

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: what are conclusions about bias or not of $Ne(t)$?)

Introduction

The inescapable reality that biological organisms live, move, and reproduce in continuous geography is usually omitted from population genetic models. However, mates tend to

live near to one another and to their offspring, leading to a positive correlation between genetic and geographic distances. This pattern of “isolation by distance” (Wright 1943), describing mean relatedness as a function of distance in flat landscapes, is one of the most widely replicated empirical findings in population genetics (Aguillon *et al.* 2017; Jay *et al.* 2012; Sharbel *et al.* 2000). Despite a long history of analytical work describing the genetics of populations distributed across continuous geography (e.g., Wright (1943); Rousset (1997); Barton *et al.* (2002, 2010); Ringbauer *et al.* (2017); Robledo-Arnuncio and Rousset (2010)), much modern work still describes geographic structure as a set of populations connected by migration (e.g., Wright 1931; Epperson 2003; Rousset and Leblois 2011; Shirk and Cushman 2014; Lundgren and Ralph 2018). For this reason, most population genetics statistics are interpreted with reference to discrete, well-mixed populations, 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. We use these simulations to test how sampling strategy interacts with geographic population structure to cause systematic variation in population genetic summary statistics typically analyzed assuming discrete population models. We then examine how the fine-scale spatial structures occurring under limited dispersal impact demographic inference from the site frequency spectrum. Lastly, we examine the impacts of evolution in continuous space on genome-wide association studies (GWAS) and identify regions of parameter space under which the results from GWAS may be misleading.

Materials and Methods

Modeling Evolution in Continuous Space

The degree to which genetic relationships are geographically correlated depends on the chance that two geographically nearby individuals are close relatives – in modern terms, by the tension between migration (the chance that one is descended from a distant location) and coalescence (the chance that they share a parent). A key early observation Wright model is that this balance is nicely summarized by the “neighborhood size”, defined to be $N_W = 4\pi\rho\sigma^2$, where σ is the mean parent–offspring distance and ρ is population density. This can be thought of as proportional to the average number of potential mates for an individual (those within distance 2σ), or the number of potential parents of a randomly chosen individual. Maruyama (1972) found that the rate of decline in genetic diversity in a two-dimensional continuous population approaches the random mating expectation when $\rho\sigma^2 > 1$, and proposed that this had the important implication that most population genetic expectations for randomly mating populations could be applied to continuously distributed populations with relatively little error.

The first approach to modeling continuously-distributed populations was to endow individuals in a Wright-Fisher model with locations in continuous space. However, since the total size of the population is constrained, this introduces interactions between arbitrarily distant individuals, which (aside from being implausible) Felsenstein (1975) showed would eventually lead to unrealistic population clumping if the species range is sufficiently large. After global population regulation, the next method for modeling spatial populations might

123 to assume the existence of a grid of discrete randomly-mating populations connected by
124 migration, thus enforcing regular population density by edict. Among many other important
125 results drawn from this class of “lattice” or “stepping stone” models, Rousset (1997) showed
126 that the slope of the a linear regression of genetic differentiation (F_{ST}) against the logarithm
127 of spatial distance is an estimate of neighborhood size. Although these grid models are
128 likely good approximations of continuous geography in many situations, these do not model
129 demographic fluctuations, and limit investigation of spatial structure below the level of the
130 deme, assumptions whose impacts are unknown. An alternative method for dealing with
131 continuous geography is a new class of coalescent models, the Spatial Lambda Fleming-Viot
132 models (Barton *et al.* 2010; Kelleher *et al.* 2014).

133 To avoid questionable assumptions, we worked with forwards-time, individual-based
134 simulations. By scaling the probability of survival in each timestep to local population density,
135 we shift reproductive output towards regions of low-density, which prevents populations
136 from clustering. Such models have been used extensively in ecological modeling but rarely in
137 population genetics, where to our knowledge previous implementations of continuous space
138 models have focused on a small number of genetic loci, which limits the ability to investigate
139 the impacts of continuous space on genome-wide genetic variation as is now routinely
140 sampled from real organisms. By simulating chromosome-scale sequence alignments and
141 complete population histories we are able to treat our simulations as real populations and
142 replicate the sampling designs and analyses commonly conducted on real genomic data.

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

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

153 At the beginning of the simulation individuals are distributed uniformly at random on
 154 a continuous, square landscape. Individuals are hermaphroditic, and each time step, each
 155 produces a Poisson number of offspring with mean $1/L$ who disperse a random, Gaussian-
 156 distributed distance away from the parent with mean zero and standard deviation σ in both
 157 the x and y coordinates, reflected to stay within the species range. Each offspring is produced
 158 with a mate selected randomly from those within distance 3σ , with probability of choosing a
 159 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

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

160 We chose this functional form so that the equilibrium population density per unit area is
 161 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

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

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

165 To isolate spatial effects from other components of the model such as overlapping genera-
 166 tions, increased variance in reproductive success, and density-dependent fitness, we also

167 implemented simulations identical to those above except that mates are selected uniformly
168 random from the population, and offspring disperse to a uniform random location on the
169 landscape. We refer to this model as the “random mating” model, in contrast to the first,
170 “spatial” model.

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

174 We ran 400 simulations for the spatial and random-mating models on a square landscape
175 of width $W = 50$ with per-unit carrying capacity $K = 5$ (census $N \approx 10,000$), average lifetime
176 $L = 4$, genome size 10^8 , recombination rate 10^{-9} , and drawing σ values from a uniform
177 distribution between 0.2 and 4. To speed up the simulations and limit memory overhead
178 we used a mutation rate of 0 in SLiM and later applied mutations to the tree sequence
179 with msprime’s `mutate` function (Kelleher *et al.* 2016). Because msprime applies mutations
180 proportionally to elapsed time, we divided the mutation rate of 10^{-8} mutations per site per
181 generation by the average generation time estimated for each value of σ (see ‘Demographic
182 Parameters’ below) to convert the rate to units of mutations per site per unit time. (We
183 verified that this procedure produced the correct number of mutations by comparing to a
184 subset of simulations with SLiM-generated mutations, which are applied only at meiosis.)
185 Simulations were run for 1.6 million timesteps (approximately $30N$ generations), or until all
186 extant individuals shared a common ancestor within the simulation (i.e., the tree sequence
187 had coalesced). (*maybe worth including a table with some basic runtime results in the supplement?*)

188 **Demographic Parameters**

189 Our demographic model includes paramters for population density (K), mean life span (L),
190 and dispersal distance (σ). However, nonlinearity of local demographic stochasticity causes
191 actual realized averages of these demographic quantitites to deviate from the specified values
192 in a way that depends on the neighborhood size. Therefore, to properly compare to theoretical
193 expectations, we empirically calculated these demographic quantities in simulations. We
194 recorded the census population size in all simulations. To estimate generation times, we stored
195 ages of the parents of every new individual born across 200 timesteps, after a 100 generation
196 burn-in, and took the mean. To estimate variance in offspring number, we tracked the number

197 of offspring for all individuals for 100 timesteps following a 100-timestep burn-in period,
198 subset the resulting table to include only the last timestep recorded for each individual, and
199 calculated the variance in number of offspring across all individuals in timesteps 50-100. All
200 calculations were performed with information recorded in the tree sequence, using pyslim
201 (<https://github.com/tskit-dev/pyslim>).

202 **Sampling**

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

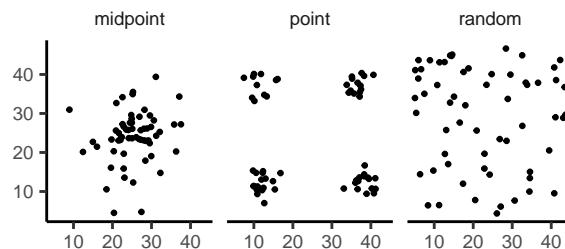


Figure 1 Example sampling maps for 60 individuals on a 50×50 landscape for midpoint, point, and random sampling strategies, respectively.

211 **Summary Statistics**

212 We calculated the site frequency spectrum and a set of 18 summary statistics (Table S1) from
213 60 diploid individuals sampled from the final generation of each simulation using the python
214 package scikit-allel (Miles and Harding 2017). Statistics included common single-population
215 summaries including mean pairwise divergence (π), inbreeding coefficient (F_{IS}), and Tajima’s
216 D , as well as the classic isolation-by-distance regression of genetic distance (D_{xy}) against the
217 logarithm of geographic distances (Rousset 1997), which we summarized as the correlation
218 coefficient between the logarithm of the spatial distance and the proportion of identical base

219 pairs across pairs of individuals.

