

PREPRINT

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

C.J. Battey<sup>\*,†</sup>, Peter L. Ralph<sup>\*,†</sup> and Andrew D. Kern<sup>\*,†</sup>

\*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 clustered samples from populations that are structured clinally over space. 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 most 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. The combination of low dispersal and clustered sampling causes demographic inference from the site frequency spectrum to infer more turbulent demographic histories, but averaged results across multiple simulations were surprisingly robust to isolation by distance. We also show that the combination of spatially autocorrelated environments and limited dispersal causes genome-wide association studies to identify 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

25

## 26 **Introduction**

28 The inescapable reality that biological organisms live, move, and reproduce in continuous  
29 geography is usually omitted from population genetic models. However, mates tend to live  
30 near to one another and to their offspring, leading to a positive correlation between genetic

Manuscript compiled: Monday 3<sup>rd</sup> June, 2019

<sup>1</sup>301 Pacific Hall, University of Oregon Dept. Biology, Institute for Ecology and Evolution. cbattey2@uoregon.edu.

<sup>†</sup>these authors co-supervised this project

31 differentiation and geographic distance. This pattern of “isolation by distance” (Wright 1943)  
32 is one of the most widely replicated empirical findings in population genetics (Aguillon *et al.*  
33 2017; Jay *et al.* 2012; Sharbel *et al.* 2000). Despite a long history of analytical work describing  
34 the genetics of populations distributed across continuous geography (e.g., Wright (1943);  
35 Rousset (1997); Barton *et al.* (2002, 2010); Ringbauer *et al.* (2017); Robledo-Arnuncio and  
36 Rousset (2010)), much modern work still describes geographic structure as a set of discrete  
37 populations connected by migration (e.g., Wright 1931; Epperson 2003; Rousset and Leblois  
38 2011; Shirk and Cushman 2014; Lundgren and Ralph 2018). For this reason, most population  
39 genetics statistics are interpreted with reference to discrete, well-mixed populations, and most  
40 empirical papers analyze variation within clusters of genetic variation inferred by programs  
41 like *STRUCTURE* (Pritchard *et al.* 2000) with methods that assume these are randomly mating  
42 units.

43 The assumption that populations are “well-mixed” has important implications for down-  
44 stream inference of selection and demography. Methods based on the coalescent (Kingman  
45 1982; Wakeley 2009) assume that the sampled individuals are a random draw from a well-  
46 mixed population that is much larger than the sample (Wakeley and Takahashi 2003). The  
47 key assumption is that the individuals of each generation are *exchangeable*, so that there is no  
48 correlation between the fate or fecundity of a parent and that of their offspring (Huillet and  
49 Möhle 2011). If dispersal or mate selection is limited by geographic proximity, this assump-  
50 tion can be violated in many ways. For instance, if mean viability or fecundity is spatially  
51 autocorrelated, then limited geographic dispersal will lead to parent–offspring correlations.  
52 Furthermore, nearby individuals will be more closely related than an average random pair, so  
53 drawing multiple samples from the same area of the landscape will represent a biased sample  
54 of the genetic variation present in the whole population (Städler *et al.* 2009).

55 Two areas in which spatial structure may be particularly important are demographic  
56 inference and genome-wide association studies (GWAS). Previous work has found that  
57 discrete population structure can create false signatures of population bottlenecks when  
58 attempting to infer demographic histories from microsatellite variation (Chikhi *et al.* 2010),  
59 statistics summarizing the site frequency spectrum (SFS) (Ptak and Przeworski 2002; Städler  
60 *et al.* 2009; St. Onge *et al.* 2012), or runs of homozygosity in a single individual (Mazet *et al.*  
61 2015). The increasing availability of whole-genome data has led to the development of many

62 methods that attempt to infer detailed trajectories of population sizes through time based on  
63 a variety of summaries of genetic data (Liu and Fu 2015; Schiffels and Durbin 2014; Sheehan  
64 *et al.* 2013; Terhorst *et al.* 2016). Because all of these methods assume that the populations being  
65 modeled are approximately randomly mating, they are likely affected by spatial biases in the  
66 genealogy of sampled individuals (Wakeley 1999), which may lead to incorrect inference of  
67 population changes over time (Mazet *et al.* 2015). However, previous investigations of these  
68 effects have focused on discrete rather than continuous space models, and the level of isolation  
69 by distance at which inference of population size trajectories become biased by structure is not  
70 well known. Here we test how two methods suitable for use with large samples of individuals  
71 – stairwayplot (Liu and Fu 2015) and SMC++ (Terhorst *et al.* 2016) – perform when applied  
72 to populations evolving in continuous space with varying sampling strategies and levels of  
73 dispersal.

