroding divergence at potentially locally adapted loci.

Increased gene flow from inland eastern California populations into northern coastal populations should also lead to the increased homogenization of allele frequencies among populations towards the present. Further, if outlier loci associated with local adaptation to tidal

247 marshes exhibit increased resilience to gene flow then we would expect less homogenization of
248 allele frequencies at these loci relative to non-outlier loci. We tested these predictions by
249 estimating the degree of correlation between historical deviation in allele frequency and change in
250 allele frequency through time for each SNP [31]. In this analysis, stronger correlations will be
251 indicative of greater homogenization due to higher levels of gene flow. For each of the four tidal
252 marsh populations, we estimated deviation from mean allele frequency of the tidal marsh and
253 eastern California populations as $\overline{p_i} - p_{ij.}$ Where $\overline{p_i}$ is the mean allele frequency of the ith SNP
254 between inland eastern California and the jth tidal marsh population and p_{ij} is the allele frequency
255 of the ith SNP in the jth tidal marsh population. Difference in allele frequency change through time
256 was estimated for each SNP in each tidal marsh population as $p_{ij}^{modern} - p_{ij}^{historical}$. We used
257 permutation tests to test for significant differences from zero (no correlation) as well as between
258 outlier and non-outlier SNPs. This showed that correlations between historic deviation and
259 temporal change in allele frequency were significantly different from zero for both the Bay Area
260 (mean correlation: 0.17; p-value=0.015) and Humboldt Bay tidal marsh populations (mean: 0.30;
261 p-value=0.001), but not significant for Morro Bay (mean: 0.16; p-value=0.08) or Newport Bay
262 (mean: 0.05; p-value: 0.33; Fig. 4a). Second, we found that outlier loci exhibit significantly more
263 positive correlations relative to non-outlier loci across all four coastal populations (Fig. 4a).
264 Finally, the degree of correlation varies significantly with the extent to which inland eastern
265 California ancestry increased in each tidal marsh population through time (r^2=0.92; p-value=0.03;
266 Fig. 4b). This result remained regardless of the threshold used to identify outlier SNPs (thresholds:
267 0.9, 0.95, 0.99th percentile of Fst divergence). In sum, ΔFst and temporal allele frequency change
268 results both indicate that the inferred increase in gene flow levels from the freshwater-adapted,

eastern California population into tidal marsh populations eroded divergence at loci potentially associated with adaptation to a high salinity environment.

DISCUSSION:

There is a pressing need to understand the ecological and demographic factors that influence population genetic responses to habitat loss and environmental change. Here, we leveraged spatially replicated, time-series of specimens to document the impacts of habitat loss in tidal marsh populations of the Savannah sparrow in California. We found that 5 of 6 tidal marsh populations exhibit evidence for declines in genetic diversity, in line with expectations that coastal populations would experience the greatest overall loss. However, the amount of tidal marsh habitat lost was not linked to patterns of genetic diversity declines (Fig. 2d). Rather, we showed that: (1) historic levels of genetic diversity in southern California were lower than northern California (Fig. 2a-c); (2) modern levels of genetic diversity were correlated with the historic extent of tidal marsh habitat in each estuary (Fig. 2e); and southern populations have likely remained small since divergence from northern populations ~250 kya (Fig. 3a). Further, despite experiencing nearly 90% tidal marsh habitat loss, populations from the Bay Area (San Francisco, San Pablo and Suisun Bays) still exhibit similar levels of genetic diversity to other northern California populations (Fig. 2c). While habitat loss and fragmentation likely played a primary role in observed declines of genetic diversity, our analyses point to more complex demographic dynamics contributing to overall patterns of changing genetic diversity among these populations.

In other species, a disconnect between habitat loss and genetic diversity declines has been attributed to a range of different demographic, life history, ecological, and behavioral factors [33-

291 34]. Many of these factors are unlikely to contribute to the observed differences in genetic diversity
292 between southern and northern tidal marshes as generation time, body size, and migratory behavior
293 are all similar across coastal populations. Differences in the degree of specialization to tidal
294 marshes could contribute to differences in genetic diversity declines. Tidal marsh habitats are the
295 stronghold for *P. s. alaudinus* populations in northern California, but this subspecies can regularly
296 be found breeding in coastal grasslands [35]. This greater niche breadth may buffer them from
297 tidal marsh habitat losses. Indeed, populations from Humboldt Bay shifted from tidal marsh into
298 pasture habitats as pasture land expanded at the expense of tidal marsh habitat [21].

299 Reduced genetic diversity declines in the Bay Area may also be due to temporal changes
300 in gene flow rates among populations. Population structure (Fig. 1c), demographic (Fig. 3; Table
301 1), and simulation analyses (Fig. 3c-d) all support a recent shift to greater immigration from eastern
302 California into the Bay Area. These results support theoretical work that habitat loss and
303 subsequent population declines can leave populations more susceptible to hybridization and gene
304 flow [36]. Further, differences in Ne between populations will influence the direction of gene flow
305 with greater gene flow expected from larger into smaller populations [37]. Our empirical results
306 and theoretical work [38-39] together suggest that habitat loss may have led to declining local
307 recruitment and population growth within the Bay Area, resulting in the Bay Area becoming a sink
308 population, sustained by migration from eastern California.

309 The impact of gene flow on the persistence of natural populations is hotly debated in the
310 conservation genetics literature [40-42]. Gene flow can be crucial for preventing or reversing
311 inbreeding depression [43-45]. Gene flow can also contribute critical standing variation for
312 adaptation and may be important for future responses to climate change [46]. However, gene flow
313 could also play a role in eliminating divergence between populations and may lead to outbreeding

314 depression [7]. Our data show that coastal populations experiencing higher levels of gene flow do
315 exhibit reduced declines in π (but not Watterson's θ) over time. However, we also show that
316 increased immigration into the Bay Area through time likely contributed to the homogenization of
317 outlier regions associated with tidal marsh adaptation (Fig. 4, 5). Our findings contrast with a
318 recent study on Trinidadian guppies, which found that allele frequency differences in outlier loci
319 were more resistant to experimentally-induced gene flow than non-outlier loci [45]. This study
320 also showed that increasing gene flow from main-stream environments led to increased fitness of
321 headwater guppy populations, but the study was performed only for a limited time frame of ~6
322 generations. It will be critical for the future management of the tidal marsh sparrow populations to
323 determine whether increased gene flow contributes to reduced fitness and outbreeding depression.
324 However, the current study remains an important natural replication of the guppy experiment
325 showing that shifting gene flow rates in response to human-induced habitat degradation can lead
326 to the greater erosion of locally adapted loci.

327 Linking outlier loci from genome scans to performance and fitness is a persistent challenge
328 in studies of non-model organisms [47]. Despite this caveat, over 40% of identified outlier loci
329 may contribute to increased osmoregulatory performance in high salinity tidal marsh environments
330 via their known roles in kidney development, function, and pathology (Supplemental Table S3).
331 Only the estrogen receptor 1 (ESR1) and IgGFc-binding protein-like (FCGBP-like) genes were
332 found to be outliers across multiple populations. The ESR1 gene in birds has primarily been linked
333 to differences in aggressive behavior [48-49]. However, ESR1 is widely expressed in the
334 mammalian kidney [50] and contributes to the generation of important osmoregulatory phenotypes
335 matching those found in tidal marsh sparrows [16; 51]. These include larger kidneys [52] and the
336 production of a more concentrated urine through the inhibition of Aquaporin-2 expression in the

collecting duct of rodent kidneys [53]. Salinity and temperature are expected to increase in northern California in the near future and could exacerbate osmoregulatory stress experienced by these sparrows [54]. If the ESR1 and other candidate genes identified in this study contribute to important osmoregulatory functions, declining allele frequencies due to gene flow may reduce the adaptive capacity of these Bay Area birds.

