Title: Genomic insights into the origins of diversity in a California endemic songbird (*Chamaea fasciata*)

Authors: Michael A. Tofflemire[1,2*], Phred M. Benham[3], Carla Cicero[3], Michelle Davila[3], Lydia Smith[3,4], Erik D. Enbody[5,6], Dino N. Barhoum[1], Rauri C. K. Bowie[3,4]†, and Kevin J. Burns[1]†

Author affiliations:

(1) Department of Biology, San Diego State University, San Diego, CA, USA

(2) Department of Evolution, Ecology and Organismal Biology, University of California, Riverside, CA, USA

(3) Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA

(4) Department of Integrative Biology, University of California, Berkeley, CA USA

(5) Biomolecular Engineering Department, University of California, Santa Cruz, CA, USA

(6) Department of Computational Biology, Cornell University, Ithaca, New York, USA.

†Joint senior authors

*Corresponding author: mtofflemire@gmail.com

Abstract

Identifying the processes shaping spatial patterns of genomic variation among species is important for conserving biodiversity. The Wrentit (*Chamaea fasciata*) is an endemic, sedentary

songbird inhabiting lowland chaparral habitats across the California Floristic Province. We used a whole genome landscape phylogeographic approach for understanding the most important drivers of spatial genomic variation (e.g., isolation by history, isolation by distance, or isolation by environment) and for identifying signatures of selection across climatic gradients. We found support for two genomic ancestral populations geographically divided by an axis running from northwest to southeast across the California Floristic Province. We also found two putative contact zones, one near the San Francisco Bay Area and another where the southern Sierra Nevada meet the southern California mountains. Admixture is minimal and confined primarily along the central California coast. Genomic diversity is higher among southern populations compared to the north, indicating a stable southern ancestral origin. Our demographic divergence models best supported a Pleistocene founder event >1.4 million years ago of the northern populations from the southern populations. Isolation by history (e.g., population structure) was identified as the most important driver of genomic variation. Nonetheless, we identified signatures of adaptive variation across the range of the Wrentit, with significant outlier loci correlating strongly with increased precipitation gradients along the northern California and Oregon coasts. These outliers mapped to genes associated with synapses, neuron development, and immune system responses. Isolation by environment and isolation by distance patterns shaped genomic diversity to some degree, but demographic history from historical colonization during the Pleistocene is the primary driver of variation across the California Floristic Province.

Keywords: Whole genome sequencing, Wrentit, *Chamaea fasciata*, California Floristic Province, landscape genomics, phylogeography, demography, ecological niche model, genotype-environment associations

Introduction

Understanding the evolutionary processes that shape the distribution of genetic variation within species is important for helping to identify and maintain biodiversity and inform conservation strategies (Aitken et al., 2024; Andrello et al., 2022; Exposito-Alonso et al., 2022; Meek et al., 2023). Spatial patterns of genomic diversification are often shaped by isolation by distance (IBD; Wright, 1943), isolation by environment (IBE; Orr & Smith, 1998), or historical isolation (Mayr, 1999), hereafter referred to as isolation by history (IBH; Moreira et al., 2020). IBD describes a pattern where gene flow between populations decreases as geographic distance increases, leading to the accumulation of genetic variation among geographically distant populations (Wright, 1943). Alternatively, IBE occurs when population differentiation is correlated closely with environmental conditions resulting from local adaptation as opposed to geographic distance (Sexton et al., 2014). Lastly, IBH reflects genetic differentiation that accumulates when

70 populations become physically isolated as a result of historical vicariance events, including
71 mountain uplift, historical seaways, or climatic events that fragment populations and reduce gene
72 flow (e.g., Bryson Jr. et al., 2016). Identifying the relative contribution of IBD, IBE, and IBH in
73 shaping the spatial distribution of genetic variation provides crucial insight into the mechanisms
74 that promote and maintain genetic diversity (e.g., Moreira et al., 2020).

75 Selective and non-selective evolutionary forces shape genome-wide patterns of genetic
76 variation (Kern & Hahn, 2018), which can be reflected in patterns of IBD, IBH, and IBE. Natural
77 selection acts on fitness-related variation, often contributing to spatial genetic patterns associated
78 with IBE (Nei et al., 2010; Orr, 2009). In contrast, non-selective forces—including genetic drift,
79 mutation, and gene flow—contribute to genome-wide differentiation through neutral
80 demographic processes and are often associated with patterns of IBD or IBH (Charlesworth et
81 al., 2003; Kimura, 1985). Reduced representation sequencing approaches such as genotype by
82 sequencing, restriction site associated DNA, and target capture sequence only a small fraction of
83 the genome, but are generally useful for exploring demographic patterns driven by neutral
84 variation (Davey & Blaxter, 2010). Signatures of IBE can also be detected in these smaller
85 datasets if gene flow is reduced enough between environments due to local adaptation, but
86 reduced representation datasets are limited in their ability to identify genes involved in local
87 adaptation (Shafer & Wolf, 2013). In contrast, whole-genome resequencing of individuals allows
88 for the detection of genome-wide variation including adaptive alleles under weak selection,
89 allowing for a more explicit test of IBE patterns shaped by climatic gradients (Forester et al.,
90 2018; Hoban et al., 2016). Critically, whole-genome sequencing offers the strongest insights into
91 selective and non-selective spatial processes, and consequently species' capacity to respond to

current and future anthropogenic-induced climate change (Aitken & Whitlock, 2013; Alberto et al., 2013; Sgrò et al., 2011; Sork et al., 2013).

The California Floristic Province is a globally recognized biodiversity hotspot in the North American West known for its high richness of terrestrial plant and vertebrate species (Myers et al., 2000), making it ideal for exploring patterns of species diversification. Spatiotemporal patterns for many of these species are proposed to have been shaped by the region's unique topography, high number of distinct ecoregions, and climatic history, which includes a series of glacial cycles during the Pleistocene (Raven & Axelrod, 1974). Emerging biogeographical patterns indicate that several ecoregions consistently harbor elevated levels of genetic and phenotypic diversity (Schierenbeck, 2014). For example, numerous taxa have shown consistent phylogeographic breaks at or near the Transverse/Peninsular Ranges in southern California, indicating a possible shared history at the community scale (Calsbeek et al., 2003; Chatzimanolis & Caterino, 2007). Recently, a growing number of whole genome landscape genomics studies of California Floristic Province vertebrate and invertebrate species have allowed for the robust assessment of their genomic diversity patterns and signatures of selection, providing important insight into preventing loss of biodiversity (Heraghty et al., 2023; Shaffer et al., 2022; Song et al., 2025; Wooldridge et al., 2024). However, genomic studies of birds in this region remain sparse, with most prior studies relying on neutral markers from mitochondrial DNA (mtDNA) or reduced-representation datasets (Alexander & Burns, 2006; Burns et al., 2007; Burns & Barhoum, 2006; Sgariglia & Burns, 2003), thus restricting their ability to infer complex processes such as historical demography and adaptive evolution across the landscape.

The Wrentit (*Chamaea fasciata*) is a non-migratory bird species endemic to the California Floristic Province and is primarily restricted to low elevation shrubland habitat

(Grinnell & Miller, 1944). It has an especially limited dispersal ability with a natal dispersal of < 400 m (Baker et al., 1995), and individuals spend their entire lives on a single territory, in some cases for over 10 years (Erickson, 1938). The biogeographic origin of the Wrentit is unique: this species is the sole North American member of a large clade of over 600 species, with its nearest relatives thousands of miles away in Asia. Previous research shows the Wrentit split from these relatives around 10 million years ago (Cai et al., 2019). Since its arrival in North America, the Wrentit has diversified both genetically and phenotypically (Bowers, 1960; Burns & Barhoum, 2006), but the factors underlying this diversification are unknown. Across the range of *C. fasciata*, strong climatic gradients in humidity have been proposed to facilitate the development of locally adapted traits among populations (Bowers, 1960; Grinnell & Miller, 1944). As a result, populations distributed across humid environments along the California coast and in Oregon exhibit darker plumage patterns, whereas populations in drier climates of southern and inland California correspond to lighter plumage coloration. Thus, patterns of phenotypic and genetic variation may be associated with environmental gradients resulting from adaptations to unique selective pressures in these regions. For example, previous work suggests variation in plumage darkness in birds may be an adaptive response to selective pressures associated with higher parasite loads in humid environments (Delhey, 2019). Five subspecies of the Wrentit are currently recognized (Clements et al., 2024) and are distinguished largely on the basis of plumage color differences across their range (Browning, 1992; Dickinson et al., 2004; Geupel & Ballard, 2023; A. R. Phillips, 1986). Briefly, these five subspecies are distributed as follows: (1) *Chamaea f. phaea*, Klamath-Siskiyou region of coastal northwestern Oregon (Osgood, 1899); (2) *Chamaea f. margra*, interior of southern Oregon in the southern Cascade Range (Browning, 1992); (3) *C. f. rufula*, chaparral in northwestern to west-central California along the northern

Coast Ranges (Ridgway, 1902); (4) *C. f. fasciata*, central coastal California from Monterey County to central San Luis Obispo County (Gambel & Ziegler, 1845); and (5) *C. f. henshawi,* inner southern Coast Range and western slope of the Sierra Nevada through coastal southern California (including the Transverse/Peninsular Ranges) to northern Baja California (Ridgway, 1882).

For this study, we obtained whole-genome sequences of *C. fasciata* individuals from across their distribution within the California Floristic Province to address questions about the processes driving spatial genetic differentiation and the importance of environmental gradients in shaping their genetic and phenotypic diversity. We leveraged thorough sampling to apply a landscape genomic approach for characterizing spatial genetic patterns across populations of *C. fasciata.* Our overarching objectives were to: (1) establish a landscape picture of genomic population structure, diversity, and gene flow, (2) investigate the demographic processes that have shaped the present-day distribution of *C. fasciata*, and (3) test for genome-environment associations to infer adaptive processes across the landscape. Using a whole-genome landscape genetic approach, we assess the relative contributions of isolation by distance (IBD), isolation by history (IBH), and isolation by environment (IBE) patterns on shaping spatial genetic variation across the landscape. We hypothesize that environmental heterogeneity has promoted the formation of locally adapted populations within *C. fasciata*, leading to IBE spatial genetic patterns. Alternatively, findings of IBD and IBH patterns would suggest that non-selective demographic processes may instead be the primary drivers of genomic variation resulting from the region's unique geological history.

Methods

Sampling

We sampled tissues from 159 individuals representing the five subspecies of *C. fasciata* from 80 localities in California and Oregon (n=151 samples from California and n=8 samples from Oregon). All tissue samples are vouchered by museum specimens (see Supplemental Table S1). DNA was extracted from tissue samples using the Qiagen DNeasy extraction kit and following the manufacturer's protocols (Qiagen Inc., Valencia, California, USA). Libraries were prepared using the Illumina Kapa preparation kit and submitted for whole genome resequencing on 3 Illumina NovaSeq platform 150 PE S4 lanes (see Benham et al. 2025a for details on wet lab methods).

Read filtering, mapping, and SNP calling

We used the snpArcher pipeline to perform adapter trimming, read mapping, and variant calling (Mirchandani et al., 2023). As part of this workflow, raw sequencing reads were quality-filtered using fastp v0.20.1 (Chen et al., 2018) with automatic adapter detection for paired-end reads. Reads were aligned to a single *C. fasciata* reference (Benham et al., 2025b; GenBank accession no. GCF_029207755.1) genome using Sentieon's optimized BWA implementation (Kendig et al., 2019) with the -M flag and batch processing size of 10,000,000 (-K 10000000), followed by sorting with maximum BAM compression. BAM files were merged with samtools (Li et al., 2009) and duplicate reads were identified using Sentieon's two-pass algorithm (LocusCollector for scoring followed by Dedup for marking). Per-sample variant calling was performed with Sentieon Haplotyper in GVCF mode using the multinomial genotype model (--genotype_model multinomial), emit and call confidence thresholds of 30, and ploidy set to 2. Joint genotyping was executed using Sentieon GVCFtyper with variant-only emission mode (--emit_mode VARIANT). Variants were filtered using GATK v4.1.8.0 VariantFiltration (McKenna et al., 2010) following best practices (Van der Auwera et al. 2013) with the following criteria:

ReadPosRankSum < -8.0 for SNPs and < -20.0 for indels, QD < 2.0, FS > 60.0 or SOR > 3.0 for SNPs, FS > 200.0 or SOR > 10.0 for indels, MQ < 40.0, MQRankSum < -12.5, and QUAL < 30.0. As part of the quality control and postprocessing of VCFs, snpArcher uses Plink2 v2.00a2.3 (Chang et al. 2015) and ADMIXTURE (Alexander et al. 2009). The workflow was orchestrated using Snakemake (Mölder et al. 2021).

We applied several additional filters including removing all indels, non-biallelic SNPs, SNPs with a minor allele frequency (maf)<0.01, SNPs with >75% missing data, and samples with <2x sequencing depth. This filtered dataset is referred to throughout as our 'high-quality' SNP dataset and includes 18,441,292 SNPs in total. For our downstream analysis of the functional impact of our SNPs, we used SnpEff v4.1 (Cingolani et al., 2012) to annotate SNPs according to the gene annotation of a single *C. fasciata* reference (GCF_029207755.1).

