Identifying genomic adaptation to local climate using a mechanistic evolutionary model

Abstract

1. Identifying genomic adaptation is key to understanding species' evolutionary responses to environmental changes. However, current methods to identify adaptive variation have two major limitations. First, when estimating genetic variation, most methods do not account for observational uncertainty in genetic data because of finite sampling and missing genotypes. Second, many current methods use phenomenological models to partition genetic variation into adaptive and non-adaptive components.

2. We address these limitations by developing a hierarchical Bayesian model that explicitly accounts for observational uncertainty and underlying evolutionary processes. The first layer of the hierarchy is the data model that captures observational uncertainty by probabilistically linking RAD-sequence data to genetic variation. The second layer is a process model that represents how evolutionary forces, such as local adaptation, mutation, migration, and drift, maintain genetic variation. The third layer is the parameter model, which incorporates our knowledge about biological processes. For example, because most loci in the genome are expected to be neutral, the environmental sensitivity coefficients are assigned a regularized prior centered at zero. Together, the three models provide a rigorous probabilistic framework to identify local adaptation in wild organisms.

3. Analysis of simulated RAD-seq data shows that our statistical model can reliably infer adaptive genetic variation. To show the real-world applicability of our method, we re-analyzed RAD-seq data (~105k SNPs) from Willow Flycatchers (*Empidonax traillii*) in the USA. We found 30 genes close to loci that showed a statistically significant association with temperature seasonality. Gene ontology suggests that several of these genes play a crucial role in egg mineralization, feather development, and the ability to withstand extreme temperatures.

4. Moreover, the data and process models can be modified to accommodate a wide range of genetic datasets (*e.g.*, pool and low coverage genome sequencing) and demographic histories (*e.g.,* range shifts) to study climatic adaptation in a wide range of natural systems.

Data and code availability: The data and code files are provided with the submission. Upon acceptance, the data and code files will be deposited in Zenodo.

Keywords: Bayesian statistics, climate adaptation, Environmental association analysis, metapopulation, and uncertainty quantification.

Introduction

Many species face a looming threat of extinction as global temperature rises (Thomas *et al.* 2004; Urban 2015). This risk is particularly concerning for species with limited dispersal capacity that are slow to track their climatic niches (Carroll *et al.* 2015). Alternatively, some species may adapt to the local climate, potentially alleviating the risk of extinction. Natural selection favors individuals with heritable phenotypes best suited to survive and reproduce in local climates, creating a geographical mosaic in frequencies of adaptive alleles congruent with climatic gradients (Hedrick, Ginevan & Ewing 1976; Hedrick 1986; Hedrick 2006). Therefore, to understand the evolutionary responses of a species to a changing climate and incorporate this knowledge to inform evolutionary management policies (Smith *et al.* 2014), we need robust methodologies to identify genes (and their function) that confer local adaptation in natural populations.

With recent advances in high-throughput sequencing and global environmental sensing technologies (Chuvieco 2020; Satam *et al.* 2023), new computational approaches have enabled researchers to identify adaptive genetic variation in wild populations. These approaches, collectively known as environmental association analysis (Rellstab *et al.* 2015; Hoban *et al.* 2016), have gained widespread popularity, which can be partly attributed to declining sequencing costs, improvements in genomic tools, detailed global climatic maps, and availability of computational resources. These advances have allowed researchers to study climatic adaptation in non-model organisms at a fine genomic resolution and large spatial extent for a reasonable cost (Frichot *et al.* 2013; De Villemereuil & Gaggiotti 2015; Fitzpatrick & Keller 2015; Wagner, Chávez-Pesqueira & Forester 2017).

Broadly, environmental association analysis involves three steps. First, sequencing data from spatially referenced individuals are used to estimate spatial variation in allele frequencies. Next, a statistical model is used to partition this genetic variation into adaptive and non-adaptive components. The adaptive variation corresponds to the response of an allele to changes in prevailing climatic conditions. The non-adaptive variation describes the allelic variation (or neutral variation) that stems from evolutionary forces other than local adaptation, including gene flow among populations, mutation, and drift. Finally, a statistical

66 criterion is used to select loci that exhibit strong adaptive genetic variation relative to non-adaptive
67 variation.

68 Despite the success of environmental association studies, current methods face several conceptual
69 and statistical challenges that limit the reliability of the inferences. Many commonly used environmental
70 association methods do not account for uncertainty in estimates of allele frequencies (see Coop *et al.* (2010)
71 and Foll and Gaggiotti (2008) for exceptions), which can introduce biases in downstream analyses. Due to
72 practical constraints, biogeographers collect and sequence only a finite number of individuals in a
73 population. The genotype of these individuals, inferred from RAD (restriction site-associated DNA)
74 sequencing (Baird *et al.* 2008), for instance, contains imprecise information regarding the underlying allele
75 frequencies that result from sampling a small subset of individuals within a population. Additionally,
76 because of low coverage and alignment errors, the true genotypes of some individuals may be missing
77 (Huang & Knowles 2016), which further reduces sample size and increases uncertainty. Failure to account
78 for these sources of uncertainty results in overly precise estimates of genetic variation, which can lead to
79 an increased number of false positives (i.e., incorrectly inferring non-adaptive loci as important).
80 Alternatively, researchers discarding data due to noisy estimates of genetic variation may lose valuable
81 (although imprecise) information, thereby increasing the rates of false negatives (i.e., inability to identify
82 an adaptive locus).

83 Another major difficulty is that many environmental association methods characterize genetic
84 variation using models, such as generalized linear models (Rellstab *et al.* 2015). Although the underlying
85 relationships between data and parameters (e.g., environment and allele frequency) in phenomenological
86 models are based on evolutionary thinking, these relationships are often weakly motivated by evolutionary
87 theory (but not always; see De Villemereuil and Gaggiotti (2015) and Günther and Coop (2013) for
88 exceptions). For example, some classical environmental association models assume an S-shape response
89 curve to relate environmental variables and allele frequency (Joost *et al.* 2007; Stucki *et al.* 2017). This
90 response curve has several attractive characteristics: it is bounded between zero and one, and has an incline
91 in the middle that emulates the response of an allele to local climatic adaptation. However, the S-shaped

response curve is not unique in these characteristics. A biogeographer can construct alternate equally plausible response curves (*e.g.*, cdf functions) and, consequently, there is no easy way to choose one over the other *a priori*.

In contrast, mechanistic models of evolution are based on the first principles of birth-death processes, and, therefore, a biogeographer can evaluate model assumptions based on the constraints imposed by the population demography and natural history of the species (Rice 2004). Mechanistic models of evolution in statistical inferences offers a unique advantage—they constrain the flow of information from data to model parameters using a theoretical understanding. This may allow biogeographers to create bespoke statistical models tailored to match the demographic history of the species (Wikle 2003; Hooten & Hefley 2019).

Over the past two decades, theoretical and computational advances in Bayesian statistics have fueled a surge in the development of probabilistic hierarchical models (Beaumont & Rannala 2004; Stadler *et al.* 2024), offering researchers new avenues to address the aforementioned challenges. These hierarchical models can account for uncertainty in estimates of genetic variation (stemming from finite sampling and sequencing) (Gompert & Buerkle 2011) and then propagate this uncertainty downstream to mechanistic process models of evolution, enabling a wide range of inferences ranging from constructing phylogenies (Suchard *et al.* 2018) to identifying migration corridors (Hanks & Hooten 2013; Marcus *et al.* 2021) to estimating growth rates of emerging viral diseases (Stadler 2010). In fact, these statistical insights are also beginning to permeate the latest generation of genotype-environment association methods (Coop *et al.* 2010; De Villemereuil & Gaggiotti 2015).

Hierarchical models help us do joint inferences that allows us to learn parameters using data, process, and parameter models (Berliner 1996). In the context of environmental association analysis, the data model describes how the unobserved allele frequencies could have led to genetic data. The process model describes the biological processes determining spatial variation in allele frequencies. The parameter model describes the knowledge the biogeographer has about the parameters before the data are collected, based on past research.