74 Spatial structure is also a major challenge for interpreting the results of genome-wide asso-  
75 ciation studies (GWAS). This is because many phenotypes of interest have strong geographic  
76 differences due to the (nongenetic) influence of environmental or socioeconomic factors,  
77 which can therefore show spurious correlations with spatially patterned allele frequencies  
78 (Bulik-Sullivan *et al.* 2015; Mathieson and McVean 2012). Indeed, two recent studies found  
79 that previous evidence of polygenic selection on human height in Europe was confounded  
80 by subtle population structure (Sohail *et al.* 2018; Berg *et al.* 2018), suggesting that existing  
81 methods to correct for population structure in GWAS are insufficient. However we have little  
82 quantitative idea of the population and environmental parameters that can be expected to  
83 lead to biases in GWAS.

84 Last, some of the most basic tools of population genetics are summary statistics like  $F_{IS}$  and  
85 Tajima's  $D$ , which are often interpreted as reflecting the influence of selection or demography  
86 on sampled populations (Tajima 1989). Statistics like Tajima's  $D$  are essentially summaries  
87 of the site frequency spectrum, which itself reflects variation in branch lengths and tree  
88 structure of the underlying genealogies of sampled individuals. Geographically limited mate  
89 choice distorts the distribution of these genealogies (Maruyama 1972; Wakeley 1999), which  
90 can affect the value of Tajima's  $D$  (Städler *et al.* 2009). Similarly, the distribution of tract  
91 lengths of identity by state among individuals contains information about not only historical  
92 demography (Harris and Nielsen 2013; Ralph and Coop 2013) and selection (Garud *et al.*

93 2015), but also dispersal and mate choice (Ringbauer *et al.* 2017; Baharian *et al.* 2016). We are  
94 particularly keen to examine how such summaries will be affected by models that incorporate  
95 continuous space, both to evaluate the assumptions underlying existing methods and to  
96 identify where the most promising signals of geography lie.

97 To study this, we have implemented an individual-based model in continuous geography  
98 that incorporates overlapping generations, local dispersal of offspring, and density-dependent  
99 survival. We simulate chromosome-scale genomic data in tens of thousands of individuals  
100 from parameter regimes relevant to common subjects of population genetic investigation such  
101 as humans and *Drosophila*, and output the full genealogy and recombination history of all  
102 final-generation individuals. We use these simulations to test how sampling strategy interacts  
103 with geographic population structure to cause systematic variation in population genetic  
104 summary statistics typically analyzed assuming discrete population models. We then examine  
105 how the fine-scale spatial structures occurring under limited dispersal impact demographic  
106 inference from the site frequency spectrum. Last, we examine the impacts of continuous  
107 geography on genome-wide association studies (GWAS) and identify regions of parameter  
108 space under which the results from GWAS may be misleading.

## 109 Materials and Methods

### 110 *Modeling Evolution in Continuous Space*

111 The degree to which genetic relationships are geographically correlated depends on the  
112 chance that two geographically nearby individuals are close relatives – in modern terms, by  
113 the tension between migration (the chance that one is descended from a distant location)  
114 and coalescence (the chance that they share a parent). A key early observation by Wright  
115 (Wright 1946) is that this balance is often nicely summarized by the “neighborhood size”,  
116 defined to be  $N_W = 4\pi\rho\sigma^2$ , where  $\sigma$  is the mean parent–offspring distance and  $\rho$  is population  
117 density. This can be thought of as proportional to the average number of potential mates for  
118 an individual (those within distance  $2\sigma$ ), or the number of potential parents of a randomly  
119 chosen individual. Empirical estimates of neighborhood size vary hugely across species – even  
120 in human populations, estimates range from 40 to over 5,000 depending on the population  
121 and method of estimation (Table 1).