Population structure

To identify population genetic structure, we filtered our high-quality SNP dataset for sites with minor allele frequency (maf) threshold >0.05, with a p-value >0.01 for exact test of Hardy-Weinberg Equilibrium and retained only autosomal scaffolds. To account for linkage disequilibrium (LD), we randomly selected SNPs separated by a minimum distance of 1 kb with 'bcftools +prune -n 1 -N rand -w 1 kb', resulting in a total of 817,234 SNPs. This pruning window was informed by our analysis of linkage disequilibrium decay using popLDdecay (Zhang et al., 2019; Figure S1). We first performed a principal component analysis (PCA) on all 159 *C. fasciata* samples with PLINK v1.9 (Chang et al., 2015). In addition, we used the software ADMIXTURE v1.3 (Alexander et al., 2009) to estimate ancestry fractions and infer the most likely number of populations in our dataset. To identify the optimal value for K, we used ADMIXTURE's cross-validation (cv) procedure with the —cv option to perform a 10-fold cv on

a range of K values (1 to 10). In addition, we performed a discriminant analysis of principal components (DAPC; Jombart et al., 2010).

We estimated genome-wide F_{ST} between genetic clusters identified from the ADMIXTURE and DAPC results using the Weir and Cockerham's F_{ST} estimator (Weir & Cockerham, 1984) in VCFtools (Danecek et al., 2011). For each pairwise comparison, we calculated non-overlapping 10 kbp windows.

For phylogenetic inference, we constructed a Maximum Likelihood (ML) tree in the IQ-TREE software (Nguyen et al., 2015). The most likely substitution model was identified using the ModelFinder software (Kalyaanamoorthy et al., 2017) in IQ-TREE. Estimates of branch support were obtained with the 1000 ultrafast bootstraps (Hoang et al., 2018). The ML Tree was rooted using a single Vinous-throated Parrotbill sequence (*Suthora webbiana;* SRR8271220).

Genomic diversity

Genome-wide nucleotide diversity (π) and Tajima's D was estimated in 10kbp non-overlapping sliding windows with VCFtools v0.1.17 (Danecek et al., 2011) for each genomic cluster identified from our structure analysis, using the --window-pi 10000 option, and the --TajimaD 10000 option, respectively. We reported genome-wide values by calculating the median across all windows. We estimated observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (F_{IS}) for each *C. fasciata* individual using PLINK v1.9 (Chang et al., 2015) with the --het option.

Diversity rate (q) and migration (m) rate estimates were visualized across the landscape using estimated effective migration surfaces (EEMS; Petkova et al., 2016). For the input data, we calculated a pairwise identity-by-state (IBS) matrix with PLINK v1.9 using our high-quality SNP dataset with sex-linked scaffolds removed. EEMS settings were run using 400 demes, a single

MCMC chain with 1×10^7 iterations, and a burn-in of 5×10^6. We checked posterior probabilities to ensure convergence.

Climatic niche modeling

To understand the role of history in shaping landscape diversity patterns in *C. fasciata*, we constructed climatic niche models to predict contemporary distributions (~1960–1990; spatial resolution of 2.5 arcminutes) in addition to historical distributions at the Last Glacial Maximum (LGM; ~21,000 years BP; 2.5 arcminutes), Last Interglacial (LIG; ~120,000 –140,000 years; 30 arc-seconds) and Mid-Holocene (12,000 years BP; 2.5 arcminutes; Watanabe et al., 2010). Models were constructed using the maximum entropy (MaxEnt; Phillips et al., 2004) algorithm with the R package dismo (Hijmans et al., 2017). We used 19 bioclimatic layers downloaded from the WorldClim database (Hijmans et al., 2005). Occurrence data from vouchered specimens was used and totaled 159 observations. Data were cleaned to remove replicate localities, which reduced the total number of observations to 80.

We ran an initial MaxEnt model using all 19 bioclimatic variables downloaded from WorldClim based on contemporary conditions. To avoid model overfitting, we excluded highly correlated variables based on Pearson's correlation coefficient (r) tests using the corrplot package in R. For each highly correlated pair of variables (e.g., $r \geq 0.90$), the variable with the larger percent contribution to the model was retained. Final models were run using default software settings and taking the average AUC (area under the receiver operating characteristic curve) values of 10 replicates. The models were then projected to present-day climate conditions and LIG, LGM, and mid-Holocene historical conditions.

Demographic history

We used the program ∂a∂i (Gutenkunst et al., 2009) to estimate demographic parameters including divergence time (*T*), changes in effective population sizes (Ne), and the timing and levels of gene flow (*m*) experienced among distinct genetic clusters of *C. fasciata.* We filtered our raw SNP dataset for the largest scaffold (JARCOQ010000001.1) and kept SNPs with <0% missing data bringing the total number of SNPs to 4,804,850. We produced a folded joint site frequency spectrum (jSFS) using the easySFS (https://github.com/isaacovercast/easySFS) Python script, projecting samples to 168 and 150 chromosomes for the northern and southern genetic clusters, respectively (proj=[168, 150]). We tested phylogeographic patterns of diversification in *C. fasciata* by fitting eight divergence models: four vicariance and four founder event scenarios. In each scenario, a single ancestral population splits into two, with fraction 's' going into population 1 (e.g., northern clade) and fraction (1-s) going into population 2 (e.g., southern genetic cluster). For both the vicariance and founder event models, we varied the timing of when migration occurs (strict isolation [SI]; isolation with continuous migration [IM]; secondary contact [SC]; and ancient migration [AM]). For our four vicariance scenarios, both s and (1-s) remain constant following the initial split. For our four founder event scenarios, the ancestral population also splits into two, with a fraction 's' going into the northern clade (the founder population) but now allowed to grow exponentially to size Ne1, while a fraction (1-s) goes into the southern genetic cluster (the mainland population) but remains constant. Choosing the northern clade as the founder population was informed by the ML tree topology (see results). To optimize parameters for each demographic model, we performed 20 optimizations with a maximum number of iterations set to 10. For each run, starting parameters were initially defined based on biologically reasonable values and perturbed using the perturb_params() function. Parameters were optimized using the Nelder-Mead method (Nelder and Mead 1965). We defined

our extrapolation grid size values as 170, 180, and 190 to ensure accuracy of the diffusion approximation. To ensure that we found the global optimum, we performed a final optimization routine with a maximum number of iterations set to 50, using as our starting parameters those found from the previous optimizations with the lowest log-likelihood score. The set of parameters with the lowest AIC score among all tested models was chosen as the best overall demographic model. To convert parameters, we used the equation $\theta = 4N_e\mu L$, where L was estimated by multiplying the total base pairs from our callable sites by the fraction of SNPs used in our analysis after projecting down. We used the average mutation rate for the Collared Flycatcher (*Ficedula albicollis;* $\mu = 2.3 \times 10^{-9}$; Smeds et al., 2016). Uncertainty in parameter estimates was assessed using 100 parametric bootstrap replicates of our original data. To convert the divergence time parameter into years, we used a generation time of 2.6 years (IUCN 2019).

Spatial drivers of genomic differentiation

To identify which spatial processes or combination of processes best explain the observed patterns of genomic variation in *C. fasciata*, we used multiple matrix regression with randomization (MMRR; Wang, 2013). We tested the extent to which genomic variation was explained by geographic distance (IBD), environmental variables (IBE), and genetic structure, which we used as a proxy for isolation by history (IBH). Pairwise genetic distances were calculated using an identity-by-state (IBS) matrix with the PLINK package, while Euclidean distance matrices were generated from sampling coordinates. A multivariate distance matrix for the six environmental variables from our RDA analysis (described in the section below) was also calculated using Euclidean distance in R. Finally, to account for the effects of genetic structure, we calculated pairwise distances using PC1 scores from our genomic PCA analysis, as PC1 captured the largest axis of genomic variation and aligned closely with the spatial structure

298 inferred from our ADMIXTURE results. For each process, we performed a multiple matrix
299 regression using the function provided by Wang et al. (2013). Significance of each model was
300 tested using 1,000 permutations. To disentangle the combined effects of each variable, we
301 calculated commonality coefficients using the yhat package (Nimon et al., 2013) in R. We
302 estimated 95% confidence intervals for each coefficient using 1,000 bootstrap replicates with the
303 boot() function (Canty & Ripley, 2017).
304 *GEA analysis*
305 To investigate genotype-environment associations (GEA), we performed a redundancy analysis
306 (RDA) using the vegan package (Oksanen et al., 2007) and a latent factor mixed modeling
307 (LFMM) approach with the LEA package (Frichot & François, 2015). RDA is a multivariate
308 linear regression method that uses constrained ordination to identify significant associations
309 between response variables (e.g., genomic data) and predictor variables (e.g., environmental
310 variables). In this study, we aimed to detect associations between our SNP dataset and 19
311 bioclimatic variables obtained from WorldClim. To account for sites in linkage disequilibrium
312 (LD), which can lead to an excess of false positives, we thinned our high-quality SNP dataset for
313 LD by randomly selecting SNPs separated by a minimum distance of 1 kb. We filtered for minor
314 allele frequency <0.05 and variants with >25% missing data. Finally, we filtered to include only
315 autosomal scaffolds. To account for covariance among different climatic variables, we conducted
316 a PCA on all 19 variables with temperature and precipitation variables run separately (e.g.,
317 Moreira & Smith, 2023). The scores from the first three principal component axes (PC1–PC3)
318 for temperature and precipitation, which represented over 80% of the combined total variation,
319 were used as explanatory variables. To account for population structure, a partial RDA (pRDA)
320 was conducted by incorporating the first two PCA axes from our structure analysis as covariates

(set K = 2). Since RDA does not allow missing data, we imputed missing genotypes by replacing missing values with the most frequently observed genotype at each SNP across all individuals. Outlier loci in the pRDA were identified using a similar approach as that proposed by Capblancq et al. (2018) and Luu et al. (2017). Briefly, we used Mahalanobis distances calculated from z-transformed SNP loadings from the first two RDA axes. Multiple testing correction was applied using the False Discovery Rate (FDR) approach (Benjamini & Hochberg, 1995), and SNPs with q-values <0.01 were considered outliers. The significance of the model, RDA axes, and climatic variables were tested with the anova.cca() function and 999 permutations.

We performed a second RDA using outlier SNPs that were previously detected with a q-value threshold < 0.01. This analysis was conducted to visualize an "adaptively enriched" genetic space (Capblancq et al., 2020). We took the scores of each environmental variable across the first two RDA axes and calculated indices to predict adaptive scores for each individual using the equation: adaptive index = $\sum a_i b_i$, where a represents the variable score along the relevant RDA axis, b is the value of the standardized environmental variable at the relevant location, and i refers to the environmental variables used in our RDA model (Steane et al., 2014). We then projected the adaptive index scores across the landscape of *C. fasciata* for the first two RDA axes.

To complement the RDA results, we performed an LFMM analysis. LFMM is a univariate linear regression method with high power and low rates of false discovery (Caye et al., 2019). We tested for significant associations between the pruned SNP dataset and the same environmental variables used in the pRDA. The latent factor (K-value) was set to K = 1 to account for population structure in *C. fasciata*, and the "ridge" method was applied for outlier detection. A significance threshold of $p < 0.05$ was used to identify candidate SNPs.

Signatures of selective sweeps

Pairwise F_{ST} sliding windows were calculated between major genomic clusters using VCFtools (Danecek et al., 2011) to detect loci under selection between structured populations. We removed SNPs with a minor allele frequency <0.05, SNPs with >25% missing data, and isolating autosomal scaffolds (Nosil et al., 2009). To explicitly examine signatures of selection across the genome at the putative expanding edge, we also calculated pairwise F_{ST} sliding windows between the Oregon coast population and all other population across *C. fasciata*. For each calculation, we used a 50kbp sliding window size and a 10kbp window step size. The significance threshold was set to 5 standard deviations (SD).

PCAdapt (Luu et al., 2017) was used as a secondary method for detecting outlier loci using the same filtered SNP dataset as the pairwise F_{ST} analysis described above. We set K = 1, following recommendations from the user manual. Outlier loci under strong selection were identified using the q-value correction for p-values, setting the alpha threshold to 0.01. SNPs at the union of both PCAdapt and F_{ST}-outlier detection methods were considered candidate SNPs under positive selection.

GO Enrichment

We assessed the functional impact of annotated genes using Gene Ontology (GO) enrichment analysis via the PANTHER classification system (https://pantherdb.org). For testing the statistical significance of enriched genes, we used Fisher's exact test with a threshold<0.05.

Results

Population structure

Our ADMIXTURE analysis support two geographically congruent genetic clusters, as K= 2 yielded the lowest cross-validation error of 0.50221 (Figure 1; Figure S2). This method clusters

367 individuals into a “northern” group (e.g., samples from the Oregon Coast south through the
368 California Coast Range to San Francisco Bay, and inland from northern California through the
369 Cascade Range and Sierra Nevada) and a “southern” group (e.g., samples from the east San
370 Francisco Bay Region and central California Coast Range to southern California). Admixed
371 individuals occur in one population in northwestern coastal California, around San Francisco
372 Bay, across the California Central Valley from the bay to the Sierra Nevada foothills, and in the
373 central California Coast. The division between the two groups occurs in a line running from the
374 San Francisco Bay southeast to where the southern Sierra Nevada approaches the Transverse
375 Range.

