

Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data

C.J. Battey^{*,1}, Peter L. Ralph* and Andrew D. Kern*

*University of Oregon Dept. Biology, Institute for Ecology Evolution

ABSTRACT Real geography is continuous, but standard models in population genetics are based on discrete, well-mixed populations. As a result many methods of analyzing genetic data assume that samples are a random draw from a well-mixed population, but are applied to real populations that have significant geographic structure. Here we use simulations of populations living in continuous geography to study the impacts of dispersal and sampling strategy on population genetic summary statistics, demographic inference, and genome-wide association studies. We find that several common summary statistics have distributions that differ substantially from that seen in well-mixed populations, especially when Wright's neighborhood size is less than 100 and sampling is spatially clustered. Standard estimates of population size history based on the site frequency spectrum become biased with lower dispersal. We also show that the combination of spatially autocorrelated environments and limited dispersal can cause genome-wide association studies to identify numerous spurious signals of genetic association with purely environmentally determined phenotypes, and that this bias is only partially corrected by regressing out principal components of ancestry. Last, we discuss the relevance of our simulation results for inference from genetic variation in real organisms.

KEYWORDS Space; Population Structure; Demography; Haplotype block sharing; GWAS

(note: address confounding of σ and neighborhood size in disco)

(note: conclusions are that $Ne(t)$ is not very biased)

Introduction

The inescapable reality that biological organisms live, move, and reproduce in continuous geography is usually omitted from population genetic models. Indeed, a near universal

rule of reproduction is that individuals mate with other nearby individuals, leading to a positive correlation between genetic and geographic distances. This pattern of “isolation by distance” (Wright 1943) is one of the most widely replicated empirical findings in population genetics (Aguillon *et al.* 2017; Jay *et al.* 2012; Sharbel *et al.* 2000), but classical mathematical models describing it (Malécot 1948) are flawed approximations of the underlying process (Felsenstein 1975; Barton *et al.* 2002). Mathematical difficulties have made further development of population models in continuous space difficult, so most models describe geographic structure as a set of populations connected by migration (e.g., Wright 1931; Epperson 2003), and most empirical papers analyze variation within clusters of genetic variation inferred by programs like *STRUCTURE* (Pritchard *et al.* 2000), effectively assuming these are randomly mating units.

The assumption that populations are “well-mixed” has important implications for downstream inference of selection and demography. Methods based on the coalescent (Kingman 1982; Wakeley 2005) assume that the sampled individuals are a random draw from a well-mixed population that is much larger than the sample (Wakeley and Takahashi 2003). The key assumption is actually that the individuals of each generation are *exchangeable*, so that there is no correlation between the fate or fecundity of a parent and that of their offspring (Huillet and Möhle 2011). If dispersal or mate selection is limited by geographic proximity, this assumption can be violated in many ways. For instance, if mean viability or fecundity is spatially autocorrelated, then limited geographic dispersal will lead to parent–offspring correlations. Furthermore, nearby individuals will be more closely related than an average random pair, so drawing multiple samples from the same area of on the landscape will represent a biased sample of the genetic variation present in the whole population. It has long been appreciated that this model misspecification subjects downstream inferences to bias, but the extent and nature of these effects remain largely uninvestigated.

For instance, nonzero values of Tajima’s D are often interpreted as reflecting evidence of selection or past population size changes (Tajima 1989). This statistic is a summary of the site frequency spectrum, which itself reflects variation in branch lengths and tree structure of the underlying genealogies of sampled individuals. Geographically limited mate choice distorts the distribution of these genealogies (Maruyama 1972), which can affect the value of Tajima’s D . Similarly, the distribution of tract lengths of identity by state among individuals contains

information about not only historical demography (Harris and Nielsen 2013; Ralph and Coop 2013) and selection (Garud *et al.* 2015), but also dispersal and mate choice (Ringbauer *et al.* 2017; Baharian *et al.* 2016). We are particularly keen to examine how such summaries will be affected by models that incorporate continuous space, both to evaluate the assumptions underlying existing methods, but also to identify where the most promising signals of geography lie.

A related issue is the spatial distribution of sampling effort – since nearby individuals are more likely to be closely related than distant ones, observed patterns of relatedness are expected to depend on the geographic sampling scheme, perhaps strongly. In addition, range edges have been observed to create complex patterns of heterozygosity in stepping-stone simulations (Neel *et al.* 2013; Shirk and Cushman 2014), but the effects of this process on many downstream inference procedures is unknown. (*must be more out there on sampling – any recs for papers to read+cite?*)

The issue of fine-scale geographic structure may have particularly important implications for genome-wide association studies (GWAS). This is because many phenotypes of interest have strong geographic differences due to the (nongenetic) influence of environmental or socioeconomic factors, which can therefore show spurious correlations with spatially patterned allele frequencies (Bulik-Sullivan *et al.* 2015; Mathieson and McVean 2012). This may be particularly important for the study of selection on polygenic traits, whose heritable genetic components are determined by many loci of weak effect, those hardest to disentangle from spurious correlations. Indeed, two recent studies found that previous evidence of polygenic selection on human height in Europe was confounded by subtle population structure (Sohail *et al.* 2018; Berg *et al.* 2018), suggesting that existing methods to correct for population structure in GWAS are insufficient. However we have little quantitative idea of the population and environmental parameters that can be expected to lead to biases in GWAS. As the scale of sequence data now available for many species allows inference of increasingly fine-scale patterns of selection and demography, understanding how and when subtle spatial structure is likely to bias results is an important task for population genetics.

To study this, we have implemented an individual-based model in continuous geography that incorporates overlapping generations, local dispersal of offspring, and density-dependent survival. We simulate chromosome-scale genomic data in tens of thousands of individuals, and output the full genealogy and recombination history of all final-generation individuals.

93 We use these simulations to test how sampling strategy interacts with geographic population
94 structure to cause systematic variation in population genetic summary statistics typically
95 analyzed assuming discrete population models. We then examine how the fine-scale spatial
96 structures occurring under limited dispersal impact demographic inference from the site
97 frequency spectrum. Lastly, we examine the impacts of evolution in continuous space on
98 genome-wide association studies (GWAS) and identify regions of parameter space under
99 which the results from GWAS may be misleading.

100 (*I think we can cut the following paragraph if we just add the last sentence above*) (agree - commenting
101 out)

102 Materials and Methods

103 **Modeling Evolution in Continuous Space**

104 (*PETER TO REVISIT THIS SECTION*)

105 The best-studied approaches to population genetics in continuous space were developed
106 by Wright *et al.* (1942); Wright (1943) and Malécot (1948), who derived expressions for genetic
107 differentiation in continuous space assuming Poisson distributed numbers of offspring and
108 independent dispersal among individuals. A key finding of Wright's model is that many
109 important aspects of continuous populations can be described in terms of "neighborhood size"
110 – the number of potential mates for an individual in a given generation, defined as $4\pi\sigma^2d$,
111 where σ is the average dispersal distance and d is population density. Maruyama (1972)
112 found that the rate of decline in genetic diversity in a 2-dimensional continuous population
113 approaches the random mating expectation when $d\sigma^2 > 1$, and proposed that this had
114 the important implication that most population genetic expectations for randomly mating
115 populations could be applied to continuously distributed populations with relatively little
116 error.

117 Though some aspects of continuous populations are well described by the Wright and
118 Malécot models, Felsenstein (1975) showed that the assumptions of independent dispersal
119 and Poisson distributed offspring that are the basis of these models are incompatible. Over
120 time, a population meeting them will clump into a small number of geographic clusters
121 occupying only a part of the available range. Although real populations are often clumped
122 on landscapes due to factors like varying habitat quality and competition among species,

123 the Wright and Malécot models produce much more extreme clumping than is observed in
124 practice and fail to account for the density-dependent declines in population growth rate that
125 are widely observed in real populations (CITE).

126 One method for modeling continuous populations is then to assume the existence of a grid
127 of discrete randomly-mating populations connected by migration, which prevents clustering
128 by forcing all regions to be occupied in every generation. Among many other important results
129 drawn from this class of “lattice” or “stepping stone” models, Rousset (1997) showed that the
130 slope of the a linear regression of genetic differentiation (F_{ST}) against the logarithm of spatial
131 distance is an estimate of neighborhood size. Though good approximations of continuous
132 structure given high dispersal, these models are not truly continuous, force a uniform realized
133 population density across landscapes, and limit investigation of spatial structure below the
134 level of the deme. An alternative method is to model the geographic spread of ancestry
135 backwards in time through a diffusion approximation – an approach that has recently made
136 significant progress in modeling both dispersal and demographic parameters (Barton *et al.*
137 2010; Kelleher *et al.* 2014; Ringbauer *et al.* 2017, 2018).

138 We took a direct approach to the clustering problem of classical forward-time models by
139 incorporating density dependence into an individual-based model similar to the analytic
140 models developed by Wright and Malécot. By scaling the probability of survival in each
141 timestep to local population density we shift reproductive output towards regions of low-
142 density and prevent populations from clustering. A similar approach was taken previously by
143 (Doebeli and Dieckmann 2003) who used an individual based model with continuous space
144 and density dependent fitness to study the probability of speciation along continuous environ-
145 mental gradients. However to our knowledge previous implementations of continuous space
146 models have focused on a small number of genetic loci as the unit of analysis, which limits
147 the ability to investigate the impacts of continuous space on genome-wide genetic variation
148 as is now routinely sampled from real organisms. By simulating chromosome-scale sequence
149 alignments and complete population histories we are able to treat our simulations as real
150 populations and replicate the sampling designs and analyses commonly conducted on real
151 genomic data.

152 **A Forward-Time Model of Evolution in Continuous Space**

153 We simulated populations using the non-Wright-Fisher module in the program SLiM v3.0
154 (Haller and Messer 2019). Each time step consists of three stages: reproduction, dispersal, and
155 mortality. To reduce the parameter space we use the same parameter, denoted σ , to modulate
156 the spatial scale of interactions at all three stages by adjusting the standard deviation of the
157 corresponding Gaussian functions. As in previous work (Wright 1943; Ringbauer *et al.* 2017),
158 σ as applied in our dispersal step is equal to the mean parent-offspring distance. A key
159 parameter we report below is Wright citeyearpar{Wright1943} “neighborhood size”, defined
160 to be $N_W = 4\pi\sigma^2\rho$ where ρ is the population density. This the approximate number of
161 individuals available for mating in our simulation.

162 At the beginning of the simulation individuals are distributed uniformly at random on
163 a continuous, square landscape. Individuals are hermaphroditic, and each time step, each
164 produces a Poisson number of offspring with mean $1/L$ who disperse a random, Gaussian-
165 distributed distance away from the parent with mean zero and standard deviation σ in both
166 the x and y coordinates, reflected to stay within the species range. Each offspring is produced
167 with a mate selected randomly from those within distance 3σ , with probability of choosing a
168 neighbor at distance x proportional to $\exp(-x^2/2\sigma^2)$.

To maintain a stable population, mortality increases with local population density. To do this we say that individuals at distance d have a competitive interaction with strength $g(d)$, where g is the Gaussian density with mean zero and standard deviation σ . Then, the sum of all competitive interactions with individual i is $n_i = \sum_j g(d_{ij})$, where d_{ij} is the distance between individuals i and j and the sum is over all neighbors within distance 3σ . Since g is a probability density, n_i is an estimate of the number of nearby individuals per unit area. Then, given a per-unit carrying capacity K , the probability of survival until the next time step for individual i is

$$p_i \min(0.95, \frac{1}{1 + n_i/(K(1 + L))}). \quad (1)$$

169 We chose this functional form so that the equilibrium population density per unit area is
170 around K , and the mean lifetime is around L .

An important step in creating any spatial model is dealing with range edges. Because local population density is used to model competition, edge or corner populations can be assigned artificially high fitness values because they lack neighbors within their interaction radius but

outside the bounds of the simulation. We approximate a decline in habitat suitability near edges by decreasing the probability of survival proportional to the square root of distance to edges in units of σ . The final probability of survival for individual i is then

$$s_i = p_i \min(1, \sqrt{x_i/\sigma}) \min(1, \sqrt{y_i/\sigma}) \min(1, \sqrt{(W - x_i)/\sigma}) \min(1, \sqrt{(W - y_i)/\sigma}) \quad (2)$$

where x_i and y_i are the spatial coordinates of individual i , and W is the width (and height) of the square habitat. This buffer roughly counteracts the increase in fitness individuals close to the edge would otherwise have.