Conclusions: Research on geographic variation has long motivated the growth of natural history collections. Consequently, museum collections now provide abundant material for documenting the impact of human-induced ecological change on temporal changes in migration, population size, and selection. Taking full advantage of this unique strength of collections across broad taxonomic scales will be essential for developing a more robust understanding of plant and animal responses to anthropogenic change and translating this knowledge into improved management strategies. Our work demonstrates the power of analyzing these replicate time series of specimens to reconstruct how habitat loss and demography have interacted over the past century to shape population genetic change. Specifically, we provide novel evidence for human-induced habitat loss transforming a locally adapted population from a source population to a sink with associated declines in divergence of locally adapted loci.

METHODS:

Sample library prep, capture, and sequencing: DNA was extracted from the toepads of 145 historic specimens and blood and tissues from 94 modern samples (74 from California and 20 from outside California; Dataset S1). Historical toepad samples were extracted using a phenol-chloroform protocol in a lab space dedicated to working with degraded DNA. Historic samples

360 were treated with the USER enzyme to mitigate C->T base misincorporation [55]. Modern samples
361 were extracted using a Qiagen DNeasy extraction kit (Valencia, California) and randomly sheared
362 using a qSonica Q800R3 Sonicator (Newtown, CT). For all samples, libraries were prepared for
363 sequencing using a Kapa Illumina Hyper Prep kit (see supplemental methods). DNA from prepared
364 libraries was captured using a custom designed exome capture array (see supplemental methods).
365 For captures, 500 ng of DNA from each library was pooled and hybridized in solution to the
366 SeqCap EZ probes following the manufacturer's protocols (Roche Sequencing Solutions, Inc,
367 Pleasanton, CA). Captured libraries were sequenced on two lanes of NovaSeq S4 and a single lane
368 of NovaSeq SP at the University of California Berkeley Vincent J. Coates Genomics Sequencing
369 Lab. De-multiplexed reads were filtered and aligned to the reference using the seqCapture pipeline
370 (https://github.com/CGRL-QB3-UCBerkeley/seqCapture/tree/master/scripts). We randomly
371 downsampled the number of reads included in bam files of the modern, higher-coverage samples
372 using the DownsampleSam function in Picard tools (https://broadinstitute.github.io/picard/).
373 Variants from resulting bam files were identified and called using the 'mpileup' and 'call' functions
374 in bcftools [56] (see supplemental methods for additional filtering, alignment, and variant calling
375 details).
376
377 *Population structure analyses:* Principal components analysis of the full, exon, and nongenic
378 datasets resulting from different filtering protocols (supplemental methods) was performed using
379 the python package scikit-allel v. 1.2.0 [57]. For all SNP datasets, we removed singletons and
380 pruned SNPs that were in linkage (r=0.25) in 2500 bp blocks. The estimated effective population
381 size (Ne) in DyStruct was set to 100,000 after preliminary analyses showed that results were robust
382 to varying the Ne setting from the smaller (25,000) to larger population sizes (500,000) previously

383 reported for this species [58]. Generation time was set to 2.2 years [59] and we binned samples
384 into eight 10-year windows spanning 48 generations from 1890 (generation 0) to 2017 (generation
385 48). We ran DyStruct three times independently for K = 1 to 12, where K is the number of pre-
386 defined population clusters. The most likely value of K was identified using a cross-validation,
387 hold-out method within the program.
388
389 *Genetic diversity and demographic analyses:* We compared historical and modern patterns of
390 migration and genetic diversity across California with Estimated Effective Migration Surfaces in
391 the program EEMS [27]. For both the modern and historic datasets, input files were generated
392 using the bed2diffs function in EEMS. We ran three independent runs of the MCMC chain. Each
393 chain started from a different random seed and ran for 2 million iterations of burn-in followed by
394 5 million iterations. A population grid density of 250 demes was selected and we adjusted the
395 qEffctProposalS2 (0.01) and qSeedsProposalS2 (0.2) settings upwards to ensure a proposal
396 acceptance rate between 10-40%. We used the R package rEEMSplots to assess model fit and
397 convergence among the three independent chains.
398 To match sequence depth between modern and historical samples, population genetic
399 statistics were estimated using a dataset with downsampled coverage in modern samples and the
400 removal of C->T, G->A mutations. We estimated Tajima's D, nucleotide diversity (π), and
401 Watterson's Theta (θ) for each exon (n=9,971 loci) and non-genic target region (n=1,024 loci) in
402 the dataset using scikit-allel v. 1.2.0 [57]. We also estimated these metrics across three and six
403 time points in Newport Bay (*P. s. beldingi*) and San Francisco Bay (*P. s. alaudinus*), respectively.
404 To estimate the magnitude of temporal change in each metric we subtracted the historic from the
405 modern value. Within each population we assessed whether significant differences existed

406 between sampling points using ANOVA and post-hoc TukeyHSD tests or t-tests depending on the
407 number of time points sampled. All statistical analyses were performed in R version 3.5.1
408 (https://www.r-project.org/).

409 We used GADMA2 [28] to identify the demographic model that best fit the three
410 dimensional site frequency spectrum (SFS) of eastern California (*P. s. nevadensis*), Bay Area (*P.*
411 *s. alaudinus*), and Newport Bay (*P. s. beldingi*) birds. We ran separate analyses for a historic
412 dataset and a modern dataset of these three populations. For separate historical and modern
413 datasets, we generated the SFS from both exon and non-genic data, however, to restrict our
414 analyses to putatively neutral loci we eliminated regions of the genome found to be under selection
415 by our analyses below. The final filtered vcf file was converted to the *Moments* input format using
416 a perl script from (https://github.com/wk8910/bio_tools/blob/master/
417 01.dadi/00.convertWithFSC/convert_vcf_to_dadi_input.pl). The final datasets included 17
418 individuals of *P. s. nevadensis,* 37 *P. s. alaudinus,* and 10 *P. s. beldingi* in the historical dataset,
419 and 13 *P. s. nevadensis,* 31 *P. s. alaudinus, and* 11 *P. s. beldingi* in the modern dataset. Within
420 GADMA2 these datasets were downprojected to include an equal number of samples for each
421 population with 10 for the modern and 11 for the historic population. The final input datasets
422 included 26,080 and 21,758 SNPs for the historical and modern data, respectively.

423 Within GADMA2, we used *Moments* [60] as an engine for local optimization. We
424 initialized the global search with a simple structured model (1,1,1) that allowed for one
425 demographic event in between divergence times of the three populations. We allowed for
426 asymmetric migration among all populations, and population size changes that were linear,
427 exponential or sudden. For each dataset, we performed 20 independent runs of the global search

428 optimization in GADMA. Since all optimizations produced a model with the same number of
429 parameters top models were identified based on their likelihood. Demographic parameter values
430 were estimated from the value of theta ($4N_e\mu L$; where L is sequence length) based on a generation
431 time of 2.2 years for the Savannah sparrow [59], an estimated germline mutation rate of 4.6e-9 for
432 another songbird species, *Ficedula albicollis* [61]. The total length of sequence (L) from which
433 SNPs were called was estimated as: total sequence length * (SNPs in SFS/ total SNPs pre-
434 filtering). This resulted in a length estimate of 6.5 Mb for the historic dataset and 4.1 Mb for the
435 modern dataset.