This structure of the hierarchical model offers several advantages in identifying genomic adaptation to climate. First, the data model enables a biogeographer to quantify observational uncertainty in estimates of genetic variation, particularly in cases where sample sizes are small and genotypes are missing. Second, these imprecise estimates of genetic variation can then be linked to a process model of evolution. This evolution model is based on the demographic history of the species and provides a mechanistic explanation of how to partition genetic variation into adaptive (response curve) and non-adaptive (neutral) components. Lastly, the parameter model can be used to incorporate knowledge from past research. For example, the theory of molecular evolution suggests that the overwhelming majority of variation in genomes is non-adaptive and stems from the interplay between mutation, drift, and migration, while only a small fraction of the variation stems from local adaptation (Kimura 1983). To incorporate this prior knowledge, a biogeographer can assign a low prior probability that the response curve has a finite slope. This prior effectively shrinks (regularizes) the adaptive genetic variation to zero for most loci, providing a systematic way to evaluate the relative contribution of climate adaptation and neutral processes in determining genetic variation.

Because of these features of the data, process, and parameter models, hierarchical models provide a rigorous probabilistic framework to identify genomic adaptation in wild organisms informed by theoretical principles of evolutionary biology using noisy genetic data. In what follows, we present a new demographic Bayesian model and test its robustness by analyzing synthetic data with known parameter values. We use the model to analyze RAD-seq data from Willow Flycatchers (*Empidonax traillii*) and identify candidate genes that may be responsible for local climatic adaptation. Although we apply our model to Willow Flycatchers, the statistical insights are general and broadly applicable to other species.

Willow Flycatchers

The Willow Flycatcher is a migratory songbird that breeds mainly in the United States and Southern Canada. The species is categorized into four distinct subspecies—the Pacific Northwestern form (*E. t. brewsteri*), Western Central form (*E. t. adastus*), Eastern form (*E. t. traillii*), and Southwestern form (*E. t.

144 *extimus*). Among them, the Southwestern form has experienced a precipitous decline in abundance, likely
145 due to the loss of riparian habitat along streams and waterways (Sedgwick 2000), which provide respite
146 during extreme temperatures. In 1995, the Southwestern form was listed as federally endangered
147 subspecies (Unitt 1987) due to its genetic, ecological, and song distinctiveness (Theimer *et al.* 2016; Ruegg
148 *et al.* 2018; Mahoney *et al.* 2020).

149 Previous landscape genomic work identified highly significant correlations between allele
150 frequencies in genes linked to thermal tolerance and the intensity of summer heat waves in the southwest
151 (Ruegg *et al.* 2018). Therefore, re-analyzing the Willow Flycatcher genome-wide genetic dataset, initially
152 examined using traditional phenomenological models, offers an ideal opportunity to apply a new
153 environmental association analysis that mechanistically accounts for genetic variation. More broadly, the
154 results have important implications for implementing management practices designed to improve the
155 genetic health of the endangered subspecies.

156

157 **Data**

158 A total of 175 Willow flycatchers were sampled at 23 one-degree squares across the continental
159 United States (Fig. 1). The sampling effort varied from 2-21 individuals per location. DNA was extracted
160 from blood and tissue samples using the QiagenTM DNeasy Blood and Tissue extraction kit and quantified
161 using the Qubit® dsDNA HS Assay kit (Thermo Fisher Scientific). Sequencing was conducted across three
162 lanes of 100 bp paired-end reads on an Illumina HiSeq 2500 at the UC Davis Genome Center. We filtered
163 SNPs using the tradeoff between discarding SNPs with low coverage and discarding individuals with
164 missing genotypes using the R package genoscapeRtools (Anderson 2019), resulting in approximately 105k
165 SNPs, of which 3 percent of the loci had a missing genotype (Ruegg *et al.* 2018). At each sampling location,
166 climate data were obtained from WorldClim, which averaged the climate between 1960 and 1990 (Hijmans
167 *et al.* 2005). Due to the high correlation between some of the top-ranked climatic variables identified in
168 Ruegg *et al.* (2018), we limited our analysis to the four least correlated climate variables (standardized) to
169 reduce collinearity: BIO 4 (temperature seasonality), BIO 5 (maximum temperature of the warmest month),

170 BIO 11 (mean temperature of the coldest quarter), and BIO 17 (Precipitation in the driest quarter). For more
171 details about the sampling design, sequencing, and bioinformatic analysis, refer to Ruegg *et al.* (2018).
172
173 **Model Formulation**
174 To identify putative loci under selection due to local adaptation, we specified a Bayesian model
175 with three hierarchical levels corresponding to the data, process, and parameter models. These three levels
176 are organized such that the output of the parameter model is the input for the process model, whose output
177 is the input for the observer model (Pagel & Schurr 2012). To facilitate a broad understanding of how the
178 statistical model works, Figure 2 illustrates the three levels of the hierarchical model, showing how these
179 levels are connected. In Figure 3, we provide a Directed Acyclic Graph (DAG) to illustrate the relationship
180 between data and parameters. In the following section, we give details on the assumptions underlying the
181 model's construction (also see the Supplementary Information for a concise summary).
182
183 *Data model*
184 We assume that Willow Flycatchers were sampled at K sites across the United States, and each bird
185 was genotyped at L independent genetic (RAD) markers (see sampling and sequencing above). We
186 characterized the genotype of the ith sampled bird at site k using g_{ilk}, which corresponds to the number of
187 copies of the reference allele at locus l. Assuming individuals have biallelic loci, g_{ilk} is a discrete variable
188 that takes values zero, one, or two. Using these individual genotype data, we define a population-level
189 genotype variable

$$y_{lk} = \sum_{i=1}^{N_{lk}} g_{ilk}, \tag{1}$$

190 that corresponds to the number of reference alleles for N_{lk} birds that were genotyped at locus l. In the data,
191 N_{lk} is always less than or equal to the number of birds sampled at a site because, for some individuals, the
192 genotype at a locus may be missing due to sequencing errors or low coverage. Assuming the birds were
193 randomly sampled, we model the variation in reference allele frequency (p_{lk}) using a binomial distribution,

$$y_{lk} \sim \text{Binomial}(2N_{lk}, p_{lk}). \tag{2}$$

Alternatively, we can also use

$$g_{ilk} \sim \text{Binomial}(2, p_{lk}), \tag{3}$$

as a data model to describe the statistical relationship between allele frequency and an individual's genotype. However, because of computational efficiency, we use equation (2). Note that the observer model accounts for the uncertainty in genetic variation that stems from finite sampling and missing genotypes.

Process model

We consider an island-metapopulation model proposed by Wright (1931) to model evolutionary dynamics. We rely on this model because it provides a parsimonious explanation of how evolutionary processes maintain genetic variation and allows for reasonable statistical inferences even when the process model is misspecified due to complexity in evolutionary dynamics (see the Model Testing section below).

We assume that the demes in the metapopulation correspond to the sites where the birds were sampled. Each deme has a population size of N_e, and the migration rate between any pair of demes is equal. We assume that the variation in allele frequencies across demes is maintained by directional and non-directional evolutionary forces. Directional forces—such as local adaptation to climate, mutation, and migration—result in changes in the mean value of allele frequency. Non-directional forces—such as genetic drift—do not change the mean value of allele frequency but create sampling variance due to finite population size. Because of this variance, the frequency of an allele cannot be determined exactly. Instead, allele frequency is characterized probabilistically using a probability distribution, $\psi(p_{lk}, t)$ (Rice 2004; Blanquart, Gandon & Nuismer 2012). Using the Fokker-Planck equation, one can show that the probability distribution of the frequency of the reference allele changes as follows:

$$\frac{\partial \psi(p_{lk}, t)}{\partial t} = -\frac{\partial}{\partial p_{lk}}[\psi(p_{lk}, t)M(p_{lk})] + \frac{1}{2}\frac{\partial^2}{\partial^2 p_{lk}}[\psi(p_{lk}, t)V(p_{lk})], \tag{4}$$

where

$$M(p_{lk}) = \underbrace{s_{lk}p_{lk}q_{lk}}_{\text{Selection}} + \underbrace{u_{2l}q_{lk} - u_{1l}p_{lk}}_{\text{Mutation}} + \underbrace{m(\overline{p_l} - p_{lk})}_{\text{Migration}} \tag{5}$$

is the rate of directional change in the allele frequency and

$$V(p_{lk}) = \frac{p_{lk}q_{lk}}{2N_e} \tag{6}$$

is the variance in the allele frequency due to non-directional (*i.e.*, drift) effects. In equations (4)-(6), $q_{lk}(= 1 - p_{lk})$ is the frequency of the alternate allele, s_{lk} is the environmentally regulated selection coefficient, u_{2l} and u_{1l} are the locus-specific forward and backward mutation rates, m is the migration rate, and $\overline{p_l}$ is the average allele frequency of the immigrants. We consider multiplicative selection dynamics (*i.e.*, no dominance), such that the relative fitness of individuals with genotype (g_{ilk}) zero, one, and two is $1 + 2s_{lk}$, $1 + s_{lk}$, and 1, respectively. Our model can also be modified to accommodate alternate forms of selection dynamics, such as frequency-dependent selection, underdominance, and heterozygote advantage.