220 Following recent studies that showed strong signals for dispersal and demography in the
221 distribution of shared haplotype block lengths (Ringbauer *et al.* 2017; Baharian *et al.* 2016),
222 we also calculated various summaries of the distribution of pairwise identical-by-state (IBS)
223 block lengths among samples. The full distribution of lengths of IBS tracts for each pair of
224 chromosomes was first calculated with a custom python function. We then calculated the first
225 three moments of this distribution (mean, variance, and skew) and the number of blocks over
226 1e6 base pairs both for each pair of individuals and for the full distribution across all pairwise
227 comparisons.

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

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

239 **Demographic Modeling**

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

248 In preliminary runs we found that inferred population histories were highly variable

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

259 **Association Studies**

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

In all simulations, the phenotype of each individual is determined by adding independent Gaussian noise with mean zero and standard deviation 10 to a mean that may depend on spatial position. We adjust the geographic pattern of mean phenotype to create spatially autocorrelated environmental influences on phenotype. In the first simulation of *nonspatial* phenotypes, the mean did not change, so that all individuals' phenotypes were drawn independently from a Gaussian distribution with mean 110 and standard deviation 10. Next, to simulate *clinal* environmental influences on phenotype, we increased the mean phenotype from 90 on the left edge of the range to 120 on the right edge (two phenotypic standard deviations). Concretely, an individual at position (x, y) in a 50×50 landscape has mean

phenotype $110 + 2x/5$. Third, we simulated a more concentrated “corner” environmental effect by setting the mean phenotype for individuals with both x and y coordinates below 20 to 130 (two standard deviations above the rest of the map). Finally, in “patchy” simulations we selected 10 random points on the map and set the mean phenotype of all individuals within three map units of each of these points to 130.

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

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

Results

Demographic Parameters

Adjusting the spatial dispersal and interaction distance, σ , has a surprisingly large effect on demographic quantities that are usually fixed in Wright-Fisher models – the generation time, census population size, and variance in offspring number. These are shown in Figure 2. This occurs because, even through the “population density” (K) and “mean lifetime” (L) parameters were the same in all simulations, the strength of stochastic effects depends strongly on σ . For instance, the population density near to individual i (denoted n_i above) is computed by averaging over roughly $N_W = 4\pi K\sigma^2$ individuals, and so has standard deviation proportional to $1/\sqrt{N_W}$ – it is more variable at lower densities. (Recall that N_W is Wright’s neighborhood

size.) Since the probability of survival is a nonlinear function of n_i , actual equilibrium densities and lifetimes differ from K and L . This is the reason that we included *random mating* simulations – where mate choice and offspring dispersal are both nonspatial – since this should preserve the random fluctuations in local population density while destroying any spatial genetic structure. We verified that random mating models retained no geographic signal by showing that summary statistics did not differ significantly between sampling regimes (Table S2), unlike in spatial models (discussed below).

There are a few additional things to note about Figure 2. First, all three quantities are non-monotone with neighborhood size. Census size largely declines as neighborhood size increases for both the spatial and random mating models. However, for spatial models this decline only begins for neighborhood size ≥ 10 . By a neighborhood sizes larger than 100, the spatial and random mating models are indistinguishable from one another, a sign that our simulations are performing as expected. Census sizes range from $\approx 14,000$ at low σ in the random mating model to $\approx 10,000$ for both models when neighborhood sizes approach 1,000.

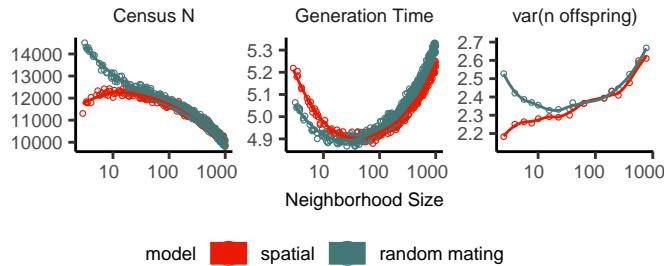


Figure 2 Genealogical parameters from spatial and random mating SLiM simulations, by neighborhood size.

Generation time similarly shows complex behavior with respect to neighborhood sizes, and varies between 5.2 and 4.9 timesteps per generation across the parameter range explored. Under both the spatial and random mating models, generation time reaches a minimum at a neighborhood size of around 50. Interestingly, under the range of neighborhood sizes that we examined, generation times between the random mating and spatial models are never quite equivalent – presumably this would cease to be the case at neighborhood sizes higher than we simulated here.

Last, we looked at the variance in number of offspring – a key parameter determining the

331 effective population size. Surprisingly, the spatial and random mating models behave quite
332 differently: while the variance in offspring number increases nearly monotonically under the
333 spatial model, the random mating model actually shows a decline in the variance in offspring
334 number until a neighborhood size ≈ 10 before it increases and eventually equals what we
335 observe in the spatial case.

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

337 Even though certain aspects of population demography depend on the scale of spatial inter-
338 actions, it still could be that population genetic variation is well-described by a well-mixed
339 population model. Indeed, mathematical results suggest that genetic variation in some spatial
340 models should be well-approximated by a Wright-Fisher population if neighborhood size is
341 large and all samples are geographically widely separated (Wilkins 2004; Zähle *et al.* 2005).
342 However, the behavior of most common population genetic summary statistics has not yet
343 been described in realistic geographic models. Moreover, as we will show, spatial sampling
344 strategies can affect summaries of variation at least as strongly as the underlying population
345 dynamics.

346 ***Site Frequency Spectra and Summaries of Diversity*** Figure 3 shows the effect of varying
347 neighborhood size and sampling strategy on the site frequency spectrum (Figure 3A) and
348 several standard population genetic summary statistics (Figure 3B). These show a significant
349 enrichment of intermediate frequency variants in comparison to the nonspatial expectation,
350 for smaller neighborhood sizes (≤ 100) that is exacerbated by midpoint and point sampling
351 of individuals (depicted in Figure 1). Reflecting this, Tajima's D is quite positive in the same
352 situations (Figure 3B). Notably, the point at which Tajima's D approaches 0 differs strongly
353 across sampling strategies – varying from a neighborhood size of roughly 50 for random
354 sampling to at least 1000 for midpoint sampling.

355 One of the most commonly used summaries of variation is Tajima's summary of nucleotide
356 divergence, $\widehat{\theta}_\pi$, (*In the figure this has no hat on; should we remove these in the text? The notation*
357 *is not standard.*) calculated as the mean density of nucleotide differences, averaged across
358 pairs of samples. As can be seen in Figure 3B, $\widehat{\theta}_\pi$ differs as much as nearly three-fold between
359 the random mating and spatial models. $\widehat{\theta}_\pi$ using each sampling strategy approaches the
360 random mating expectation at its own rate, but by a neighborhood size of around 100 all