436 *Coalescent simulations:* We performed a series of coalescent simulations in msprime 1.2.0 [29] to
437 confirm whether recent changes in migration rates between coastal and interior populations of the
438 Savannah sparrow could explain inferences from DyStruct. We used the historical best-fit
439 GADMA model to parameterize our demographic model in GADMA. In total we performed 25
440 separate simulations that included all combinations of five different timings of migration rate
441 changes (5,10,15,20,25 generations) and five different migration rates from eastern California to
442 the Bay Area (0.00015, 0.0015,0.015,0.15, 0.5). We ran a final simulation that involved no change
443 in migration rate from the historical best-fit demographic model. For each simulation we sampled
444 25 individuals from the Bay Area, Newport Bay, and eastern California at 50 generations before
445 the present and another 25 individuals from each of the populations at the present time. This
446 sampling scheme closely resembles the earliest and latest sampling times for DyStruct analyses on
447 observed data. For each of the 26 simulations we ran 60,000 replicates of the demographic model
448 sampling from a sequence length of 650 to approximate the ascertainment scheme of our capture
449 dataset. For each of the 60,000 replicated tree sequences mutations were added using default
450 parameters in msprime and a mutation rate of 4.6e-9 for passerine birds [61]. To generate a

genotype matrix from these simulated mutations we randomly selected a single variant from each of the 60,000 replicates that was present in 2 or more individuals. The resulting genotype matrix for each simulation was analyzed in DyStruct to test which parameter combinations most closely approximate observed changes in ancestry proportions. We ran Dystruct for each simulation with k=3 and the pop size parameter set to 70,000.

Spatial and temporal selection analyses: We identified putative outlier loci using the downsampled dataset with C->T and G->A sites removed, a maf filter of 0.05, and only retained sites with no missing data. We divided the dataset into five populations: (1) freshwater-adapted populations of eastern California (*P. s. nevadensis*); (2) Humboldt Bay estuary (*P. s. alaudinus*); (3) San Francisco Bay estuary (*P. s. alaudinus*); (4) Morro Bay estuary (*P. s. alaudinus*); and (5) Newport Bay estuary (*P. s. beldingi*). Each of these five populations were further divided into historic (sampled 1900-1920) and modern (sampled after 2000) datasets. First, we estimated Hudson's Fst [62] and Dxy between salt and freshwater-adapted populations for 5Kb windows with a step size of 2.5 Kb following approaches used to analyze other exome datasets [63]. Third, we employed a Latent Factor Mixed Models approach (LFMM; [30]) to identify SNPs with significant associations with mean salinity estimates from a previously published dataset [17]. Loci were designated as outliers if the 5kb region was in the top 5% of Fst divergence, top 5% of Dxy divergence, and contained a SNP with a p-value <0.05 based on LFMM analyses for each of the four populations. We compared temporal changes in Fst divergence for each 5kb window by subtracting historical Fst from modern Fst divergence in each locus to estimate ΔFst. We used t-tests to compare whether outlier versus non-outlier regions exhibited significantly different changes in Fst through time. We performed coalescent simulations in msprime 1.2.0 [29] as above

474 with the historic best-fit GADMA2 model used to parameterize simulations. We simulated a
475 demographic history with no change in the demographic parameters over the past one hundred
476 years and one with an increase in migration rate to 0.15 from eastern California to the Bay Area at
477 15 generations ago. We simulated a sequence of length 100 Mb and estimated Fst divergence in
478 sliding windows identically to the empirical dataset. Measures of historical Fst divergence, and
479 delta Fst were estimated for both the simulated Bay Area and Newport Bay population. These
480 simulated estimates were compared to observed patterns of Fst divergence using ANOVA and
481 Tukey post-hoc tests in r.

482 To assess the degree of correlation between historical deviations in mean allele frequency
483 and the direction of allele frequency change through time. We estimated allele frequencies using
484 VCFtools [64]. We classified SNPs as outliers if they exceeded the 99th percentile of Fst
485 divergence. We compared Pearson's correlation of the outlier SNPs to non-outlier SNPs by
486 performing a permutation test where we randomly sampled an equal number of non-outlier to
487 outlier SNPs and calculated correlations 1000 times. We then compared whether the non-outlier
488 loci were both statistically different from 0 (no correlation) or to the correlation estimated for the
489 outlier loci. We used changes in the proportion of eastern California ancestry found in each of the
490 four coastal populations from the DyStruct analysis to approximate temporal increases in gene
491 flow.

492

493 **ACKNOWLEDGEMENTS:**

494 We thank the following museums and individuals for providing access to their collections for
495 this study: Museum of Vertebrate Zoology (Carla Cicero); California Academy of Sciences
496 (Maureen Flannery, Jack Dumbacher); San Diego Museum of Natural History (Philip Unitt); Los

Angeles County Museum of Natural History (Allison Shultz, Kimball Garrett); American Museum of Natural History (Paul Sweet, Tom Trombone); Field Museum of Natural History (John Bates, Shannon Hackett, Ben Marks); University of Washington Burke Museum (Sharon Burkes); University of Montana Zoological Museum (Angela Hornesby). Personnel at the California Department of Fish and Wildlife, California State Parks, and United States Fish and Wildlife Service assisted with access to federal and state lands. Finally, we thank Elizabeth Beckman, Ammon Corl, Mackenzie Kirchner-Smith, Jackie Childers, Rosa Jimenez, Dan Wait, and Cynthia Wang-Claypool for useful comments on the manuscript. This study was supported by an NSF PRFB fellowship (#1812282) to PMB.

REFERENCES:

[1] Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., Rylands, A. B., Konstant, W. R., … Hilton-Taylor, C. (2002). Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology*, *16*(4), 909–923. https://doi.org/10.1046/j.1523-1739.2002.00530.x

[2] Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

[3] Forester, B. R., Beever, E. A., Darst, C., Szymanski, J., & Funk, W. C. (2022). Linking evolutionary potential to extinction risk: applications and future directions. *Frontiers in Ecology and the Environment*, *20*(9), 507–515. https://doi.org/10.1002/fee.2552

[4] Díaz, S., Zafra-Calvo, N., Purvis, A., Verburg, P. H., Obura, D., Leadley, P., … Zanne, A. E. (2020). Set ambitious goals for biodiversity and sustainability. *Science*, *370*(6515), 411–413. https://doi.org/10.1126/science.abe1530

[5] Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., … Zess, E. (2022). Genetic diversity loss in the Anthropocene. *Science*, *377*(6613), 1431–1435. https://doi.org/10.1126/science.abn5642

[6] Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology and Evolution*, *30*(1), 42–49. https://doi.org/10.1016/j.tree.2014.10.009

[7] Edmands, S. (2007). Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, *16*(3), 463–475. https://doi.org/10.1111/j.1365-294X.2006.03148.