At the time of sampling, the allele frequency distribution is assumed to be in quasi-equilibrium with climate conditions. To obtain this equilibrium distribution, also known as the stationary distribution (ψ_s), we set $\partial\psi(p_{lk}, t)/\partial t = 0$. The stationary distribution of the allele frequency can be expressed in terms of evolutionary parameters as follows (Blanquart, Gandon & Nuismer 2012):

$$\psi_s(p_{lk}|\tilde{s}_{lk}, \mu_l, \kappa_l) = \frac{e^{\tilde{s}_{lk}p_{lk}}}{{}_1F_1(\mu_l\kappa_l, \kappa_l, \tilde{s}_{lk})}\mathrm{Beta}(p_{lk}|\,\mu_l, \kappa_l), \tag{7}$$

where $\tilde{s}_{lk} = 4N_e s_{lk}$, ${}_1F_1$ is the hypergeometric confluent function, and $\mu_l = (u_{2l} + m\overline{p_l})/(u_{1l} + u_{2l} + m)$ and $\kappa_l = 4N_e(u_{1l} + u_{2l} + m)$ are the mean and precision parameters of the beta distribution. To incorporate climate adaptation, we assume that the selection coefficient, $\tilde{s}_{lk}$, is regulated by local environmental conditions. For simplicity, we assume a linear relationship between the selection coefficient and *E* standardized environmental variables:

$$\tilde{s}_{lk} = \alpha_l + \sum_{j=1}^{E} \beta_{lj} e_{jk}. \tag{8}$$

However, based on the natural history of the species, a biogeographer may consider alternative functional forms to describe the relationship between the selection coefficient and environmental variables. In equation (8), e_{jk} is the jth environmental variable (standardized) in deme k, β_{lj} is the selection coefficient's

sensitivity to variation in the jth environmental variable, and α_l can be interpreted either as the selection coefficient at mean environmental conditions or a fixed value of selection coefficient arising from adaptive processes other than local climatic adaptation.

The stationary distribution in equation (7) has several notable features. The response curve of an adaptive allele (the relationship between climate and mean allele frequency),

$$\mathbb{E}[p_{lk}] = \mu_l \frac{{}_1F_1(1+\mu_l\kappa_l, 1+\kappa_l, \tilde{s}_{lk})}{{}_1F_1(\mu_l\kappa_l, \kappa_l, \tilde{s}_{lk})}, \tag{9}$$

is bounded between zero and one, and its shape is determined by evolutionary parameters with biological interpretation (Blanquart, Gandon & Nuismer 2012). When selection is absent, the beta distribution captures the non-adaptive genetic variation,

$$\psi_s(p_{lk}|0, \mu_l, \kappa_l) = \mathrm{Beta}(p_{lk}|\, \mu_l, \kappa_l). \tag{10}$$

Therefore, to account for the joint contribution of adaptive and non-adaptive evolutionary forces in determining genetic variation, we use the stationary distribution to statistically model evolutionary dynamics,

$$p_{lk} \sim \frac{e^{\tilde{s}_{lk}p_{lk}}}{{}_1F_1(\mu_l\kappa_l, \kappa_l, \tilde{s}_{lk})} \mathrm{Beta}(\mu_l, \kappa_l). \tag{11}$$

For fast and memory-efficient implementation of the statistical model, we combine the data (Eq. 2) and process (Eq. 11) models by marginalizing over p_{lk},

$$y_{lk} \sim \int_0^1 \mathrm{Binomial}(y_{lk}|2N_{lk}, p_{lk})\psi_s(p_{lk}|\tilde{s}_{lk}, \mu_l, \kappa_l)\, dp_{lk}, \tag{12}$$

which results in an integrated likelihood that relates genotype counts to evolutionary dynamics as follows:

$$y_{lk} \sim \frac{{}_1F_1(\mu_l\kappa_l + y_{lk}, 2N_{lk} + \kappa_l, \tilde{s}_{lk})}{{}_1F_1(\mu_l\kappa_l, \kappa_l, \tilde{s}_{lk})} \mathrm{BetaBinomial}(2N_{lk}, \mu_l, \kappa_l). \tag{13}$$

Note that the integrated likelihood in equation (13) accounts for the uncertainty that stems from noisy genetic data and the stochastic nature of evolutionary dynamics.

Parameter model

Finally, we assign priors to parameters in equation (13) based on our prior scientific knowledge. We assign $\mu_l \sim \text{Uniform}(0,1)$ and $\kappa_l \sim \text{Normal}^+(0,5)$ priors in equation (13). Due to degeneracy of the likelihood (Eq. 11, see Fig. S1), the mean of the beta distribution (μ_l) and selection coefficient at the mean environment (α_l) are non-identifiable (see Supplementary Information for more details). We alleviate this by fixing $\alpha_l = 0$ and re-interpret μ_l as the mean of the baseline allele frequency distribution (or neutral distribution) resulting from evolutionary forces other than adaptation to variation in local climate around the mean. Next, because we expect that most of the genetic variation arises due to neutral processes (Kimura 1983), we use a regularized horseshoe shrinkage parameter model (Piironen & Vehtari 2017) for sensitivity coefficients, β_{lj}. The horseshoe prior shrinks most of the β_{lj} to zero, allowing a fraction of the coefficients to take non-zero values. The horseshoe model achieves this using global (τ) and local ($\tilde{\lambda}_{lj}$) shrinkage parameters and is expressed as

$$\beta_{lj} \sim \text{Normal}\left(0\,, \tau\tilde{\lambda}_{lj}\right), \tag{14}$$

where

$$\tilde{\lambda}_{lj} = \frac{c\lambda_{lj}}{\sqrt{c^2 + \tau^2\lambda_{lj}^2}}. \tag{15}$$

To control global shrinkage, we specify a half-Cauchy prior,

$$\tau \sim \text{Cauchy}^+\left(0, \frac{f}{\sqrt{K}LE}\right), \tag{16}$$

where f is a free variable, which we set to 20. Based on the recommendations of Piironen and Vehtari (2017), we tested a range of values of f and found that the statistical results are robust to orders of magnitude of variation in f (see Model Testing section below). Note that the global sparsity (controlled by τ) increases with the number of SNPs (L) and environmental variables (E). On the other hand, the local shrinkage parameter, $\tilde{\lambda}_{lj}$, allows some coefficients to escape shrinkage because of a heavy-tail prior, $\lambda_{lj} \sim \text{Cauchy}^+(0,1)$. However, these large-valued coefficients are weakly shrunk with an implied regularization of $\text{Normal}(0, c)$ with $c^2 \sim \text{InvGamma}(2,4)$. We adopt an inverse-gamma prior because the

prior means for both c^2 (which controls adaptive dynamics via sensitivity coefficients) and κ_l (which governs neutral dynamics) are similar in magnitude, but the heavy right tail of the inverse-gamma distribution allows adaptive loci to attain large sensitivity coefficients. These features of the local and global shrinkage parameters effectively induce a spike and slab prior (Mitchell & Beauchamp 1988), which prevent numerical pathologies when large effect loci are weakly identifiable.