To isolate spatial effects from other components of the model such as overlapping generations, increased variance in reproductive success, and density-dependent fitness, we also implemented simulations identical to those above except that mates are selected uniformly random from the population, and offspring disperse to a uniform random location on the landscape. We refer to this model as the “random mating” model, in contrast to the first, “spatial” model.

We stored the full genealogy and recombination history of final-generation individuals as tree sequences (Kelleher *et al.* 2018), as implemented in SLiM (Haller *et al.* 2019). Scripts for figures and analyses are available at <https://github.com/petrelharp/spaceness>.

We ran 400 simulations for the spatial and random-mating models on a square landscape of width $W = 50$ with per-unit carrying capacity $K = 5$ (census $N \approx 10,000$), average lifetime $L = 4$, genome size 10^8 , recombination rate 10^{-9} , and drawing σ values from a uniform distribution between 0.2 and 4. To speed up the simulations and limit memory overhead we used a mutation rate of 0 in SLiM and later applied mutations to the tree sequence with msprime’s `mutate` function (Kelleher *et al.* 2016). Because msprime applies mutations proportionally to elapsed time, we divided the mutation rate of 10^{-8} mutations per site per generation by the average generation time estimated for each value of σ (see ‘Demographic Parameters’ below) to convert the rate to units of mutations per site per unit time. (We verified that this procedure produced the correct number of mutations by comparing to a subset of simulations with SLiM-generated mutations, which are applied only at meiosis.) Simulations were run for 1.6 million timesteps (approximately $30N$ generations), or until all extant individuals shared a common ancestor within the simulation (i.e., the tree sequence had coalesced). (*maybe worth including a table with some basic runtime results in the supplement?*)

197 ***Demographic Parameters***

198 Our demographic model includes paramters for population density (K), mean life span (L),
199 and dispersal distance (σ). However, nonlinearity of local demographic stochasticity causes
200 actual realized averages of these demographic quantitites to deviate from the specified values
201 in a way that depends on the neighborhood size. Therefore, to properly compare to theoretical
202 expectations, we empirically calculated these demographic quantities in simulations. We
203 recorded the census population size in all simulations. To estimate generation times, we stored
204 ages of the parents of every new individual born across 200 timesteps, after a 100 generation
205 burn-in, and took the mean. To estimate variance in offspring number, we tracked the number
206 of offspring for all individuals for 100 timesteps following a 100-timestep burn-in period,
207 subset the resulting table to include only the last timestep recorded for each individual, and
208 calculated the variance in number of offspring across all individuals in timesteps 50-100. All
209 calculations were performed with information recorded in the tree sequence, using pyslim
210 (<https://github.com/tskit-dev/pyslim>).

211 ***Sampling***

212 Our model records the genealogy and sequence variation of the complete population, but in
213 real data, genotypes are only observed from a relatively small number of sampled individuals.
214 We modeled three sampling strategies similar to common data collection methods in empirical
215 genetic studies (Figure 1). “Random” sampling selects individuals at random from across
216 the full landscape, “point” sampling selects individuals proportional to their distance from
217 four equally spaced points on the landscape, and “midpoint” sampling selects individuals in
218 proportion to their distance from the middle of the landscape. Downstream analyses were
219 repeated across all sampling strategies.

220 ***Summary Statistics***

221 We calculated the site frequency spectrum and a set of 18 summary statistics (Table S1) from
222 60 diploid individuals sampled from the final generation of each simulation using the python
223 package scikit-allel (Miles and Harding 2017). Statistics included common single-population
224 summaries including mean pairwise divergence (π), inbreeding coefficient (F_{IS}), and Tajima’s
225 D , as well as the classic isolation-by-distance regression of genetic distance (D_{xy}) against the
226 logarithm of geographic distances (Rousset 1997), which we summarized as the correlation

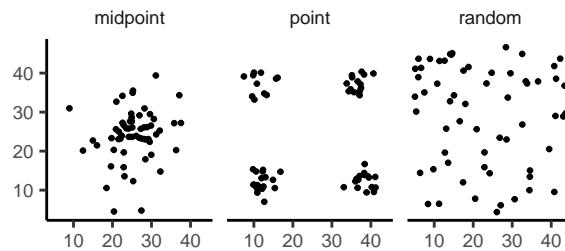


Figure 1 Example sampling maps for 60 individuals on a 50×50 landscape for midpoint, point, and random sampling strategies, respectively.

227 coefficient between the logarithm of the spatial distance and the proportion of identical base
228 pairs across pairs of individuals.

229 Following recent studies that showed strong signals for dispersal and demography in the
230 distribution of shared haplotype block lengths (Ringbauer *et al.* 2017; Baharian *et al.* 2016),
231 we also calculated various summaries of the distribution of pairwise identical-by-state (IBS)
232 block lengths among samples. The full distribution of lengths of IBS tracts for each pair of
233 chromosomes was first calculated with a custom python function. We then calculated the first
234 three moments of this distribution (mean, variance, and skew) and the number of blocks over
235 $1e6$ base pairs both for each pair of individuals and for the full distribution across all pairwise
236 comparisons.

237 We then estimated correlation coefficients between spatial distance and each moment of
238 the pairwise IBS tract distribution. Because more closely related individuals on average share
239 longer haplotype blocks we expect that spatial distance will be negatively correlated with
240 mean haplotype block length, and that this correlation will be strongest (i.e. most negative)
241 when dispersal is low. The variance, skew, and count of long haplotype block statistics are
242 meant to reflect the relative length of the right (upper) tail of the distribution, which represents
243 the frequency of long haplotype blocks so should reflect recent demographic events (Chapman
244 and Thompson 2002).

245 The effects of sampling on summary statistic estimates were summarized by testing for
246 differences in mean (ANOVA, (R Core Team 2018)) and variance (Levene's test, (Fox and
247 Weisberg 2011)) across sampling strategies for each summary statistic.

248 ***Demographic Modeling***

249 We fit single-population demographic models to the site frequency spectra of 20 individuals
250 from each spatial SLiM simulation with the program Stairwayplot (Liu and Fu 2015). This
251 analysis was replicated across random, point, and midpoint sampling strategies. Site fre-
252 quency spectra used for input data were calculated in scikit-allel (Miles and Harding 2017),
253 and 100 bootstrap replicates were generated for each simulation by resampling over sites.
254 (*what were bootstrap replicates used for? if anything, need to say more precisely what you mean here*)
255 Models were fit across all bootstrap replicates using default settings in Stairwayplot and the
256 median estimate of N_e per generation was used to represent the output of each simulation.

257 In preliminary runs we found that inferred population histories were highly variable
258 even when simulating under a coalescent model, suggesting that some of the differences in
259 demographic estimates for spatial models are caused by the behavior of the optimization
260 algorithm rather than bias in the SFS caused by spatial mate choice and dispersal. To separate
261 these effects we ran 100 coalescent simulations with constant population size 6.1×10^{-3} (the
262 mean N_e of random-mating SLiM models estimated from Θ_π) and fit stairwayplot models
263 using the same script as for our spatial models. All coalescent simulations were performed
264 using msprime (Kelleher *et al.* 2016). We then calculated the standard deviation of inferred
265 N_e in each stairwayplot model to summarize the degree of fluctuation around the simulated
266 population size, and asked if standard deviations were higher in spatial relative to coalescent
267 models with a one-tailed t-test.

268 ***Association Studies***

269 To assess the degree to which spatial structure confounds GWAS we simulated four types of
270 nongenetic phenotype variation for 1000 randomly sampled individuals in each spatial SLiM
271 simulation and conducted a linear regression GWAS with principal components as covariates
272 in PLINK (Purcell *et al.* 2007). SNPs with a minor allele frequency less than 0.5% were excluded
273 from this analysis. Phenotype values were set to vary by two standard deviations across the
274 landscape in a rough approximation of the variation seen in height across Europe, which has
275 recently been found to be confounded with population structure in large scale GWAS (Berg
276 *et al.* 2018; Sohail *et al.* 2018). Conceptually our approach is similar to that taken in (Mathieson
277 and McVean 2012), though here we model fully continuous spatial variation and compare

278 GWAS output across a range of dispersal distances.

279 In all simulations, the phenotype of each individual is determined by adding independent
280 Gaussian noise with mean zero and standard deviation 10 to a mean that may depend on
281 spatial position. We adjust the geographic pattern of mean phenotype to create spatially
282 autocorrelated environmental influences on phenotype. In the first simulation of *nonspatial*
283 phenotypes, the mean did not change, so that all individuals' phenotypes were drawn inde-
284 pendently from a Gaussian distribution with mean 110 and standard deviation 10. Next, to
285 simulate *clinal* environmental influences on phenotype, we increased the mean phenotype
286 from 90 on the left edge of the range to 120 on the right edge (two phenotypic standard
287 deviations). Concretely, an individual at position (x, y) in a 50×50 landscape has mean
288 phenotype $110 + 2x/5$. Third, we simulated a more concentrated "corner" environmental
289 effect by setting the mean phenotype for individuals with both x and y coordinates below 20
290 to 130 (two standard deviations above the rest of the map). Finally, in "patchy" simulations we
291 selected 10 random points on the map and set the mean phenotype of all individuals within
292 three map units of each of these points to 130.

293 We performed principal components analysis (PCA) using scikit-allel (Miles and Harding
294 2017) on the matrix of derived allele counts by individual for each simulation. SNPs were
295 first filtered to remove strongly linked sites by calculating LD between all pairs of SNPs in
296 a 200-SNP moving window and dropping one of each pair of sites with an R^2 over 0.1. The
297 LD-pruned allele count matrix was then centered and all sites scaled to unit variance when
298 conducting the PCA, following recommendations in (Patterson *et al.* 2006).

299 We ran linear-model GWAS both with and without the first 10 principal components as
300 covariates in PLINK and summarized results across simulations by counting the number of
301 SNPs with p -value below 0.05 after adjusting for an expected false positive rate of less than 5%
302 (Benjamini and Yekutieli 2001). We also examined p values for systemic inflation by estimating
303 the expected values from a uniform distribution (because no SNPs were used when generating
304 phenotypes), plotting observed against expected values for all simulations, and summarizing
305 across simulations by finding the mean σ value in each region of quantile-quantile space.
306 Results from all analyses were summarized and plotted with the 'ggplot2' (Wickham 2016)
307 and "cowplot" (Wilke 2019) packages in R (R Core Team 2018).

308 **Results**

309 **Genealogical Parameters**

310 Having implemented our continuous space forward time population genetic simulation, the
311 first thing we were interested in examining was the effect of spatial parameters on population
312 parameters such as the generation time, the census population size, and the variance in
313 offspring number. In contrast to intuition based on standard population genetic models, in the
314 context of our non-Wright-Fisher model basic quantities like the generation time depend on life
315 history parameters such as the dispersal distance and carrying capacity of the landscape. We
316 first examined how the population size, generation time, and variance in number of offspring
317 vary with neighborhood size, which itself is a function of σ (Figure 2). There are a few things
318 to note. First all three quantities are non-linear with respect to neighborhood size. Census
319 size largely declines as neighborhood size increases for both the spatial and random mating
320 models. However for spatial models this decline only begins for neighborhood size ≥ 10 .
321 By a neighborhood size ≥ 100 the spatial and random mating models are indistinguishable
322 from one another, a sign that our simulations are performing as expected. Census sizes range
323 from $\approx 14,000$ at low σ in the random mating model to $\approx 10,000$ for both models when
324 neighborhood sizes approach 1,000.

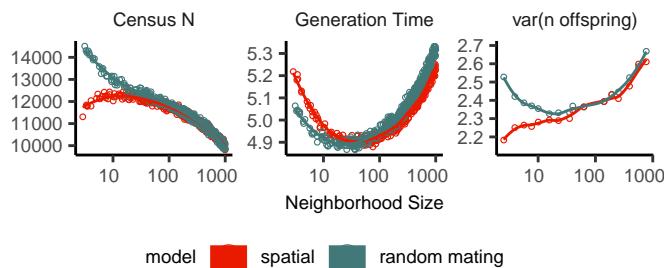


Figure 2 Genealogical parameters from spatial and random mating SLiM simulations, by neighborhood size.

325 Generation times similarly show complex behavior with respect to neighborhood sizes, and
326 vary across the parameter range explored between 5.2 and 4.9 timesteps per generation. To
327 stress this point further, the generation time here varies because under our non-Wright-Fisher
328 dynamics individuals are free to reproduce more than once over their lifespan and the lifespan
329 of an individual is only indirectly determined by the parameterization of density-dependent

330 competition the model. Under both the spatial and random mating regimes generation time
331 reaches a nadir at a neighborhood size of ≈ 50 . Interestingly under the range of neighborhood
332 sizes that we examined, generation times between the random mating and spatial models are
333 never quite equivalent – presumably this would cease to be the case at neighborhood sizes
334 higher than we simulated here.