376 Our principal component analysis (PCA) similarly recovered two major clusters along
377 PC1 (Figure 2a; Figure S3). The first three principal components (PCs) represented 24.01%
378 (PC1), 9.43% (PC2), and 8.62% (PC3) of the total variation in the dataset. Interpolation of PC1
379 scores across the landscape yielded a similar break between the northern and southern clusters
380 running from the San Francisco Bay region to the boundary between the southern Sierra Nevada
381 and the Transverse Range (Figure 2b). The average genome-wide F_{ST} between northern and
382 southern genomic clusters was 0.0734, indicating a moderate level of differentiation. Our DAPC
383 results showed a similar clustering pattern as both ADMIXTURE and PCA.

384 The ML phylogenetic tree constructed from IQ-TREE showed a topology consistent with
385 our ADMIXTURE and PCA results (Figure 3). Individuals from the northern cluster form a
386 strongly supported monophyletic group, whereas individuals from the southern cluster are a
387 paraphyletic assemblage formed by successive basal splits. The clustering of all southern
388 individuals near the root likely indicates a southern ancestry, with the northern cluster being
389 more recently derived in the evolutionary history of *C. fasciata*. This pattern of paraphyly might

390 also be influenced by incomplete lineage sorting. Other well supported geographic groups
391 include most samples from the Northern California Coast Ranges, Oregon Coast, Northern
392 California Coast, Central Valley, and the Sierra Nevada and Sierra Nevada Foothills (Figure 3).
393 All samples from the Oregon Coast and several samples from the Klamath ranges clustered
394 together and largely correspond with the *C. f. phaea* subspecies. Additionally, all samples from
395 the Central California Coast population formed a monophyletic group and likely correspond with
396 the *C. f. fasciata* subspecies. Similarly, most of the samples from the Northern California Coast
397 populations (representing *C. f. rufula*) clustered together. In contrast, *C. f. henshawi* – which is
398 found from Baja California north through the Sierra Nevada foothills into northern California -
399 spans both sides of the most significant genetic break identified with our genomic data. We only
400 have one sample representing the *C. f. margra* subspecies in Oregon. This individual was not
401 part of the exclusive clade of the other Oregon samples that all represent *C. f. phaea,* a finding
402 consistent with the idea that *C. f. margra* is a separate subspecies from *C. f. phaea*.
403 *Genomic diversity*
404 Estimated effective migration surface (EEMS) results performed well, qualitatively reaching
405 convergence. Migration rate estimates (m) revealed significantly higher connectivity and less
406 fragmentation among individuals across southern California compared to those from the north
407 (Figure 4a). Migration rate estimates are low ($\log(m) < 0$) where the putative contact zone occurs
408 between northern and southern genetic clusters near the San Francisco Bay Area. Likewise,
409 significantly low $\log(m)$ values ($P > 0.95$) occur between individuals from Southern California
410 and the Sierra Nevada Foothills. EEMS diversity rates (q) show significantly high q values ($P >$
411 0.95) among samples across the southern California ecoregions, including the central California
412 Coast, southern California mountains and valleys, and Colorado Desert (Figure 4b), coinciding

413 with samples from the southern genetic cluster. Conversely, significantly low log(q) estimates (P
414 > 0.95) were found among individuals across Coastal Oregon, Northern California Coast, and the
415 Sierra Nevada ecoregions.

416 Genome-wide nucleotide diversity (π) estimates were higher for the southern genetic
417 cluster compared to the northern cluster (Table 1). F-statistics for inbreeding coefficient (F_{IS}) and
418 Tajima's D (D) estimates were highest among individuals within the northern cluster compared
419 to the southern cluster (Table 1), but Tajima's D was still positive in both lineages, possibly
420 reflecting a population currently in decline or balancing selection if significantly different from
421 zero. Observed heterozygosity (Ho) was highest among individuals of the southern cluster (Table
422 1).

423 *Climatic niche models*

424 Climatic niche models constructed with MaxEnt performed well, with mean AUC values > 0.9
425 across all models (Figure 5). Precipitation of the coldest quarter (BIO19; 41.2%) and
426 precipitation seasonality (BIO15; 33.7%) represented the bioclimatic variables with highest
427 percent contributions. Paleodistributions of the Last Interglacial show suitable habitat at higher
428 latitudes during warming events compared to the Last Glacial Maximum (Figure 5a, b).
429 Projections of the model onto the mid-Holocene (Figure 5c) and contemporary distribution
430 (Figure 5d) showed a pattern of expansion of highly suitable habitat back into northern latitudes
431 and interior regions during the most recent glacial warming event. The climate model projected
432 onto the contemporary distribution showed high overlap with currently described range
433 boundaries of *C. fasciata* and a high probability of occurrence along coastal environments.

434 *Demographic history*

435 The best demographic model inferred with ∂a∂i for *C. fasciata* was the founder event with
436 continuous asymmetric migration (Table 2). After converting parameters into absolute estimates
437 and calculating 95% confidence intervals (CI) from 100 bootstrap replicates, the best model
438 estimated an ancestral effective population size (Ne) of ~682,071 individuals (95% bootstrap CI:
439 664,000-721,000). The initial divergence time estimate between the northern and southern
440 genetic clusters is dated to ~1,428,050 years ago (1,359,800-1,534,000). The founder event
441 experienced by the island population (i.e., northern cluster) was estimated to have undergone a
442 severe reduction in Ne compared to the ancestral population, with an estimated Ne of 4,983
443 individuals (3,710-4,970). Following exponential growth, the current estimated Ne is ~293,137
444 individuals (280,000-297,000). Conversely, the southern genetic cluster shows a significantly
445 higher estimated Ne of ~677,088 individuals (659,000-716,000). Finally, migration rates were
446 higher in the north-to-south direction compared to south-to-north, with the number of
447 migrants/generations estimated to be 4.3033 (4.13-4.50) and 0.0288 individuals (0.02-0.03),
448 respectively.
449 *Important drivers of spatial genomic variation*
450 The regression model explained 80.16% of the overall genetic variation (Table 3). The pure
451 population structure model, which we used as a proxy for patterns of isolation by history (IBH),
452 explained the largest proportion of overall genetic variation (47.55%) out of all variables tested
453 (Table 3). Geographic distance (IBD) and structure together ranked second in explaining overall
454 genetic variation (34.39%). Interestingly, pure IBD explained the lowest of our 'pure' models,
455 explaining only 0.01% of total genetic variation, whereas pure environment explains 1.60%,
456 which is more than an order higher than that explained by IBD alone. Given this pattern, the
457 proportion of variance explained by geographic distance is being driven primarily by historical

458 population structure. Thus, environmental factors play a larger role than geographic distance in
459 explaining genetic variation, yet history remains the dominant process shaping genetic variation
460 in *C. fasciata*.

461 *GEA analysis*

462 We identified a clear climatic gradient in precipitation and temperature across the range of *C.*
463 *fasciata* (Figure S4). The pRDA full model was found to be highly significant (P(> F) between
464 0 and 0.001). Each of the six environmental terms used as explanatory variables in this model
465 were also found to be highly significant and accounted for 2.926113% of the total variation.
466 While the total explained variation is relatively low, this is expected as we have a high number of
467 loci expected to follow neutrality and not be driven explicitly by selection. We identified a total
468 of 3,194 outlier SNPs deviating significantly from the background distribution of SNP loadings
469 (FDR<1%). The most important variable driving genomic variation was temperature PC1
470 (associated with 1343 candidate loci), followed by precipitation PC3 (1096 candidate loci), and
471 precipitation PC2 (327 candidate loci; see Table S2 for bioclimatic variables contributing most
472 strongly to each retained PC axis). Our results from the "adaptively enriched" genetic space
473 show that environmental variables related to temperature PC1 and precipitation PC3 loaded most
474 strongly on the RDA1 axis, separating populations from the Oregon Coast, Klamath Mountains,
475 and Northern California Coast from those further south and inland in the Southern California
476 Mountains and Valleys and the Sierra Nevada Foothills (Figure 6).

477 Our LFMM analysis identified a total of 6,295 outlier SNPs associated with precipitation
478 and temperature variables (Figure S5). Precipitation variables accounted for the largest number
479 of SNP associations (6,223), followed by temperature (72), indicating precipitation variables as
480 important environmental factors. For precipitation, most candidate SNPs were associated with

PC1 (4,871), followed by PC3 (1,333), and PC2 (19). For temperature, most candidate SNPs were associated with PC1 (44), followed by PC3 (24) , and PC2 (4).

A total of 505 candidate genes overlap between both the pRDA and LFMM analyses, and 412 uniquely mapped to genes recognized by the gene ontology (GO) enrichment database. The GO enrichment test found these to be significantly associated with a substantial number of biological processes, including those related to synaptic signaling (GO:0099536), nervous system development (GO:0007399), and immune system responses (GO:0002376). The processes with the largest fold enrichment were maintenance of presynaptic active zone structure (GO:0048790), presynaptic active zone organization (GO:1990709), synaptic vesicle priming (GO:0016082), cranial nerve morphogenesis (GO:0021602), exocytic process (GO:0140029), and positive T cell selection (GO:0043368; Table S3).

Signatures of selective sweeps

Our PCAdapt analysis identified 318 outlier SNPs, while 1,323 outlier windows were detected across all pairwise F_{ST} sliding window comparisons. We found 205 SNPs overlapped between the two methods, providing strong evidence that these putative candidate loci are under selection. Additionally, we identified 25 SNPs from the GEA analysis that overlapped with these outlier windows. Outlier windows associated with population differentiation between the northern and southern genetic groups were concentrated in multiple regions across the genome and showed clear divergence peaks and clusters of windows sharing elevated F_{ST}, a pattern often associated with recent selective sweeps (Figure 7a, b). Among these regions, several annotated genes were identified: NADSYN1, DHCR7, RAB3GAP2, SMARCA4, DPP7, CDKN1B, and MORC2 (see Table 4 for full list of annotated genes). Similarly, pairwise F_{ST} comparisons of samples at the putative expanding edge in Oregon consistently detected well-defined divergence peaks at

several genomic regions, which mapped outlier SNPs to the USO1 and TTN genes (Figure 7c, d, Figure S6). We describe the functional roles of these genes in our discussion below.

Discussion

Spatial patterns of genomic differentiation and landscape diversity in the Wrentit are best explained by historical colonization

Multiple lines of evidence support two genetic ancestral populations of *Chamaea fasciata* geographically divided along an axis stretching northwest to southeast across the California Floristic Province (Figures 1, 2). The northern genetic cluster consists of individuals from Oregon south to coastal ecoregions north of the San Francisco Bay Area, extending east into the interior of northern California, and south across the Sierra Nevada. Conversely, the southern genetic cluster includes individuals restricted to coastal ecoregions south of the San Francisco Bay, extending southeast across the Peninsular Ranges to northern Baja California, Mexico. The strongest signals of genetic discontinuity occur near the San Francisco Bay Area and where the southern Sierra Nevada meets the Transverse Range. The San Francisco Bay Area coincides with a single putative contact zone with moderate levels of admixture. Admixture was not detected in a potential second contact zone between the southern Sierra Nevada and southern California populations, but additional sampling is necessary.

A southern origin of current-day Wrentit populations within chapparal habitats is supported by multiple lines of evidence, including our ML tree and greater genetic diversity in the south. Our demographic model estimating divergence patterns supports a Pleistocene colonization event experienced by the northern clade with continuous asymmetric migration since diverging from a southern ancestral population >1.4 million years ago (Table 2). Historical dispersal likely occurred in the south to north direction across chaparral habitats in the California

Floristic Province during the Pleistocene, initially leading to a severe population bottleneck (Ne=~4,983) but with later demographic recovery (Ne=~293,137). Similarly, Burns & Barhoum (2006) found evidence for a southern origin for Wrentit populations based on phylogeographic assessment of mtDNA. Their haplotype network and phylogenetic tree are consistent with the north-south pattern found in this study, and they also inferred a long-distance dispersal event from southern individuals into chapparal habitats in Oregon following Pleistocene warming. These patterns are consistent with the long-term persistence of the chaparral ecosystem in California since its establishment during the late Miocene (Raven & Axelrod 1978). The onset of the Pleistocene brought cycles of glaciation events that caused major disruptions to California native plant distributions (Axelrod 1981), making a dispersal event of *C. fasciata* entirely plausible.