Model Testing

To assess the robustness of our statistical method, we fit the model to synthetic data that are similar in characteristics to the Willow Flycatcher genomic data, and the generative process only partially resembles the process model considered in our statistical model. This simulation allows us to test the sensitivity of our method to model misspecifications and test its applicability to biologically realistic datasets.

We generated the synthetic genotype data in three stages (see Supplementary code for simulation details). In the first stage, we simulated selection dynamics. We considered diploid individuals with a genome size of one thousand independent loci. Each locus in the genome was assigned a random non-zero α_l. We randomly selected fifteen loci that contribute to local adaptation. Each of the fifteen loci was randomly paired with one of the four environmental variables, and the corresponding selection coefficient (s_{lk}) was calculated using equation (8). We used real environmental values (standardized) to calculate s_{lk} to preserve the correlation structure between environmental variables.

In the second stage, we simulated metapopulation dynamics with demes equal to the number of sampling sites in the real data. We sampled random variables from the stationary distribution of allele frequencies by simulating the following stochastic differential equation (Korolev *et al.* 2010):

$$dp_{lk} = \left[\underbrace{s_{lk}p_{lk}q_{lk}}_{\text{Selection}} + \underbrace{u_{2l}q_{lk} - u_{1l}p_{lk}}_{\text{Mutation}} + \sum_{n=1}^{K} \underbrace{e^{-\rho d_{nk}}(p_{ln} - p_{lk})}_{\text{Migration}}\right] dt + \underbrace{\sqrt{p_{lk}q_{lk}/2N_e}}_{\text{Drift}}\, dB, \quad (17)$$

294 where s_{lk} is the selection coefficient obtained from stage one, B is the standard Brownian motion, d_{nk} is
295 the distance between demes n and k, and ρ^{-1} is the dispersal length scale. The above equation is an Itô
296 representation of the Fokker-Planck equation presented earlier (Eq. 4), with a small modification: migration
297 rate depends on the distance between demes (Fig. S2).

298 In the third stage, we simulated genotypes. In each deme, we simulated genotypes for the same
299 number of individuals as in the real data. For each individual, we generated a genotype at a locus by
300 sampling from the distribution $\text{Binomial}(2, p_{lk})$ (Eq. 3), where p_{lk} is the reference allele frequency
301 obtained from the second stage. We randomly selected three percent of the loci and treated them as missing.

302 To test if the statistical model can identify loci that contribute to local adaptation in synthetic data,
303 we fit the model using Stan programming language (Carpenter *et al.* 2017) to obtain a posterior sample of
304 sensitivity coefficients, β_{lj}, using Hamiltonian Monte Carlo (Neal 2011). For each pair of loci and
305 environmental variables, we computed the probability that the posterior distribution of the sensitivity
306 coefficient included zero. If this probability is less than a threshold value of 0.05 (p_{th}), we inferred that the
307 locus may have contributed to local adaptation for the corresponding environmental variable.

308 To conduct robustness checks, we generated ten synthetic datasets for nine parameter regimes (a
309 total of 90 datasets; see Supplementary Information for parameter values used to generate synthetic
310 datasets). These parameter combinations represent three mutation and migration regimes—high, medium,
311 and low—relative to local adaptation. For each regime, we calculated the false-negative and false-positive
312 rates. We found that our model performs best in the low mutation-migration regime, where the genetic
313 variation is more sensitive to climatic variation (Tables S1 and S2). However, this regime also yields higher
314 rates of false positives. In contrast, in the high mutation-migration regime, neutral evolutionary forces play
315 a larger role, and consequently, the statistical model results in high false negative rates.

316 For illustration, the Manhattan plot in Fig. 4A shows the result from a simulation corresponding to
317 the low mutation and medium migration regime. The y-axis corresponds to the negative log probability that
318 the posterior distribution of sensitivity coefficients included zero (Wang *et al.* 2022). The points above $y =$

319 2.99 (the black line) have $p_{th} < 0.05$. These points are characterized by a ring and solid center. The color
320 of the center represents the true environmental variable associated with the loci, and the color of the ring
321 represents the statistically inferred adaptive environmental variable. Note that, in this simulation run, the
322 model incorrectly identified the corresponding environmental variable for two of the nine loci (points above
323 the black line with a green ring and purple center). This feature can be explained by a negative correlation
324 (-0.72) between environmental variables, BIO 4 (green) and BIO 11 (purple). As a result, the statistical
325 model used one of the environmental variables as a substitute for another (Rellstab *et al.* 2015). We also
326 re-analyzed the data in Fig. 4A by varying f over four orders of magnitude (2^0 to 2^{14}). We find that our
327 results are nearly identical to results in Fig. 4A (see Fig. S3 and S4), suggesting that our analysis is robust
328 to a wide variation in choice of priors for τ.

329 Next, we re-analyzed the synthetic datasets using latent factor mixed model (LFMM; Frichot *et al.*
330 2013) which was used to analyze Willow Flycatcher data in Ruegg *et al.* (2018). We conducted simulations
331 using two versions of the synthetic dataset. In the first version, we used individual genotypes, and, in the
332 second version, we used raw allele frequencies (*i.e.*, $p_{lk} = 0.5\, y_{lk}/N_{lk}$) as input. Our results suggest that
333 LFMM yielded much higher false-negative and false-positive rates (Tables S3-S6).

334

335 **Data Analysis**

336 Analyzing genomic data from Willow Flycatchers revealed 47 significant loci, all except one were
337 associated with BIO 4 (Fig. 4B and Fig. S5). A closer inspection of these loci suggests that many of them
338 cluster together on the genome (indicated by vertical gray lines in Fig. 4B). These clustering patterns are
339 unlikely to happen by random chance. Alternatively, the observed clustering patterns can be explained by
340 gene hitchhiking (Smith & Haigh 1974). The allele frequencies at a locus under selection and its
341 neighboring neutral loci rise or fall in unison due to physical linkage on the chromosome. Consequently,
342 when a gene is under selection, neutral polymorphic sites close to the gene experience environmentally
343 regulated pseudo-selection (Barton 1998).

Using the annotated genome of Willow Flycatchers, 36 of the candidate loci were found within or nearby (within 25kb) 30 named genes (Table S7). Twenty-three of these genes are characterized, and 20 have functional roles in chicken (*Gallus gallus*) that span 8 gene ontology categories. The majority of genes cluster in 4 categories: 5 genes are involved in catalytic activity, 5 genes have binding functionality, 3 have transcription regulatory activity, and 3 have transporter activity. A closer investigation into several genes shows that EDIL3 plays a role in the egg mineralization process in Aves (Le Roy *et al.* 2021), PCDH1 is involved in feather development (Lin, Wang & Redies 2013), and GRIK2 acts as a thermoreceptor conferring sensitivity to cold temperatures in mice (Cai *et al.* 2024).

The above analysis required 32 hours to complete on a machine with a 24-core 3.6 GHz processor. By subsampling the genomic data, we also estimated the model run time for $E = 4$ and $L = 2^8, 2^9, \ldots,$ and 2^{17} (Fig. S6).

Discussion

Adaptation to climate is ubiquitous and will continue to play a major role in maintaining biodiversity in the Anthropocene (Thompson 2013). However, many commonly used methods for identifying genomic adaptation are limited because they often lack the ability to handle noisy genetic data and do not formally account for the demographic history of the species. Our mechanistic Bayesian model addresses these limitations. The important aspects of our approach comprise the following features.

First, we proposed a data model that probabilistically links RAD-seq data to genetic variation (Eq. 2). This probabilistic link allowed us to quantify uncertainty in genetic variation that arises due to finite sampling and missing genotypes. Consequently, the data model properly accounts for uncertainty in the estimation of genetic variation and faithfully propagates the available information in raw genetic data to the process model for downstream analysis (Hobbs & Hooten 2025).