335 Last we looked at the effect of neighborhood size on the variance in number of offspring in
336 the population – a key parameter determining the effective population size. Surprisingly the
337 spatial and random mating model behave quite differently with respect to the neighborhood
338 size: while the variance in offspring number increases monotonically, or nearly so, under the
339 spatial model, the random mating model actually shows a decline in the variance in offspring
340 number until a neighborhood size ≈ 10 before it increases and eventually equals what we
341 observe in the spatial case.

342 ***Impacts of Continuous Space on Population Genetic Summary Statistics***

343 One of the main goals of this study is to examine the effect of continuous space on canonical
344 summaries of population genetic variation. Indeed we have little knowledge to date about
345 how some of our most beloved population genetic summary statistics might be affected by
346 such realistic models. Moreover, as we will show, sampling strategies of individuals with
347 respect to space can affect summaries of variation at least as strongly as the underlying
348 population dynamics.

349 ***Site Frequency Spectra and Summaries of Diversity*** In Figure 3 we examine the effect of
350 varying neighborhood size and sampling strategy on the site frequency spectrum (Figure 3A)
351 and several canonical population genetic summary statistics (Figure 3B). For populations in
352 continuous space we observe a significant skew in the SFS for smaller neighborhood sizes
353 (≤ 100) that is exacerbated by mid point and point sampling of individuals (see Figure
354 1). As expected these effects are strongest for the smallest neighborhood sizes examined
355 and are consistently in the direction of an enrichment of intermediate frequency variants in
356 comparison to the standard neutral expectation. This can be seen succinctly in the response
357 of Tajima's D to variation in neighborhood sizes and sampling regime shown in Figure 3B,
358 whereby we see D turning quite positive at small neighborhood sizes. This effect is particularly
359 strong for the midpoint and point sampling regimes, but also occurs under random sampling.

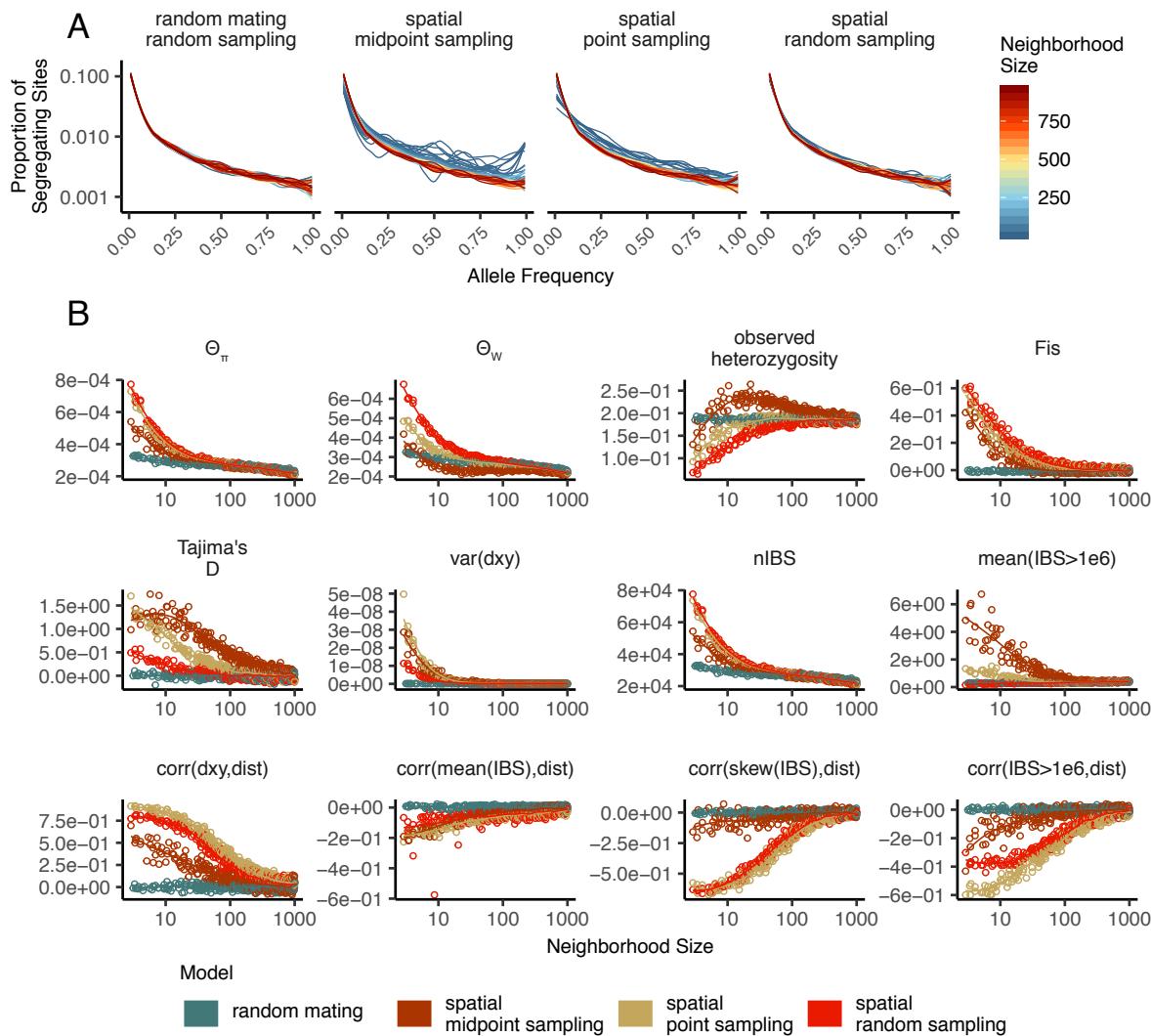


Figure 3 Site frequency spectrum (A) and summary statistic distributions (B) by sampling strategy and neighborhood size.

360 Notably, the point at which Tajima's D approaches 0 differs strongly across sampling strategies
361 – varying from a neighborhood size of ≈ 50 for random sampling to ≥ 1000 for midpoint
362 sampling.

363 One of the most commonly used summaries of variation is Tajima's estimator of the neutral
364 mutation rate $\hat{\theta}_\pi$. As can be seen in Figure 3B $\hat{\theta}_\pi$ can vary as much as nearly three-fold for the
365 smallest neighborhood sizes between the random mating and spatial models depending on
366 the sampling strategies. Each sampling strategy approaches the random mating expectation
367 at its own rate, but by neighborhood size of ≈ 100 all models are equivalent. These results are
368 to be expected as at small neighborhood sizes the time to the most recent common ancestor for
369 the population should be pushed back– i.e. it takes longer for lineages to coalesce if dispersal
370 is limited. Interestingly the rate of approach to random mating expectations across sampling
371 strategies is reversed relative to that observed in Tajima's D – midpoint sampling reaches
372 random mating expectations at neighborhood sizes ≈ 50 while random sampling is inflated
373 until neighborhood size /approx 100. In effect this means that any estimates of population
374 size based on expected heterozygosity from populations with limited dispersal should be
375 treated with caution.

376 Patterns from observed heterozygosity and its derivative F_{IS} also depend heavily on
377 neighborhood size under spatial models as well as the sampling scheme. F_{IS} is inflated above
378 the expectation across most of the parameter space examined and across all sampling strategies.
379 This effect is caused by a deficit of heterozygous individuals in low-dispersal simulations
380 – a continuous-space version of the Wahlund effect (Wahlund 1928). Indeed for random
381 sampling under the spatial model F_{IS} does not approach the random mating equivalent until
382 neighborhood sizes of nearly 1000. The patterns for raw observed heterozygosity are more
383 complex. Under midpoint sampling observed heterozygosity is inflated even over the random
384 mating expectation, as a result of the a higher proportion of heterozygotes occurring in the
385 middle of the landscape (Figure SS3). This squares with a report from Shirk and Cushman
386 (2014) who observed a similar excess of heterozygosity in the middle of the landscape when
387 simulating under a lattice model.

388 Trends in pairwise haplotype block sharing parallel those in allele-frequency-based diversity
389 estimates (Figure 3, Supplementary Figure S1). At low dispersal the distribution of IBS block
390 lengths in a set of samples is shifted towards smaller values with respect to the random mating

391 expectation– resulting in lower means and fewer long IBS blocks. The variance and skew
392 of the distribution of haplotype block lengths are only minorly affected by neighborhood
393 size in our simulations when calculated across all pairs of individuals; however, they are
394 strongly dependent on sampling regimes. For example, the number of long haplotype blocks
395 declines as neighborhood size increases under midpoint sampling but changes very little
396 across neighborhood sizes under point or random sampling. Thus sampling strategies with
397 respect to geography will affect conclusions drawn from haplotype length distributions quite
398 dramatically.

399 **Correlations of summary statistics with geographic distance** Correlating population genetic
400 summaries such as F_{ST} against geographic distance has shown great utility in empirical
401 population genetics (Rousset 1997). As we know the exact locations of individuals that are
402 sampled in our simulated populations we have examined the relationship of geographic
403 distance among samples with a number of our summary statistics (Figure 3 and Figure S1).
404 Starting with a measure of population differentiation, D_{xy} , we observe a positive correlation
405 with distance that declines as dispersal increases, as expected under the theory developed by
406 (Rousset 1997) and others. This relationship varies across sampling strategies of course, with
407 the weakest correlations observed for midpoint sampling. Moreover there is clearly strong
408 signal for neighborhood size in the strength of the correlation between D_{xy} and geographic
409 distance.

410 We next turn our attention to the effect of geographic distance on haplotype block length
411 sharing. As in (Ringbauer *et al.* 2017) and (Baharian *et al.* 2016) we found that the pairwise
412 distribution of haplotype block lengths is more strongly left-skewed under limited dispersal.
413 This is reflected in negative correlation coefficients between spatial distance and the mean,
414 variance, skew, and count of long blocks from the pairwise distribution of identical-by-
415 state block lengths (Figure 3 and Figure S1). Of these summaries the mean of the IBS tract
416 length distribution is only weakly affected by neighborhood size, likely because it is heavily
417 influenced by the small number of very long IBS tracts. In contrast the count of long IBS blocks
418 and the skew of the pairwise IBS block distribution are strongly dependent on distance among
419 individuals, and the magnitude of this correlation declines predictably with neighborhood
420 size. In all spatial correlations random and point sampling are similarly correlated with space
421 across neighborhood sizes, but midpoint sampling causes weaker correlations because it

422 incorporates less genetic and geographic distance than the full sample.

423 **Effects of Sampling on Random Mating and Spatial Models** To summarize the effect of sam-
424 pling strategy on genetic diversity in random mating and spatial models we also asked if
425 summary statistics varied significantly across sampling regimes. In table S2 we show that
426 for spatial models the mean and variance of nearly all summary statistics was significantly
427 different across sampling strategies (Table S2) for spatial models, but not for random mating
428 models. In summary, sampling strategy can be safely ignored when the population is ran-
429 domly mating but will shape estimates of genetic diversity in any population with limited
430 dispersal.

431 **Effects of Space on Demographic Inference**

432 One of the most important uses for population genetic data is inferring demographic history
433 of populations. As demonstrated above, the site frequency spectrum varies across neigh-
434 borhood sizes and sampling strategies. Does this variation lead to different inferences of
435 past population sizes? To ask this we inferred population size histories from samples drawn
436 from our simulated populations using a popular software package that uses the SFS as its
437 information, Stairwayplot (Liu and Fu 2015).

438 In Figure 4 we show inferred population size histories from our simulations as a function of
439 neighborhood size binned in to one of four categories. In general we observe that demographic
440 models from Stairwayplot tended to infer patterns of ancient population increases and recent
441 declines when neighborhood sizes were below 20 under all sampling strategies (Figure 4). This
442 is consistent with our observations of the SFS from which Stairwayplot is doing its inference.
443 Inflated past population sizes were seen in both point and random sampling, demonstrating
444 that the relatively minor shift in the site frequency spectrum observed among sampling
445 regimes is enough to alter demographic estimates. More alarmingly, inference of severe
446 population bottlenecks was common at neighborhood sizes under 100 for midpoint and point
447 sampling strategies. Above neighborhood sizes of 100 the average inferred demography across
448 all simulations was relatively accurate, with minor fluctuations slightly above the expected
449 variance Ne . While that is so individual model fits were highly variable and often inferred
450 five-fold or greater population fluctuations even in high-dispersal simulations. To test whether
451 the variation we observed in inferred demographic histories from Stairwayplot was the result

452 of spatial effects in our simulations rather than the behavior of the optimization routine we
453 compared standard deviations of inferred population sizes in each sampling/neighborhood
454 size bin with those returned by an equilibrium coalescent simulation. Were these to be equal
455 we could assume that the variation were purely as to be expected under the noise inherent
456 in the genealogical process. Instead we found that standard deviations were significantly
457 greater in all sampling strategies for neighborhood sizes under 20 and for midpoint sampling
458 with neighborhood sizes 20-100, but all other sets performed similarly to the coalescent model
459 (Table S3). In summary, spatial mate choice and dispersal causes strong bias in SFS-based
460 demographic estimates for neighborhood sizes below 20 or when sampling is clustered, but
461 otherwise any biases are within the range of variability regularly inferred by Stairwayplot.
462 This underscores the fact that some *a priori* knowledge about the population dynamics at play
463 will be important to interpreting results of demographic estimation routines.