We found that historical processes (IBH; i.e., population structure) best explained genomic differentiation among *C. fasciata*, explaining nearly half of the overall variation. While our historical climate models do not coincide with the estimated divergence time of >1.4 million years ago between our northern and southern groups, our models provide a proxy for understanding broader climatic shifts that occurred across lowland habitats in this region during Pleistocene glaciation events, suggesting the important role that history played in shaping spatial genetic variation of *C. fasciata*. Climate models of historical changes in Mediterranean climates reconstructed for the Last Interglacial (~120,000-140,000 years ago), LGM (~21,000 years ago) and mid-Holocene (~12,000 years ago) showed severe habitat modification and fragmentation across the California Floristic Province. During glaciation events (e.g., Last Glacial Maximum), suitable habitats for *C. fasciata* were restricted primarily to southern California and along coastal ecoregions but expanded into northern latitudes during subsequent glacial warming (e.g., Last

550 Interglacial, mid-Holocene). Assuming *C. fasciata* individuals tracked changes in suitable habitat
551 during Pleistocene warming events, our climate models are consistent with post-glacial range
552 expansion by a southern ancestral source population, followed by historical colonization into
553 chaparral habitats across the northern California coast and Oregon. Consistent with this result,
554 nucleotide diversity (π) is higher in the southern genetic group (~0.32%) than in the northern
555 group (~0.28%) (Figure 4; Table 1), a sign of population stability in southern California over
556 time. While continuous asymmetric migration following dispersal into northern California was
557 inferred from our ∂a∂i analysis, discontinuities in historical distributions may have limited gene
558 flow and facilitated moderate genetic differentiation among populations (F_{ST}=0.0734). Moreover,
559 the formation of the Carquinez strait ~600 ka near the San Franscisco Bay roughly corresponds
560 to the location of the northern and southern genetic clusters (Sarna-Wojcicki et al., 1985) and is
561 often attributed to many phylogenetic breaks among taxa from this region, possibly playing a
562 role in limiting gene flow. Alternatively, genetic drift resulting from founder events is often
563 strong enough to counteract the effects of gene flow (Excoffier et al., 2009).
564 Isolation by distance (IBD) and IBH together represented the second most important
565 variable set, explaining 34.39% of genetic variation. The pure IBD model only explains 0.01% of
566 the overall genetic variation, suggesting IBD is largely insignificant once accounting for
567 population structure. In other words, IBD may be more important in shaping variation within
568 each genetic group, but history is more important in shaping overall genetic diversity between
569 genetic clusters. Secondary dispersal by the northern clade throughout the Oregon Coast,
570 Northern California Coast, and Sierra Nevada is likely, given that these populations have notably
571 low nucleotide diversity estimates and each form well-supported clades on our ML tree. This is
572 consistent with a recent range expansion or secondary demographic bottleneck at the range

periphery following postglacial dispersal. MaxEnt climate models confirm that suitable habitat in these regions emerged only after the Last Glacial Maximum, supporting recent range expansion and IBD patterns at the range edge.

Finally, environmental differences only explained a small portion of the genetic variation, with the total contribution of IBE ~1.60%. Still, IBE represents a significantly larger proportion than a pure IBD model, indicating that the environment alone has some influence on shaping genetic patterns within *C. fasciata* as opposed to strictly neutral demographic processes. Moreover, we identified a clear climatic gradient in precipitation and temperature across the range of *C. fasciata* (Figure S4). Despite these environmental differences, IBH is still significantly more important in explaining overall genetic variation than either IBE or IBD. Overall, historical factors resulting from Pleistocene glacial/interglacial events likely played an important role in shaping the IBH pattern of genetic variation in *C. fasciata*, and is a common trend found in vertebrate species across in this region (see Calsbeek et. al., 2003).

Climatic gradient in precipitation drives selection on traits related to cognitive behavior, immune system processes, and environmental stressors

Climatic gradients play an important role in shaping adaptive genomic variation within *C. fasciata*, with our GEA analyses showing precipitation variables dominating as a driver of this variation. The 'adaptively enriched' genetic space revealed the most important axes of adaptive differentiation separates coastal and northern populations from those with southern and interior distributions. Most outlier SNPs are significantly associated with the Oregon population, loading strongly with variables associated with wetter climates. Conversely, outliers associated with the Sierra Nevada and southern California populations correspond with drier, warmer climates. As a

result, we hypothesize that these are locally adapted loci driving genomic variation across environmental gradients in precipitation and humidity between coastal and interior habitats.

Candidate regions identified from the GEA analysis were enriched with genes associated with synaptic structure and neuron development processes and represented the majority of significant GO terms. Given that both synapse structure and neuron axogenesis can modify brain connectivity and signaling, variations in these pathways are likely to influence behavior cognition, learning, and responses to stimuli (Grant 2016; Lam et al., 2017). We suspect movement into novel environments, for example, by Wrentit populations along the Oregon and Northern California coast, may plausibly lead to selection on genes acting on synapse and neuronal functions, as adaptations to unique environments often comes with behavioral and cognitive shifts to new stimuli (e.g., Mueller et al., 2020).

Candidate genes enriched for immune system functions were also identified. We hypothesize individuals associated with elevated humidity along the coast are under a unique set of selective pressures compared to those found within the interior where the environment is substantially drier. For example, one such selective pressure along coastal environments is the cost of dealing with elevated pathogens, diseases, and/or parasites (Becker et al., 2012; Dunn et al., 2010; Guernier et al., 2004; Moyer et al., 2002; Sehgal et al., 2011). As a consequence, adaptations associated with higher immune competence (e.g., immunocompetence) to combat these pressures along the coast is likely (Horrocks et al., 2015). Moreover, we found enrichment for biological processes related to exocytosis (GO:0006887), cell junction assembly (GO:0034329), establishment of cell polarity (GO:0007163), and regulation of cellular component organization (GO:0051128) were also enriched. This finding demonstrates the

breadth and complexity of biological processes involved and the potential for polygenic effects in shaping adaptive variation across this climatic gradient.

We identified signatures of selection at the putative expanding edge in coastal Oregon localities using pairwise F_{ST}-outlier outlier detection methods. Across all pairwise F_{ST} comparisons with Oregon Coast samples to all other population, we consistently found strong F_{ST} peaks at regions for the genes USO1 (which was flanked by G2BP2) and TTN, both of which were also identified by the GEA analysis. Further, G2BP2, a protein involved in regulating RNA and protein metabolism in endothelial cells, is associated directly with stress response in organisms (Lloyd, 2016). Specifically, G3BP2 responds to environmental stressors (e.g., inflammation) through the formation of stress granules (Lloyd, 2016). Elevated stressors at the leading edge along the Oregon coast is likely, given that expansion into new environments often exposes individuals to novel selective pressures. Similarly, TTN is involved in skeletal muscle formation, responsible for building the protein known as titin, a molecule involved in structural support and elasticity (Loescher et al., 2022). Consequently, this gene region could reflect a recent adaptation in response to range expansion into considerably wetter and cooler climates at this latitudinal extreme by helping increase muscle mobility and blood flow to extremities.

Finally, candidate SNPs linked to population differentiation between the northern and southern genetic groups were identified through a combined approach using F_{ST} sliding windows and PCAdapt. We identified several significant candidate regions across the genome with qualitatively distinct divergence peaks and dense clusters of SNPs sharing high F_{ST} values, indicative of a recent selective sweep and 'hitchhiking' of variants in close proximity (Cruickshank & Hahn, 2014). Significant outlier windows ($p<0.00001$) included a coding region of the NADSYN1 gene and a close flanking upstream region near the DHCR7 gene. Previous

studies in humans found a strong correlation between vitamin D related loci (e.g., NADSYN1/DHCR7) with geographic changes in environmental ultraviolet B radiation (UVB) (Jones et al., 2020). Lower exposure to UVB in the northern range may be an adaptation by the northern lineage as they expanded their range following divergence. Several candidate genes associated with immune system responses were also identified and include DPP7 (a serine protease) and RNF216 (an E3 ubiquitin ligase), both shown to play a role in innate immune signaling (Chang et al., 2025; He et al., 2024). Finally, we found strong F_{ST} divergence for the MORC2 gene which is associated with responses to elevated levels of *Toxoplasma gondii* that is common in high humidity areas, with lower prevalence in dryer and warmer regions (Peymani et al., 2025).

Conclusions and future directions

Phenotypic variation in plumage coloration across the range of *C. fasciata* has been the primary basis of subspecies boundary descriptions. While we identified genetic patterns that matched with some subspecies boundaries, they did not correspond to the main genetic division we identified. This finding could reflect the tracking of phenotype with environmental variation while genome-wide patterns are showing historical divergence. Thus, we expected to find melanin related genes in association with environmental variation and in line with observed phenotypic variation (e.g., Ducrest et al., 2008; Hoekstra & Nachman, 2003; Mundy, 2005). While no gene ontology terms associated with melanin production were found to be enriched from our selection analyses described above, a single melanogenic gene, oculocutaneous albinism II (OCA2), was identified by our genotype-environment association outlier analysis. Previous studies have shown OCA2, a melanosomal transmembrane protein, is involved in melanin deposition and can directly influence plumage and skin color in birds (Zhang et al.,

2023). Despite neuronal and synaptic processes related to brain development representing the largest contribution of enrichment terms, the association of OCA2 with climatic variation across the landscape suggests environmental selection may contribute to geographic patterns of plumage variation. Exploring whether this variation is aligned closely with plumage color variation goes beyond the scope of our study, but we suggest future work on *C. fasciata* could benefit from more explicitly examining the relationship between genotype, plumage color, and climate.

Author Contribution

Michael A. Tofflemire, Phred M. Benham, Rauri C. K. Bowie, and Kevin J. Burns designed the study. Kevin J. Burns, Dino N. Barhoum, Phred M. Benham, Carla Cicero, and Rauri C. K. Bowie collected and prepared the majority of vouchered specimens used in the study. Lydia L. Smith and Michelle M. Davila performed DNA extractions, library preparation, and sequencing. Erik D. Enbody performed the bioinformatics processing of raw sequencing reads. Michael A. Tofflemire analyzed the data and wrote the original draft. All authors contributed to reviewing and editing the manuscript.

Acknowledgments:

We thank the following museums and staff that provided tissues for this project: Museum of Vertebrate Zoology at UC Berkeley, San Diego State University Biodiversity Museum, Natural History Museum of Los Angeles County, Museum of Wildlife and Fish Biology at UC Davis,

and the San Diego Natural History Museum. This work was supported by the California Conservation Genomics Project with funding provided to the University of California by the State of California, State Budget Act of 2019 (UC Award ID RSI-19-690224). We thank Michael W. Nachman for his contributions in securing funding for the California Conservation Genomics Project. For comments on the manuscript, we thank J. Alderson, S. Hood, E. Funk, R. Sedano-Cruz, A. Shultz, and A. Tappe.

Conflicts of Interest

The authors declare no conflict of interest.

Data Accessibility Statement

All raw sequencing reads have been deposited to NCBI Genbank and will be made publicly available upon publication of our manuscript. The BioProject associated with our study can be found under accession PRJNA1008185. All analysis scripts used in this study are available on GitHub: https://github.com/mtofflemire/chamaea_wgs

References

Aitken, S. N., Jordan, R., & Tumas, H. R. (2024). Conserving evolutionary potential: Combining landscape genomics with established methods to inform plant conservation. *Annual Review of Plant Biology*, *75*(1), 707–736. https://doi.org/10.1146/annurev-arplant-070523-044239

Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 367–388. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., Lefèvre, F., Lenormand, T., Yeaman, S., Whetten, R., & Savolainen, O. (2013). Potential for evolutionary responses to climate change – evidence from tree populations. *Global Change Biology*, *19*(6), 1645–1661. https://doi.org/10.1111/gcb.12181

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Alexander, M. P., & Burns, K. J. (2006). Intraspecific phylogeography and adaptive divergence in the White-headed Woodpecker. *The Condor*, *108*(3), 489–508.

Andrello, M., D'Aloia, C., Dalongeville, A., Escalante, M. A., Guerrero, J., Perrier, C., Torres-Florez, J. P., Xuereb, A., & Manel, S. (2022). Evolving spatial conservation prioritization with intraspecific genetic data. *Trends in Ecology & Evolution*, *37*(6), 553–564. https://doi.org/10.1016/j.tree.2022.03.003

Axelrod, D. I. (1981). Holocene climatic changes in relation to vegetation disjunction and speciation. *The American Naturalist*, *117*(6), 847–870

730 Baker, M., Nur, N., & Geupel, G. R. (1995). Correcting biased estimates of dispersal and
731 survival due to limited study area: Theory and an application using Wrentits. *The*
732 *Condor*, *97*(3), 663–674. https://doi.org/10.2307/1369175
733 Becker, A.-C., Rohen, M., Epe, C., & Schnieder, T. (2012). Prevalence of endoparasites in stray
734 and fostered dogs and cats in Northern Germany. *Parasitology Research*, *111*(2), 849–
735 857. https://doi.org/10.1007/s00436-012-2909-7
736 Benham, P. M., C. Cicero, M. M. Davila, E. D. Enbody, K. S. Miller, A. J. Shultz, L. L. Smith,
737 M. W. Nachman and R. C. K. Bowie (2025a). Patterns of genetic diversity within three
738 California Quail species are best explained by climate and landscape changes. *Molecular*
739 *Ecology* **34**(20): e70093.
740 Benham, P. M., Cicero, C., Burns, K. J., Escalona, M., Beraut, E., Fairbairn, C. W., Seligmann,
741 W. E., Marimuthu, M. P. A., Nguyen, O., Chumchim, N., Toffelmier, E., Nachman, M.
742 W., & Bowie, R. C. K. (2025b). A highly contiguous genome assembly for the wrentit
743 (*Chamaea fasciata*), the sole representative of the babbler radiation in the Americas.
744 *Journal of Heredity*, esaf063. https://doi.org/10.1093/jhered/esaf063
745 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and
746 powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*
747 *(Methodological)*, *57*(1), 289–300.
748 Bowers, D. E. (1960). Correlation of variation in the Wrentit with environmental gradients. *The*
749 *Condor*, *62*(2), 91–120. https://doi.org/10.2307/1365675
750 Browning, M. R. (1992). A new subspecies of Chamaea fasciata (Wrentit) from Oregon (Aves:
751 Timaliinae). *Proceedings of the Biological Society of Washington*, *105*(3), 414–419.