Second, we used a metapopulation process model to partition estimated genetic variation into adaptive and non-adaptive components (Blanquart, Gandon & Nuismer 2012). The metapopulation model can be implemented using a beta distribution to characterize neutral variation (Eq. 10) and hypergeometric

370 confluent functions to model adaptive variation (Eq. 9). Our process model is motivated by theoretical
371 principles from evolutionary biology. Thus, we can evaluate the model based on the underlying
372 assumptions.

373 Finally, previous work shows that most of the genetic variation in wild populations is maintained
374 by non-adaptive evolutionary forces (Kimura 1983). We incorporated this prior knowledge into our
375 statistical model using a regularized horseshoe parameter model, which shrinks most of the selection
376 coefficients to zero. The level of shrinkage is controlled by a global shrinkage parameter that introduces
377 sparsity in sensitivity coefficients (Piironen & Vehtari 2017) (Eq. 14). Thus, the parameter model allowed
378 us to partition genetic variation into adaptive (driven by environmental variation) and non-adaptive (driven
379 by neutral processes) components.

380 To test our statistical model, we conducted simulations to assess the model's performance on ninety
381 synthetic RAD-seq datasets generated by simulating bird genomes of 1000 loci (Eq. 17), which emulate the
382 characteristics of genetic data from Willow Flycatchers. Our statistical method performed best in the low
383 mutation-migration regime, and the performance decreased in the high mutation-migration regime (Table
384 S1 and S2). These results are consistent with our theoretical expectations. When the mutation and migration
385 rates are low, genetic variation is more sensitive to variation in climate. However, in the high mutation-
386 migration regime, neutral processes play a bigger role in maintaining genetic variation. As a result, it is
387 more challenging to identify weakly adaptive loci.

388 Figure 4A shows results from a simulation from the low mutation and medium migration regime.
389 The model correctly identified nine out of fifteen adaptive loci. However, we found that out of the nine
390 correctly identified loci, two loci were incorrectly paired with their corresponding environmental variable
391 due to a -0.72 correlation between BIO 4 and BIO 11. This highlights that biogeographers may need to
392 exercise caution when interpreting statistical results or use uncorrelated predictors. The statistically inferred
393 environmental variable might be correlated with the true environmental variable responsible for local
394 adaptation (Rellstab *et al.* 2015). Nevertheless, the synthetic simulations suggest that our statistical model
395 can be used with genetic datasets from wild populations.

396 Indeed, we identified 30 genes located within the 25 kb region flanking 47 significant loci in RAD-
397 seq data from Willow Flycatchers, most of which were associated with temperature seasonality (Fig. 4B
398 and S5). Some of these genes include EDIL3, PCDH1, and GRIK2, which play a role in the egg
399 mineralization process (Le Roy *et al.* 2021), feather development (Lin, Wang & Redies 2013), and act as a
400 thermoreceptor conferring sensitivity to cold temperatures (Cai *et al.* 2024), respectively. This suggests that
401 temperature fluctuations may be a key driver of local adaptation in the species, providing evidence that
402 standing genetic variation in Willow Flycatchers could alleviate or buffer extinction risk due to increasing
403 temperature variability predicted by climate projections (Olonscheck *et al.* 2021).

404 Although these genes differ from those identified in Ruegg *et al.* (2018), these differences are not
405 surprising because our hierarchical model is fundamentally different from traditional environmental
406 association analysis in terms of quantifying uncertainty while estimating genetic variation and using an
407 evolutionary process model to explain sources of genetic variation. Our synthetic data simulations confirm
408 that these differences may play an important role in inferences; re-analyzing synthetic data using LFMM
409 resulted in substantially higher rates of false positives and false negatives (Table S3-S6).

410 In addition to the conceptual advantages offered by our mechanistic statistical model, synthetic
411 simulations highlight several practical scenarios where our hierarchical model may be better suited to
412 analyze genomic data. Some of these scenarios include low sample size (between 1-4 individuals), wide
413 dispersion in the number of individuals sampled at various locations, and a large fraction of missing
414 genotypes. We can address these data scenarios by quantifying genetic variation probabilistically and
415 propagating the corresponding uncertainty in downstream analyses. However, there are some
416 computational tradeoffs: our statistical analysis required 32 hours to analyze 105k SNPs, and, as a result,
417 our approach may require substantial computational resources to analyze genomic datasets with millions of
418 SNPs.

419 Despite these computational requirements, the hierarchical structure of the model provides new
420 opportunities to improve statistical inferences. For example, the hierarchical model is flexible and can
421 integrate a wide range of process and data models, allowing biogeographers to better understand the

422 evolutionary responses of a species to a changing climate while dramatically reducing the cost of analysis.
423 The cost of sequencing varies along three axes—the number of sequenced sites in the genome, the depth of
424 sequencing effort, and the number of sampled individuals. Typically, in RAD sequencing, a small
425 proportion of sites in the genome are sequenced deeply to identify true genotypes, which are subsequently
426 used to estimate allele frequencies (Baird *et al.* 2008). Although RAD sequencing is a cost-effective
427 protocol to obtain a reduced representation of the genome, it fails to characterize a large proportion of
428 genetic variation in the genome that could be potentially adaptive (Lowry *et al.* 2017). New sequencing
429 protocols, such as low-depth whole genome sequencing (Alex Buerkle & Gompert 2013) and pool
430 sequencing (Gautier *et al.* 2013), are emerging as attractive alternatives because they provide a wider
431 genomic coverage without increasing cost or sacrificing statistical power. The key idea behind these
432 approaches is to redistribute the same resources to sequence a larger sample of individuals and a greater
433 proportion of the genome (Lou *et al.* 2021). This reduction in sequencing effort (per individual per locus)
434 increases uncertainty in the estimates of individual genotypes. But, instead of discarding SNPs due to low
435 coverage, one can model the true genotype as a parameter and propagate the corresponding uncertainty to
436 inform population-level allele frequencies. In effect, these approaches sacrifice certainty in individual
437 genotypes to gain genomic coverage while keeping the cost and genetic information constant. To leverage
438 information from these sequencing protocols, the data model in the Bayesian hierarchy can be modified to
439 account for genotype uncertainty in estimating allele frequencies.

440 Another potential avenue to improve inference is to construct alternate process models
441 incorporating a wider range of demographic histories to correct for model misspecification. For example,
442 most species, including Willow Flycatchers, are geographically structured, and, as such, the migration rates
443 depend on the distance between demes. Closer demes exchange more migrants than demes that are far apart.
444 This may create genetic patterns, such as isolation by distance (Fig. S2) and genetic swamping, that cannot
445 be adequately captured by an unstructured metapopulation model (Eq. 11). Because of these
446 misspecifications, our method less sensitive in identifying adaptive loci in weak selection and high
447 migration regimes. One possible resolution to this problem is to use the stationary distribution

corresponding to the structured evolutionary dynamics as the process model (see Constable & McKane 2015 for an analytical solution for the stationary distribution of a structured metapopulation). Our process model may also be limited in non-equilibrium settings, such as range shifts due to rapidly changing climate or land use practices. In these cases, the quasi-stationary assumption in the process model is unlikely to hold because of transient evolutionary dynamics and eco-evolutionary feedback. Consequently, one may want to consider a probabilistic process model that jointly describes spatiotemporal changes in abundance and genetic variation (Schurr *et al.* 2012; Polechová & Barton 2015). To inform these joint process models, researchers will require temporal and spatial sampling of abundances and genomes, which can be obtained from population surveys (Pardieck *et al.* 2020) and museum collections (Payne & Sorenson 2002), respectively. These improvements in data and process models may allow biogeographers to forecast species' response to global change using mechanistic models of ecology and evolution.