[8] Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., Labarbera, K., Wommack, E. A., … Nachman, M. W. (2016). Natural history collections as windows on evolutionary processes. *Molecular Ecology*, *25*(4), 864–881. https://doi.org/10.1111/mec.13529

[9] Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: A review of methods, challenges, applications, and future directions. *Molecular Ecology Resources*, (July), 1755-0998.13066. https://doi.org/10.1111/1755-0998.13066

[10] Lang, P. L. M., Willems, F. M., Scheepens, J. F., Burbano, H. A., & Bossdorf, O. (2019). Using herbaria to study global environmental change. *New Phytologist*, *221*(1), 110–122. https://doi.org/10.1111/nph.15401

[11] Benham, P. M., & Bowie, R. C. K. (2022). Natural history collections as a resource for conservation genomics: Understanding the past to preserve the future. *Journal of Heredity*, (December 2022), 1–18. https://doi.org/10.1093/jhered/esac066

[12] Wheelwright, N.T., Rising, J.D. (2008). Savannah Sparrow (*Passerculus sandwichensis*). *in* Birds of North American Online (A. Poole, Ed.). Cornell Lab of Ornithology, Ithaca, NY.

[13] Cade, T., & Bartholomew, G. A. (1959). Sea-water and salt utilization by Savannah sparrows. *Physiological Zoology*, *32*(4), 230–238. Retrieved from http://www.jstor.org/stable/30155400 .

[14] Poulson, T. L., & Bartholomew, G. A. (1962). Salt balance in the Savannah Sparrow. *Physiological Zoology*, *35*, 109–119.

[15] Walsh, J., Benham, P. M., Deane-Coe, P. E., Arcese, P., Butcher, B. G., Chan, Y. L., … Lovette, I. J. (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters*, 1–15. https://doi.org/10.1002/evl3.126

[16] 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). https://doi.org/10.1111/evo.13886

[17] Benham, P. M., & Bowie, R. C. K. (2021). The influence of spatially heterogeneous anthropogenic change on bill size evolution in a coastal songbird. *Evolutionary Applications*, (September), 1–18. https://doi.org/10.1111/eva.13144

[18] Marshall, J., & Dedrick, K. (1994). Endemic Song Sparrows and Yellowthroats of San Francisco Bay. *Studies in Avian Biology*, *15*(15), 316-327.

[19] Brophy, L. S., Greene, C. M., Hare, V. C., Holycross, B., Lanier, A., Heady, W. N., … Dana, R. (2019). *Insights into estuary habitat loss in the western United States using a new method for mapping maximum extent of tidal wetlands*. *PLoS ONE* (Vol. 14). https://doi.org/10.1371/journal.pone.0218558

[20] Zembal, R. & Hoffmann, S.M. (2010). A survey of the Belding's Savannah sparrow (*Passerculus sandwichensis beldingi*) in California 2010. Calif. Dep. Fish and Game, Wildlife Branch, Nongame Wildlife Program Report 2010-10, Sacramento, CA 17 pp.

[21] Fitton, S.D. (2008). Bryant's Savannah Sparrow (*Passerculus sandwichensis alaudinus*). Shuford, W.D. & Gardali, T., eds. California Bird Species of Special Concern: A ranked assessment of species, subspecies, and distinct populations of birds of immediate conservation concern in California. *Studies of Western Birds 1.* Western Field Ornithologists, Camarillo, California, and California Department of Fish and Game, Sacramento.

[22] Feng, S., Fang, Q., Barnett, R., Li, C., Han, S., Kuhlwilm, M., … Zhang, G. (2019). The Genomic Footprints of the Fall and Recovery of the Crested Ibis. *Current Biology*, *29*(2), 340-349.e7. https://doi.org/10.1016/j.cub.2018.12.008

[23] van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Current Biology*, *29*(1), 165-170.e6. https://doi.org/10.1016/j.cub.2018.11.055

[24] Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*(1), 185–202. https://doi.org/10.1111/mec.13304

[25] Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., … Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLOS Genetics*, *15*(5), e1008119. https://doi.org/10.1371/journal.pgen.1008119

[26] Joseph, T. A., & Pe'er, I. (2019). Inference of Population Structure from Time-Series Genotype Data. *American Journal of Human Genetics*, *105*(2), 317–333. https://doi.org/10.1016/j.ajhg.2019.06.002

[27] Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100. https://doi.org/10.1038/ng.3464

[28] Noskova, E., Abramov, N., Iliutkin, S., Sidorin, A., Dobrynin, P., & Ulyantsev, V. (2022). GADMA2: more efficient and flexible demographic inference from genetic data. *BioRxiv*, (Ld), 2022.06.14.496083. Retrieved from https://doi.org/10.1101/2022.06.14.496083%0Ahttps://www.biorxiv.org/content/10.1101/2022.06.14.496083v1%0Ahttps://www.biorxiv.org/content/10.1101/2022.06.14.496083v1.abstract

[29] Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., … Kelleher, J. (2022). Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, *220*(3). https://doi.org/10.1093/genetics/iyab229

[30] Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

[31] Gompert, Z., Springer, A., Brady, M., Chaturvedi, S., & Lucas, L. K. (2021). Genomic time-series data show that gene flow maintains high genetic diversity despite substantial genetic drift in a butterfly species. *Molecular Ecology*, (April), 1–18. https://doi.org/10.1111/mec.16111

[32] Powell, A. N. 2006. Are southern California's fragmented saltmarshes capable of sustaining endemic bird populations? *Studies in Avian Biology*, (32), 198-204.

[33] Welch, A. J., Wiley, A. E., James, H. F., Ostrom, P. H., Stafford, T. W., & Fleischer, R. C. (2012). Ancient DNA reveals genetic stability despite demographic decline: 3,000 years of population history in the endemic Hawaiian petrel. In *Molecular Biology and Evolution* (Vol. 29, pp. 3729–3740). https://doi.org/10.1093/molbev/mss185

[34] Lino, A., Fonseca, C., Rojas, D., Fischer, E., & Ramos Pereira, M. J. (2019). A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. *Mammalian Biology*, *94*, 69–76. https://doi.org/10.1016/j.mambio.2018.09.006

[35] Grinnell, J., Miller, A.H. (1944). The distribution of the birds of California. Pacific Coast Avifauna 27: 1-608.

[36] Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*, *9*(7), 892–908. https://doi.org/10.1111/eva.12367

[37] Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*(8), 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

[38] Dias, P. C. (1996). Sources and sinks in population biology. *Trends in Ecology and Evolution*, *11*(8), 326–330. https://doi.org/10.1016/0169-5347(96)10037-9

[39] Gaggiotti, O. E., & Smouse, P. E. (1996). Stochastic Migration and Maintenance of Genetic Variation in Sink Populations. *American Naturalist*, *147*(6), 919–945.