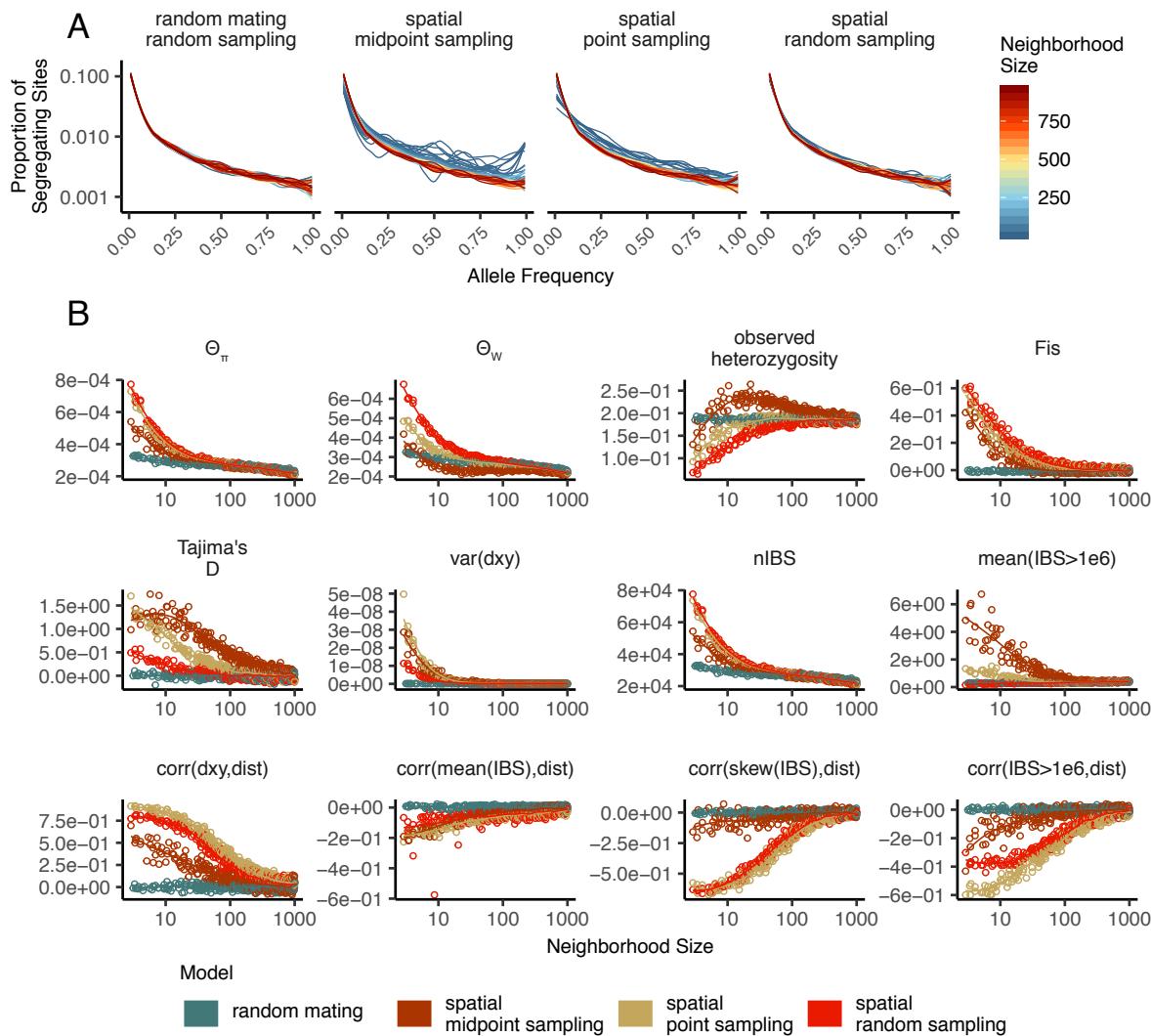


Figure 3 Site frequency spectrum (A) and summary statistic distributions (B) by sampling strategy and neighborhood size.

models are equivalent. (*To interpret this we need to compare to census size over variance in offspring number.*) The differences between spatial and random mating simulations are much greater than expected from differences in census size divided by variance in offspring number. This likely occurs because $\hat{\theta}_\pi$ is a measure of mean time to most recent common ancestor between two samples, and at small values of σ , the time for dispersal to mix ancestry across the range exceeds the mean nonspatial coalescent time. (For instance, at the smallest value of $\sigma = 0.2$, the range is 250 dispersal distances wide, and since the location of a diffusively moving lineage after k generations has variance $k\sigma^2$, it takes around $250^2 = 62500$ generations to mix across the range, which is roughly ten times larger than the random mating effective population size.) Interestingly, the effect of sampling strategy is reversed relative to that observed in Tajima's D – midpoint sampling reaches random mating expectations around neighborhood size 50, while random sampling is inflated until around neighborhood size 100.

Values of observed heterozygosity and its derivative F_{IS} also depend heavily on neighborhood size under spatial models as well as the sampling scheme. F_{IS} is inflated above the expectation across most of the parameter space examined and across all sampling strategies. This effect is caused by a deficit of heterozygous individuals in low-dispersal simulations – a continuous-space version of the Wahlund effect (Wahlund 1928). Indeed, for random sampling under the spatial model, F_{IS} does not approach the random mating equivalent until neighborhood sizes of nearly 1000. On the other hand, the dependency of raw observed heterozygosity on neighborhood size is not monotone. Under midpoint sampling observed heterozygosity is inflated even over the random mating expectation, as a result of the a higher proportion of heterozygotes occurring in the middle of the landscape (Figure S3). This squares with a report from Shirk and Cushman (2014) who observed a similar excess of heterozygosity in the middle of the landscape when simulating under a lattice model.

(*revisit this paragraph*) Trends in pairwise haplotype block sharing parallel those in allele-frequency-based diversity estimates (Figure 3, Supplementary Figure S1). At low dispersal the distribution of IBS block lengths in a set of samples is shifted towards smaller values with respect to the random mating expectation – resulting in lower means and fewer long IBS blocks. The variance and skew of the distribution of haplotype block lengths are only minorly affected by neighborhood size in our simulations when calculated across all pairs of individuals; however, they are strongly dependent on sampling regimes. For example, the

392 number of long haplotype blocks declines as neighborhood size increases under midpoint
393 sampling but changes very little across neighborhood sizes under point or random sampling.
394 Thus sampling strategies with respect to geography will affect conclusions drawn from
395 haplotype length distributions quite dramatically.

396 **Correlations of summary statistics with geographic distance** Correlating population genetic
397 summaries such as F_{ST} against geographic distance has shown great utility in empirical popu-
398 lation genetics (Rousset 1997). As we know the exact locations of individuals that are sampled
399 in our simulated populations we have examined the relationship of geographic distance
400 between samples with a number of summary statistics (Figures 3 and S1). Mean density of
401 nucleotide differences between individuals, D_{xy} , is positively correlated with the geographic
402 distance between the individuals, and the strength of this correlation declines as dispersal
403 increases, as expected under theory (Wright 1943; Rousset 1997). This relationship varies
404 across sampling strategies, with the weakest correlations observed for midpoint sampling,
405 perhaps due to a dearth of long-distance comparisons.

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

419 **Spatial distribution of allele copies** (*insert bit describing that plot here*)

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

427 (*revisit with smc++*) Figure 4 shows inferred population size histories, grouped by neighbor-
428 hood size and sampling strategy. In general, Stairwayplot tends to infer ancient population
429 increases and recent declines when neighborhood sizes were below 20 under all sampling
430 strategies (Figure 4). This is consistent with our observations of the SFS from which Stairway-
431 plot is doing its inference. Inflated past population sizes were seen in both point and random
432 sampling, demonstrating that the relatively minor shift in the site frequency spectrum ob-
433 served among sampling regimes is enough to alter demographic estimates. More alarmingly,
434 inference of severe population bottlenecks was common at neighborhood sizes under 100
435 for midpoint and point sampling strategies. Above neighborhood sizes of 100 the average
436 inferred demography across all simulations was relatively accurate, with minor fluctuations
437 slightly above the expected variance Ne . While that is so individual model fits were highly
438 variable and often inferred five-fold or greater population fluctuations even in high-dispersal
439 simulations. We compared this range of variation to Stairwayplot results run on coalescent
440 simulations with constant population size, and found that the noisiest results ($N_W < 20$, or
441 midpoint sampling and $N_W < 100$) were noisier than expected, but remaining simulation
442 results were consistent with Stairwayplot's behavior on data from random mating models
443 (Table S3). In summary, spatial mate choice and dispersal causes strong bias in SFS-based
444 demographic estimates for neighborhood sizes below 20 or when sampling is clustered, but
445 otherwise any biases are within the range of variability regularly inferred by Stairwayplot.
446 This underscores the fact that some *a priori* knowledge about the population dynamics at play
447 will be important to interpreting results of demographic estimation routines.

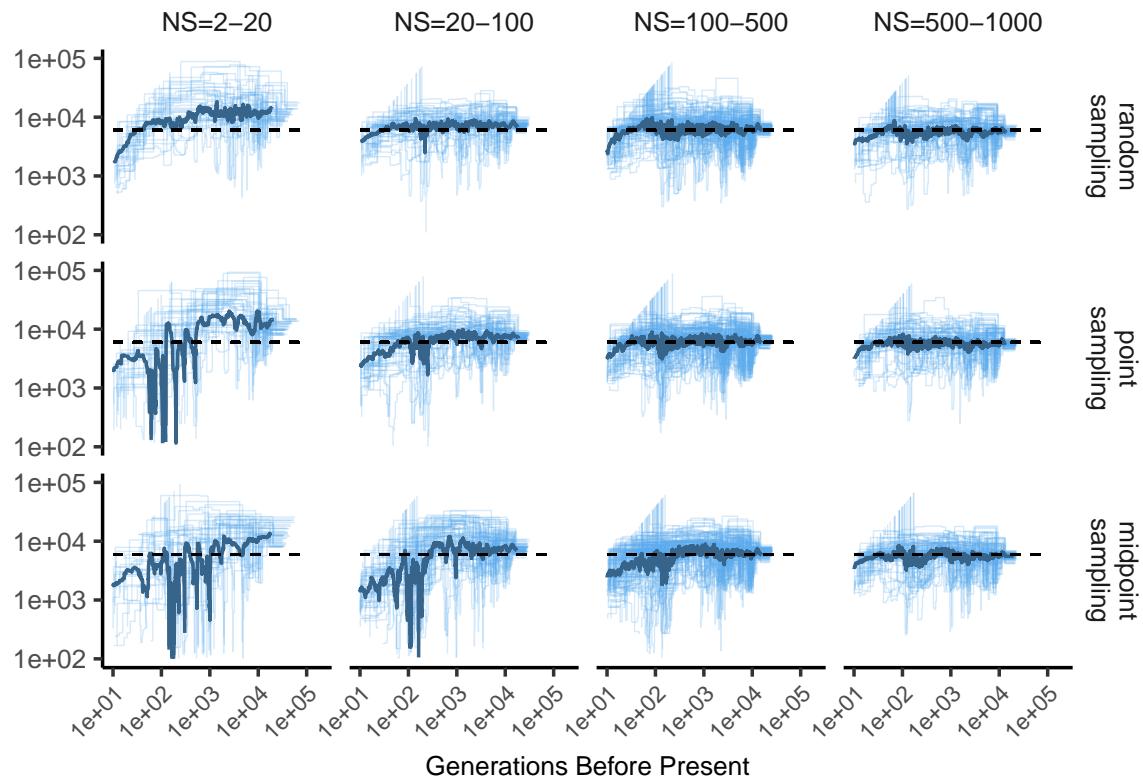


Figure 4 Inferred demographic histories for spatial SLiM simulations from Stairwayplot, by sampling scheme and neighborhood size (NS) range. The thick line is a rolling mean and thin lines are individual model fits. Dashed horizontal lines are the average N_e across random-mating SLiM models estimated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