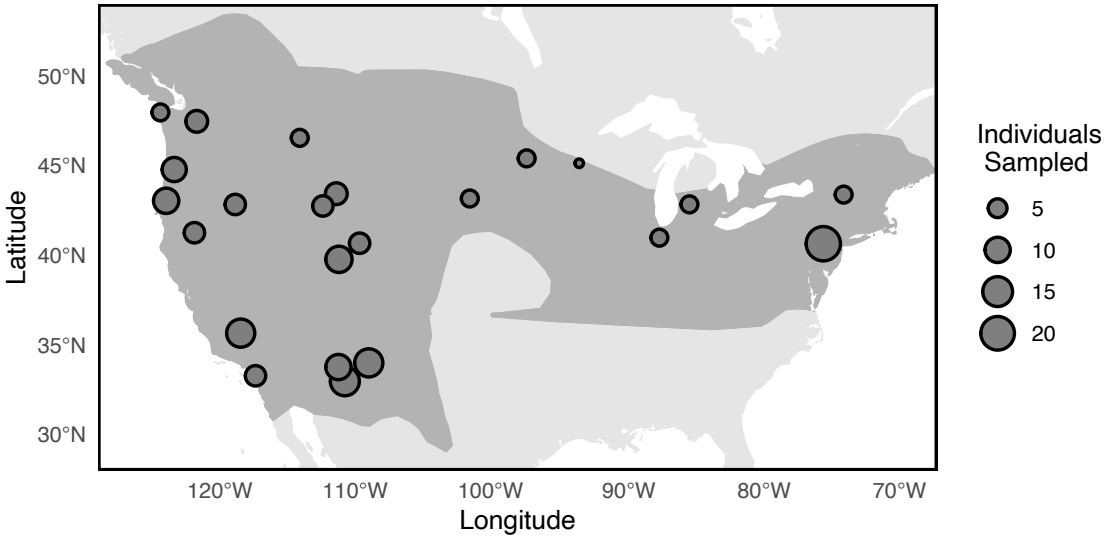

Figure 1: Breeding range of Willow Flycatchers (dark grey region) and 23 sampling locations. The size of the points is proportional to the number of sampled individuals in our study, which varied between 2 and 21 individuals.

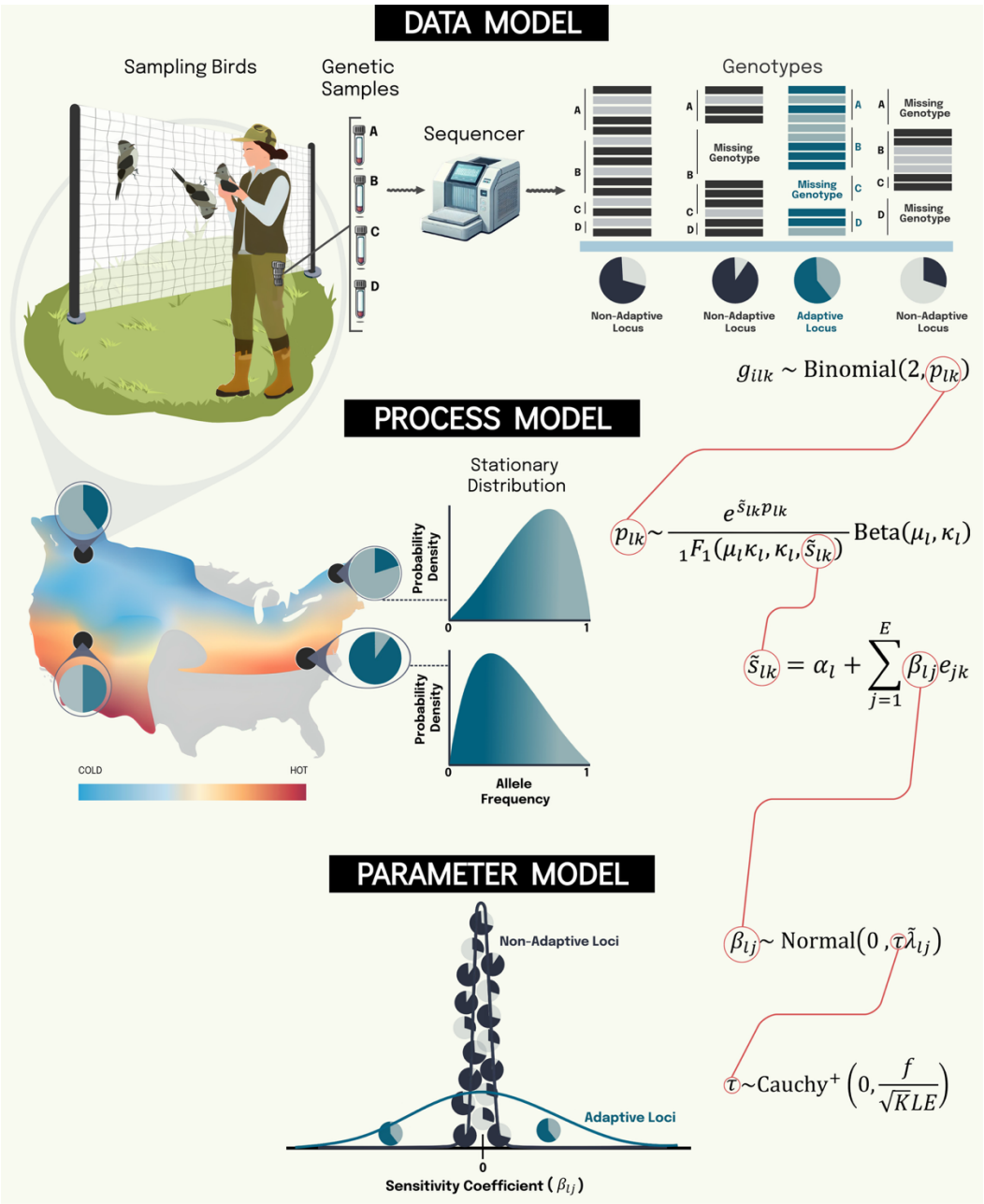

Figure 2: A conceptual diagram describing the hierarchical structure of the Bayesian model. The data model (*top*) links the individuals' genotype (g_{ilk}) to allele frequency (p_{lk}) at a locus l in patch k (Eq. 3). The process model (*middle*) describes the stationary distribution of allele frequency resulting from evolutionary forces, such as mutation, migration, drift, and local adaptation. (Eq. 8 and 11). Finally, the parameter model (*bottom*) allows us to regularize sensitivity coefficients (β_{lj}) using prior knowledge from molecular evolutionary theory (Eqs. [14], [15], and [16]).

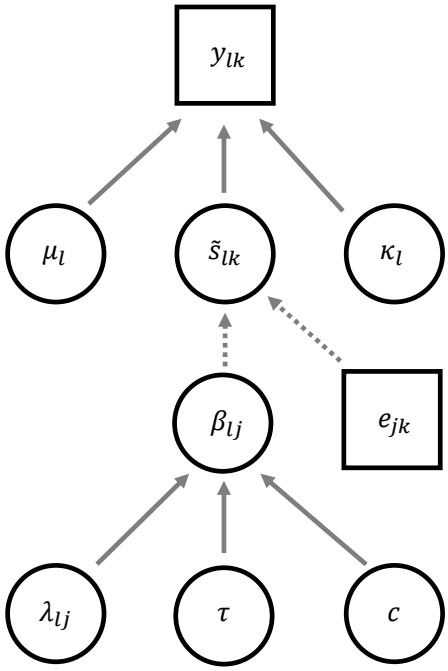

Figure 3: Directed Acyclic Graph (DAG) of the hierarchical Bayesian model. The squares represent data (genotype counts and environmental variables), the circles represent parameters, the dashed arrows represent deterministic relationships, and the solid arrows represent stochastic relationships.