[40] Tallmon, D. A., Luikart, G., & Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology and Evolution*, *19*(9), 489–496. https://doi.org/10.1016/j.tree.2004.07.003

[41] Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy, R. C., Ralls, K., Dudash, M. R., & Fenster, C. B. (2011). Predicting the probability of outbreeding depression. *Conservation Biology*, *25*(3), 465–475. https://doi.org/10.1111/j.1523-1739.2011.01662.x

[42] Bell, D. A., Robinson, Z. L., Funk, W. C., Fitzpatrick, S. W., Allendorf, F. W., Tallmon, D. A., & Whiteley, A. R. (2019). The Exciting Potential and Remaining Uncertainties of Genetic Rescue. *Trends in Ecology and Evolution*, *34*(12), 1070–1079. https://doi.org/10.1016/j.tree.2019.06.006

[43] Westemeier, R. L., Brawn, J. D., Simpson, S. A., Esker, T. L., Jansen, R. W., Walk, J. W., … Paige, K. N. (1998). Tracking the long-term decline and recovery of an isolated population. *Science*, *282*(5394), 1695–1698. https://doi.org/10.1126/science.282.5394.1695

[44] Hogg, J. T., Forbes, S. H., Steele, B. M., & Luikart, G. (2006). Genetic rescue of an insular population of large mammals. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1593), 1491–1499. https://doi.org/10.1098/rspb.2006.3477

[45] Fitzpatrick, S. W., Bradburd, G. S., Kremer, C. T., Salerno, P. E., Angeloni, L. M., & Funk, W. C. (2020). Genomic and Fitness Consequences of Genetic Rescue in Wild Populations. *Current Biology*, *30*(3), 517-522.e5. https://doi.org/10.1016/j.cub.2019.11.062

[46] Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177. https://doi.org/10.1038/s41559-018-0777-y

[47] Storz, J. F., & Cheviron, Z. A. (2021). Physiological Genomics of Adaptation to High-Altitude Hypoxia. *Annual Review of Animal Biosciences*, *9*(1), 1–23. https://doi.org/10.1146/annurev-animal-072820-102736

[48] Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., … Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

[49] Merritt, J. R., Grogan, K. E., Zinzow-Kramer, W. M., Sun, D., Ortlund, E. A., Yi, S. V., & Maney, D. L. (2020). A supergene-linked estrogen receptor drives alternative phenotypes in a polymorphic songbird. *Proceedings of the National Academy of Sciences*, *117*(35), 202011347. https://doi.org/10.1073/pnas.2011347117

[50] Jelinsky, S. A., Harris, H. A., Brown, E. L., Flanagan, K., Zhang, X., Tunkey, C., … Evans, M. J. (2003). Global transcription profiling of estrogen activity: Estrogen receptor α regulates gene expression in the kidney. *Endocrinology*, *144*(2), 701–710. https://doi.org/10.1210/en.2002-220728

[51] Goldstein, D. L. (2006). Osmoregulatory biology of saltmarsh passerines. *Terrestrial Vertebrates of Tidal Marshes: Evolution, Ecology, and Conservation*, (32), 110–118.

669 [52] Lovegrove, A. S., Sun, J., Gould, K. A., Lubahn, D. B., Korach, K. S., Lane, P. H., …
670 Estrogen, P. H. L. (2004). How Genes Influence Renal Function and Disease Estrogen
671 receptor-mediated events promote sex-specific diabetic glomerular hypertrophy. *Blood*,
672 *2196*, 586–591.
673 [53] Cheema, M. U., Irsik, D. L., Wang, Y., Miller-Little, W., Hyndman, K. A., Marks, E. S., …
674 Norregaard, R. (2015). Estradiol regulates AQP2 expression in the collecting duct: A novel
675 inhibitory role for estrogen receptor α. *American Journal of Physiology - Renal Physiology*,
676 *309*(4), F305–F317. https://doi.org/10.1152/ajprenal.00685.2014
677 [54] Callaway, J.C., Parker, V.T., Vasey, M.C., & Schile, L.M. (2007). Emerging issues for the
678 restoration of tidal marsh ecosystems in the context of predicted climate change. *Madroño,*
679 54: 234-248.

680

681 [55] Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2009).
682 Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA.
683 *Nucleic Acids Research*, *38*(6), 1–12. https://doi.org/10.1093/nar/gkp1163

684 [56] Li H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping
685 and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987-
686 93.
687 [57] Miles, A. et al. 2020 Scikit-Allel: https://zenodo.org/record/4759368#.YQ2eqC2cYWo

688 [58] Benham, P. M., & Cheviron, Z. A. (2019). Divergent mitochondrial lineages arose within a
689 large, panmictic population of the Savannah sparrow (*Passerculus sandwichensis*) .
690 *Molecular Ecology*, 1765–1783. https://doi.org/10.1111/mec.15049

[59] Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., … Butchart, S. H. M. (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, *00*(0), 1–10. https://doi.org/10.1111/cobi.13486

[60] 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. https://doi.org/10.1534/genetics.117.200493

[61] Smeds, L., Qvarnstrom, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, 1211–1218. https://doi.org/10.1101/gr.204669.116

[62] Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting FST: The impact of rare variants. *Genome Research*, *23*(9), 1514–1521. https://doi.org/10.1101/gr.154831.113

[63] Schweizer, R. M., Jones, M. R., Bradburd, G. S., Storz, J. F., Senner, N. R., Wolf, C., & Cheviron, Z. A. (2021). Broad Concordance in the Spatial Distribution of Adaptive and Neutral Genetic Variation across an Elevational Gradient in Deer Mice. *Molecular Biology and Evolution*, *38*(10), 4286–4300. https://doi.org/10.1093/molbev/msab161