448 **GWAS**

449 To ask what confounding effects spatial genetic variation might have on genome-wide associa-
450 tion studies we performed GWAS on our simulations using phenotypes that were determined
451 solely by the environment – so, any SNP showing statistically significant correlation with
452 phenotype is a false positive. As expected, spatial autocorrelation in the environment causes
453 spurious associations across much of the genome if no correction for genetic relatedness
454 among samples is performed (Figure 5). This effect is particularly strong for clinal and corner
455 environments, for which the lowest dispersal levels cause over 60% of SNPs in the sample
456 to return significant associations. Patchy environmental distributions, which is less strongly
457 spatially correlated (Figure 5A), cause fewer false positives overall but still produce spurious
458 associations at roughly 10% of sites at the lowest neighborhood sizes. Notably, no simulations
459 with nonspatial environments returned more than one significant association (Figure 5C),
460 demonstrating that this effect is caused specifically by the interaction of population structure
461 and spatial variation in the environment rather than by population structure itself.

462 The confounding effects of geographic structure are well-known, and it is common practice
463 to control for this by including principal components (PCs) as covariates to control for these
464 effects. This mostly works in our simulations – after doing this, the vast majority of SNPs
465 no longer surpass a 5% FDR significance threshold. However, a substantial number of SNPs
466 – up to 1.5% of SNPs – still surpass this threshold (and thus would be false positives in a
467 GWAS), especially under “corner” and “patchy” environmental distributions (Figure 5C).
468 At neighborhood sizes larger than 500, up to 0.31% of SNPs were significant for corner
469 and clinal environments. Given an average of 132,000 SNPs across simulations after MAF
470 filtering, this translates to up to 382 false-positive associations. In most cases the p values for
471 these associations were significant after FDR correction but would not pass the threshold for
472 significance under the more conservative Bonferroni correction.

473 Clinal environments cause an interesting pattern in false positives after PC correction:
474 at low neighborhood sizes the correction removes nearly all significant associations, but at
475 neighborhood sizes above ≈ 250 the proportion of significant SNPs increases to up to 0.4%
476 (Figure 5). This may be due to a loss of descriptive power of the PCs – as neighborhood size
477 increases, the total proportion of variance explained by the first 10 PC axes declines from
478 roughly 10% to 4% (Figure 5B). Essentially, PCA seems unable to effectively summarize the

479 weak population structure present in large-neighborhood simulations, but these populations
480 continue to have enough spatial structure to create significant correlations between genotypes
481 and the environment. A similar process can also be seen in the corner phenotype distribution,
482 in which the count of significant SNPs initially declines as neighborhood size increases and
483 then increases at approximately the point at which the proportion of variance explained by
484 PCA approaches its minimum.

485 Figure 5D shows quantile-quantile plots that show the degree of genome-wide inflation of
486 test statistics in PC-corrected GWAS across all simulations and environmental distributions.
487 For clinal environments, $-\log_{10}(p)$ values are most inflated when neighborhood sizes are
488 large, consistent with the pattern observed in the count of significant associations after
489 PC regression. In contrast corner and patchy environments cause the greatest inflation
490 in $-\log_{10}(p)$ at neighborhood sizes < 100 , which likely reflects the inability of PCA to
491 account for fine-scale structure caused by very limited dispersal. Finally, we observed that PC
492 regression appears to overfit to some degree for all phenotype distributions, visible in Figure
493 5D as points falling below the 1:1 line.

494 (*Add a few manhattan plots to the supplement and refer to them somewhere.*)

495 Discussion

496 In this study, we have used efficient forward time population genetic simulations to describe
497 the myriad influence of continuous geography on genetic variation. In particular, we examine
498 how three main types of downstream empirical inference are affected by unmodeled spatial
499 population structure – 1) population genetic summary statistics, 2) inference of population
500 size history, and 3) genome-wide association studies (GWAS). As discussed above, space often
501 matters (and sometimes dramatically), both because of how samples are arranged in space,
502 and because of the inherent patterns of relatedness established by geography.

503 Effects of Dispersal

504 Limited dispersal inflates effective population size, creates correlations between genetic and
505 spatial distances, and introduces strong distortions in the site frequency spectrum that are
506 reflected in a positive Tajima's D (Figure 3). At the lowest dispersal distances, this can increase
507 genetic diversity threefold relative to random-mating expectations. These effects are strongest
508 when neighborhood sizes are below 100, but in combination with the effects of nonrandom

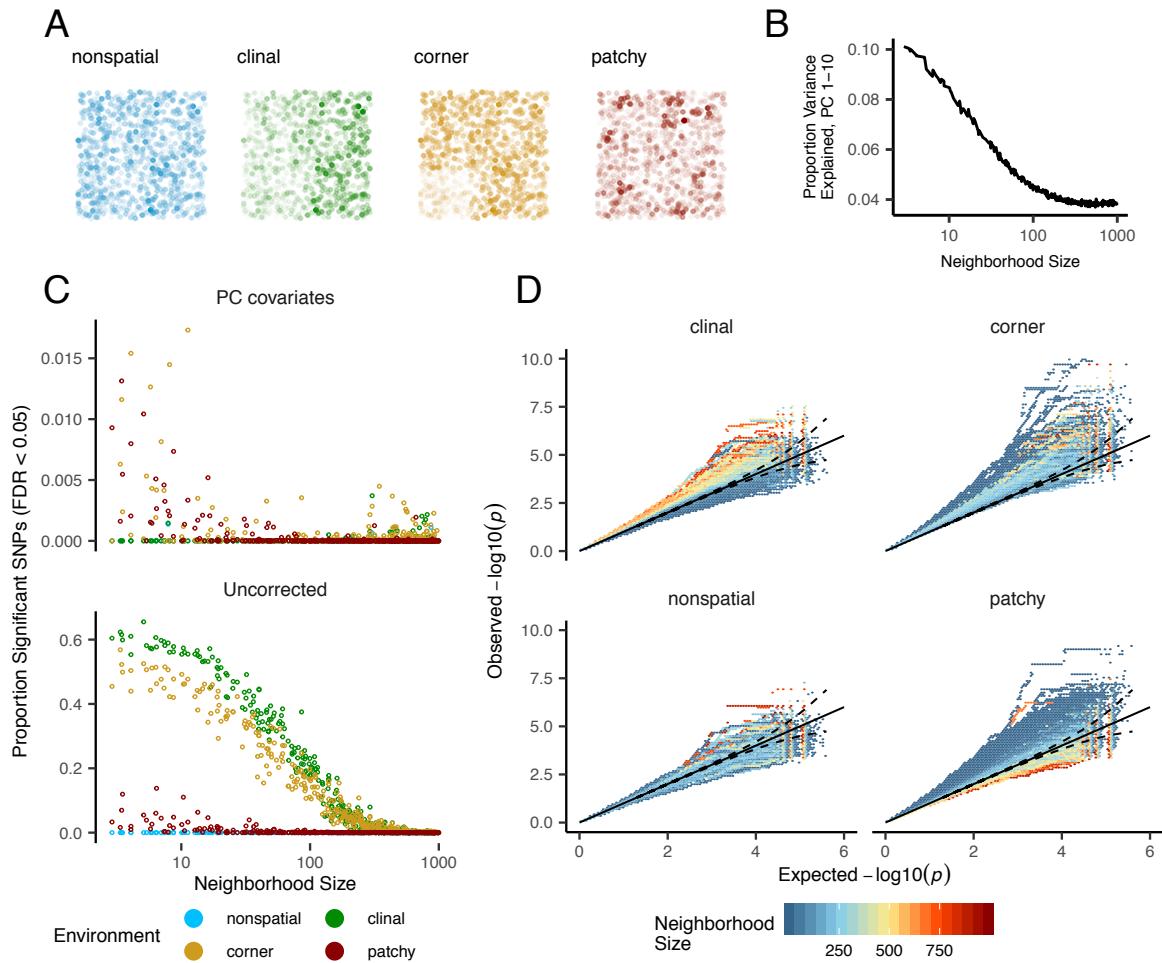


Figure 5 Impacts of spatially varying environments and isolation by distance on linear regression GWAS. Simulated quantitative phenotypes are determined only by an individual's location and the spatial distribution of environmental factors. In A we show the phenotypes and locations of sampled individuals 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. (suggestion: (a) log-scale the y axes in (C) and add another y-axis on the right showing (rough number of human genome SNPs above 5% times this percentage) so we can see absolute numbers)

509 sampling they can persist up to neighborhood sizes of at least 1000 (e.g., inflation in Tajima's
510 D and observed heterozygosity under midpoint sampling). Under random sampling the
511 general pattern is similar to expectations of the original analytic model of Wright (1943), which
512 predicts that populations with neighborhood sizes under 100 will differ substantially from
513 random mating, while those above 10,000 will be nearly indistinguishable from panmixia.