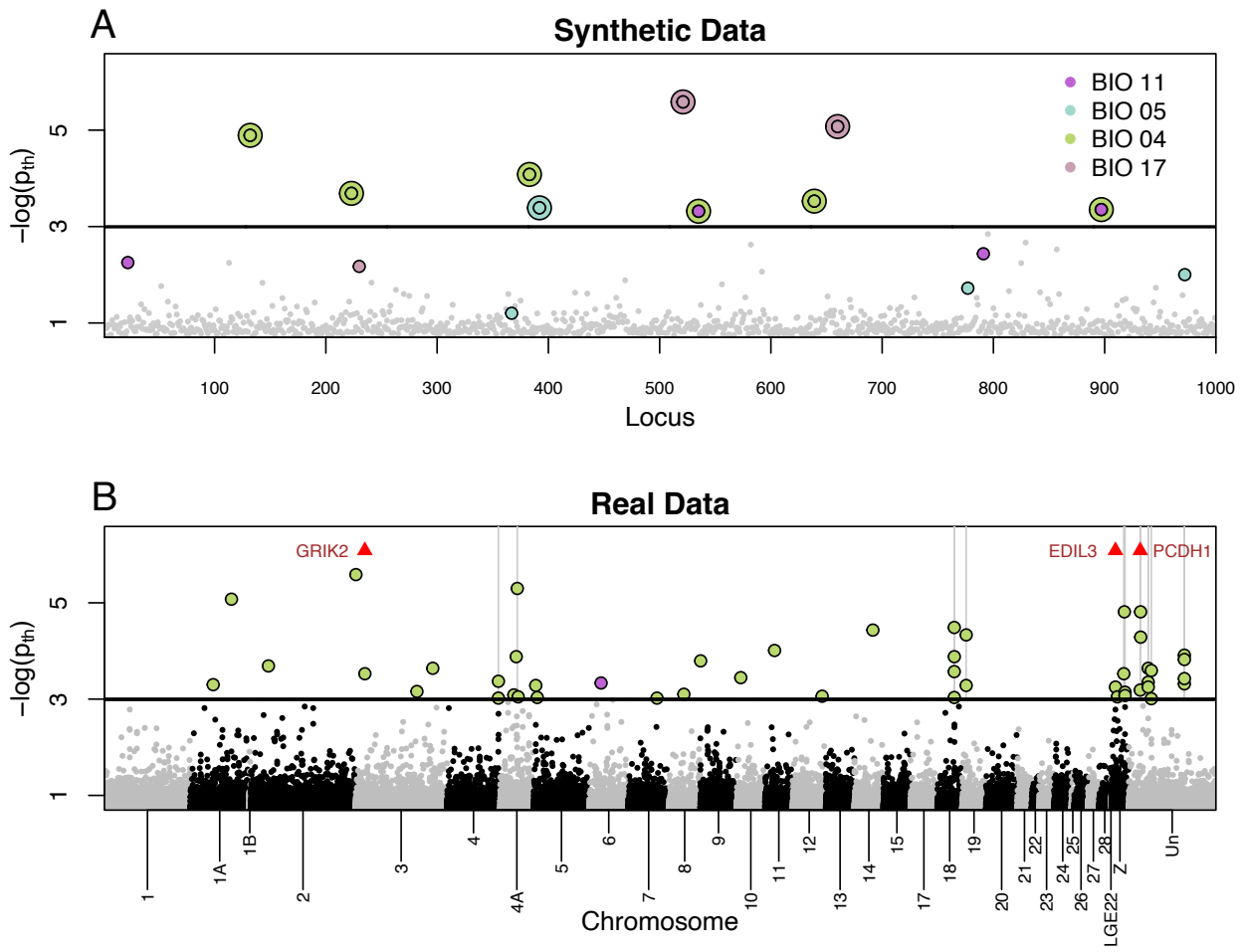

Figure 4: Manhattan plot showing negative log probability ($-\log(p_{th})$) that the posterior distribution of sensitivity coefficients (β_{li}) includes zero for synthetic (A) and real data (B). In both plots, points above the black horizontal line have p_{th} less than 0.05. (A) For synthetic data, we denote points above the black line using an outer ring and a center. The color of the center (ring) corresponds to the true (statistically inferred) environmental variable responsible for local adaptation. (B) For real data, we denote points above the black line with a solid center, with its color corresponding to the statistically inferred environmental variable responsible for local adaptation (also see Fig. S5). The vertical gray lines represent physically linked loci with statistically significant sensitivity coefficients (see Supplementary Information for parameter values used to simulate synthetic data in Fig. 4A)

References

Alex Buerkle, C. & Gompert, Z. (2013) Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology,* **22,** 3028-3035.

Anderson, E. (2019) genoscapeRtools: Tools for building migratory bird genoscapes. *R package version 0.1. 0. https://rdrr. io/github/eriqande/genoscapeRtools*.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE,* **3,** e3376.

Barton, N.H. (1998) The effect of hitch-hiking on neutral genealogies. *Genetics Research,* **72,** 123-133.

Beaumont, M.A. & Rannala, B. (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics,* **5,** 251-261.

Berliner, L.M. (1996) Hierarchical Bayesian Time Series Models. pp. 15-22. Springer Netherlands, Dordrecht.

Blanquart, F., Gandon, S. & Nuismer, S. (2012) The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology,* **25,** 1351-1363.

Cai, W., Zhang, W., Zheng, Q., Hor, C.C., Pan, T., Fatima, M., Dong, X., Duan, B. & Xu, X.S. (2024) The kainate receptor GluK2 mediates cold sensing in mice. *Nature Neuroscience,* **27,** 679-688.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017) Stan: A probabilistic programming language. *Journal of Statistical Software,* **76,** 1–32.

Carroll, C., Lawler, J.J., Roberts, D.R. & Hamann, A. (2015) Biotic and climatic velocity identify contrasting areas of vulnerability to climate change. *PLoS ONE,* **10,** e0140486.

Chuvieco, E. (2020) *Fundamentals of Satellite Remote Sensing: An Environmental Approach*. CRC press, Boca Raton.

Constable, G.W. & McKane, A.J. (2015) Stationary solutions for metapopulation Moran models with mutation and selection. *Physical Review E,* **91,** 032711.

512 Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. (2010) Using environmental correlations to
513 identify loci underlying local adaptation. *Genetics,* **185,** 1411-1423.
514 De Villemereuil, P. & Gaggiotti, O.E. (2015) A new FST-based method to uncover local adaptation using
515 environmental variables. *Methods in Ecology and Evolution,* **6,** 1248-1258.
516 Fitzpatrick, M.C. & Keller, S.R. (2015) Ecological genomics meets community-level modelling of
517 biodiversity: Mapping the genomic landscape of current and future environmental adaptation.
518 *Ecology Letters,* **18,** 1-16.
519 Foll, M. & Gaggiotti, O. (2008) A genome-scan method to identify selected loci appropriate for both
520 dominant and codominant markers: A Bayesian perspective. *Genetics,* **180,** 977-993.
521 Frichot, E., Schoville, S.D., Bouchard, G. & François, O. (2013) Testing for associations between loci and
522 environmental gradients using latent factor mixed models. *Molecular Biology and Evolution,* **30,**
523 1687-1699.
524 Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P.,
525 Kerdelhué, C. & Estoup, A. (2013) Estimation of population allele frequencies from next-
526 generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology,* **22,**
527 3766-3779.
528 Gompert, Z. & Buerkle, C.A. (2011) A hierarchical Bayesian model for next-generation population
529 genomics. *Genetics,* **187,** 903-917.
530 Günther, T. & Coop, G. (2013) Robust identification of local adaptation from allele frequencies. *Genetics,*
531 **195,** 205-220.
532 Hanks, E.M. & Hooten, M.B. (2013) Circuit theory and model-based inference for landscape connectivity.
533 *Journal of the American Statistical Association,* **108,** 22-33.
534 Hedrick, P.W. (1986) Genetic polymorphism in heterogeneous environments: A decade later. *Annual*
535 *Review of Ecology and Systematics,* **17,** 535-566.
536 Hedrick, P.W. (2006) Genetic polymorphism in heterogeneous environments: The age of genomics. *Annual*
537 *Review of Ecology and Systematics,* **37,** 67-93.