464 **GWAS**

465 To ask what confounding effects spatial genetic variation might have on genome-wide associa-
466 tion studies we performed GWAS on our simulations using phenotypes that were determined
467 solely by the environment. In general we found that for simulations with limited dispersal,
468 i.e. neighborhood size < 100, spatial variation in the environment causes GWAS to infer
469 significant associations with purely environmental phenotypes at over 25% of sampled SNPs
470 if no correction for genetic relatedness among samples is performed (Figure 5). This effect is
471 particularly strong for clinal and corner environments, for which the lowest dispersal levels
472 cause over 60% of SNPs in the sample to return significant associations. Patchy environmental
473 distributions, which result in fewer extreme phenotype values in the sample (Figure 5A),
474 cause fewer false-positives overall but still produce spurious associations at roughly 10% of
475 sites at the lowest neighborhood sizes. Notably no simulations with nonspatial environments
476 returned more than one significant association (Figure 5C), demonstrating that this effect
477 is caused specifically by the interaction of population structure and spatial variation in the
478 environment rather than by population structure itself.

479 When PCA positions are included as covariates to control for population structure in
480 GWAS the vast majority of SNPs no longer surpass a 5% FDR significance threshold, but up to
481 1.5% of SNPs are significantly associated at low dispersal distances under corner and patchy

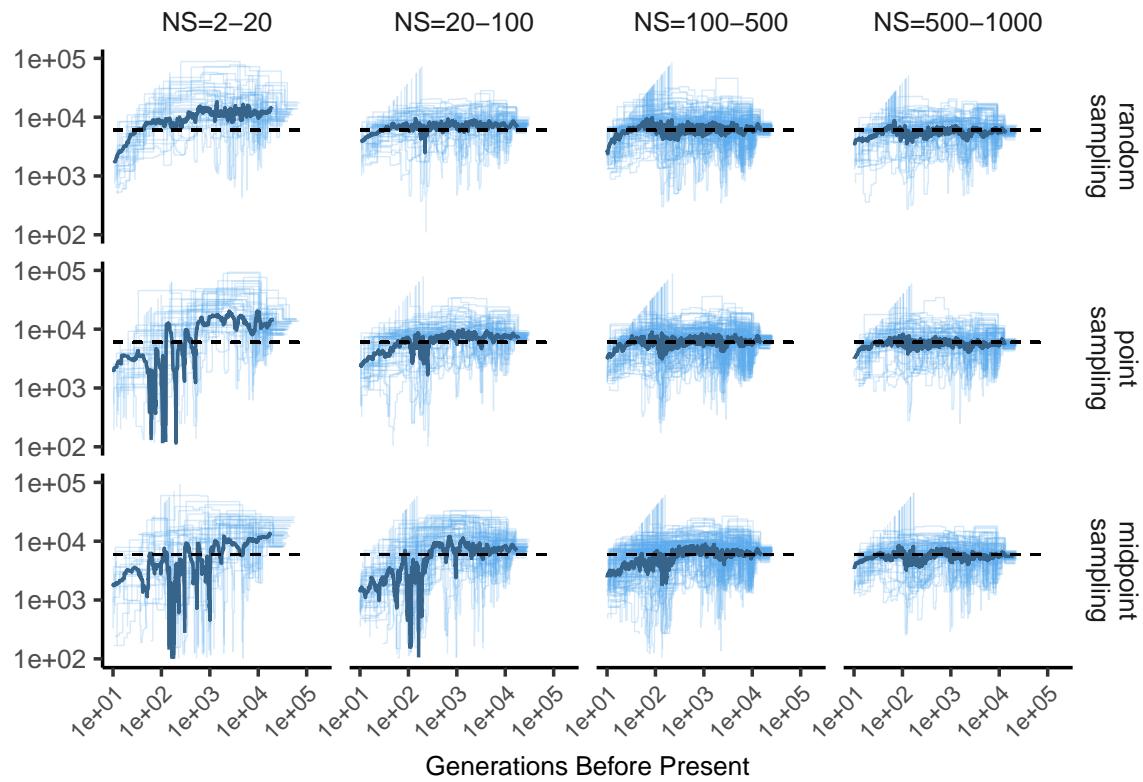


Figure 4 Inferred demographic histories for spatial SLiM simulations from Stairwayplot, by sampling scheme and neighborhood size (NS) range. The thick line is a rolling mean and thin lines are individual model fits. Dashed horizontal lines are the average N_e across random-mating SLiM models estimated from θ_π .

482 environmental distributions (Figure 5C). At neighborhood sizes > 500 up to 0.31% of SNPs
483 were significant for corner and clinal environments. Given an average of 132,000 SNPs across
484 simulations after MAF filtering, this translates to up to 382 false-positive associations. In
485 most cases the p values for these associations were significant after FDR correction but would
486 not pass the threshold for significance under the more conservative Bonferroni correction,
487 suggesting that the most strongly associated variants from many GWAS of mono- or oligogenic
488 traits are robust to this variety of stratification bias.

489 Clinal environments cause an interesting pattern in false positives after PC correction:
490 at low neighborhood sizes the correction removes nearly all significant associations, but at
491 neighborhood sizes above ≈ 250 the proportion of significant SNPs increases to up to 0.4%
492 (Figure 5. This appears to reflect a loss of descriptive power in the PCA – as neighborhood
493 size increases, the total proportion of variance explained by the first 10 PC axes declines from
494 roughly 0.1 to 0.04 (Figure 5B). Essentially, PCA seems unable to effectively summarize the
495 weak population structure present in large-neighborhood simulations, but these populations
496 continue to have enough spatial structure to create significant correlations between genotypes
497 and the environment. A similar process can also be seen in the corner phenotype distribution,
498 in which the count of significant SNPs initially declines as neighborhood size increases and
499 then increases at approximately the point at which the proportion of variance explained by
500 PCA approaches its minimum.

501 In Figure 5D we present quantile-quantile plots that show the degree of genome-wide
502 inflation of test statistics in PC-corrected GWAS across all simulations and environmental
503 distributions. For clinal environments $-\log_{10}(p)$ values are most inflated when neighborhood
504 sizes are large, consistent with the pattern observed in the count of significant associations
505 after PC regression. In contrast corner and patchy environments cause the greatest inflation in
506 $-\log_{10}(p)$ at neighborhood sizes < 100, which likely reflects the inability of PCA to account for
507 fine-scale structure caused by very limited dispersal. Finally, we observed that PC regression
508 appears to cause some degree of overcorrection for all phenotype distributions, visible in
509 Figure 5D as points falling below the 1:1 line.

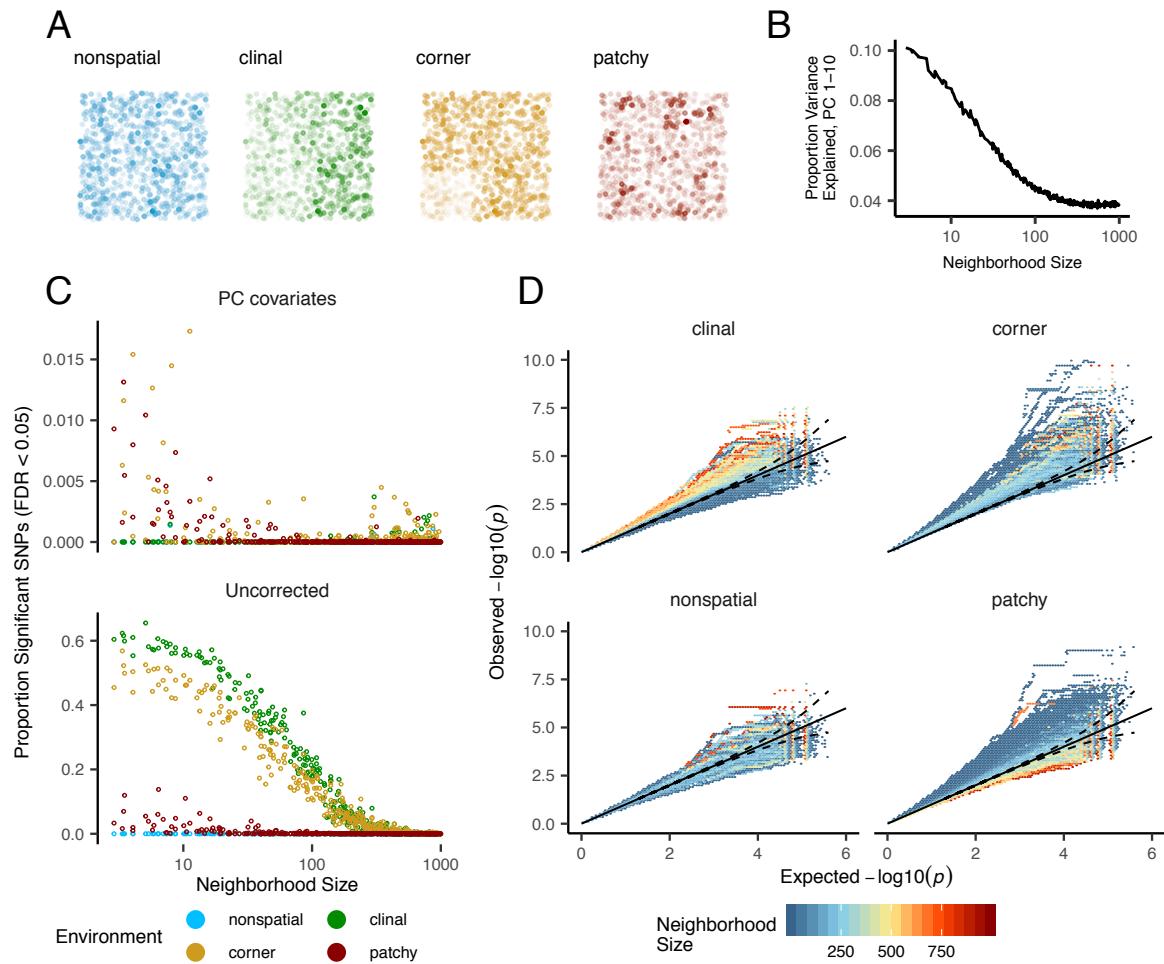


Figure 5 Impacts of spatially varying environments and isolation by distance on linear regression GWAS. Simulated quantitative phenotypes are determined only by an individual's location and the spatial distribution of environmental factors. In A we show the phenotypes and locations of sampled individuals under four environmental distributions, with transparency scaled to phenotype. As neighborhood size increases a PCA explains less of the total variation in the data (B). Spatially correlated environmental factors cause false positives at a large proportion of SNPs, which is partially but not entirely corrected by adding PC positions as covariates (C). Quantile-quantile plots in D show inflation of $-\log_{10}(p)$ after PC correction across all simulations and environmental distributions, with colors scaled by the median neighborhood size in each region of q-q space.

510 **Discussion**

511 Patterns of genetic variation are influenced by the fact that organisms are more likely, on
512 average, to reproduce with others of their species that are geographically proximate, in
513 particular when dispersal distances are short. While historically some analytical progress
514 has been made in describing the influence of continuous space on population genetics (e.g.
515 Wright (1943); Rousset (1997); Ringbauer *et al.* (2017); Barton *et al.* (2010)), the theoretical work
516 is challenging and often untenable for realistic biological models. Instead we here use efficient
517 forward time population genetic simulations to describe the myriad influence of space on
518 genetic variation. In particular we examine three axes of variation across three different
519 sampling strategies – 1) population genetic summary statistics, 2) inference of population
520 size history, and 3) the consequences on genome-wide association studies (GWAS). We are
521 in particular interested in asking how our empirical inferences from data might be affected
522 by spatial processes. As we show below the answers seems to be - often space matters, both
523 because of how populations are sampled with respect to space, and because of the inherent
524 dispersal properties of those populations.