514 The patterns observed in sequence data reflect the effects of space on the underlying
515 genealogy. Nearby individuals coalesce rapidly under limited dispersal and so are connected
516 by short branch lengths, while distant individuals take much longer to coalesce than they
517 would under random mating. Mutation and recombination events in our simulation both
518 occur at a constant rate along branches of the genealogy, so the genetic distance and number
519 of recombination events separating sampled individuals simply gives a noisy picture of the
520 genealogies connecting them. Tip branches (i.e., branches subtending only one individual) are
521 then relatively short, and branches in the middle of the genealogy connecting local groups of
522 individuals relatively long. These patterns then create the biases we observed in the SFS – the
523 lowest frequency bins are deflated by the short branch lengths connecting nearby individuals,
524 while mid-frequency bins are inflated by the long branches connecting local groups.

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

539 **Effects of Sampling**

540 One of the most important differences between random mating and spatial models is the
541 effect of sampling: in a randomly mating population the spatial distribution of sampling effort
542 has no effect on estimates of genetic variation, but when dispersal is limited sampling strategy
543 can compound spatial patterns in the underlying genealogy and create pervasive impacts
544 on all downstream genetic analyses. In most species, the difficulty of traveling through all
545 parts of a species range and the inefficiency of collecting single individuals at each sampling
546 site means that most studies follow something closest to the “point” sampling strategy we
547 simulated, in which multiple individuals are sampled from nearby points on the landscape.
548 For example, in ornithology a sample of 10 individuals per species per locality is a common
549 target when collecting for natural history museums. In classical studies of *Drosophila* variation
550 the situation is considerably worse, in which a single orchard might be extensively sampled.

551 When sampling is clustered at points on a landscape and dispersal is limited, the sampled
552 individuals will be more closely related than a random set of individuals. Average coalescence
553 times of individuals collected at a locality will then be more recent and branch lengths shorter
554 than expected by analyses assuming random mating. This leads to fewer mutations and
555 recombination events occurring since their last common ancestor, causing a random set of
556 individuals to share longer average IBS tracts and have fewer nucleotide differences. For some
557 data summaries, such as Tajima’s D , Watterson’s Θ , or the correlation coefficient between
558 spatial distance and the count of long haplotype blocks, this can result in large differences in
559 estimates between random and point sampling (Figure 3). Inferring underlying demographic
560 parameters from these summary statistics – unless the nature of the sampling is somehow
561 taken into account – will be subject to bias if sampling is not random across the landscape.

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

sample. This increase in observed heterozygosity appears to reflect the effects of range edges, which are a fundamental facet of spatial genetic variation. 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).

In summary, empirical researchers should collect individuals from across as much of the species' range as practical, choosing samples at a variety of spatial scales. Many summary statistics are designed for well-mixed populations, and so provide different insights into genetic variation when applied to different subsets of the population. Applied to a cluster of samples, 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 obtained from a wider distribution of samples. Comparing the results of analyses conducted on all individuals versus those limited to single individuals per locality can provide an informative contrast.

589 **Demography**

Classical results in population genetics collapse many aspects of life history into a single compound parameter, effective population size (N_e). Inferring effective population size back through time is now a common goal of population genomic analyses and an important step in establishing baseline expectations of genetic variation, e.g., 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 spectrum, Stairwayplot (Liu and Fu 2015), is relatively robust to variation in dispersal distance when sampling is random and neighborhood size is over 20. However, non-random sampling, and particularly midpoint sampling, causes the method to infer inflated estimates of past population sizes and a series of recent bottlenecks (Figure 4). All sampling strategies lead to inflated ancient and deflated recent N_e when neighborhood

600 sizes were less than 20.

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

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

613 **GWAS**

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

623 Stepping back from specific methods, the general goal is to identify alleles that are more
624 correlated with the phenotype of interest than one would expect due to chance. Both phe-
625 notype and genotype are often spatially autocorrelated (for different reasons), which makes
626 spurious correlations a major obstacle, both by creating false positives and reducing power to
627 identify truly causal alleles (Price *et al.* 2006; Yu *et al.* 2005; Young *et al.* 2018; Mathieson and
628 McVean 2012; Kang *et al.* 2008, 2010; Bulik-Sullivan *et al.* 2015). It may be that the success of
629 quantitative genetics in agriculture is in part because crop and animal breeding operations

630 can be conducted so that the environment is identical (or nearly so) or randomized across
631 populations. However, most natural populations are structured by a combination of limited
632 dispersal and geographic barriers, and nearly all environments vary over space.

633 Incorporating PC positions as covariates in the analysis (Price *et al.* 2006) is designed to
634 address this difficulty by regressing out a baseline level of “average” differentiation. In
635 essence, a PC-corrected GWAS asks “what regions of the genome are more associated with
636 this phenotype than the average genome-wide association observed across populations?” In
637 our simulations, we observed that this procedure can fail under a variety of circumstances. If
638 dispersal is limited and environmental variation is clustered in space (i.e., corner or patchy
639 distributions in our simulations), PCA positions fail to capture the fine-scale spatial structure
640 required to remove all signals of association. Conversely, as dispersal is increased, PCA loses
641 power to describe population structure before spatial mixing breaks down the relationship
642 between genotype and the environment. These effects were observed with all spatially
643 correlated environmental patterns, but were particularly pronounced if environmental effects
644 are concentrated in one region, as was also found by Mathieson and McVean (2012). (*need*
645 *to highlight a statement like this more*) As a result we can expect to see several thousand weak
646 false-positive associations in a PC-corrected GWAS conducted on a human-sized genome in
647 species with neighborhood sizes up to at least 1000.

648 This suggests some caveats for interpretation of GWAS results. Very few of the spurious
649 associations we identified would be significant at a conservative Bonferroni-adjusted *p*-value
650 cutoff, suggesting that most of the very strong signals of association observed in studies
651 of mono- or oligogenic traits are robust to these concerns. Further, the most dramatic *p*-
652 value inflations occurred at neighborhood sizes below 100 – smaller than most recent human
653 populations (but see below for further discussion of empirical cases). However, as recently
654 identified in studies of genotype associations for human height in Europe (Berg *et al.* 2018;
655 Sohail *et al.* 2018), PC regression GWAS in modern human populations does leave residual
656 signal of population structure in large-scale GWAS of polygenic traits. (*Is there any indication*
657 *that mixed models do better about this? (Where's the “for precisely this reason” come from?)*) Indeed,
658 studies in strongly structured species like *Arabidopsis* have long relied on more sophisticated
659 mixed model approaches to correcting for population structure for precisely this reason
660 (Aranzana *et al.* 2005; Sasaki *et al.* 2015).

661 A second point that has received less attention in the literature is the issue of overfitting in
662 GWAS. If a truly causal allele segregates at different frequencies in different populations, then
663 correcting for population structure in a regression analysis will result in an underestimate of its
664 effect size. Though our simulations had no causal alleles, we observed some evidence of this
665 effect in the distribution of *p*-values across the genome (Figure 5D): after PC regression many
666 analyses resulted in *p*-values falling below their expected values from a uniform distribution.
667 This is consistent with a recent empirical study of heritability in human height and body mass
668 index, which found that increasing the number of PC axes used as covariates caused the total
669 proportion of variance explained by SNPs to decline from ≈ 0.8 to ≈ 0.75 (Wainschtein *et al.*
670 2019).