[64] Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

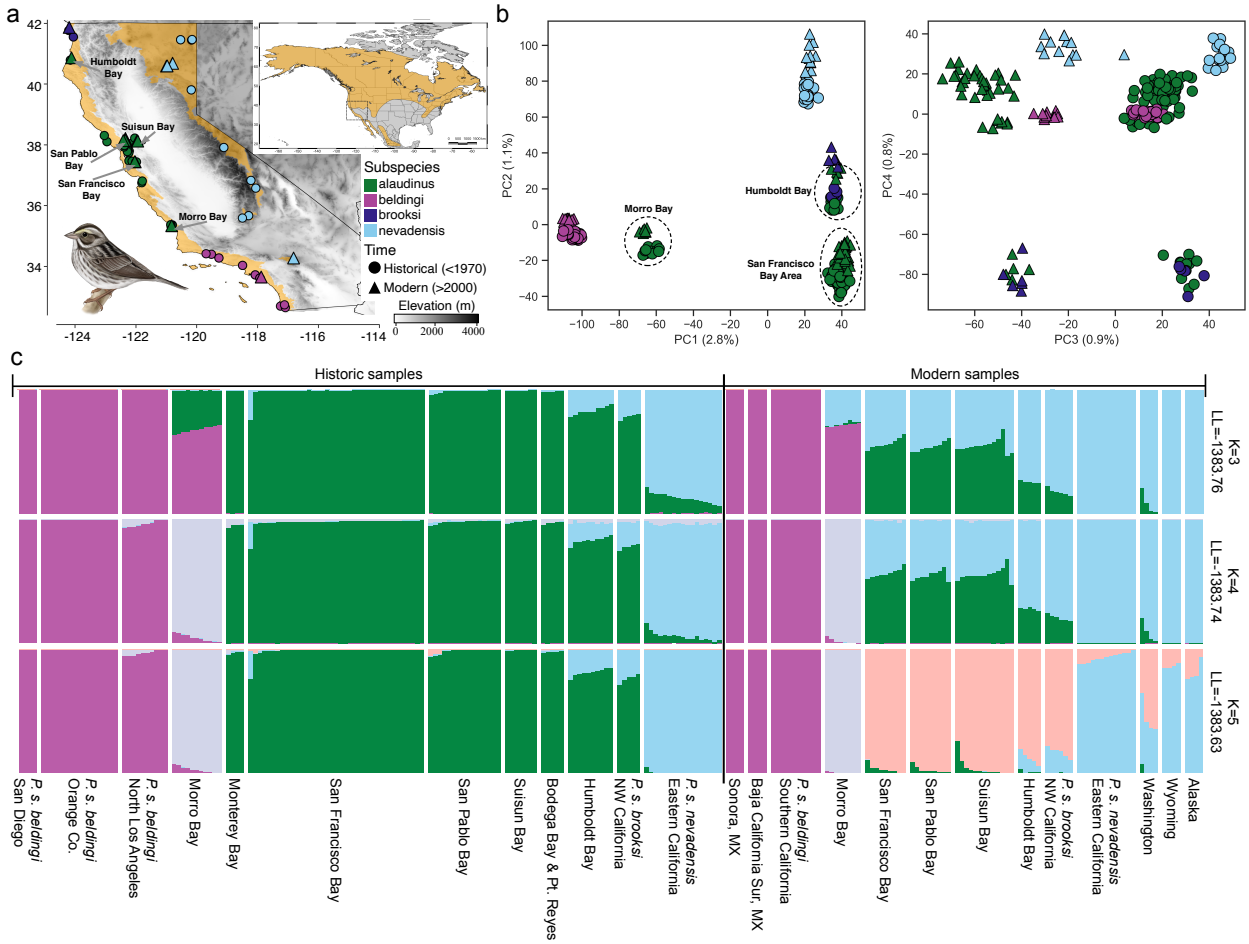

Figure 1: Spatial and temporal patterns of population structure within California populations of the Savannah sparrow (*Passerculus sandwichensis*). **(a)** California breeding distribution of the Savannah sparrow (ochre). Circles represent historical (pre-1970) sampling localities of specimens sequenced in the study, and triangles represent modern (post-2000) sampling localities. Inset shows the extent of the entire North American breeding distribution of the Savannah sparrow in ochre. **(b)** Principal components analysis of the full exome capture dataset (60,765 SNPs). Left panel shows PC1 (2.8% variance explained) versus PC2 (1.1% variance explained) and right panel shows PC3 (0.9% variance explained) versus PC4 (0.8% variance explained). Colors correspond to different subspecies as illustrated in Fig. 1a. Dashed circles denote different clusters corresponding to different populations within *P. s. alaudinus.* **(c)** Results from DyStruct analysis. Runs with clusters specified as K=3-5 all exhibited very similar log-likelihoods and are illustrated here. Colors correspond to different K clusters. Artwork of *P. s. beldingi* individual courtesy of Jillian Nichol Ditner

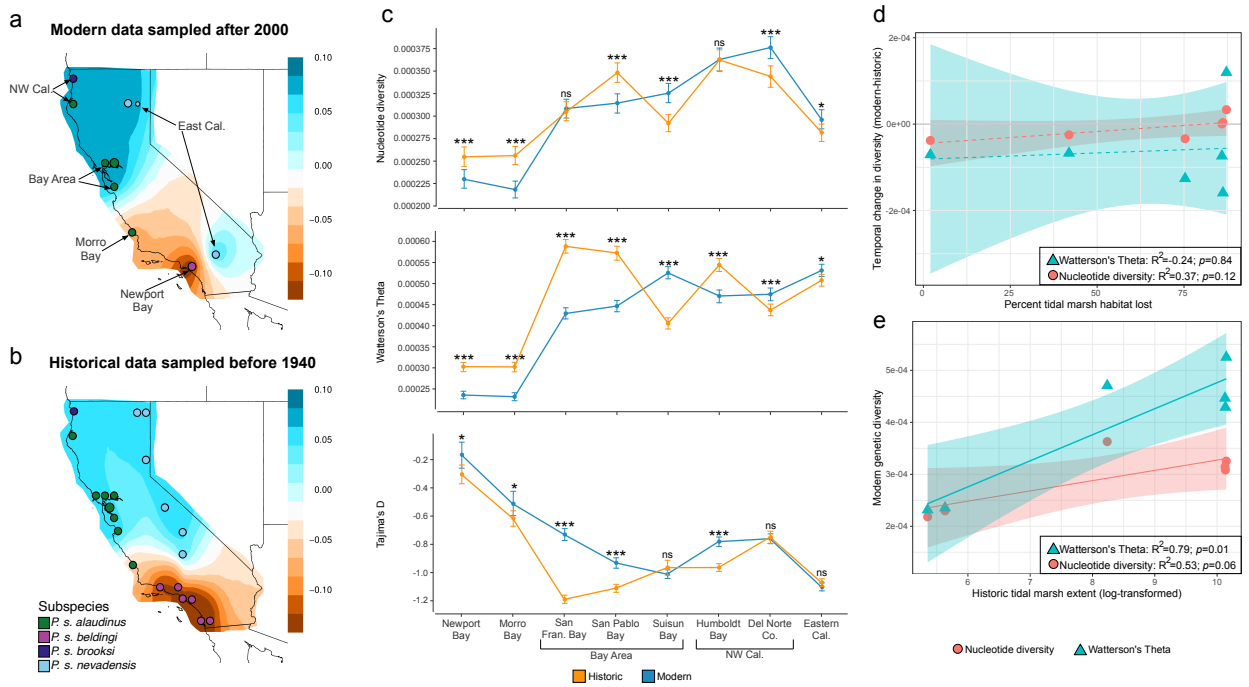

Figure 2: Spatial and temporal patterns of genetic diversity within California Savannah sparrows. **(a)** Modern variation in genetic diversity across California estimated using the Estimated Effective Migration Surfaces (EEMS) program. Contour plots vary from low (dark brown) to high (dark blue) posterior mean diversity rates. **(b)** Historic variation in genetic diversity across California estimated using EEMS. **(c)** Plots comparing historical (orange) and modern (blue) estimates of nucleotide diversity, Watterson's theta, and Tajima's D across all populations. For both **b** & **c** plots are based on a downsampled dataset with C->T and G->A mutations removed. Asterisks in **b** & **c** denote p-values based on t-tests between time groups (ns: non-significant; *p<0.05; **p<0.01; ***: p<0.0001). **(d & e)** Relationship between the amount of tidal marsh habitat lost or available in each of six estuaries and population genetic statistics. Values on y-axis have all been scaled and centered to visualize on same axis. **(d)** The amount of tidal marsh lost (%) over past century is not significantly related to temporal changes in nucleotide diversity (red circle) or Watterson's theta (blue square) **(e)** Historically, the total amount of tidal marsh habitat present in each estuary significantly predicts modern levels of Watterson's theta, while nucleotide diversity is marginally non-significant. Contemporary levels of tidal marsh habitat did not predict levels of genetic diversity (Supplemental Table S2).