122 The first approach to modeling continuously distributed populations was to endow indi-

123 individuals in a Wright-Fisher model with locations in continuous space. However, since the total  
124 size of the population is constrained, this introduces interactions between arbitrarily distant  
125 individuals, which (aside from being implausible) was shown by Felsenstein (1975) to eventually  
126 lead to unrealistic population clumping if the range is sufficiently large. Another method  
127 for modeling spatial populations is to assume the existence of a grid of discrete randomly  
128 mating populations connected by migration, thus enforcing regular population density by  
129 edict. Among many other important results drawn from this class of “lattice” or “stepping  
130 stone” models, Rousset (1997) showed that the slope of the linear regression of genetic dif-  
131 ferentiation ( $F_{ST}$ ) against the logarithm of spatial distance is an estimate of neighborhood  
132 size. Although these grid models may be good approximations of continuous geography  
133 in many situations, they do not model demographic fluctuations, and limit investigation of  
134 spatial structure below the level of the deme, assumptions whose impacts are unknown. An  
135 alternative method for dealing with continuous geography is a new class of coalescent models,  
136 the Spatial Lambda Fleming-Viot models (Barton *et al.* 2010; Kelleher *et al.* 2014).

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

#### 147 **A Forward-Time Model of Evolution in Continuous Space**

148 We simulated populations using the non-Wright-Fisher module in the program SLiM v3.1  
149 (Haller and Messer 2019). Each time step consists of three stages: reproduction, dispersal, and  
150 mortality. To reduce the parameter space we use the same parameter, denoted  $\sigma$ , to modulate  
151 the spatial scale of interactions at all three stages by adjusting the standard deviation of the  
152 corresponding Gaussian functions. As in previous work (Wright 1943; Ringbauer *et al.* 2017),

153  $\sigma$  is equal to the mean parent-offspring distance.

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

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

161 We chose this functional form so that the equilibrium population density per unit area is  
162 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  $\sigma$ . The final probability of survival for individual  $i$  is then

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

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

166 To isolate spatial effects from other components of the model such as overlapping gener-

167 ations, increased variance in reproductive success, and density-dependent fitness, we also  
168 implemented simulations identical to those above except that mates are selected uniformly  
169 at random from the population, and offspring disperse to a uniform random location on the  
170 landscape. We refer to this model as the “random mating” model, in contrast to the first,  
171 “spatial” model.

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

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

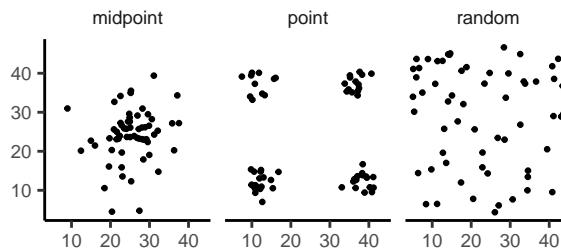
### 187 ***Demographic Parameters***

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

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

201 **Sampling**

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



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

210 **Summary Statistics**

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

219 Following recent studies that showed strong signals for dispersal and demography in the  
220 distribution of shared haplotype block lengths (Ringbauer *et al.* 2017; Baharian *et al.* 2016), we  
221 also calculated various summaries of the distribution of pairwise identical-by-state (IBS) block  
222 lengths among sampled chromosomes. The full distribution of lengths of IBS tracts for each  
223 pair of chromosomes was first calculated with a custom python function. We then calculated  
224 the first three moments of this distribution (mean, variance, and skew) and the number of  
225 blocks over  $10^6$  base pairs both for each pair of individuals and for the full distribution across  
226 all pairwise comparisons.

227 We then estimated correlation coefficients between spatial distance and each moment of  
228 the pairwise IBS tract distribution. Because more closely related individuals on average share  
229 longer haplotype blocks we expect that spatial distance will be negatively correlated with  
230 mean haplotype block length, and that this correlation will be strongest (i.e., most negative)  
231 when dispersal is low. The variance, skew, and count of long haplotype block statistics are  
232 meant to reflect the relative length of the right (upper) tail of the distribution, which represents  
233 the frequency of long haplotype blocks, and so should reflect recent demographic events  
234 (Chapman and Thompson 2002). For a subset of simulations, we also calculated cumulative  
235 distributions for IBS tract lengths across pairs of distant ( $> 48$  map units) and nearby ( $< 2$   
236 map units) individuals. Last, we examined the relationship between allele frequency and the  
237 spatial dispersion of an allele by calculating the average distance among individuals carrying  
238 each derived allele in a set of simulations representing a range of neighborhood sizes.