671 In summary, spatial covariation in population structure and the environment confound the
672 interpretation of GWAS *p*-values, and correction using principal components is insufficient to
673 fully separate these signals for polygenic traits under a variety of environmental and popu-
674 lation parameter regimes. Other GWAS methods may be less sensitive to this confounding,
675 but there is no obvious reason that this should be so. One approach to estimating the degree
676 of bias in GWAS caused by population structure is LD score regression (Bulik-Sullivan *et al.*
677 2015). Though this approach appears to work well in practice, its interpretation is not always
678 straightforward and it is likely biased by the presence of linked selection (Berg *et al.* 2018).
679 We suggest a straightforward alternative for species in which the primary axes of population
680 differentiation is space (note this is likely not the case for many modern human populations):
681 run a GWAS with spatial coordinates as phenotypes and check for *p*-value inflation or signifi-
682 cant associations. If significant associations with sample locality are observed after correcting
683 for population structure through PC regression or a kinship matrix, the method is sensitive
684 to false positives induced by spatial structure. This is essentially the approach taken in our
685 “clinal” model (though we add normally distributed noise to our phenotypes). Of course, it is
686 possible that genotypes indirectly affect individual locations by adjusting organismal fitness
687 and thus habitat selection across spatially varying environments, but we believe that this
688 hypothesis should be tested against a null of stratification bias inflation rather than accepted
689 as true based on GWAS results. (*what do you think of the second half here?*) (*love it*)

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)

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

691 How strong is the spatial patterning of genetic variation in extant species? To get an idea of this,
692 we gathered estimates of neighborhood size from a range of organisms (Table 1). Although this
693 sample is almost certainly biased towards small-neighborhood species (because few studies
694 have quantified neighborhood size in species with very high dispersal or population density),
695 we find that neighborhood sizes in the range we simulated are fairly common across a range of
696 taxa. At the extreme low end of empirical neighborhood size estimates we see some flowering
697 plants, large mammals, and colonial insects like ants. Species such as this have neighborhood
698 size estimates small enough that spatial processes are likely to strongly influence inference.
699 These include some human populations such as the Gainj- and Kalam-speaking people of
700 Papua New Guinea, for whom the estimated neighborhood sizes in Rousset (1997) range
701 from 40 to 410 depending on the method of estimation. Many more species occur in a middle
702 range of neighborhood sizes between 100 and 1000 – a range in which spatial processes play a
703 minor role in our analyses under random spatial sampling but are important when sampling
704 of individuals in space is clustered. Last, many species likely have neighborhood sizes much
705 larger than we simulated, including modern humans in northeastern Europe (Ringbauer *et al.*
706 2017). If these species' ranges and population densities have been constant for a long period of
707 time (unlike humans), demographic inference methods that assume well-mixed populations
708 may show minimal bias from spatial effects.

709 ***Future Directions and Limitations***

710 As we have shown, a large number of population genetic summary statistics contain informa-
711 tion about spatial population processes. We imagine that combinations of such summaries
712 might be sufficient for the construction of supervised machine learning regressors (e.g.,
713 Schrider and Kern (2018)) for the accurate estimation of dispersal from genetic data. Indeed,
714 Ashander *et al.* (2018) found that inverse interpolation on a vector of summary statistics
715 provided a powerful method of estimating dispersal distances. Expanding this approach to
716 include the haplotype-based summary statistics studied here and applying machine learning
717 regressors built for general inference of nonlinear relationships from high-dimensional data
718 may allow precise estimation of spatial parameters under a range of complex models.

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

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

734 Though our continuous space simulation allows incorporation of realistic demographic and
735 spatial processes and is much faster than previous individual-based models, it is inevitably
736 limited by the computational burden of tracking tens or hundreds of thousands of individuals
737 in every generation. In particular, computations required for mate selection and spatial
738 competition caled as the number of individuals within a three- σ radius increases. The reverse-
739 time model of continuous space evolution described by Barton *et al.* (2010) and implemented by
740 Kelleher *et al.* (2014) allows exploration of parameter regimes with population and landscape
741 sizes more directly comparable to empirical cases like humans. Alternatively, implementation
742 of parallelized simulations may allow progress with forwards-time simulations.

743 A mixed approach can combine forward- and reverse-time models, as recently done for
744 a continuous-space Wright-Fisher model by Lotterhos (2019) and a simulation with linked
745 selection by Buffalo and Coop (2019). This allows us to generate short runs of large, realistic
746 simulations in forward time in SLiM (Haller and Messer 2019) and then “finish” the simula-
747 tions as a coalescent simulation in msprime (Kelleher *et al.* 2016), combining the two using an
748 operation called “recapitation” (Haller *et al.* 2019) As we have seen in this work, a nontrivial
749 step in this approach is scaling the variance in reproductive output and generation time across
750 forward- and reverse-time methods. Further development of our understanding of how to

751 merge forward- and reverse-time models is a promising avenue for future research that will
752 be necessary for scaling continuous-space simulations to millions or billions of individuals.

753 Finally, we believe that the difficulties in correcting for population structure in continuous
754 populations using principal components analysis or similar decompositions is a difficult
755 issue, well worth considering on its own. How can we best avoid spurious correlations while
756 correlating genetic and phenotypic variation without underpowering the methods? Perhaps
757 optimistically, we posit that process-driven descriptions of ancestry and/or more generalized
758 unsupervised methods may be able to better account for carry out this task.

759 **Data Availability**

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

761 **Acknowledgements**

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

765 **Literature Cited**

766 Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, *et al.*, 2017 Deconstruct-
767 ing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics
768 **13**: 1–27.

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

771 Antolin, M. F., B. V. Horne, M. D. Berger, Jr., A. K. Holloway, J. L. Roach, *et al.*, 2001 Effec-
772 tive population size and genetic structure of a piute ground squirrel (*Spermophilus mollis*)
773 population. Canadian Journal of Zoology **79**: 26–34.

774 Aranzana, M. J., S. Kim, K. Zhao, E. Bakker, M. Horton, *et al.*, 2005 Genome-wide association
775 mapping in arabidopsis identifies previously known flowering time and pathogen resistance
776 genes. PLoS genetics **1**: e60; e60–e60.

- 777 Ashander, J., P. Ralph, E. McCartney-Melstad, and H. B. Shaffer, 2018 Demographic inference
778 in a spatially-explicit ecological model from genomic data: a proof of concept for the mojave
779 desert tortoise. *bioRxiv* .
- 780 Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, *et al.*, 2016 The great
781 migration and african-american genomic diversity. *PLOS Genetics* **12**: 1–27.
- 782 Barton, N. H., F. Depaulis, and A. M. Etheridge, 2002 Neutral evolution in spatially continuous
783 populations. *Theoretical Population Biology* **61**: 31–48.
- 784 Barton, N. H., J. Kelleher, and A. M. Etheridge, 2010 A new model for extinction and recolo-
785 nization in two dimensions: Quantifying phylogeography. *Evolution* **64**: 2701–2715.
- 786 Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing
787 under dependency. *The Annals of Statistics* **29**: 1165–1188.
- 788 Berg, E. E. and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. *Evolution*
789 **49**: 110–120.
- 790 Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, *et al.*, 2018
791 Reduced signal for polygenic adaptation of height in uk biobank. *bioRxiv* .
- 792 Buffalo, V. and G. Coop, 2019 The linked selection signature of rapid adaptation in temporal
793 genomic data. *bioRxiv* .
- 794 Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 Ld score regression
795 distinguishes confounding from polygenicity in genome-wide association studies. *Nature*
796 *Genetics* **47**: 291 EP –.
- 797 Campbell, D. R. and J. L. Dooley, 1992 The spatial scale of genetic differentiation in a
798 hummingbird-pollinated plant: Comparison with models of isolation by distance. *The*
799 *American Naturalist* **139**: 735–748.
- 800 Chapman, N. H. and E. A. Thompson, 2002 The effect of population history on the lengths of
801 ancestral chromosome segments. *Genetics* **162**: 449–458.
- 802 Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, *et al.*, 1993 Direct
803 and indirect estimates of neighborhood and effective population size in a tropical palm,
804 *astrocaryum mexicanum*. *Evolution* **47**: 75–87.
- 805 Epperson, B., 2003 *Geographical Genetics*. Monographs in Population Biology, Princeton Uni-
806 versity Press.
- 807 Felsenstein, J., 1975 A pain in the torus: Some difficulties with models of isolation by distance.