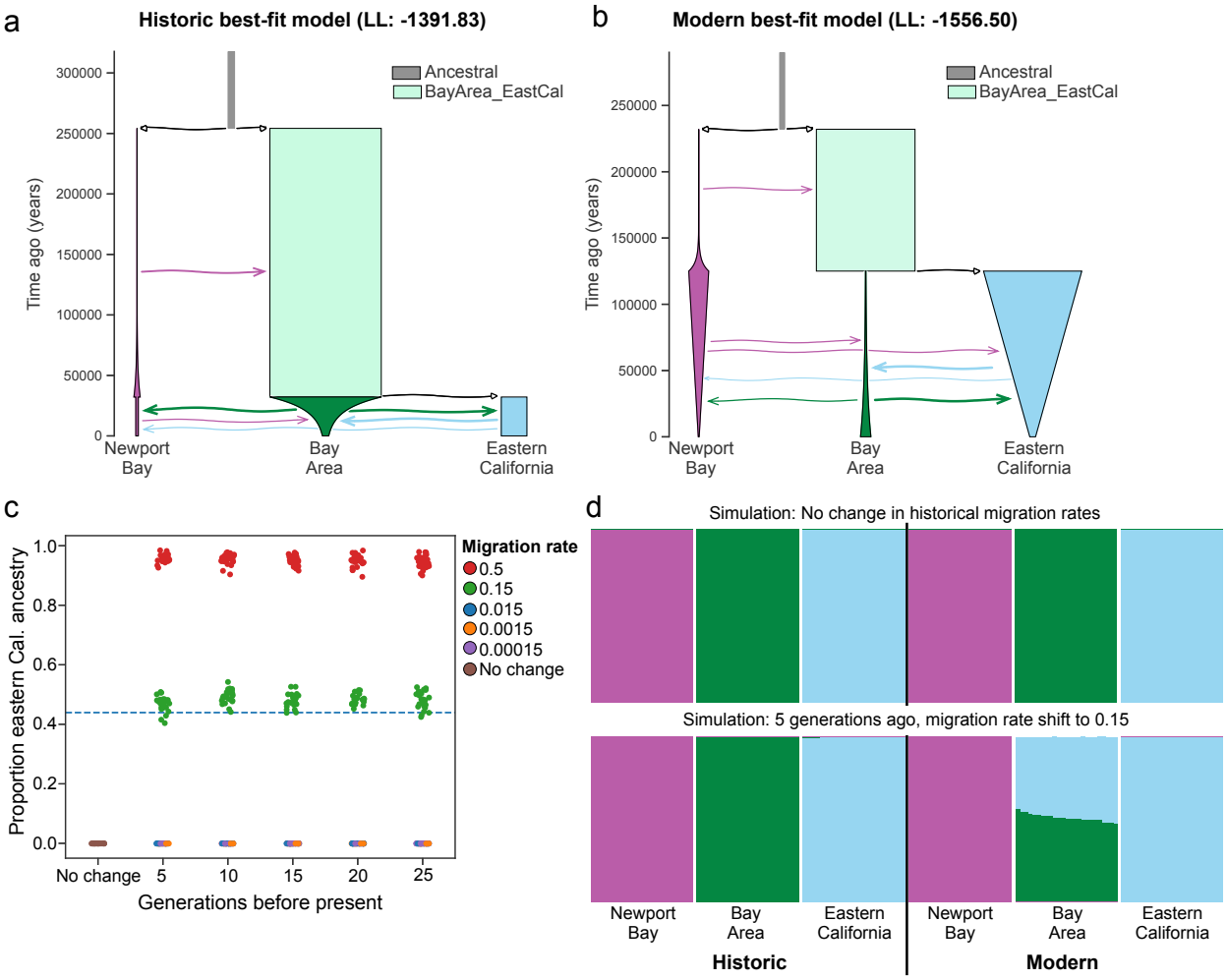

Figure 3: Temporal change in migration rates within California Savannah sparrows. (a) Historic best-fit model estimated with GADMA2. Y-axis shows time of demographic events, width of bars represents effective population size for each population, and arrows signify migration rates from source to destination population forward in time. Arrows are colored by source population and represent continuous migration. (b) Representation of best-fit model to the modern dataset in GADMA2. (c) Coalescent simulation results where the historic best-fit model (Fig. 3a) was used as the baseline demographic history with migration rates changing at varying generations before the present (x-axis) to different rates (colors in legend) from eastern California to the Bay Area. Simulated datasets were analyzed within Dystruct and the proportion of eastern California ancestry in the modern Bay Area population was estimated (y-axis). Dashed blue line represents the average observed eastern California ancestry found in the modern Bay Area population (see Fig. 1c). Migration rates <0.15 did not result in any changes to proportion of eastern California ancestry. (d) Barplots showing ancestry proportions estimated for simulated datasets in Dystruct. Top plot shows the results based on simulation of the historic demographic model (Fig. 3a) with no change in migration rate. Bottom plot shows results from simulated dataset where migration rate shifted 5 generations ago from a proportion of 0.00015 immigrants per generation to a proportion of 0.15 immigrants per generation from eastern California into the Bay Area population.

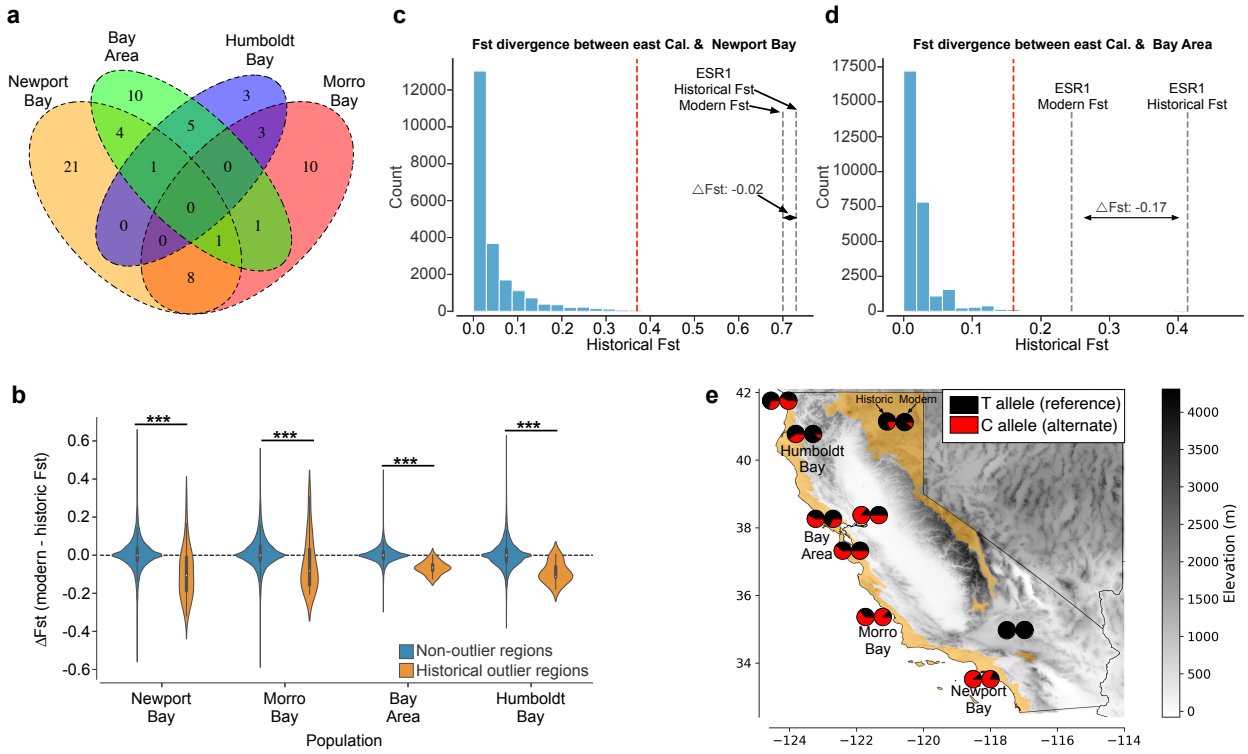

Figure 4: Temporal variation in Fst divergence estimated for genes found to be outliers in Fst, Dxy, and LFMM analyses. **(a)** Venn diagram showing overlap in genes with outlier SNPS across the four tidal marsh populations. **(b)** Change in Fst divergence between eastern California and four tidal marsh populations of Savannah sparrow. ΔFst for each locus was estimated as the difference between historic and modern Fst divergence with positive values indicating increasing Fst divergence through time. Asterisks denote p-values from t-tests of ΔFst between outlier and non-outlier SNPs (ns: non-significant; *p<0.05; **p<0.01; ***: p<0.0001). **(c-d)** Distribution of per SNP Fst estimates for historical comparison between Newport Bay (*P. s. beldingi*) and Bay Area (*P. s. alaudinus*) with eastern California (*P. s. nevadensis*). Red dashed line denotes the mean 99th percentile threshold for calling outlier SNPs. Gray dashed lines represent modern and historic Fst estimates for a C->T mutation in the 3' UTR region of the final exon of the Estrogen Receptor 1 gene (ESR1). **(e)** Spatial variation in allele frequencies for the ESR1 T->C SNP across nine sampling localities with paired modern and historic sampling. Pie charts on left illustrate historical allele frequencies and pie charts on the right modern allele frequencies.