Bryson Jr., R. W., Savary, W. E., Zellmer, A. J., Bury, R. B., & McCormack, J. E. (2016). Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. *Molecular Ecology*, *25*(15), 3731–3751. https://doi.org/10.1111/mec.13707

Burns, K. J., Alexander, M. P., Barhoum, D. N., & Sgariglia, E. A. (2007). Statistical assessment of congruence among phylogeographic histories of three avian species in the California Floristic Province. *Ornithological Monographs*, *63*, 96–109. https://doi.org/10.2307/40166901

Burns, K. J., & Barhoum, D. N. (2006). Population-level history of the wrentit (Chamaea fasciata): Implications for comparative phylogeography in the California Floristic Province. *Molecular Phylogenetics and Evolution*, *38*(1), 117–129. https://doi.org/10.1016/j.ympev.2005.07.008

Cai, T., Cibois, A., Alström, P., Moyle, R. G., Kennedy, J. D., Shao, S., Zhang, R., Irestedt, M., Ericson, P. G. P., Gelang, M., Qu, Y., Lei, F., & Fjeldså, J. (2019). Near-complete phylogeny and taxonomic revision of the world's babblers (Aves: Passeriformes). *Molecular Phylogenetics and Evolution*, *130*, 346–356. https://doi.org/10.1016/j.ympev.2018.10.010

Calsbeek, R., Thompson, J. N., & Richardson, J. E. (2003). Patterns of molecular evolution and diversification in a biodiversity hotspot: The California Floristic Province. *Molecular Ecology*, *12*(4), 1021–1029. https://doi.org/10.1046/j.1365-294X.2003.01794.x

Canty, A., & Ripley, B. (2017). Package 'boot.' *Bootstrap Functions. CRAN R Proj*.

Capblancq, T., Luu, K., Blum, M. G., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233.

775 Capblancq, T., Morin, X., Gueguen, M., Renaud, J., Lobreaux, S., & Bazin, E. (2020). Climate-
776 associated genetic variation in Fagus sylvatica and potential responses to climate change
777 in the French Alps. *Journal of Evolutionary Biology*, *33*(6), 783–796.
778 https://doi.org/10.1111/jeb.13610
779 Caye, K., Jumentier, B., Lepeule, J., & François, O. (2019). LFMM 2: Fast and accurate
780 inference of gene-environment associations in genome-wide studies. *Molecular Biology*
781 *and Evolution*, *36*(4), 852–860. https://doi.org/10.1093/molbev/msz008
782 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015).
783 Second-generation PLINK: Rising to the challenge of larger and richer datasets.
784 *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8
785 Chang, J., Niu, Y., Zhou, S., Zhu, W., Zhang, Z., Xiu, H., Shang, K., Feng, Q., & Wei, Y.
786 (2025). DPP7 promotes fatty acid β-oxidation in tumor-associated macrophages and
787 determines immunosuppressive microenvironment in colorectal cancer. *Int J Biol Sci*,
788 *21*(14), 6305–6325. https://doi.org/10.7150/ijbs.117909
789 Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The effects of genetic and
790 geographic structure on neutral variation. *Annual Review of Ecology, Evolution, and*
791 *Systematics*, *34*(1), 99–125. https://doi.org/10.1146/annurev.ecolsys.34.011802.132359
792 Chatzimanolis, S., & Caterino, M. S. (2007). Toward a better understanding of the "Transverse
793 range break": Lineage diversification in southern California. *Evolution*, *61*(9), 2127–
794 2141.
795 Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ
796 preprocessor. *Bioinformatics*, *34*(17), i884–i890.
797 https://doi.org/10.1093/bioinformatics/bty560

798 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., &
799 Ruden, D. M. (2012). A program for annotating and predicting the effects of single
800 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster
801 strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92.
802 Clements, J., Rasmussen, T., Schulenberg, T., Iliff, M., Fredericks, T., Gerbracht, J., Lepage, D.,
803 Spencer, A., Billerman, S., Sullivan, B., Smith, M., & Wood, C. (2024). *The*
804 *eBird/Clements checklist of Birds of the World: V2024. Downloaded from*
805 *https://www.birds.cornell.edu/clementschecklist/download/*.
806 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of
807 speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*,
808 *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796
809 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R.
810 E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and
811 VCFtools. *Bioinformatics*, *27*(15), 2156–2158.
812 Davey, J. W., & Blaxter, M. L. (2010). RADSeq: Next-generation population genetics. *Briefings*
813 *in Functional Genomics*, *9*(5–6), 416–423. https://doi.org/10.1093/bfgp/elq031
814 Delhey, K. (2019). A review of Gloger's rule, an ecogeographical rule of colour: Definitions,
815 interpretations and evidence. *Biological Reviews*, *94*(4), 1294–1316.
816 https://doi.org/10.1111/brv.12503
817 Dickinson, E. C., Bahr, N., Dowsett, R., Pearson, D., Remsen, V., Roselaar, C., & Schodde, D.
818 (2004). *The Howard and Moore complete checklist of birds of the world.*

819 Ducrest, A.-L., Keller, L., & Roulin, A. (2008). Pleiotropy in the melanocortin system,
820 coloration and behavioural syndromes. *Trends in Ecology & Evolution*, *23*(9), 502–510.
821 https://doi.org/10.1016/j.tree.2008.06.001
822 Dunn, R. R., Davies, T. J., Harris, N. C., & Gavin, M. C. (2010). Global drivers of human
823 pathogen richness and prevalence. *Proceedings: Biological Sciences*, *277*(1694), 2587–
824 2595.
825 Erickson, M. M. (1938). *Territory, annual cycle, and numbers in a population of Wren-tits*
826 *(Chamaea fasciata)*. University of California Press.
827 https://cir.nii.ac.jp/crid/1971149384743534135
828 Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual*
829 *Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501.
830 https://doi.org/10.1146/annurev.ecolsys.39.110707.173414
831 Exposito-Alonso, M., Booker, T. R., Czech, L., Gillespie, L., Hateley, S., Kyriazis, C. C., Lang,
832 P. L. M., Leventhal, L., Nogues-Bravo, D., Pagowski, V., Ruffley, M., Spence, J. P.,
833 Toro Arana, S. E., Weiß, C. L., & Zess, E. (2022). Genetic diversity loss in the
834 Anthropocene. *Science*, *377*(6613), 1431–1435. https://doi.org/10.1126/science.abn5642
835 Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for
836 detecting multilocus adaptation with multivariate genotype–environment associations.
837 *Molecular Ecology*, *27*(9), 2215–2233. https://doi.org/10.1111/mec.14584
838 Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association
839 studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-
840 210X.12382

841 Gambel, W., & Ziegler, D. (1845). Descriptions of new and little known birds, collected in upper
842 California. *Proceedings of the Academy of Natural Sciences of Philadelphia*, *2*(10), 263–
843 272.
844 Geupel, G. R., & Ballard, G. (2023). Wrentit (Chamaea fasciata), version 1.1. *Birds of the World*.
845 https://doi.org/10.2173/bow.wrenti.01.1
846 Grant, S. G. N. (2016). The molecular evolution of the vertebrate behavioural repertoire.
847 *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1685).
848 https://doi.org/10.1098/rstb.2015.0051
849 Grinnell, J., & Miller, A. H. (1944). The distribution of the birds of California. *Condor*, *88*(4),
850 10.
851 Guernier, V., Hochberg, M. E., & Guégan, J.-F. (2004). Ecology drives the worldwide
852 distribution of human diseases. *PLoS Biology*, *2*(6), e141.
853 https://doi.org/10.1371/journal.pbio.0020141
854 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring
855 the joint demographic history of multiple populations from multidimensional SNP
856 frequency data. *PLoS Genetics*, *5*(10), e1000695.
857 https://doi.org/10.1371/journal.pgen.1000695
858 He, Z., Li, W., Zhang, M., Huang, M., Chen, Z., Zhao, X., Ding, Y., Zhang, J., Zhao, L., & Jiao,
859 P. (2024). RNF216 inhibits the replication of H5N1 Avian Influenza Virus and regulates
860 the RIG-I signaling pathway in ducks. *The Journal of Immunology*, *213*(2), 187–203.
861 https://doi.org/10.4049/jimmunol.2300540
862 Heraghty, S. D., Jackson, J. M., & Lozier, J. D. (2023). Whole genome analyses reveal weak
863 signatures of population structure and environmentally associated local adaptation in an

important North American pollinator, the bumble bee Bombus vosnesenskii. *Molecular Ecology*, *32*(20), 5479–5497. https://doi.org/10.1111/mec.17125

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). dismo: Species distribution modeling. *R Package Version*, *1*(4), 1–1.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397. https://doi.org/10.1086/688018

Hoekstra, H. E., & Nachman, M. W. (2003). Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, *12*(5), 1185–1194. https://doi.org/10.1046/j.1365-294X.2003.01788.x

Horrocks, N. P. C., Hegemann, A., Ostrowski, S., Ndithia, H., Shobrak, M., Williams, J. B., Matson, K. D., & Tieleman, B. I. (2015). Environmental proxies of antigen exposure explain variation in immune investment better than indices of pace of life. *Oecologia*, *177*(1), 281–292.

IUCN (International Union for Conservation of Nature). (2019). Guidelines for the IUCN Red List categories and criteria. Version 13. IUCN, Gland, Switzerland.

887 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components:
888 A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*,
889 94. https://doi.org/10.1186/1471-2156-11-94
890 Jones, P., Lucock, M., Chaplin, G., Jablonski, N. G., Veysey, M., Scarlett, C., & Beckett, E.
891 (2020). Distribution of variants in multiple vitamin D-related loci (DHCR7/NADSYN1,
892 GC, CYP2R1, CYP11A1, CYP24A1, VDR, RXRα and RXRγ) vary between European,
893 East-Asian and Sub-Saharan African-ancestry populations. *Genes & Nutrition*, *15*, 5.
894 https://doi.org/10.1186/s12263-020-00663-3
895 Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Haeseler, A. von, & Jermiin, L. S. (2017).
896 ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*,
897 *14*(6), 587–590. https://doi.org/10.1038/nmeth.4285
898 Kendig, K. I., Baheti, S., Bockol, M. A., Drucker, T. M., Hart, S. N., Heldenbrand, J. R.,
899 Hernaez, M., Hudson, M. E., Kalmbach, M. T., Klee, E. W., Mattson, N. R., Ross, C. A.,
900 Taschuk, M., Wieben, E. D., Wiepert, M., Wildman, D. E., & Mainzer, L. S. (2019).
901 Sentieon DNASeq variant calling workflow demonstrates strong computational
902 performance and accuracy [Technology and Code]. *Frontiers in Genetics*, *Volume 10 -*
903 *2019*. https://doi.org/10.3389/fgene.2019.00736
904 Kern, A. D., & Hahn, M. W. (2018). The neutral theory in light of natural selection.
905 *Molecular Biology and Evolution*, *35*(6), 1366–1371.
906 https://doi.org/10.1093/molbev/msy092
907 Kimura, M. (1985). *The neutral theory of molecular evolution*. Cambridge university press.
908 Lam, M., Trampush, J. W., Yu, J., Knowles, E., Davies, G., Liewald, D. C., Starr, J. M.,

Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., DeRosse, P., Lundervold, A. J., Steen, V. M., Espeseth, T., Räikkönen, K., Widen, E., Palotie, A.,…Lencz, T. (2017). Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Reports*, *21*(9), 2597–2613. https://doi.org/10.1016/j.celrep.2017.11.028

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lloyd, R. E. (2016). Enterovirus control of translation and RNA granule stress responses. *Viruses*, *8*(4), 93. https://doi.org/10.3390/v8040093

Loescher, C. M., Hobbach, A. J., & Linke, W. A. (2022). Titin (TTN): From molecule to modifications, mechanics, and medical significance. *Cardiovascular Research*, *118*(14), 2903–2918. https://doi.org/10.1093/cvr/cvab328

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77. https://doi.org/10.1111/1755-0998.12592

Mayr, E. (1999). *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meek, M. H., Beever, E. A., Barbosa, S., Fitzpatrick, S. W., Fletcher, N. K., Mittan-Moreau, C. S., Reid, B. N., Campbell-Staton, S. C., Green, N. F., & Hellmann, J. J. (2023). Understanding local adaptation to prepare populations for climate change. *BioScience*, *73*(1), 36–47. https://doi.org/10.1093/biosci/biac101

Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., & Sackton, T. B. (2023). A fast, reproducible, high-throughput variant calling workflow for population genomics. *Molecular Biology and Evolution*, *41*(1), msad270. https://doi.org/10.1093/molbev/msad270

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., van Dyken, P. C., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Vieira, F. G., Meesters, C., Lee, S., Twardziok, S. O., Kanitz, A., VanCampen, J., Malladi, V., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*, 33. https://doi.org/10.12688/f1000research.29032.3

Moreira, L. R., Hernandez_Ba_os, B. E., & Smith, B. T. (2020). Spatial predictors of genomic and phenotypic variation differ in a lowland Middle American bird (Icterus gularis). *Molecular ecology*, *29*(16), 3084–3101. https://doi.org/10.1111/mec.15536

Moreira, L. R., & Smith, B. T. (2023). Convergent genomic signatures of local adaptation across a continental-scale environmental gradient. *Science Advances*, *9*(20), eadd0560. https://doi.org/doi:10.1126/sciadv.add0560

Moyer, B. R., Drown, D. M., & Clayton, D. H. (2002). Low humidity reduces ectoparasite pressure: Implications for host life history evolution. *Oikos*, *97*(2), 223–228. https://doi.org/10.1034/j.1600-0706.2002.970208.x

Mueller, J. C., Carrete, M., Boerno, S., Kuhl, H., Tella, J. L., & Kempenaers, B. (2020). Genes

acting in synapses and neuron projections are early targets of selection during urban colonization. *Molecular ecology*, *29*(18), 3403–3412. https://doi.org/https://doi.org/10.1111/mec.15451

Mundy, N. I. (2005). A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1573), 1633. https://doi.org/10.1098/rspb.2005.3107

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nei, M., Suzuki, Y., & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, *11*(1), 265–289. https://doi.org/10.1146/annurev-genom-082908-150129'

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal* **7**(4): 308–313.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nimon, K., Oswald, F., & Roberts, J. K. (2013). Yhat: Interpreting regression effects. *R Package Version*, *2*(0).