- 808 The American Naturalist **109**: 359–368.
- 809 Fisher, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh **52**: 399–433.
- 810
- 811 Fox, J. and S. Weisberg, 2011 *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition.
- 812
- 813 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in north american drosophila melanogaster show signatures of soft sweeps. PLOS Genetics **11**: 1–32.
- 814
- 815
- 816 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Molecular Ecology Resources **19**: 552–566.
- 817
- 818
- 819 Haller, B. C. and P. W. Messer, 2019 SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model .
- 820
- 821 Harris, K. and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared haplotype lengths. PLOS Genetics **9**: 1–20.
- 822
- 823 Huillet, T. and M. Möhle, 2011 On the extended Moran model and its relation to coalescents with multiple collisions. Theoretical Population Biology pp. –.
- 824
- 825 Jay, F., P. Sjödin, M. Jakobsson, and M. G. Blum, 2012 Anisotropic Isolation by Distance: The Main Orientations of Human Genetic Differentiation. Molecular Biology and Evolution **30**: 513–525.
- 826
- 827
- 828 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component model to account for sample structure in genome-wide association studies. Nature Genetics **42**: 348 EP –.
- 829
- 830
- 831 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, *et al.*, 2008 Efficient control of population structure in model organism association mapping. Genetics **178**: 1709–1723.
- 832
- 833 Kelleher, J., A. Etheridge, and N. Barton, 2014 Coalescent simulation in continuous space: Algorithms for large neighbourhood size. Theoretical Population Biology **95**: 13 – 23.
- 834
- 835 Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and genealogical analysis for large sample sizes. PLoS Comput Biol **12**: 1–22.
- 836
- 837 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for fast population genetics simulation. PLOS Computational Biology **14**: 1–21.
- 838

- 839 Kerster, H. W., 1964 Neighborhood size in the rusty lizard, *sceloporus olivaceus*. *Evolution* **18**:
840 445–457.
- 841 Khera, A. V., M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, *et al.*, 2018 Genome-wide poly-
842 genic scores for common diseases identify individuals with risk equivalent to monogenic
843 mutations. *Nature Genetics* **50**: 1219–1224.
- 844 Kingman, J., 1982 The coalescent. *Stochastic Processes and their Applications* **13**: 235 – 248.
- 845 Liu, X. and Y.-X. Fu, 2015 Exploring population size changes using snp frequency spectra.
846 *Nature Genetics* **47**: 555 EP –.
- 847 Lotterhos, K. E., 2019 The effect of neutral recombination variation on genome scans for
848 selection. *G3: Genes, Genomes, Genetics* .
- 849 Lundgren, E. and P. L. Ralph, 2018 Are populations like a circuit? The relationship between
850 isolation by distance and isolation by resistance. *bioRxiv* .
- 851 Maruyama, T., 1972 Rate of decrease of genetic variability in a two-dimensional continuous
852 population of finite size. *Genetics* **70**: 639–651.
- 853 Mathieson, I. and G. McVean, 2012 Differential confounding of rare and common variants in
854 spatially structured populations. *Nature Genetics* **44**: 243 EP –.
- 855 Miles, A. and N. Harding, 2017 *cgh/scikit-allel*: v1.1.8.
- 856 Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, *et al.*, 2013 Estimation of effec-
857 tive population size in continuously distributed populations: there goes the neighborhood.
858 *Heredity* **111**: 189 EP –.
- 859 Pamilo, P., 1983 Genetic differentiation within subdivided populations of *formica* ants. *Evolu-
860 tion* **37**: 1010–1022.
- 861 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLOS
862 Genetics* **2**: 1–20.
- 863 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, *et al.*, 2006 Principal
864 components analysis corrects for stratification in genome-wide association studies. *Nature
865 Genetics* **38**: 904 EP –.
- 866 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using
867 multilocus genotype data. *Genetics* **155**: 945–959.
- 868 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.*, 2007 Plink: A tool set for
869 whole-genome association and population-based linkage analyses. *The American Journal*

- 870 of Human Genetics **81**: 559 – 575.
- 871 R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for
872 Statistical Computing, Vienna, Austria.
- 873 Ralph, P. and G. Coop, 2010 Parallel adaptation: One or many waves of advance of an
874 advantageous allele? *Genetics* **186**: 647–668.
- 875 Ralph, P. and G. Coop, 2013 The geography of recent genetic ancestry across Europe. *PLoS
876 Biol* **11**: e1001555.
- 877 Ringbauer, H., G. Coop, and N. H. Barton, 2017 Inferring recent demography from isolation
878 by distance of long shared sequence blocks. *Genetics* **205**: 1335–1351.
- 879 Robledo-Arnuncio, J. J. and F. Rousset, 2010 Isolation by distance in a continuous population
880 under stochastic demographic fluctuations. *Journal of Evolutionary Biology* **23**: 53–71.
- 881 Rousset, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under
882 isolation by distance. *Genetics* **145**: 1219–1228.
- 883 Rousset, F. and R. Leblois, 2011 Likelihood-based inferences under isolation by distance:
884 Two-dimensional habitats and confidence intervals. *Molecular Biology and Evolution* **29**:
885 957–973.
- 886 Sasaki, E., P. Zhang, S. Atwell, D. Meng, and M. Nordborg, 2015 "missing" g x e variation
887 controls flowering time in *arabidopsis thaliana*. *PLOS Genetics* **11**: 1–18.
- 888 Schemske, D. W. and P. Bierzychudek, ??? Spatial differentiation for flower color in the desert
889 annual *linanthus parryae*: Was wright right? *Evolution* **61**: 2528–2543.
- 890 Schrider, D. R. and A. D. Kern, 2018 Supervised machine learning for population genetics: A
891 new paradigm. *Trends in Genetics* **34**: 301 – 312.
- 892 Sharbel, T. F., B. Haubold, and T. Mitchell-Olds, 2000 Genetic isolation by distance in *ara-*
893 *bidopsis thaliana*: biogeography and postglacial colonization of europe. *Molecular Ecology*
894 **9**: 2109–2118.
- 895 Shirk, A. J. and S. A. Cushman, 2014 Spatially-explicit estimation of wright's neighborhood
896 size in continuous populations. *Frontiers in Ecology and Evolution* **2**: 62.
- 897 Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, *et al.*, 2018 Signals of polygenic
898 adaptation on height have been overestimated due to uncorrected population structure in
899 genome-wide association studies. *bioRxiv* .
- 900 Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA poly-

- 901 morphism. *Genetics* **123**: 585–595.
- 902 Visscher, P. M. and M. E. Goddard, 2019 From r.a. fisher's 1918 paper to gwas a century later.
903 *Genetics* **211**: 1125–1130.
- 904 Wahlund, S., 1928 Zusammensetzung von populationen und korrelationserscheinungen vom
905 standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**: 65–106.
- 906 Wainschtein, P., D. P. Jain, L. Yengo, Z. Zheng, , *et al.*, 2019 Recovery of trait heritability from
907 whole genome sequence data. *bioRxiv* .
- 908 Wakeley, J., 2005 *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village,
909 CO.
- 910 Wakeley, J. and T. Takahashi, 2003 Gene genealogies when the sample size exceeds the effective
911 size of the population. *Mol Biol Evol* **20**: 208–213.
- 912 Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- 913 Wilke, C. O., 2019 *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package
914 version 0.9.4.
- 915 Wilkins, J. F., 2004 A separation-of-timescales approach to the coalescent in a continuous
916 population. *Genetics* **168**: 2227–2244.
- 917 Wright, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97.
- 918 Wright, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- 919 Wright, S., 1946 Isolation by distance under diverse systems of mating. *Genetics* **31**: 336.
- 920 Young, A. I., M. L. Frigge, D. F. Gudbjartsson, G. Thorleifsson, G. Bjornsdottir, *et al.*, 2018
921 Relatedness disequilibrium regression estimates heritability without environmental bias.
922 *Nature Genetics* **50**: 1304–1310.
- 923 Young, H. J., 1988 Neighborhood size in a beetle pollinated tropical aroid: effects of low
924 density and asynchronous flowering. *Oecologia* **76**: 461–466.
- 925 Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, *et al.*, 2005 A unified mixed-model
926 method for association mapping that accounts for multiple levels of relatedness. *Nature
927 Genetics* **38**: 203 EP –.
- 928 Zähle, I., J. T. Cox, and R. Durrett, 2005 The stepping stone model. II. Genealogies and the
929 infinite sites model. *Ann. Appl. Probab.* **15**: 671–699.