538 Hedrick, P.W., Ginevan, M.E. & Ewing, E.P. (1976) Genetic polymorphism in heterogeneous
539 environments. *Annual Review of Ecology and Systematics,* **7,** 1-32.
540 Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005) Very high resolution interpolated
541 climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal*
542 *Meteorological Society,* **25,** 1965-1978.
543 Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K.,
544 Storfer, A. & Whitlock, M.C. (2016) Finding the genomic basis of local adaptation: pitfalls,
545 practical solutions, and future directions. *The American Naturalist,* **188,** 379-397.
546 Hobbs, N.T. & Hooten, M.B. (2025) *Bayesian Models: A Statistical Primer for Ecologists, 2nd Edition*.
547 Princeton University Press, Princeton, New Jersey.
548 Hooten, M.B. & Hefley, T.J. (2019) *Bringing Bayesian Models to Life*. CRC Press, Boca Raton, Florida,
549 USA.
550 Huang, H. & Knowles, L.L. (2016) Unforeseen consequences of excluding missing data from next-
551 generation sequences: Simulation study of RAD sequences. *Systematic Biology,* **65,** 357-365.
552 Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G. & Taberlet, P. (2007) A spatial
553 analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics
554 approach to adaptation. *Molecular Ecology,* **16,** 3955-3969.
555 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge
556 UK.
557 Korolev, K.S., Avlund, M., Hallatschek, O. & Nelson, D.R. (2010) Genetic demixing and evolution in
558 linear stepping stone models. *Reviews of Modern Physics,* **82,** 1691-1718.
559 Le Roy, N., Stapane, L., Gautron, J. & Hincke, M.T. (2021) Evolution of the avian eggshell
560 biomineralization protein toolkit–new insights from multi-omics. *Frontiers in Genetics,* **12,**
561 672433.
562 Lin, J., Wang, C. & Redies, C. (2013) Expression of multiple delta-protocadherins during feather bud
563 formation. *Gene Expression Patterns,* **13,** 57-65.

564 Lou, R.N., Jacobs, A., Wilder, A.P. & Therkildsen, N.O. (2021) A beginner's guide to low-coverage whole
565 genome sequencing for population genomics. *Molecular Ecology,* **30,** 5966-5993.
566 Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F. & Storfer, A. (2017)
567 Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for
568 genome scans of adaptation. *Molecular Ecology Resources* **17,** 142–152.
569 Mahoney, S.M., Reudink, M.W., Pasch, B. & Theimer, T.C. (2020) Song but not plumage varies
570 geographically among Willow Flycatcher *Empidonax traillii* subspecies. *Journal of Avian Biology,*
571 **51**.
572 Marcus, J., Ha, W., Barber, R.F. & Novembre, J. (2021) Fast and flexible estimation of effective migration
573 surfaces. *Elife,* **10,** e61927.
574 Mitchell, T.J. & Beauchamp, J.J. (1988) Bayesian variable selection in linear regression. *Journal of the*
575 *American Statistical Association,* **83,** 1023-1032.
576 Neal, R.M. (2011) MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* (eds S.
577 Brooks, A. Gelman, G. Jones & X.-L. Meng), pp. 2. CRC Press, New York.
578 Olonscheck, D., Schurer, A.P., Lücke, L. & Hegerl, G.C. (2021) Large-scale emergence of regional changes
579 in year-to-year temperature variability by the end of the 21st century. *Nature Communications,* **12,**
580 7237.
581 Pagel, J. & Schurr, F.M. (2012) Forecasting species ranges by statistical estimation of ecological niches
582 and spatial population dynamics. *Global Ecology and Biogeography,* **21,** 293—304.
583 Pardieck, K.L., Ziolkowski, D., Lutmerding, M., Aponte, V. & Hudson, M.R. (2020) North American
584 Breeding Bird Survey Dataset 1966—2019 (ed. U.S.G.S.).
585 Payne, R.B. & Sorenson, M.D. (2002) Museum collections as sources of genetic data. *Bonner Zoologische*
586 *Beiträge,* **51,** 97-104.
587 Piironen, J. & Vehtari, A. (2017) Sparsity information and regularization in the horseshoe and other
588 shrinkage priors. *Electronic Journal of Statistics,* **11,** 5018–5051.

589 Polechová, J. & Barton, N.H. (2015) Limits to adaptation along environmental gradients. *Proceedings of*
590 *the National Academy of Sciences,* **112,** 6401-6406.
591 Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. & Holderegger, R. (2015) A practical guide to
592 environmental association analysis in landscape genomics. *Molecular Ecology,* **24,** 4348-4370.
593 Rice, S.H. (2004) *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer Associates,
594 MA.
595 Ruegg, K., Bay, R.A., Anderson, E.C., Saracco, J.F., Harrigan, R.J., Whitfield, M., Paxton, E.H. & Smith,
596 T.B. (2018) Ecological genomics predicts climate vulnerability in an endangered southwestern
597 songbird. *Ecology Letters,* **21,** 1085-1096.
598 Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra,
599 A.K. & Das, G. (2023) Next-generation sequencing technology: current trends and advancements.
600 *Biology,* **12,** 997.
601 Schurr, F.M., Pagel, J., Cabral, J.S., Groeneveld, J., Bykova, O., O'Hara, R.B., Hartig, F., Kissling, W.D.,
602 Linder, H.P., Midgley, G.F., Schröder, B., Singer, A. & Zimmermann, N.E. (2012) How to
603 understand species' niches and range dynamics: a demographic research agenda for biogeography.
604 *Journal of Biogeography,* **39,** 2146-2162.
605 Sedgwick, J.A. (2000) Willow Flycatcher (*Empidonax traillii*). *The Birds of North America,* **533**.
606 Smith, J.M. & Haigh, J. (1974) The hitch-hiking effect of a favourable gene. *Genetics Research,* **23,** 23-35.
607 Smith, T.B., Kinnison, M.T., Strauss, S.Y., Fuller, T.L. & Carroll, S.P. (2014) Prescriptive evolution to
608 conserve and manage biodiversity. *Annual Review of Ecology, Evolution, and Systematics,* **45,** 1-
609 22.
610 Stadler, T. (2010) Sampling-through-time in birth–death trees. *Journal of Theoretical Biology,* **267,** 396-
611 404.
612 Stadler, T., Magnus, C., Vaughan, T., Barido-Sottani, J., Bošková, V., Huisman, J. & Pečerska, J. (2024)
613 *Decoding Genomes: From Sequences to Phylodynamics*. ETH Zurich, Zurich.

614 Stucki, S., Orozco-terWengel, P., Forester, B.R., Duruz, S., Colli, L., Masembe, C., Negrini, R., Landguth,
615 E., Jones, M.R. & Consortium, N. (2017) High performance computation of landscape genomic
616 models including local indicators of spatial association. *Molecular Ecology Resources,* **17,** 1072-
617 1089.
618 Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J. & Rambaut, A. (2018) Bayesian
619 phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution,* **4,** vey016.
620 Theimer, T.C., Smith, A.D., Mahoney, S.M. & Ironside, K.E. (2016) Available data support protection of
621 the Southwestern Willow Flycatcher under the Endangered Species Act. *The Condor:*
622 *Ornithological Applications,* **118,** 289-299.
623 Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.,
624 De Siqueira, M.F., Grainger, A. & Hannah, L. (2004) Extinction risk from climate change. *Nature,*
625 **427,** 145-148.
626 Thompson, J.N. (2013) *Relentless Evolution*. University of Chicago Press, Chicago.
627 Unitt, P. (1987) *Empidonax traillii extimus*: an endangered subspecies. *Western Birds,* **18,** 137-162.
628 Urban, M.C. (2015) Climate change. Accelerating extinction risk from climate change. *Science,* **348,** 571-
629 573.
630 Wagner, H.H., Chávez-Pesqueira, M. & Forester, B.R. (2017) Spatial detection of outlier loci with Moran
631 eigenvector maps. *Molecular Ecology Resources,* **17,** 1122-1135.
632 Wang, J., Yu, J., Lipka, A.E. & Zhang, Z. (2022) Interpretation of Manhattan plots and other outputs of
633 genome-wide association studies. *Genome-Wide Association Studies* (eds D. Torkamaneh & F.
634 Belzile), pp. 63-80. Springer, New York.
635 Wikle, C.K. (2003) Hierarchical Bayesian models for predicting the spread of ecological processes.
636 *Ecology,* **84,** 1382—1394.
637 Wright, S. (1931) Evolution in Mendelian populations. *Genetics,* **16,** 97-159.
638
639