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*(3), 375–402.

977 Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., & Stevens, M. (2007). The vegan package.
978 *Community Ecology Package*, *10*(631–637), 719.
979 Orr, A. H. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, *10*(8),
980 531–539. https://doi.org/10.1038/nrg2603
981 Orr, M. R., & Smith, T. B. (1998). Ecology and speciation. *Trends in Ecology & Evolution*,
982 *13*(12), 502–506. https://doi.org/10.1016/S0169-5347(98)01511-0
983 Osgood, W. H. (1899). Chamaea fasciata and its subspecies. *Proceedings of the Biological*
984 *Society of Washington*, *13*, 41–42.
985 Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with
986 estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.
987 https://doi.org/10.1038/ng.3464
988 Peymani, F., Ebihara, T., Smirnov, D., Kopajtich, R., Ando, M., Bertini, E., Carrozzo, R.,
989 Diodato, D., Distelmaier, F., Fang, F., Ghezzi, D., Hempel, M., Iwanicka-Pronicka, K.,
990 Klopstock, T., Stenton, S. L., Lamperti, C., Liu, Z., Murtazina, A., Okamoto, Y., …
991 Prokisch, H. (2025). Pleiotropic effects of MORC2 derive from its epigenetic signature.
992 *Brain*, awaf159. https://doi.org/10.1093/brain/awaf159
993 Phillips, A. R. (1986). *The Known Birds of North and Middle America: Hirundinidae to*
994 *Mimidae; Certhiidae*. AR Phillips.
995 Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species
996 distribution modeling. *Twenty-First International Conference on Machine Learning -*
997 *ICML '04*, 83. https://doi.org/10.1145/1015330.1015412
998 Raven, P. H., & Axelrod, D. I. (1974). Angiosperm biogeography and past continental
999 movements. *Annals of the Missouri Botanical Garden*, *61*(3), 539–673.

Raven, P. H. & D. I. Axelrod (1978). Origin and relationships of the California flora, Univ of California Press.

Ridgway, R. (1882). *Descriptions of two new thrushes from the United States*. http://hdl.handle.net/10088/12434

Ridgway, R. (1902). *The birds of North and Middle America: A descriptive catalogue of the higher groups, genera, species, and subspecies of birds known to occur in North America, from the Arctic lands to the Isthmus of Panama, the West Indies and other islands of the Caribbean sea, and the Galapagos Archipelago* (No. 50). US Government Printing Office. https://doi.org/10.5962/bhl.title.54021

Sarna-Wojcicki, A. M., C. E. Meyer, H. R. Bowman, N. T. Hall, P. C. Russell, M. J. Woodward and J. L. Slate (1985). Correlation of the Rockland ash bed, a 400,000-year-old stratigraphic marker in northern California and western Nevada, and implications for middle Pleistocene paleogeography of central California. *Quaternary Research* 23(2), 236–257.

Schierenbeck, K. A. (2014). *Phylogeography of California: An introduction*. University of California Press.

Sehgal, R. N. M., Buermann, W., Harrigan, R. J., Bonneaud, C., Loiseau, C., Chasar, A., Sepil, I., Valkiūnas, G., Iezhova, T., Saatchi, S., & Smith, T. B. (2011). Spatially explicit predictions of blood parasites in a widely distributed African rainforest bird. *Proceedings: Biological Sciences*, *278*(1708), 1025–1033.

Sexton, J. P., Hangartner S. B., & Hoffmann A. A. (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68(1): 1–15.

1022 Sgariglia, K. A., & Burns, K. J. (2003). Phylogeography of the California Thrasher (Toxostoma
1023 redivivum) based on bested-clade analysis of mitochondrial-DNA variation. *The Auk*,
1024 *120*(2), 346–361. https://doi.org/10.2307/4090187
1025 Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for
1026 conserving biodiversity under climate change. *Evolutionary Applications*, *4*(2), 326–337.
1027 https://doi.org/10.1111/j.1752-4571.2010.00157.x
1028 Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P.,
1029 Gold, M., Harrigan, R. J., Hodges, S., Luckau, T. K., Miller, C., Oliveira, D. R., Shaffer,
1030 K. E., Shapiro, B., Sork, V. L., & Wang, I. J. (2022). Landscape genomics to enable
1031 conservation actions: The California Conservation Genomics Project. *Journal of*
1032 *Heredity*, *113*(6), 577–588. https://doi.org/10.1093/jhered/esac020
1033 Shafer, A. B., & Wolf, J. B. (2013). Widespread evidence for incipient ecological speciation: a
1034 meta-analysis of isolation-by-ecology. *Ecology Letters*, *16*(7), 940–950.
1035 Smeds L., Qvarnström A., & Ellegren H. (2016). Direct estimate of the rate of germline
1036 mutation in a bird. *Genome Research,* 26(9): 1211–1218.
1037 Song, M. J., Freund, F., Tribble, C. M., Toffelmier, E., Miller, C., Bradley Shaffer, H., Li, F.-W.,
1038 & Rothfels, C. J. (2025). The nitrogen-fixing fern Azolla has a complex microbiome
1039 characterized by varying degrees of cophylogenetic signal. *American Journal of Botany*,
1040 *112*(3), e70010. https://doi.org/10.1002/ajb2.70010
1041 Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., & Neale, D. (2013). Putting the
1042 landscape into the genomics of trees: Approaches for understanding local adaptation and
1043 population responses to changing climate. *Tree Genetics & Genomes*, *9*(4), 901–911.

Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., & Byrne, M. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, *23*(10), 2500–2513. https://doi.org/10.1111/mec.12751

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1110), 11.10.11–11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, *67*(12), 3403–3411.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., & Sekiguchi, M. (2010). Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate*, *23*(23), 6312–6335.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

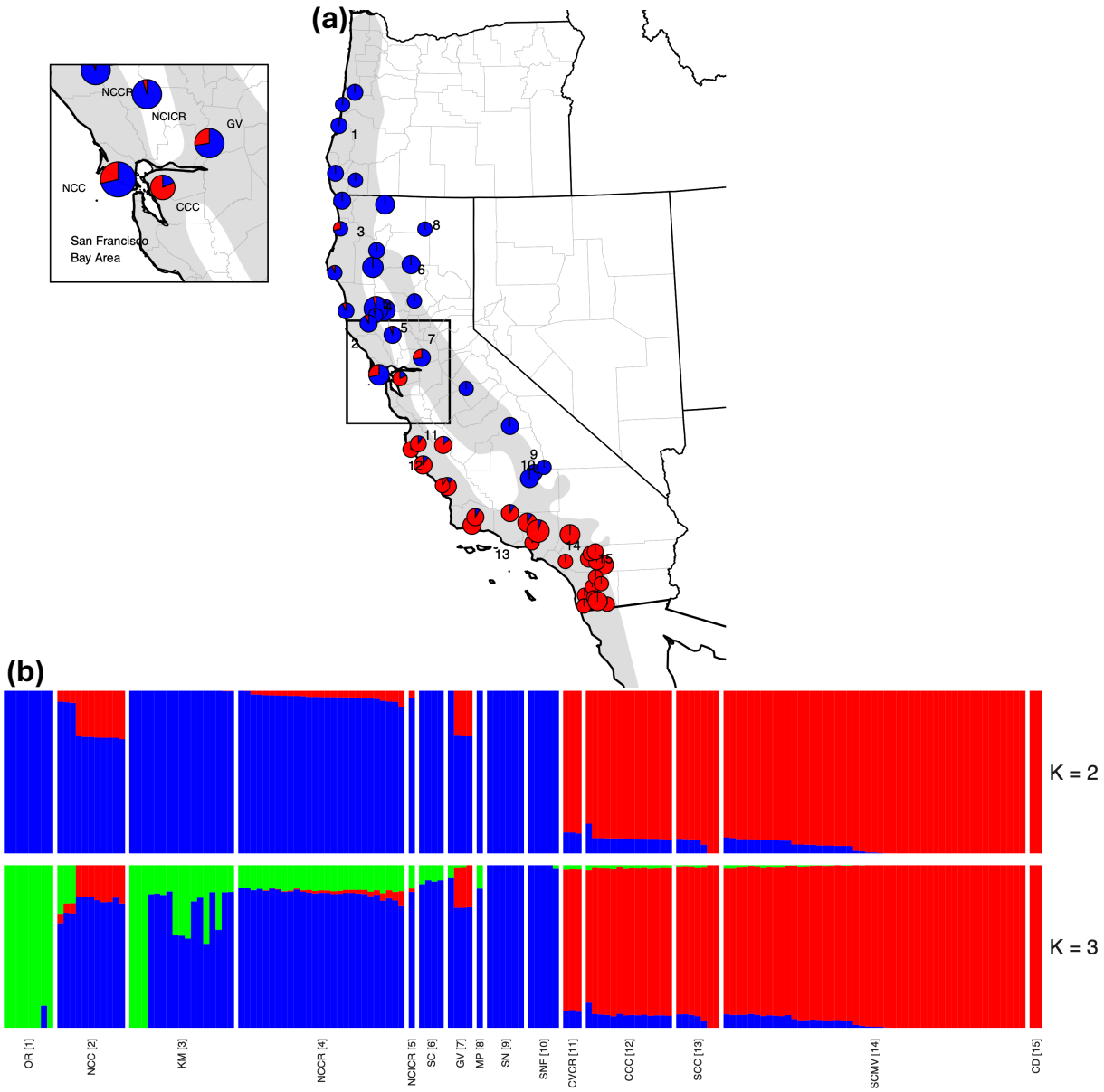
Wooldridge, B., Orland, C., Enbody, E., Escalona, M., Mirchandani, C., Corbett-Detig, R., Kapp, J. D., Fletcher, N., Cox-Ammann, K., Raimondi, P., & Shapiro, B. (2024). Limited genomic signatures of population collapse in the critically endangered Black Abalone (Haliotis cracherodii). *Molecular Ecology*, e17362.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138. https://doi.org/10.1093/genetics/28.2.114

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., & Yang, T.-L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, *35*(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zhang, X., Qi, Y., Pang, Y., Yuan, B., & Li, X. (2023). Identification of polymorphisms in the HERC2-OCA2 gene locus and their association with feather color in quail. *The Journal of Poultry Science*, *60*(2), Article 2023013. https://doi.org/10.2141/jpsa.2023013

Tables and Figures



1095

1096 **Figure 1**. (a) Map of ADMIXTURE results showing the estimated proportion of ancestral alleles

1097 at each locality based on the K=2 model (best model). (b) ADMIXTURE bar plots for K=2 and

1098 K=3 genetic ancestries. Each bar represents individual ancestry proportions for individual

1099 samples and are grouped by locality. OR=Oregon coast; NCC=Northern California coast;

1100 KM=Klamath Mountains; NCCR=Northern California coast ranges; NCICR=Northern

1101 California Interior coast ranges; SC=Southern Cascades; GV=Great Valley; MP=Modoc Plateau;

SN=Sierra Nevada; SNF=Sierra Nevada Foothills; CVCR=Central valley coast ranges; CCC=Central California coast; SCC=Southern California coast; SCMV=Southern California mountains and valleys; CD=Colorado desert. Wrentit distribution polygon shapefile requested from birdlife.org.