525 **Effects of Dispersal**

526 Limited dispersal inflates effective population size, creates correlations between genetic and
527 spatial distances, and introduces subtle biases in the site frequency spectrum that are reflected
528 in a positive Tajima's D (Figure 3). At the extreme low end of dispersal distance this can result
529 in an up to three-fold increase in genetic diversity relative to random-mating expectations.
530 These effects are strongest when neighborhood sizes are below 100, but in combination with
531 the effects of nonrandom sampling they can persist up to neighborhood sizes of at least
532 1000 (e.g. inflation in Tajima's D and observed heterozygosity under midpoint sampling).
533 Under random sampling the general pattern is similar to expectations of the original analytic
534 model of Wright (1943), which predicts that populations with neighborhood sizes under
535 100 will differ substantially from random mating, while those above 10,000 will be nearly
536 indistinguishable from panmixia.

537 The patterns observed in sequence data reflect the effects of space on the underlying
538 genealogy. Nearby individuals coalesce rapidly under limited dispersal and so are connected
539 by short branch lengths, while distant individuals take much longer to coalesce than they

would under random mating. Mutation and recombination events in our simulation both occur at a constant rate along branches of the genealogy, so the genetic distance and number of recombination events separating two individuals is simply a noisy estimate of the branch lengths connecting them. These genealogical patterns also relate directly to the site frequency spectrum. In our simulations we observed that groups of nearby individuals tend to coalesce rapidly, while coalescence among groups at opposite ends of the landscape takes much longer than under random mating. Tip branches (i.e. branches subtending only one individual) are then relatively short, and branches in the middle of the genealogy connecting local groups of individuals relatively long. These patterns then create the biases we observed in the SFS – the lowest frequency bins are deflated by the short branch lengths connecting nearby individuals, while mid-frequency bins are inflated by the long branches connecting local groups.

The genealogical patterns introduced by limited dispersal are particularly apparent in the distribution of haplotype block lengths (Figure 3). This is because identical-by-state tract lengths reflect the impacts of two processes acting along the branches of the underlying genealogy – both mutation and recombination – rather than just mutation as is the case when looking at the site frequency spectrum or related summaries. This means that the pairwise distribution of haplotype block lengths carries with it important information about genealogical variation in the population, and correlation coefficients between moments of the this distribution and geographic location contain signal similar to the correlations between F_{ST} or D_{xy} and space (Rousset 1997). Indeed this basic logic underlies two recent studies explicitly estimating dispersal from the distribution of shared haplotype block lengths (Ringbauer *et al.* 2017; Baharian *et al.* 2016). Conversely, because haplotype-based measures of demography are particularly sensitive to variation in the underlying genealogy, inference approaches that assume random mating when analyzing the distribution of shared haplotype block lengths are likely to be strongly affected by spatial processes.

Effects of Sampling

One of the most important differences between random mating and spatial models is the effect of sampling: in a randomly mating population the spatial distribution of sampling effort has no effect on estimates of genetic variation, but when dispersal is limited sampling strategy can compound spatial patterns in the underlying genealogy and create pervasive

570 impacts on all downstream genetic analyses. As expected we found that random sampling
571 provides the most accurate summary of genetic diversity across the landscape. However
572 this strategy is often impractical for empirical studies. In reality the difficulty of traveling
573 through all parts of a species range and the inefficiency of collecting single individuals at
574 each sampling site means that most studies follow something closer to the “point” sampling
575 strategy we simulated, in which multiple individuals are sampled from nearby points on the
576 landscape. For example, in ornithology a sample of 10 individuals per species per locality is a
577 common target when collecting for natural history museums. In classical studies of *Drosophila*
578 variation the situation is considerably worse, in which a single orchard might be sampled
579 with baited traps for instance.

580 When sampling is clustered at points on a landscape and dispersal is limited, the sampled
581 individuals will be more closely related than a random set of individuals. Average coalescence
582 times of individuals collected at a locality will then be more recent and branch lengths shorter
583 than expected by analyses assuming random mating. This leads to fewer mutations and
584 recombination events occurring since their last common ancestor, causing a random set of
585 individuals to share longer average IBS tracts and have fewer nucleotide differences. For some
586 data summaries, such as Tajima’s D , Watterson’s Θ , or the correlation coefficient between
587 spatial distance and the count of long haplotype blocks, this can result in large differences in
588 estimates between random and point sampling (Figure 3). Inferring underlying demographic
589 parameters from these summary statistics – for example, estimating dispersal distance as the
590 slope of a regression of F_{ST} against the logarithm of spatial distance (Rousset 1997) – may then
591 be subject to bias if sampling is not random across the landscape.

592 However, the largest sampling effects we observed occurred in our “midpoint” sampling
593 strategy. This model is meant to reflect a bias in sampling effort towards the middle of a species’
594 range. In empirical studies this sampling strategy could arise if, for example, researchers
595 choose to sample the center of the range and avoid range edges to maximize probability of
596 locating individuals during a short field season. Because midpoint sampling provides limited
597 spatial resolution it dramatically reduces the magnitude of observed correlations between
598 spatial and genetic distances. More surprisingly, midpoint sampling also leads to strongly
599 positive Tajima’s D and an inflation in the proportion of heterozygous individuals in the
600 sample. This increase in observed heterozygosity appears to reflect the effects of range edges,

which are a fundamental facet of spatial genetic variation that have often been ignored by analytic approaches focusing on infinite toroidal landscapes (Felsenstein 1975). If individuals move randomly in a finite two-dimensional landscape then regions in the middle of the landscape receive migrants from all directions while those on the edge receive no migrants from at least one direction. The average number of new mutations moving into the middle of the landscape is then higher than the number moving into regions near the range edge, leading to higher heterozygosity and lower inbreeding coefficients (F_{IS}) away from range edges. Though here we used only a single parameterization of fitness decline at range edges we believe this is a general property of non-infinite landscapes as it has also been observed in previous studies simulating under lattice models (Neel *et al.* 2013; Shirk and Cushman 2014). For empirical studies this suggests that sampling only the middle of the landscape will produce a biased view of many aspects of the genetic variation in a population.

In summary, empirical researchers should collect individuals in as random a manner as practical and should not bias their sample towards the middle of a species' range. When sampling is clustered, summary statistics based on segregating sites (e.g. Watterson's Θ and Tajima's D), heterozygosity, or the distribution of long haplotype blocks, can be expected to depart significantly from what would be observed under random sampling. Comparing the results of analyses conducted on all individuals versus those limited to single individuals per locality may help to reveal any effect of sampling bias if sufficient samples are available. Alternatively, sampling strategy could be incorporated directly into a simulation-based inferential framework by, for example, setting the distribution of samples in a simulation to mimic that used for the empirical target case. This could be readily achieved in a supervised machine learning or approximate Bayesian setting (CITES).

624 **Demography**

Classical population genetic models collapse many elements of life history variation into a single parameter, N_e , which is then taken to reflect the degree of variation present in the population when modeling the effects of selection or migration. Inferring N_e in the past is now a common goal of population genomic analyses and an important step in establishing baseline expectations of genetic variation when searching for signals of selection. Here we found that one method of inference of historic N_e based on genome-wide estimates of the site frequency

631 spectrum, Stairwayplot (Liu and Fu 2015), is relatively robust to variation in dispersal distance
632 when sampling is random and neighborhood size is over 20. However, non-random sampling,
633 and particularly midpoint sampling, causes the method to infer inflated estimates of past
634 population sizes and a series of recent bottlenecks (Figure 4). All sampling strategies lead to
635 inflated ancient and deflated recent N_e when neighborhood sizes were less than 20.

636 These predictions match the biases visible in the raw site frequency spectrum (Figure
637 3) – the deficit of low frequency alleles corresponds to the recent bottlenecks while the
638 inflation of mid-frequency alleles corresponds to the high ancestral N_e . Though we found
639 that Stairwayplot is a noisy estimator of equilibrium demography in general, there was no
640 significant bias in demographic estimates for any sampling strategy for neighborhood sizes
641 over 100. Thus many existing analyses are likely robust to biases in inferred N_e caused by
642 limited dispersal in continuous landscapes. However barriers to dispersal will likely lead to
643 higher levels of differentiation than we simulated here, and may mimic those seen at the low
644 end of continuous dispersal we simulated.

645 (*could be another paragraph here discussing the relationship between N_e , the distribution of coales-
646 cence times, and dispersal.) (wonder if it is worth revisiting MSMC or similar now that we know more
647 about how to run the method?)*

648 **GWAS**

649 Over the last twenty years genome-wide association studies (GWAS) have identified tens of
650 thousands of correlations between genetic variation and phenotypes, both in humans and
651 other species. This technique is increasingly applied to questions of human health through
652 methods like polygenic risk scores that sum the effect sizes estimated from GWAS to predict
653 an individual's phenotype or disease risk (Khera *et al.* 2018). The most common approach
654 to GWAS, which we followed in our analyses of simulated data here, is to regress the count
655 of derived alleles at a site against individual phenotypes, taking the slope of this regression
656 as an estimate of the effect of the allele on the phenotype. As recently reviewed by Visscher
657 and Goddard (2019), this is exactly the approach outlined by Fisher (1918) at the dawn of
658 quantitative genetics.

Stepping back from the mechanics of GWAS specifically, the approach of quantitative genetics is to decompose the variance in phenotypes into environmental and genetic effects,

e.g.

$$P = G + E \quad (3)$$

$$\text{var}(P) = \text{var}(G) + \text{var}(E) + \text{cov}(G, E) \quad (4)$$

659 For GWAS in structured populations, the bias identified in many previous studies (Price *et al.* 2006; Yu *et al.* 2005; Young *et al.* 2018; Mathieson and McVean 2012; Kang *et al.* 2008, 2010; Bulik-
660 Sullivan *et al.* 2015) in which test statistics reflect population structure rather than phenotype
661 association arises because of the presence of a positive covariance between genotypic and
662 environmental variation (the third term above). When the environment and genotype covary
663 across space their effects are confounded. Note this does not require interaction between
664 genotype and environment (so-called GxE effects), which will introduce an additional source
665 of bias. Here we refer simply to the fact that the allele frequency at a site will often covary
666 with the environment when both breeding structure and the environment vary over space.
667 To some degree the success of quantitative genetics in fields like agriculture is likely due to
668 the absence of this covariation – crop and animal breeding operations can be conducted so
669 that the environment is identical (or nearly so) or randomized across populations. However
670 in natural populations this situation is extremely unlikely. Most populations are structured
671 by a combination of limited dispersal and geographic barriers, and nearly all environments
672 vary over space. GWAS in natural populations is forced to confront the confounding effects of
673 population structure and the environment directly.

675 Incorporating PC positions as covariates in the analysis (Price *et al.* 2006) is designed to
676 address this difficulty by regressing out a baseline level of “average” differentiation. However
677 while this approach is quite useful, it does not truly separate the confounding signals of
678 environment and spatially varying genotypes. In essence with a PC-corrected GWAS we
679 are asking “what regions of the genome are more associated with this phenotype than the
680 average genome-wide association observed across populations?” In our simulations we
681 observed that this procedure can fail under a variety of circumstances. If dispersal is limited
682 and environmental variation is clustered in space (i.e. corner or patchy distributions in
683 our simulations), PCA positions fail to capture the fine-scale spatial structure required to
684 remove all signals of association. Conversely when dispersal is high we found that PCA
685 loses power to describe population structure before the spatial scale of dispersal breaks

686 down the relationship between genotype and the environment. These effects were observed
687 in all spatially varying environmental distributions, but were particularly pronounced for
688 concentrated environmental effects in one region, as was also found in Mathieson and McVean
689 (2012). As a result we can expect to see several thousand weak false-positive associations in
690 a PC-corrected GWAS conducted on a human-sized genome in species with neighborhood
691 sizes up to at least 1000.

692 This does not mean that GWAS is not useful, but does put some limits on the extent of
693 valid interpretation. Very few of the associations we identified would be significant at a
694 conservative Bonferroni-adjusted *p*-value cutoff, suggesting that most of the very strong
695 signals of association signals observed in studies of mono- or oligogenic traits are robust to
696 stratification bias. Further, the most dramatic effects of stratification inflation we observed
697 occurred at neighborhood sizes below 100 – smaller than the vast majority of modern human
698 populations (but see below for further discussion of empirical cases). However, as recently
699 identified in studies of genotype associations for human height in Europe (Berg *et al.* 2018;
700 Sohail *et al.* 2018), PC regression GWAS in modern human populations does leave residual
701 signal of population structure in large-scale GWAS of polygenic traits. Indeed, studies in
702 strongly structured species like *Arabidopsis* have long relied on more sophisticated mixed
703 model approaches to correcting for population structure for precisely this reason (Aranzana
704 *et al.* 2005; Sasaki *et al.* 2015).