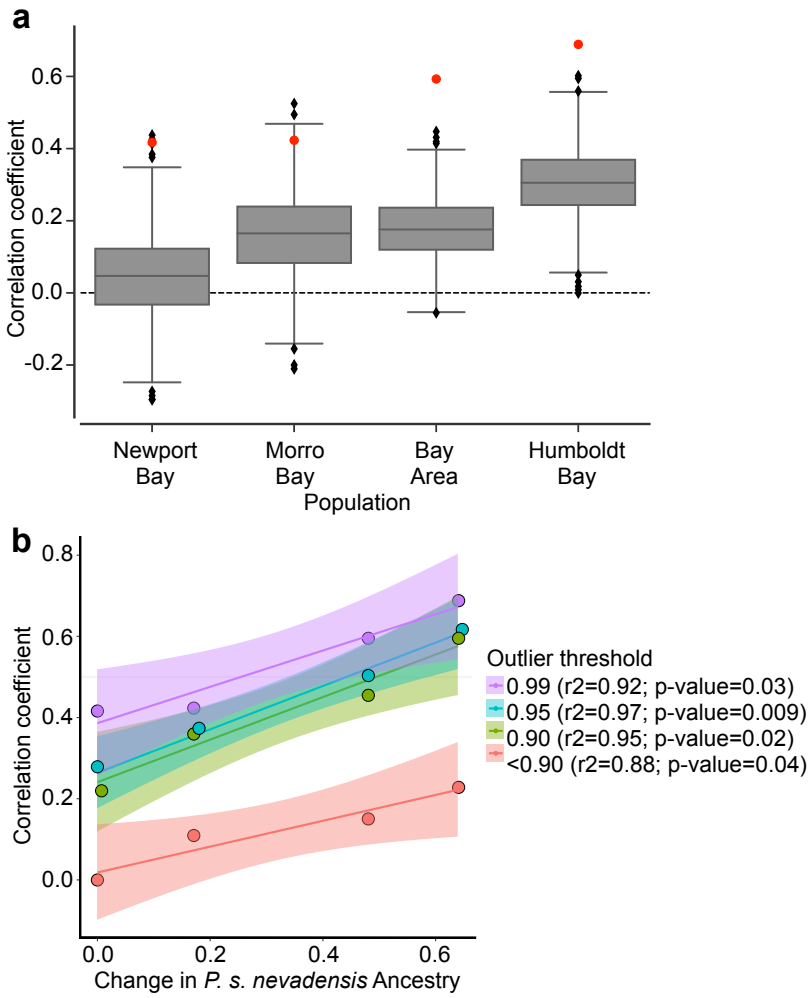

Figure 5: (a) Correlations between the historical deviation from mean allele frequency and direction of temporal change in each tidal marsh population. More positive correlations indicate greater changes in allele frequency towards the mean due to higher levels of gene flow. Red circles show observed correlation for outlier SNPs and gray box plots indicate the null distribution from 1000 permutations of non-outlier SNPs. **(b)** Regression analyses showing relationship between change in the proportion of ancestry from *P. s. nevadensis* populations in each tidal marsh population and the correlation coefficient for outlier SNPs shown in (a). Ancestry proportions from DyStruct analysis (Fig. 1c). Different colors represent SNPs inferred to be outliers from different percentile thresholds with red points showing data from non-outlier SNPs.

Table 1: Parameter estimates for the top five best-fit models for the historic and modern datasets. Demographic models and parameters were optimized in GADMA2. We report parameters for only the most recent time slice. Models are ordered by likelihood from left to right. For both time periods three populations were measured Newport Bay (n), Bay Area (b), and East Cal (e).

			Historic		
LL	-1391.83	-1393.98	-1397.1	-1398.1	-1405.83
Tsp1 (years)	254,219	238,351	192,733	243,921	3525,73
Tsp2 (years)	32,164	48,431	69,673	51,533	18,578
NewportBay Ne	6,675	9,341	13,180	8,878	6,618
BayArea Ne	16,755	38,319	23,688	35,013	30,276
EastCal Ne	68,438	84,073	102,343	77,387	54,259
n->b	1.2E-05	2.6E-08	2.6E-05	3.3E-05	0
b->n	1.1E-04	8.5E-05	4.4E-05	2.8E-08	1.2E-04
n->e	0	7.5E-06	1.4E-05	0.0E+00	0
e->n	1.2E-05	0	2.7E-06	7.2E-05	0
e->b	1.6E-04	5.9E-05	9.8E-05	1.1E-04	1.1E-04
b->e	1.3E-04	1.1E-04	1.1E-04	1.1E-04	1.5E-04
b->e mig per gen	8.90	9.25	11.26	8.51	8.14
e->b mig per gen	2.68	2.28	2.32	3.85	3.33
			Modern		
LL	-1556.5	-1583.13	-1585.11	-1585.43	-1595.09
Tsp1 (years)	232,053	276,701	310,210	313,395	343,840
Tsp2 (years)	125,040	31,727	5,803	27,926	3,694
NewportBay Ne	4,923	3,091	3,290	3,149	2,379
BayArea Ne	57,309	41,061	37,157	426,253	54,707
EastCal Ne	31,638	212,932	37,029	17,969	34,174
n->b	4.2E-05	1.5E-05	5.2E-05	2.6E-05	7.2E-05
b->n	4.5E-05	8.4E-05	1.0E-04	6.1E-05	1.2E-04
n->e	2.6E-05	1.9E-20	0	5.9E-06	0
e->n	6.8E-09	2.6E-05	1.9E-05	7.7E-05	0
e->b	1.0E-04	1.0E-04	2.4E-08	1.5E-04	0
b->e	1.1E-04	4.0E-05	0	3.1E-05	0
b->e mig per gen	3.48	8.52	0.00	0.56	0.00
e->b mig per gen	5.73	4.11	0.0009	63.94	0.00

Table 2: Mean Fst divergence between eastern California and each of four tidal marsh populations of the Savannah sparrow in both the historic and modern datasets.

Population	historic sample size	modern sample size	Number SNPs	Historic mean Fst (sd)	Modern mean Fst (sd)
Newport Bay	10	11	4495	0.098 (0.008)	0.099 (0.009)
Morro Bay	11	8	5408	0.054 (0.003)	0.062 (0.006)
Bay Area	31	37	12021	0.015 (0.001)	0.013 (0.001)
Humboldt Bay	10	5	6466	0.014 (0.001)	-0.002 (0.002)