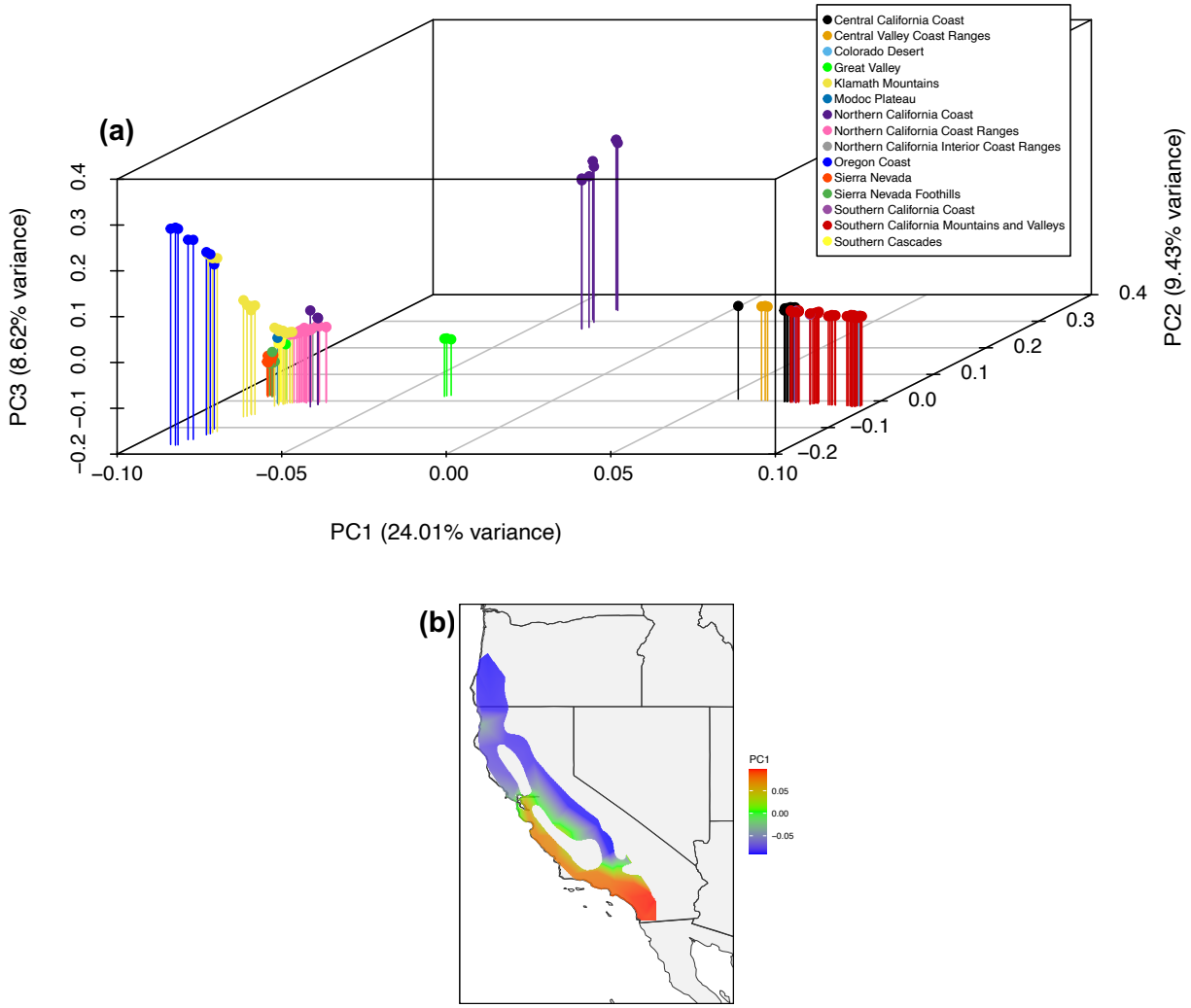


Figure 2. (a) Genomic principal component analysis (PCA) based on all 159 *C. fasciata* individuals. PCA of the first three principal components (e.g., PC1-PC3). Colors represent population assignments of each individual (see legend). (b) Map of interpolated PC1 scores.

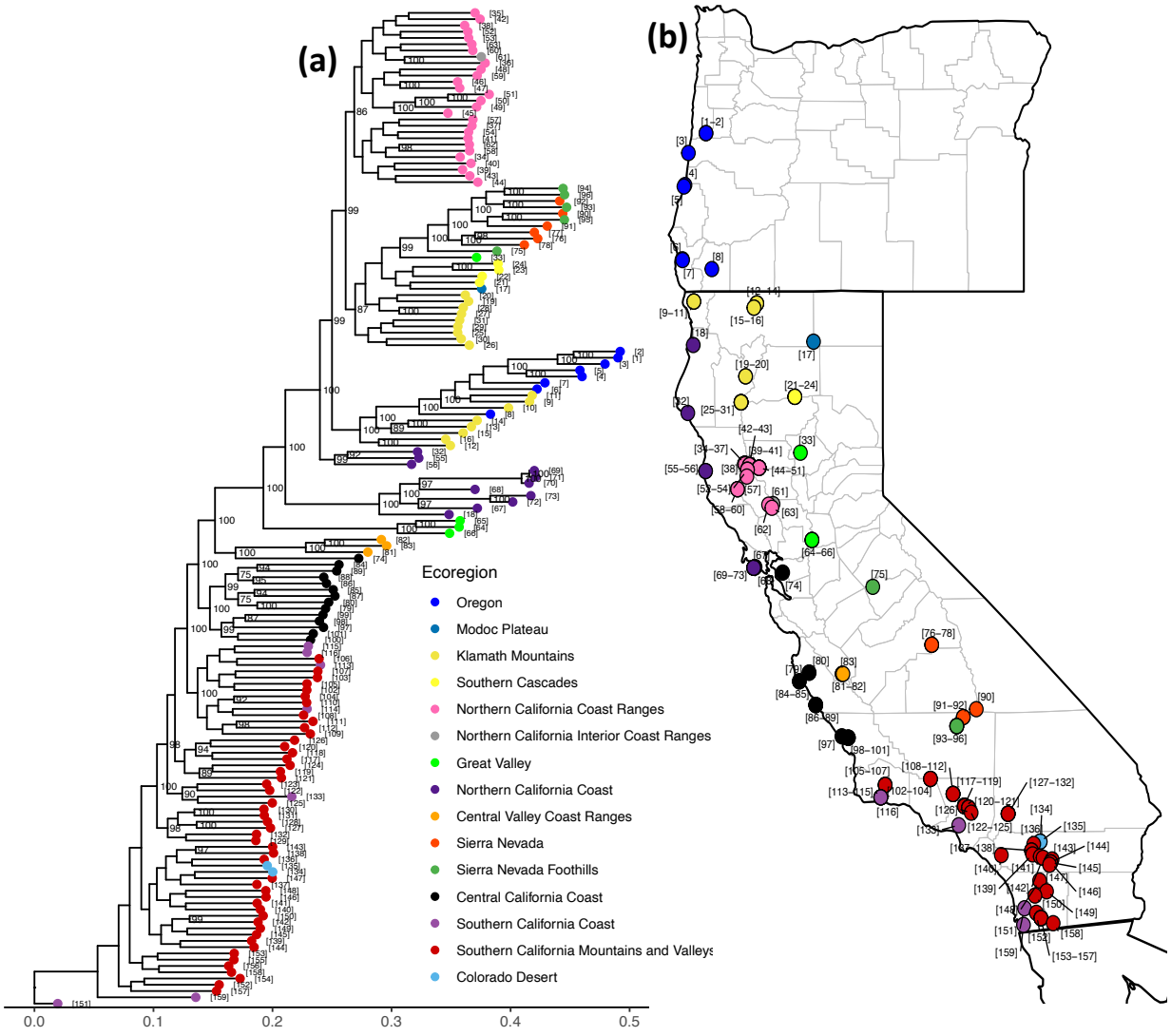
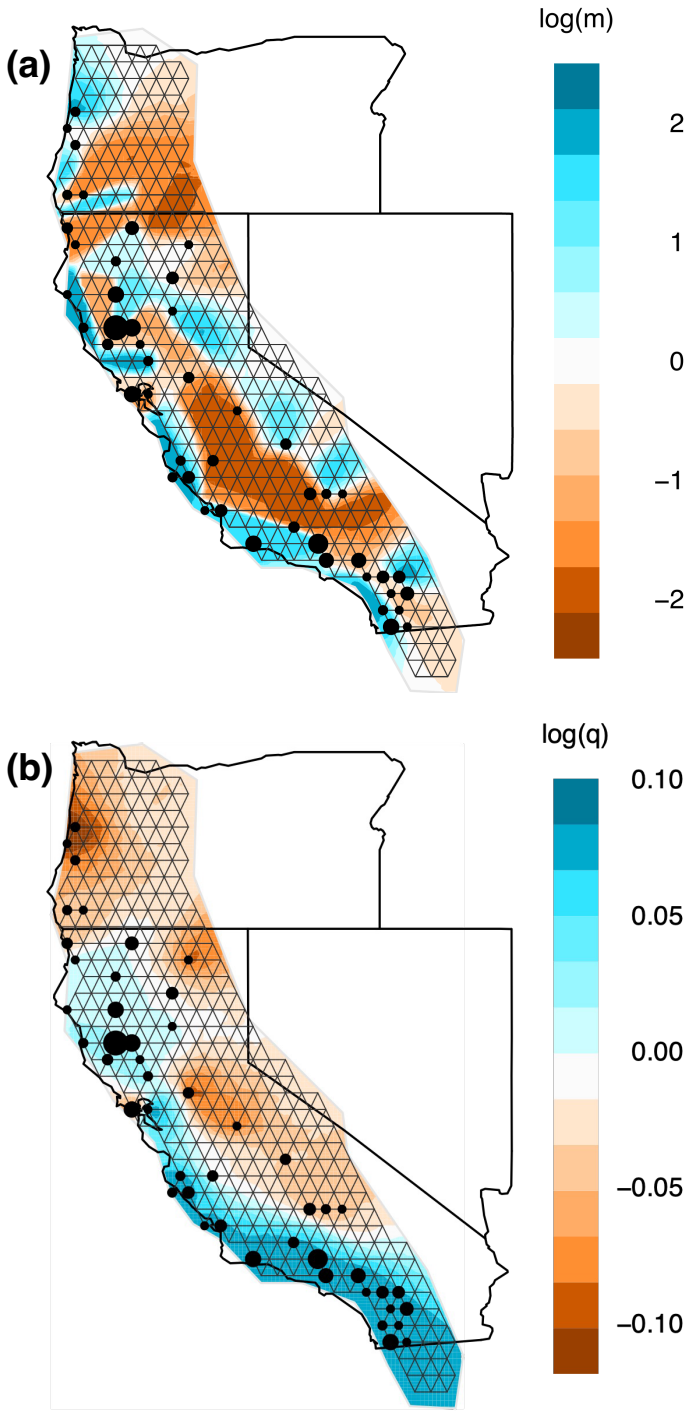


Figure 3. Maximum likelihood (ML) tree of all *C. fasciata* individuals. (a) ML tree rooted with a single *Suthora webbiana* sample and color-coded based on ecoregion association (see legend). Bootstrap support >75 is shown at nodes. (b) Map of sampling localities with numbers next to points corresponding with numbered samples on ML tree.

1119



1120

1121 **Figure 4**. Estimated effective migration surface (EEMS) results based on the high-quality SNP

1122 dataset. (a) migration (*m*) rate and (b) diversity (*q*) rate estimates rate estimates. Orange colors

denote significantly lower-than expected values, while blue colors represent significantly higher-than-expected values. Points indicate sampling across the grid for a deme size of 400.

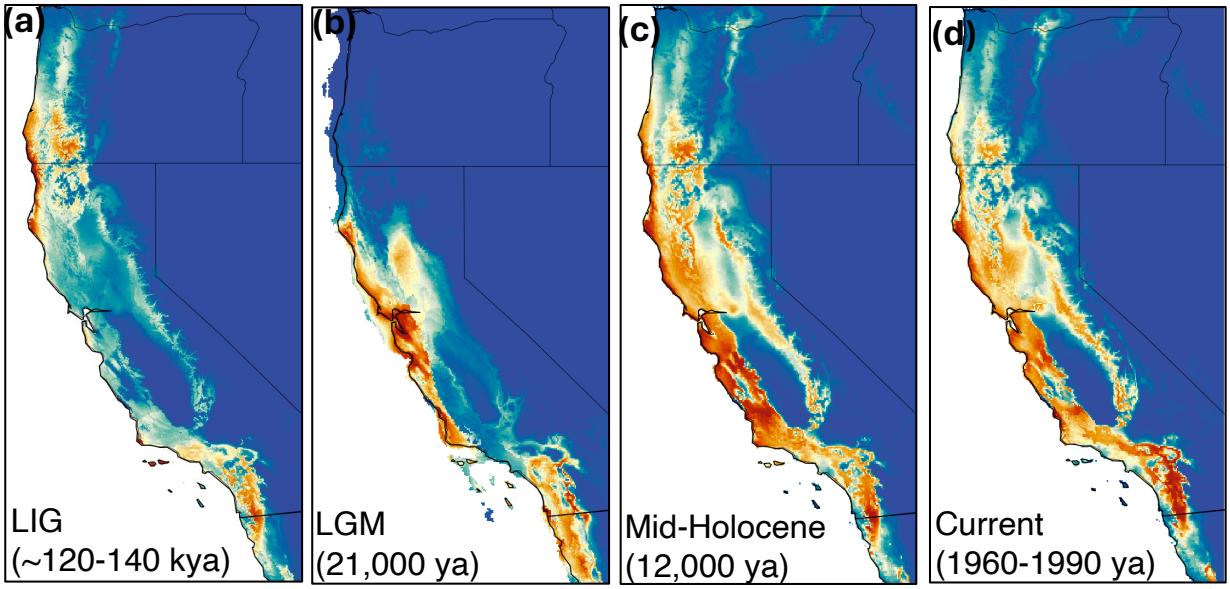


Figure 5. Climate models using MaxEnt based on sampling points from the genomic dataset. Panels show current and historical distributions modeled using the Worldclim bioclimatic variables. Historical distributions of (a) the Last Interglacial, (b) Last Glacial Maximum, (c) Mid-Holocene, and (d) Predicted current distribution of all *C. fasciata* samples.