930 **Supplementary Figures and Tables**

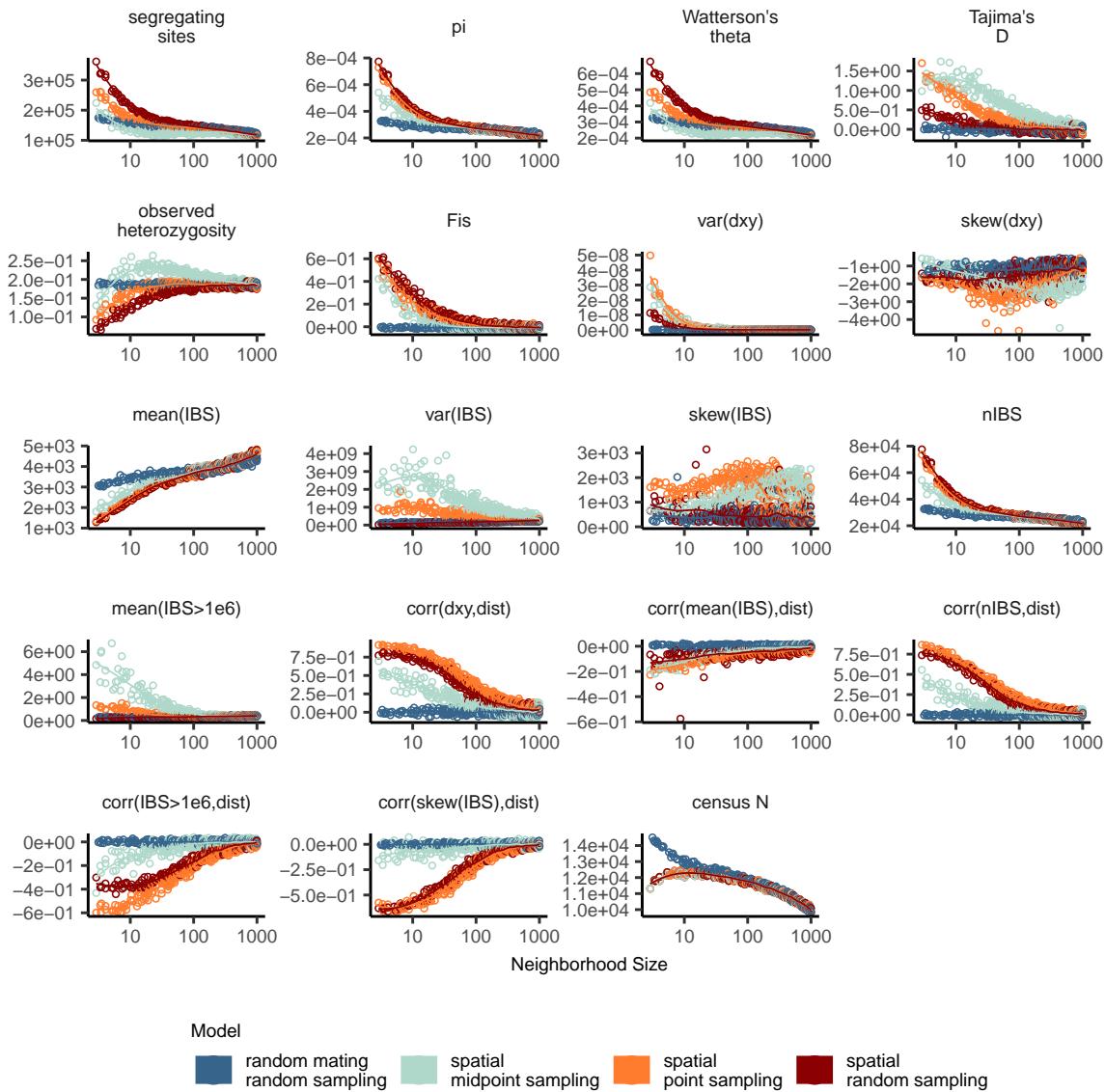


Figure S1 Change in summary statistics by neighborhood size and sampling scheme calculated from simulated sequence data of 60 individuals.

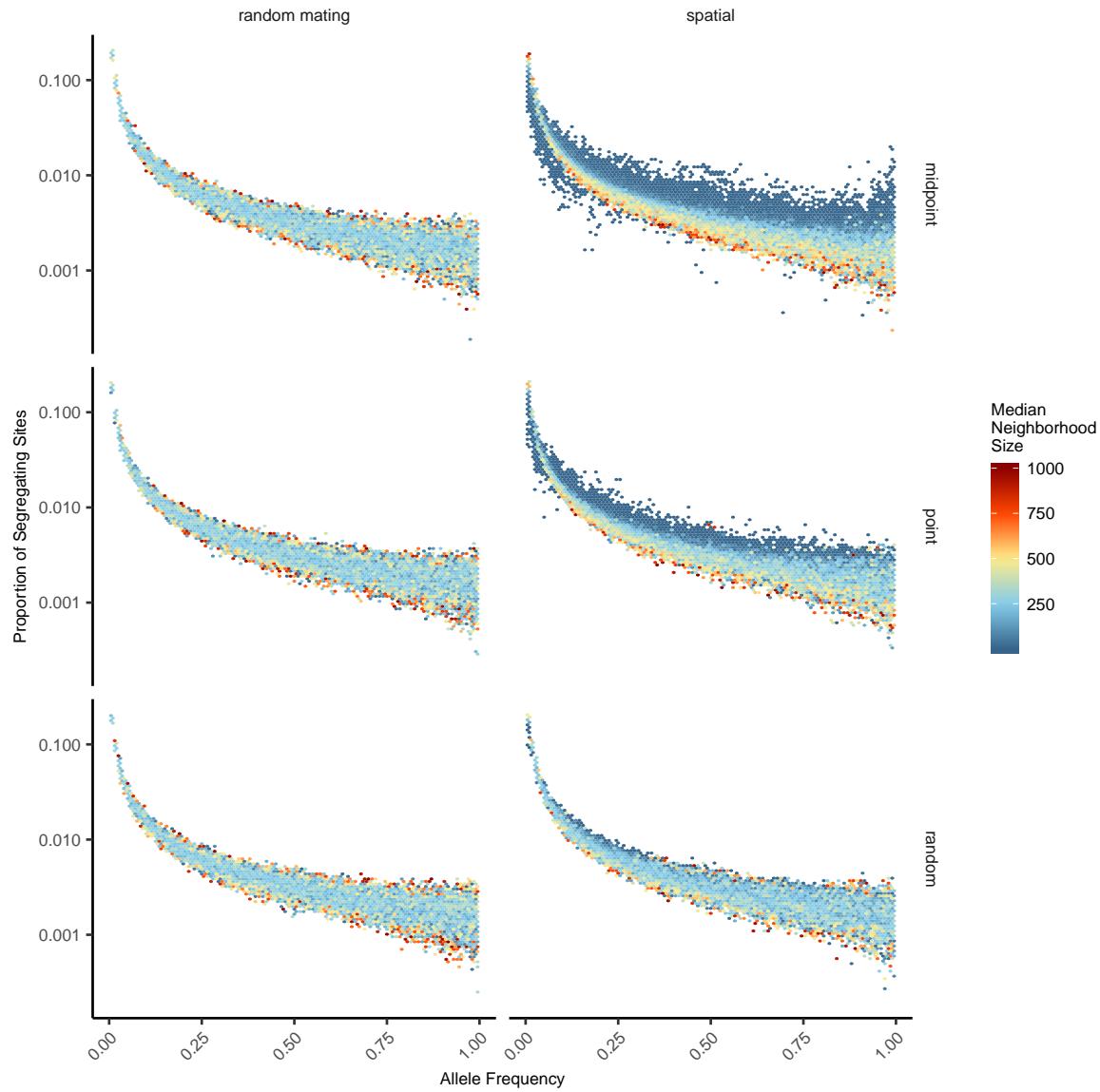


Figure S2 Site frequency spectra for random mating and spatial SLiM models under all sampling schemes.

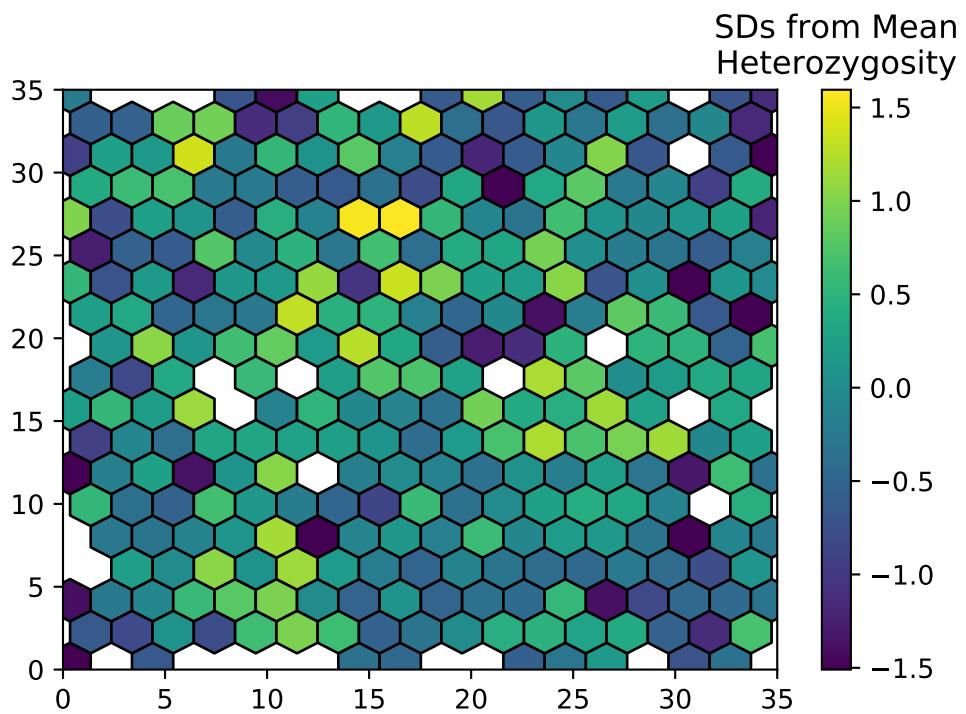


Figure S3 Normalized mean observed heterozygosity by location across 200 randomly-sampled individuals

Table 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 > 1×10^6 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 $\log_{10}(spatial\ distance)$

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