705 A second point that has received less attention in the literature is the issue of overcorrection
706 in GWAS. If a truly causal allele segregates at different frequencies in different populations,
707 then correcting for population structure in a regression analysis will result in an underestimate
708 of effect sizes. Though our simulations had no causal alleles, we observed some evidence of
709 this effect in the distribution of *p*-values across the genome (Figure 5D): after PC regression
710 many analyses resulted in *p*-values falling below their expected values from a uniform
711 distribution. This result is consistent with a recent empirical study of heritability in human
712 height and body mass index, which found that increasing the number of PC axes used as
713 covariates caused the total proportion of variance explained by SNPs to decline from ≈ 0.8
714 to ≈ 0.75 (Wainschtein *et al.* 2019). Indeed SNPs with minor alleles frequencies of 0.001 -
715 0.01, which are expected to reflect fine scale population structure (Mathieson and McVean
716 2012; Novembre and Slatkin 2009), are estimated to explain *negative* proportions of the total

717 phenotypic variance in (Wainschtein *et al.* 2019) (*i don't understand what the authors mean here—*
718 *how can we have a negative variance?*). Searching for genetic associations of polygenic traits that
719 vary systematically across but not within populations through existing GWAS approaches is
720 then unlikely to be successful: the signals are fully confounded, and new analytic methods
721 or experiments controlling for variation in the environment will be necessary to rigorously
722 identify causal variants.

723 In summary, spatial covariation in population structure and the environment confound
724 the interpretation of GWAS *p*-values, and correction using principal components is insuf-
725 ficient to fully separate these signals for polygenic traits under a variety of environmental
726 and population parameter regimes. How more sophisticated mixed-model methods would
727 perform under our simulations is an interesting question that we plan to pursue in a future
728 study, but statistical methods can only take us so far in the absence of controlled environ-
729 ments. One currently popular approach to estimating the degree of bias in GWAS caused by
730 population structure is LD score regression (Bulik-Sullivan *et al.* 2015). Though this approach
731 appears to work well in practice, its interpretation is not always straightforward and it is
732 likely biased by the presence of linked selection (Berg *et al.* 2018). We suggest a straightfor-
733 ward alternative for species in which the primary axes of population differentiation is space
734 (note this is likely not the case for many modern human populations): run a GWAS with
735 spatial coordinates as phenotypes and check for *p*-value inflation or significant associations.
736 If significant associations with sample locality are observed after correcting for population
737 structure through PC regression or a kinship matrix, the structure corrections are insufficient.
738 This is essentially the approach taken in our “clinal” model (though we also include normally
739 distributed variation in our phenotypes). Of course it is possible that genotypes indirectly
740 affect individual locations by adjusting organismal fitness and thus habitat selection across
741 spatially varying environments, but we believe that this hypothesis should be tested against a
742 null of stratification bias inflation rather than accepted as true based on GWAS results. (*what*
743 *do you think of the second half here?*)

744 **Where are natural populations on this spectrum?**

745 For how much of the tree of life do spatial patterns circumscribe genomic variation? In Table
746 1 we gathered estimates of neighborhood size from a range of organisms to get an idea of

Table 1 Neighborhood size estimates from empirical studies.

Species	Description	Neighborhood Size	Method	Citation
<i>Ipomopsis aggregata</i>	flowering plant	12.60 - 37.80	Genetic	(Campbell and Dooley 1992)
<i>Borreria frutescens</i>	salt marsh plant	20 - 30	Genetic+Survey	(Antlfinger 1982)
<i>Oreamnos americanus</i>	mountain goat	36 - 100	Genetic	(Shirk and Cushman 2014)
<i>Homo sapiens</i>	Gainj- and Kalam-speaking people, Papua New Guinea	40 - 213	Genetic	(Rousset 1997)
<i>Formica sp.</i>	colonial ants	50 - 100	Genetic	(Pamilo 1983)
<i>Astrocaryum mexicanum</i>	palm tree	102 - 895	Genetic+survey	(Eguiarte <i>et al.</i> 1993)
<i>Spermophilus mollis</i>	ground squirrel	204 - 480	Genetic+Survey	(Antolin <i>et al.</i> 2001)
<i>Sceloporus olivaceus</i>	lizard	225 - 270	Survey	(Kerster 1964)
<i>Dieffenbachia longispatha</i>	beetle-pollinated colonial herb	227 - 611	Survey	(Young 1988)
<i>Homo sapiens</i>	Gainj- and Kalam-speaking people, Papua New Guinea	410	Survey	(Rousset 1997)
<i>Quercus laevis</i>	Oak tree	> 440	Genetic	(Berg and Hamrick 1995)
<i>Drosophila pseudoobscura</i>	fruit fly	500 - 1,000	Survey+Crosses	(Wright 1946)
<i>Homo sapiens</i>	POPRES data NE Europe	1,342 - 5,425	Genetic	(Ringbauer <i>et al.</i> 2017)
<i>Bebicium vittatum</i>	intertidal snail	240,000	Survey	(Rousset 1997)
<i>Bebicium vittatum</i>	intertidal snail	360,000	Genetic	(Rousset 1997)

747 how likely dispersal is to play an important role in patterns of variation. Though this sample
748 is almost certainly biased towards small-neighborhood species (because few studies have
749 quantified neighborhood size in species with very high dispersal or population density), we
750 find that neighborhood sizes in the range we simulated are fairly common across a range of
751 taxa. At the extreme low end of empirical neighborhood size estimates we see some flowering
752 plants, large mammals, and colonial insects like ants. Species such as this have neighborhood
753 size estimates small enough that spatial processes are likely to strongly influence inference.
754 These include some human populations such as the Gainj- and Kalam-speaking people of
755 Papua New Guinea, in which the estimated neighborhood sizes in (Rousset 1997) range from
756 40 to 410 depending on the method of estimation. Many more species occur in a middle range
757 of neighborhood sizes between 100 and 1000 – a range in which spatial processes play a minor
758 role in our analyses under random spatial sampling but are important when sampling of
759 individuals in space is clustered. Last, many species likely have neighborhood sizes much
760 larger than we simulated, including modern humans in NE Europe (Ringbauer *et al.* 2017).
761 For these species demographic inference and summary statistics are likely to reflect minimal
762 bias from spatial effects as long as dispersal is truly continuous across the landscape. While
763 that is so we caution that association studies in which the effects of population structure are
764 confounded with spatial variation in the environment are still sensitive to dispersal even at
765 these large neighborhood sizes.

766 **Future Directions and Limitations**

767 As we have shown, a large number of population genetic summary statistics contain in-
768 formation about spatial population processes. We imagine that combinations of such sum-
769maries might be sufficient for the construction of supervised machine learning regressors (e.g.
770 Schrider and Kern (2018)) for the accurate estimation of dispersal from genetic data. Indeed
771 Ashander *et al.* (2018) found that inverse interpolation on a vector of summary statistics
772 provided a powerful method of estimating dispersal distances. Expanding this approach to
773 include the haplotype-based summary statistics studied here and applying machine learning
774 regressors built for general inference of nonlinear relationships from high-dimensional data
775 may allow precise estimation of spatial parameters under a range of complex models.

776 One complication in the inference of any spatial demographic parameter is the balance

777 between local and global process. Many species are structured locally by limited dispersal,
778 but also contain deeply divergent lineages in different regions that reflect signals of ancient
779 episodes of geographic isolation or strong barriers to dispersal. Gene flow upon secondary
780 contact of two previously isolated lineages should create clinal patterns similar to isolation by
781 distance, and it will be difficult to determine when inferred dispersal parameters are reflecting
782 recent demographic process versus the historic patterns of geographic isolation. In addition,
783 spatially varying selection will create allele frequency variation over space that may mimic
784 isolation by distance. Indeed, a series of field studies described in Schemske and Bierzychudek
785 (????) found that in Wright's original empirical example of isolation by distance, the flowering
786 plant *Linanthus parryae*, patterns of flower color differentiation over space primarily reflect
787 temporal and spatial variation in selection rather than limited dispersal. Studies simulating
788 selection and dispersal interacting in space (e.g. Ralph and Coop (2010)) and testing for
789 identifiability of inferred dispersal or selective parameters may offer new insight into the
790 extent of our ability to accurately infer evolutionary processes in real systems.

791 Though our continuous space simulation allows incorporation of realistic demographic and
792 spatial processes and is much faster than previous individual-based models, it is inevitably
793 limited by the computational burden of tracking tens of thousands of individuals in every
794 generation. In particular the calculations required for our mate selection and competition steps
795 involve summarizing distances across all pairs of individuals and so scale very poorly (*how*
796 *exactly does this scale? $O(N^2)$?*) as the number of individuals within a three- σ radius increases.
797 In part the issue of runtime scaling is a function of the spatial process itself – under very
798 limited dispersal we observed that coalescence requires over $30N$ generations, so forward-
799 time methods must be run for a very long time to create a complete genealogy as underlies all
800 real genome sequence data. The reverse-time model of continuous space evolution described
801 in Barton *et al.* (2010) and implemented in Kelleher *et al.* (2014) may allow exploration of
802 parameter regimes with population and landscape sizes more directly comparable to empirical
803 cases like humans. However, incorporating selection or other processes into such models will
804 be difficult.

805 A mixed approach may be possible by combining forward- and reverse-time models, as
806 was recently done for a continuous-space Wright-Fisher model in Lotterhos (2019) and for a
807 simulation with linked selection in Buffalo and Coop (2019). This would allow us to generate

808 short runs of complex, realistic simulations in forward time in SLiM (Haller and Messer 2019)
809 and then “finish” the simulations as a coalescent simulation in msprime (Kelleher *et al.* 2016).
810 However a significant difficulty in this “recapitation” approach is scaling the variance in
811 reproductive output and generation time across forward- and reverse-time methods. Further
812 development of our understanding of how to merge forward- and reverse-time models is
813 a promising avenue for future research that will be necessary for scaling continuous-space
814 simulations to millions or billions of individuals.

815 Finally we believe that the difficulties in correcting for population structure in continuous
816 populations using principal components analysis or similar decompositions is a difficult issue,
817 well worth considering on its own – what does it mean to correct for population structure in a
818 world without discrete demes? What are the boundaries between proper corrections between
819 ancestry and underpowering the search for local genetic variation? We posit that there is
820 progress yet to be made in deriving either process driven descriptions of ancestry or more
821 generalized unsupervised methods that can better account for shared relatedness, say in the
822 context of GWAS, for populations that are structured over space. (*this paragraph is a bit out*
823 *there. rein me in*)

824 **Data Availability**

825 Scripts used for all analyses and figures are available at <https://github.com/petrelharp/spaceness>.

826 **Acknowledgements**

827 We thank folks for reading and thinking about this manuscript. We thank the Hearth for
828 having such good nearby coffee. Falling Sky Brewing was key to the success of this research
829 endeavor. CJB and ADK were supported by NIH award R01GM117241.

830 **Literature Cited**

831 Aguiñal, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, *et al.*, 2017 Deconstruct-
832 ing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics
833 **13**: 1–27.

834 Antlfinger, A. E., 1982 Genetic neighborhood structure of the salt marsh composite, *Borrichia*
835 *frutescens*. Journal of Heredity **73**: 128–132.

- 836 Antolin, M. F., B. V. Horne, M. D. Berger, Jr., A. K. Holloway, J. L. Roach, *et al.*, 2001 Effective
837 population size and genetic structure of a piute ground squirrel (*Spermophilus mollis*)
838 population. Canadian Journal of Zoology **79**: 26–34.
- 839 Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, *et al.*, 2005 Genome-wide association
840 mapping in arabidopsis identifies previously known flowering time and pathogen resistance
841 genes. PLoS genetics **1**: e60; e60–e60.
- 842 Ashander, J., P. Ralph, E. McCartney-Melstad, and H. B. Shaffer, 2018 Demographic inference
843 in a spatially-explicit ecological model from genomic data: a proof of concept for the mojave
844 desert tortoise. bioRxiv .
- 845 Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, *et al.*, 2016 The great
846 migration and african-american genomic diversity. PLOS Genetics **12**: 1–27.
- 847 Barton, N. H., F. Depaulis, and A. M. Etheridge, 2002 Neutral evolution in spatially continuous
848 populations. Theoretical Population Biology **61**: 31–48.
- 849 Barton, N. H., J. Kelleher, and A. M. Etheridge, 2010 A new model for extinction and recolonization
850 in two dimensions: Quantifying phylogeography. Evolution **64**: 2701–2715.
- 851 Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing
852 under dependency. The Annals of Statistics **29**: 1165–1188.
- 853 Berg, E. E. and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. Evolution
854 **49**: 110–120.
- 855 Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, *et al.*, 2018 Reduced signal for polygenic adaptation of height in uk biobank. bioRxiv .
- 856 Buffalo, V. and G. Coop, 2019 The linked selection signature of rapid adaptation in temporal
858 genomic data. bioRxiv .
- 859 Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 Ld score regression
860 distinguishes confounding from polygenicity in genome-wide association studies. Nature
861 Genetics **47**: 291 EP –.
- 862 Campbell, D. R. and J. L. Dooley, 1992 The spatial scale of genetic differentiation in a
863 hummingbird-pollinated plant: Comparison with models of isolation by distance. The
864 American Naturalist **139**: 735–748.
- 865 Chapman, N. H. and E. A. Thompson, 2002 The effect of population history on the lengths of
866 ancestral chromosome segments. Genetics **162**: 449–458.