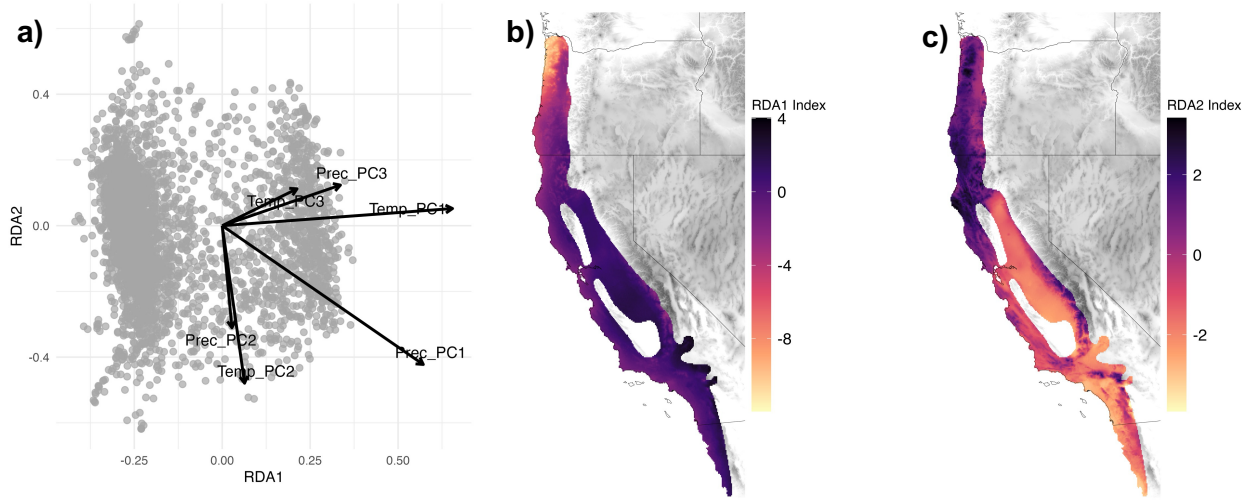


Figure 6. Genotype-environment association results for the partial redundancy analysis (pRDA) based on the 'adaptively enriched' genetic landscape. The plot represents RDA scores for sites (i.e., SNPs) plotted for RDA1 versus RDA2 axes based on associations of SNPs with the dimensionally reduced 19 bioclimatic variables. The model was conditioned for population structure by including the first two PC axes (e.g., PC1 and PC2) from our PCA structure analysis. Variable loadings are shown on both plots, with length and direction of arrows indicated. The pRDA model was found to be highly significant based on 999 permutations, and all variables were also found to be significant based on 999 permutations. Maps are plotted to show projection of scores from the adaptive index calculated for (b) RDA1 and (c) RDA2.

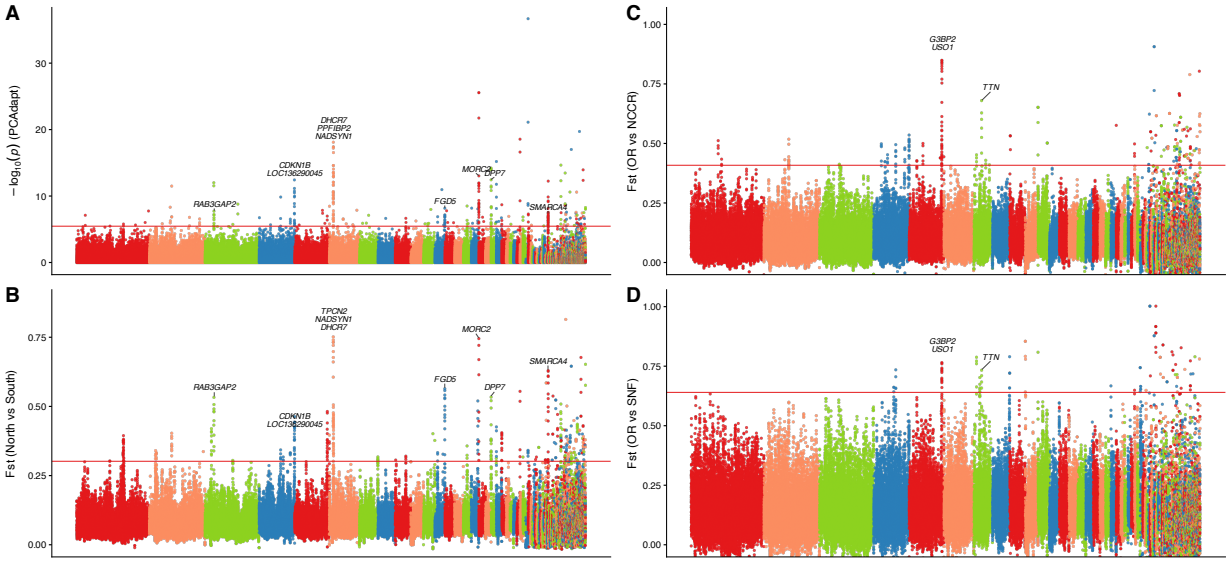


Figure 7. Manhattan plots based on windowed F_{ST} -outlier and PCAdapt outlier detection results. (a) PCAdapt results, with candidate genes highlighted. (b) Windowed F_{ST} comparison between the northern and southern genetic clusters, highlighting the same candidate genes as in panel (a). (c) Windowed F_{ST} comparison between Oregon and Northern California Coast (OR-NCCR) populations. (d) Windowed F_{ST} comparison between Oregon and Sierra Nevada (OR-SNF) populations. Candidate outlier SNPs with their associated genes are highlighted in both panels.

Table 1. Genome-wide estimates of nucleotide diversity (π), Tajima's D, observed heterozygosity, expected heterozygosity, and inbreeding coefficient (F).

Population	$\boldsymbol{\pi}$	**Tajima's** D	**Observed Het**	**Expected Het**	**Inbreeding Coefficient** F
North	0.00280522	0.54630700	0.149450	0.182740	0.182316
South	0.00317606	0.34048100	0.177041	0.182817	0.031684

Table 2. Demographic model best-fitting parameter raw estimates for each scenario tested in dadi. Nu2=current population size (Ne) of founder population following demographic expansion; s=Ne of founder population following divergence; m12=migration rate from pop1 to pop2; m21=migration rate from pop2 to pop1; t1=time of initial divergence; t2=time of end of ancient migration or start of secondary contact; theta=population-scaled mutation rate; AIC=Alkain information criterion.

Demographic model	**nu2**	**s**	**m12**	**m21**	**t1**	**t2**	**theta**	**AIC**
Founder event with ancient asymmetric migration	7.0932534	0.15513943	1.34410156	0.74331171	0.01089201	0.01367831	71245 0.733	13140 18.66
Founder event with continuous asymmetric migration	4.30E-01	7.31E-03	7.87E+00	8.67E+00	4.03E-01	na	88333 1.485	37776 7.306
Founder event with no migration	7.32234861	0.03656615	na	na	0.01707698	na	70802 9.084	10689 56.66
Founder event followed by secondary contact with asymmetric migration	3.74E+00	8.07E-03	8.05E+00	9.27E+00	2.94E-01	4.88E-01	92348 3.235	38582 6.574
Vicariance event with ancient asymmetric migration	na	0.29675939	0.34895878	0.21514596	0.01012544	0.01011346	76893 1.537	16995 12.31
Vicariance event with continuous asymmetric migration	na	0.24405045	4.67763869	9.97423897	0.82506907	na	93972 8.748	10832 55.1
Vicariance event with no migration	na	0.27172629	na	na	0.01842637	na	77200 1.313	16898 71.74

Vicariance event followed by secondary contact and asymmetric migration	na	0.3468 7854	3.0476 4208	9.1319 1145	0.0283 9705	1.6200 6253	91055 5.178	10763 17.84

Table 3. Commonality analysis results for population structure (IBH), geographic distance (IBD), and environment (IBE), including 95% confidence intervals.

Component (Model)	**Coefficient (95% CI)**	**% Total**
Pure Geography (genVec ~ geoVec | ecoVec + pc1Vec)	0.0001 (0, 0.0002)	0.01
Pure Environment (genVec ~ ecoVec | geoVec + pc1Vec)	0.0128 (0.0106, 0.0150)	1.60
Pure Structure (genVec ~ pc1Vec | geoVec + ecoVec)	0.3812 (0.3649, 0.3974)	47.55
Geography and Environment (genVec ~ geoVec + ecoVec | pc1Vec)	0.0046 (0.0036, 0.0058)	0.58
Geography and Structure (genVec ~ geoVec + pc1Vec | ecoVec)	0.2756 (0.2612, 0.2904)	34.39
Environment and Structure (genVec ~ ecoVec + pc1Vec | geoVec)	-0.0103 (-0.0119, -0.0087)	-1.29

Geography, Environment, and Structure (genVec ~ geoVec + ecoVec + pc1Vec)	0.1376 (0.1262, 0.1479)	17.17
Total	**0.8016 (0.7902, 0.8127)**	**100.00**

Table 4. Candidate genes containing PCAdapt outlier SNPs within north–south F_{ST} outlier windows. RDA and LFMM indicate additional gene-level support from genotype–environment association analyses.

Scaffold	**Gene**	**Description**	**Methods**
JARCOQ010000005.1	*CDK17*	cyclin dependent kinase 17	F_{ST}, PCAdapt, LFMM
JARCOQ010000005.1	*CDKN1B*	cyclin dependent kinase inhibitor 1B	F_{ST}, PCAdapt
JARCOQ010000046.1	*DCAF15*	DDB1 and CUL4 associated factor 15	F_{ST}, PCAdapt
JARCOQ010000007.1	*DHCR7*	7-dehydrocholesterol reductase	F_{ST}, PCAdapt
JARCOQ010000020.1	*DPP7*	dipeptidyl peptidase 7	F_{ST}, PCAdapt, LFMM

JARCOQ010000005.1	*EP300*	E1A binding protein p300	F_{ST}, PCAdapt, RDA
JARCOQ010000013.1	*FGD5*	FYVE, RhoGEF and PH domain containing 5	F_{ST}, PCAdapt, LFMM, RDA
JARCOQ010000007.1	*GAS2*	growth arrest specific 2	F_{ST}, PCAdapt, LFMM
JARCOQ010000010.1	*IPO13*	importin 13	F_{ST}, PCAdapt
JARCOQ010000046.1	*KHSRP*	KH-type splicing regulatory protein	F_{ST}, PCAdapt
JARCOQ010000005.1	LOC136290045	uncharacterized LOC136290045	F_{ST}, PCAdapt, LFMM
JARCOQ010000013.1	LOC136296481	uncharacterized LOC136296481	F_{ST}, PCAdapt, LFMM, RDA
JARCOQ010000013.1	LOC136296493	uncharacterized LOC136296493	F_{ST}, PCAdapt, LFMM, RDA
JARCOQ010000017.1	LOC136298499	centrosome-associated protein CEP250-like	F_{ST}, PCAdapt
JARCOQ010000020.1	LOC136299483	ficolin-1-like	F_{ST}, PCAdapt
JARCOQ010000046.1	LOC136304011	low-density lipoprotein receptor-like	F_{ST}, PCAdapt
JARCOQ010000046.1	LOC136304053	uncharacterized LOC136304053	F_{ST}, PCAdapt
JARCOQ010000046.1	LOC136304061	uncharacterized LOC136304061	F_{ST}, PCAdapt
JARCOQ010000007.1	*LUZP2*	leucine zipper protein 2	F_{ST}, PCAdapt
JARCOQ010000020.1	*MAN1B1*	mannosidase alpha class 1B member 1	F_{ST}, PCAdapt
JARCOQ010000018.1	*MORC2*	MORC family CW-type zinc finger 2	F_{ST}, PCAdapt
JARCOQ010000007.1	*NADSYN1*	NAD synthetase 1	F_{ST}, PCAdapt
JARCOQ010000013.1	*NR2C2*	nuclear receptor subfamily 2 group C member 2	F_{ST}, PCAdapt, LFMM, RDA
JARCOQ010000007.1	*PPFIBP2*	PPFIA binding protein 2	F_{ST}, PCAdapt
JARCOQ010000022.1	*PRPS1*	phosphoribosyl pyrophosphate synthetase 1	F_{ST}, PCAdapt
JARCOQ010000003.1	*RAB3GAP2*	RAB3 GTPase activating non-catalytic protein subunit 2	F_{ST}, PCAdapt
JARCOQ010000020.1	*RABL6*	RAB, member RAS oncogene family like 6	F_{ST}, PCAdapt

JARCOQ010000006.1	*RASSF6*	Ras association domain family member 6	F_{ST}, PCAdapt
JARCOQ010000016.1	*RNF216*	ring finger protein 216	F_{ST}, PCAdapt
JARCOQ010000046.1	*SMARCA4*	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	F_{ST}, PCAdapt
JARCOQ010000007.1	*SVIP*	small VCP interacting protein	F_{ST}, PCAdapt
JARCOQ010000007.1	*TPCN2*	two pore segment channel 2	F_{ST}, PCAdapt
JARCOQ010000020.1	*UAP1L1*	UDP-N-acetylglucosamine pyrophosphorylase 1 like 1	F_{ST}, PCAdapt

1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278

1279
1280