- 867 Doebeli, M. and U. Dieckmann, 2003 Speciation along environmental gradients. *Nature* **421**:
868 259 EP –.
- 869 Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, *et al.*, 1993 Direct
870 and indirect estimates of neighborhood and effective population size in a tropical palm,
871 *astrocaryum mexicanum*. *Evolution* **47**: 75–87.
- 872 Epperson, B., 2003 *Geographical Genetics*. Monographs in Population Biology, Princeton Uni-
873 versity Press.
- 874 Felsenstein, J., 1975 A pain in the torus: Some difficulties with models of isolation by distance.
875 *The American Naturalist* **109**: 359–368.
- 876 Fisher, R. A., 1918 The correlation between relatives on the supposition of mendelian inheri-
877 tance. *Transactions of the Royal Society of Edinburgh* **52**: 399–433.
- 878 Fox, J. and S. Weisberg, 2011 *An R Companion to Applied Regression*. Sage, Thousand Oaks CA,
879 second edition.
- 880 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in
881 north american *drosophila melanogaster* show signatures of soft sweeps. *PLOS Genetics* **11**:
882 1–32.
- 883 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence
884 recording in SLiM opens new horizons for forward-time simulation of whole genomes.
885 *Molecular Ecology Resources* **19**: 552–566.
- 886 Haller, B. C. and P. W. Messer, 2019 SLiM 3: Forward Genetic Simulations Beyond the
887 Wright–Fisher Model .
- 888 Harris, K. and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared
889 haplotype lengths. *PLOS Genetics* **9**: 1–20.
- 890 Huillet, T. and M. Möhle, 2011 On the extended Moran model and its relation to coalescents
891 with multiple collisions. *Theoretical Population Biology* pp. –.
- 892 Jay, F., P. Sjödin, M. Jakobsson, and M. G. Blum, 2012 Anisotropic Isolation by Distance: The
893 Main Orientations of Human Genetic Differentiation. *Molecular Biology and Evolution* **30**:
894 513–525.
- 895 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component
896 model to account for sample structure in genome-wide association studies. *Nature Genetics*
897 **42**: 348 EP –.

- 898 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, *et al.*, 2008 Efficient control
899 of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- 900 Kelleher, J., A. Etheridge, and N. Barton, 2014 Coalescent simulation in continuous space:
901 Algorithms for large neighbourhood size. *Theoretical Population Biology* **95**: 13 – 23.
- 902 Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and ge-
903 nealogical analysis for large sample sizes. *PLoS Comput Biol* **12**: 1–22.
- 904 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for
905 fast population genetics simulation. *PLOS Computational Biology* **14**: 1–21.
- 906 Kerster, H. W., 1964 Neighborhood size in the rusty lizard, *sceloporus olivaceus*. *Evolution* **18**:
907 445–457.
- 908 Khera, A. V., M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, *et al.*, 2018 Genome-wide poly-
909 genic scores for common diseases identify individuals with risk equivalent to monogenic
910 mutations. *Nature Genetics* **50**: 1219–1224.
- 911 Kingman, J., 1982 The coalescent. *Stochastic Processes and their Applications* **13**: 235 – 248.
- 912 Liu, X. and Y.-X. Fu, 2015 Exploring population size changes using snp frequency spectra.
913 *Nature Genetics* **47**: 555 EP –.
- 914 Lotterhos, K. E., 2019 The effect of neutral recombination variation on genome scans for
915 selection. *G3: Genes, Genomes, Genetics* .
- 916 Malécot, G., 1948 Les mathématiques de l'hérédité .
- 917 Maruyama, T., 1972 Rate of decrease of genetic variability in a two-dimensional continuous
918 population of finite size. *Genetics* **70**: 639–651.
- 919 Mathieson, I. and G. McVean, 2012 Differential confounding of rare and common variants in
920 spatially structured populations. *Nature Genetics* **44**: 243 EP –.
- 921 Miles, A. and N. Harding, 2017 *cgg/h/scikit-allel*: v1.1.8.
- 922 Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, *et al.*, 2013 Estimation of effec-
923 tive population size in continuously distributed populations: there goes the neighborhood.
924 *Heredity* **111**: 189 EP –.
- 925 Novembre, J. and M. Slatkin, 2009 Likelihood-based inference in isolation-by-distance models
926 using the spatial distribution of low-frequency alleles. *Evolution* **63**: 2914–2925.
- 927 Pamilo, P., 1983 Genetic differentiation within subdivided populations of *formica* ants. *Evolu-*
928 *tion* **37**: 1010–1022.

- 929 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. PLOS
930 Genetics **2**: 1–20.
- 931 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, *et al.*, 2006 Principal
932 components analysis corrects for stratification in genome-wide association studies. Nature
933 Genetics **38**: 904 EP –.
- 934 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using
935 multilocus genotype data. Genetics **155**: 945–959.
- 936 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.*, 2007 Plink: A tool set for
937 whole-genome association and population-based linkage analyses. The American Journal
938 of Human Genetics **81**: 559 – 575.
- 939 R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for
940 Statistical Computing, Vienna, Austria.
- 941 Ralph, P. and G. Coop, 2010 Parallel adaptation: One or many waves of advance of an
942 advantageous allele? Genetics **186**: 647–668.
- 943 Ralph, P. and G. Coop, 2013 The geography of recent genetic ancestry across Europe. PLoS
944 Biol **11**: e1001555.
- 945 Ringbauer, H., G. Coop, and N. H. Barton, 2017 Inferring recent demography from isolation
946 by distance of long shared sequence blocks. Genetics **205**: 1335–1351.
- 947 Ringbauer, H., A. Kolesnikov, D. L. Field, and N. H. Barton, 2018 Estimating barriers to gene
948 flow from distorted isolation-by-distance patterns. Genetics **208**: 1231–1245.
- 949 Rousset, F., 1997 Genetic differentiation and estimation of gene flow from f-statistics under
950 isolation by distance. Genetics **145**: 1219–1228.
- 951 Sasaki, E., P. Zhang, S. Atwell, D. Meng, and M. Nordborg, 2015 "missing" g x e variation
952 controls flowering time in arabidopsis thaliana. PLOS Genetics **11**: 1–18.
- 953 Schemske, D. W. and P. Bierzychudek, ???? Spatial differentiation for flower color in the desert
954 annual linanthus parryae: Was wright right? Evolution **61**: 2528–2543.
- 955 Schrider, D. R. and A. D. Kern, 2018 Supervised machine learning for population genetics: A
956 new paradigm. Trends in Genetics **34**: 301 – 312.
- 957 Sharbel, T. F., B. Haubold, and T. Mitchell-Olds, 2000 Genetic isolation by distance in ara-
958 bidopsis thaliana: biogeography and postglacial colonization of europe. Molecular Ecology
959 **9**: 2109–2118.

- 960 Shirk, A. J. and S. A. Cushman, 2014 Spatially-explicit estimation of wright's neighborhood
961 size in continuous populations. *Frontiers in Ecology and Evolution* **2**: 62.
- 962 Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, *et al.*, 2018 Signals of polygenic
963 adaptation on height have been overestimated due to uncorrected population structure in
964 genome-wide association studies. *bioRxiv* .
- 965 Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA poly-
966 morphism. *Genetics* **123**: 585–595.
- 967 Visscher, P. M. and M. E. Goddard, 2019 From r.a. fisher's 1918 paper to gwas a century later.
968 *Genetics* **211**: 1125–1130.
- 969 Wahlund, S., 1928 Zusammensetzung von populationen und korrelationserscheinungen vom
970 standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**: 65–106.
- 971 Wainschtein, P., D. P. Jain, L. Yengo, Z. Zheng, , *et al.*, 2019 Recovery of trait heritability from
972 whole genome sequence data. *bioRxiv* .
- 973 Wakeley, J., 2005 *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village,
974 CO.
- 975 Wakeley, J. and T. Takahashi, 2003 Gene genealogies when the sample size exceeds the effective
976 size of the population. *Mol Biol Evol* **20**: 208–213.
- 977 Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- 978 Wilke, C. O., 2019 *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package
979 version 0.9.4.
- 980 Wright, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97.
- 981 Wright, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- 982 Wright, S., 1946 Isolation by distance under diverse systems of mating. *Genetics* **31**: 336.
- 983 Wright, S., T. Dobzhansky, and W. Hovanitz, 1942 Genetics of natural populations. vii. the
984 allelism of lethals in the third chromosome of drosophila pseudoobscura. *Genetics* **27**:
985 363–394.
- 986 Young, A. I., M. L. Frigge, D. F. Gudbjartsson, G. Thorleifsson, G. Bjornsdottir, *et al.*, 2018
987 Relatedness disequilibrium regression estimates heritability without environmental bias.
988 *Nature Genetics* **50**: 1304–1310.
- 989 Young, H. J., 1988 Neighborhood size in a beetle pollinated tropical aroid: effects of low
990 density and asynchronous flowering. *Oecologia* **76**: 461–466.

Table S1 Summary statistics calculated on simulated genotypes.

Statistic	Description
Θ_{pi}	Mean of the distribution of pairwise genetic differences
Θ_{W}	Effective population size based on segregating sites
Segregating Sites	Total number of segregating sites in the sample
Tajima's D	Difference in Θ_{pi} and Θ_{W} over its standard deviation
Observed Heterozygosity	Proportion of heterozygous individuals in the sample
F_{IS}	Wright's inbreeding coefficient $1 - H_e / H_o$
$var(D_{xy})$	Variance in the distribution of pairwise genetic distances
$skew(D_{xy})$	Skew of the distribution of pairwise genetic distances
$mean(IBS)$	Mean of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$var(IBS)$	Variance of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$skew(IBS)$	Skew of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$nIBS$	Number of IBS tracts with length > 2bp across all pairs of individuals.
$mean(IBS > 1e6)$	Mean number of IBS tracts over 1×10^6 bp per pair across all pairs in the sample.
$corr(D_{xy}, dist)$	Pearson correlation between genetic distance and $\log_{10}(spatial\ distance)$
$corr(mean(IBS), dist)$	Pearson correlation between the mean of the IBS tract distribution for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(nIBS, dist)$	Pearson correlation between the number of IBS tracts for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(IBS > 1e6, dist)$	Pearson correlation between the number of IBS tracts $> 1 \times 10^6$ bp for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(skew(IBS), dist)$	Pearson correlation between the skew of the distribution of pairwise haplotype block lengths for each pair of samples and $\log_{10}(spatial\ distance)$

991 Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, *et al.*, 2005 A unified mixed-model
 992 method for association mapping that accounts for multiple levels of relatedness. *Nature*
 993 *Genetics* **38**: 203 EP –.

994 **Supplementary Figures and Tables**

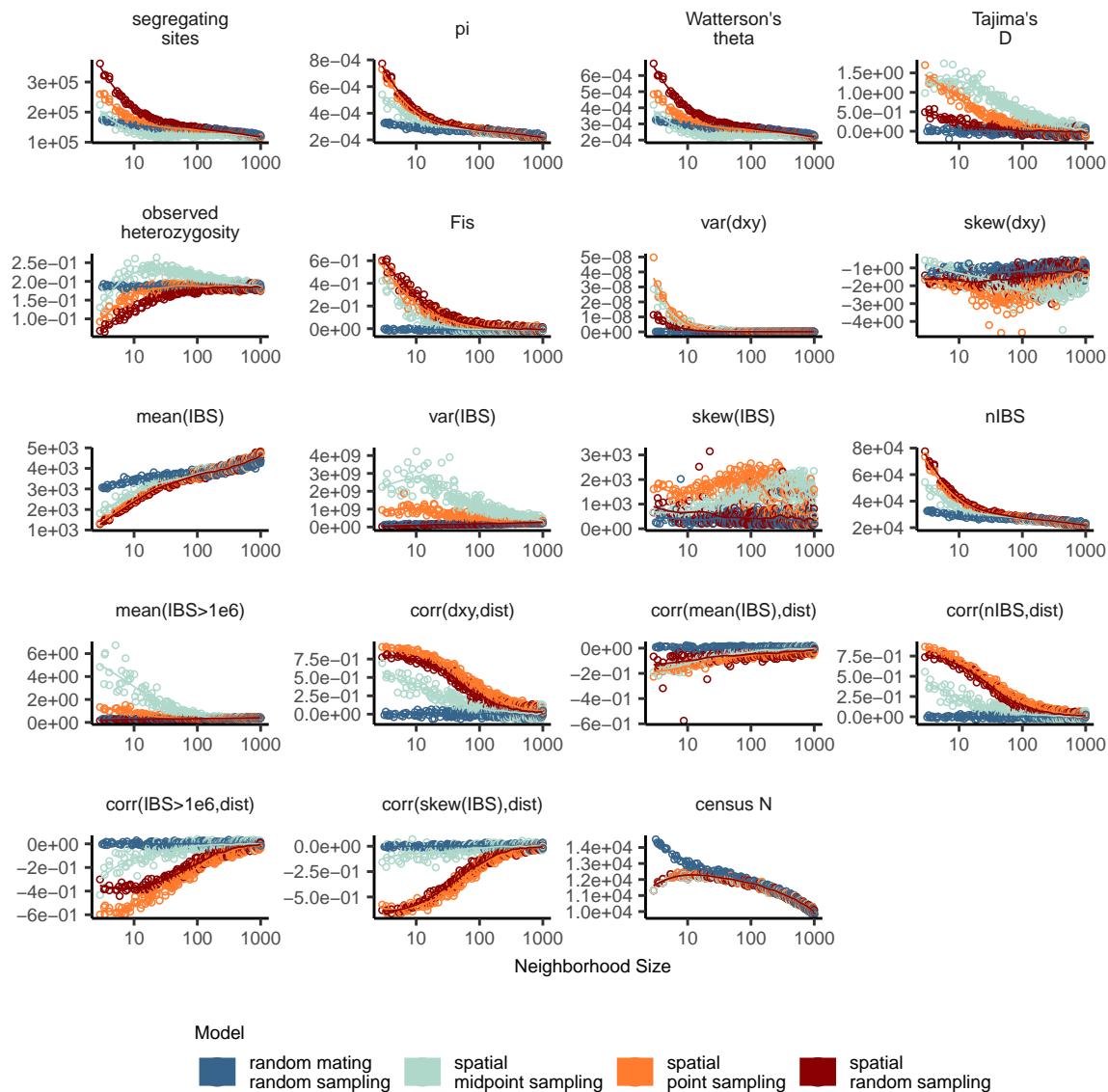


Figure S1 Change in summary statistics by neighborhood size and sampling scheme calculated from simulated sequence data of 60 individuals.

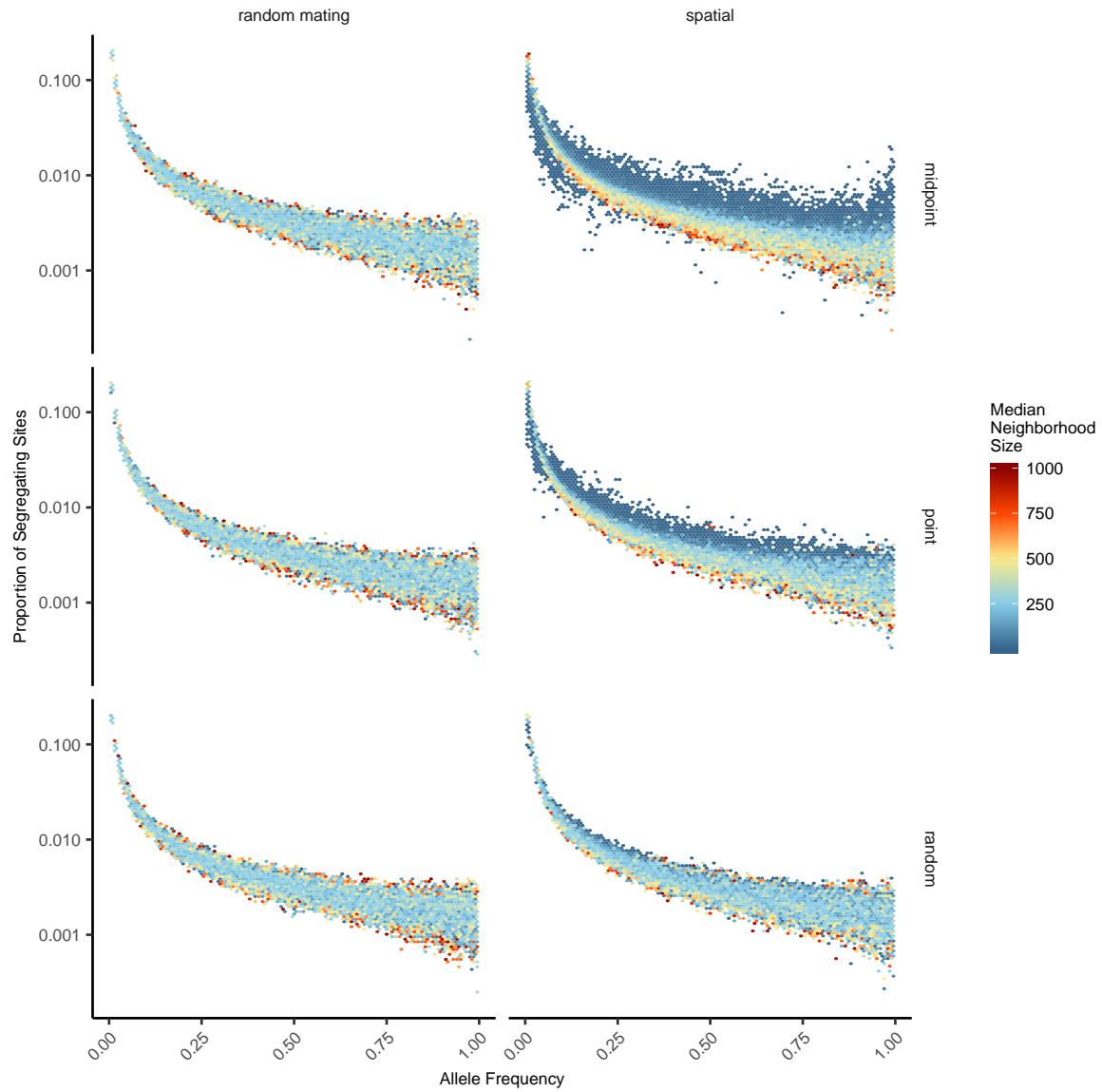


Figure S2 Site frequency spectra for random mating and spatial SLiM models under all sampling schemes.

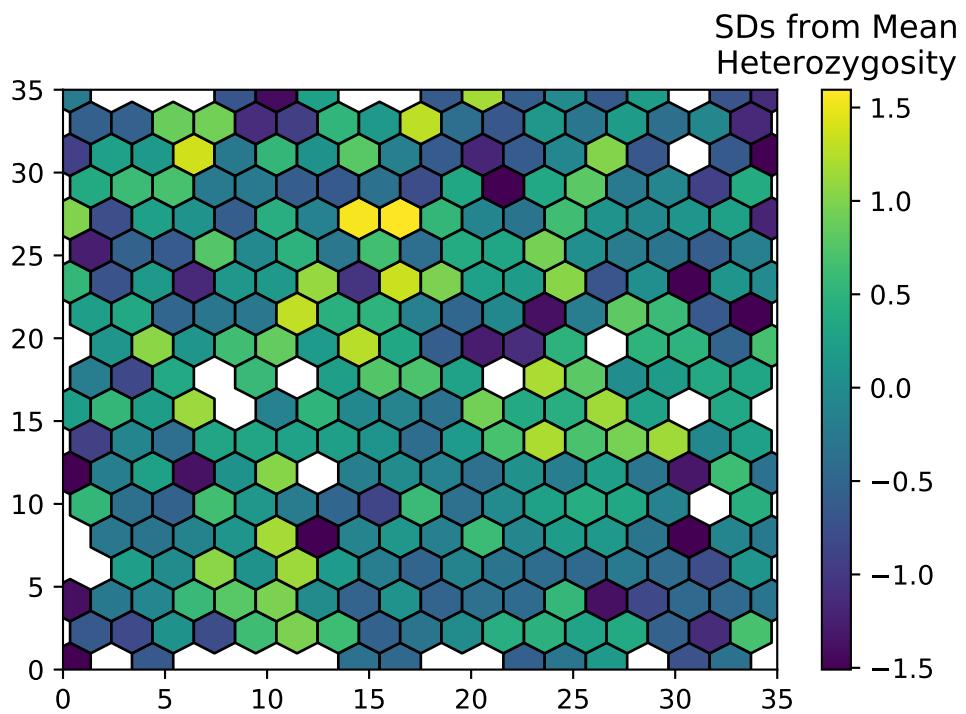


Figure S3 Normalized mean observed heterozygosity by location across 200 randomly-sampled individuals

Table S2 Anova and Levene's test p values for differences by sampling strategy. Bolded values are rejected at $\alpha = 0.05$

variable	model	p(equal means)	p(equal variance)
segsites	random mating	0.998190	0.980730
pi	random mating	0.997750	0.996450
thetaW	random mating	0.998190	0.980730
tajD	random mating	0.879690	0.188770
het_o	random mating	0.531540	0.433230
fis	random mating	0.474790	0.785730
gen_dist_mean	random mating	0.997770	0.996510
gen_dist_var	random mating	0.283630	0.647240
gen_dist_skew	random mating	0.958320	0.260750
gen_sp_corr	random mating	0.601980	0.000000
ibs_mean	random mating	0.997960	0.997730
ibs_var	random mating	0.486450	0.399490
ibs_skew	random mating	0.117980	0.069770
ibs_blocks_per_pair	random mating	0.997680	0.996570
ibs_blocks_over_1e6_per_pair	random mating	0.834870	0.888730
ibs_mean_spat_corr	random mating	0.073270	0.308420
ibs_1e6blocks_spat_corr	random mating	0.268440	0.002100
ibs_skew_spat_corr	random mating	0.396920	0.000620
ibs_blocks_spat_corr	random mating	0.581090	0.000000
segsites	spatial	0.000000	0.000000
pi	spatial	0.026510	0.013440
thetaW	spatial	0.000000	0.000000
tajD	spatial	0.000000	0.000000
het_o	spatial	0.000000	0.000000
fis	spatial	0.000000	0.000120
gen_dist_mean	spatial	0.025390	0.012910
gen_dist_var	spatial	0.004970	0.006230
gen_dist_skew	spatial	0.000000	0.000000
gen_sp_corr	spatial	0.000000	0.000000
ibs_mean	spatial	0.272400	0.114250
ibs_var	spatial	0.000000	0.000000
ibs_skew	spatial	0.000000	0.000000
ibs_blocks_per_pair	spatial	0.033920	0.016640
ibs_blocks_over_1e6_per_pair	spatial	0.000000	0.000000
ibs_mean_spat_corr	spatial	0.000000	0.590540
ibs_1e6blocks_spat_corr	spatial	0.000000	0.000000
ibs_skew_spat_corr	spatial	0.000000	0.000000
ibs_blocks_spat_corr	spatial	0.000000	0.000000

Table S3 T-test results comparing standard deviations of inferred N_e between spatial and coalescent models, by neighborhood size (NS) and sampling strategy. p is the probability that spatial models have higher standard deviations.

sampling	NS range	t	df	p
random	2-20	4.2572	41.6166	0.0001
random	20-100	-1.8473	171.9905	0.9668
random	100-500	-2.1297	164.3864	0.9827
random	500-1000	-3.9681	147.0497	0.9999
point	2-20	7.0802	44.3615	0.0000
point	20-100	-0.2038	169.3799	0.5806
point	100-500	-2.4945	152.5000	0.9932
point	500-1000	-3.8329	162.6443	0.9999
midpoint	2-20	5.9253	59.5462	0.0000
midpoint	20-100	3.8940	171.7005	0.0001
midpoint	100-500	-2.2764	139.5221	0.9878
midpoint	500-1000	-3.2223	165.0792	0.9992