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

## 242 **Demographic Modeling**

243 To assess the impacts of continuous spatial structure on demographic inference we inferred  
244 population size histories for all simulations using two approaches: stairwayplot (Liu and  
245 Fu 2015) and SMC++ (Terhorst *et al.* 2016). Stairwayplot fits its model to a genome-wide  
246 estimate of the SFS, while SMC++ also incorporates linkage information. For both methods  
247 we sampled 20 individuals from all spatial simulations using random, midpoint, and point  
248 sampling strategies.

249 As recommended by its documentation, we used stairwayplot to fit models with multiple  
250 bootstrap replicates drawn from empirical genomic data, and took the median inferred  $N_e$  per  
251 unit time as the best estimate. We calculated site frequency spectra with scikit-allel (Miles and  
252 Harding 2017), generated 100 bootstrap replicates per simulation by resampling over sites,  
253 and fit models for all bootstrap samples using default settings.

254 For SMC++, we first output genotypes as VCF with msprime and then used SMC++'s  
255 standard pipeline for preparing input files assuming no polarization error in the SFS. We  
256 used the first individual in the VCF as the "designated individual" when fitting models, and  
257 allowed the program to estimate the recombination rate during optimization. We fit models  
258 using the 'estimate' command rather than the now recommended cross-validation approach  
259 because our simulations had only a single contig.

260 To evaluate the performance of these methods we binned simulations by neighborhood  
261 size, took a rolling median of inferred  $N_e$  trajectories across all model fits in a bin for each  
262 method and sampling strategy. We also examined how varying levels of isolation by distance  
263 impacted the variance of  $N_e$  estimates by calculating the standard deviation of  $N_e$  from each  
264 best-fit model and plotting these against neighborhood size.

265 **Association Studies**

266 To assess the degree to which spatial structure confounds GWAS we simulated four types of  
267 nongenetic phenotype variation for 1000 randomly sampled individuals in each spatial SLiM  
268 simulation and conducted a linear regression GWAS with principal components as covariates  
269 in PLINK (Purcell *et al.* 2007). SNPs with a minor allele frequency less than 0.5% were excluded  
270 from this analysis. Phenotype values were set to vary by two standard deviations across the  
271 landscape in a rough approximation of the variation seen in height across Europe (Turchin  
272 *et al.* 2012; Garcia and Quintana-Domeque 2006, 2007). Conceptually our approach is similar  
273 to that taken by Mathieson and McVean (2012), though here we model fully continuous spatial  
274 variation and compare GWAS output across a range of dispersal distances.

275 In all simulations, the phenotype of each individual is determined by adding independent  
276 Gaussian noise with mean zero and standard deviation 10 to a mean that may depend on  
277 spatial position. We adjust the geographic pattern of mean phenotype to create spatially  
278 autocorrelated environmental influences on phenotype. In the first simulation of *nonspatial*

environments, 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 100 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  $50 \times 50$  landscape has mean phenotype  $100 + 2x/5$ . 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 120 (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 120.

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  $\sigma$  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).

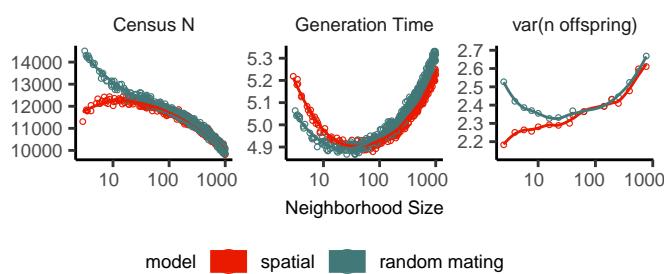
## 304 **Results**

### 305 **Demographic Parameters**

306 Adjusting the spatial dispersal and interaction distance,  $\sigma$ , has a surprisingly large effect on  
307 demographic quantities that are usually fixed in Wright-Fisher models – the generation time,  
308 census population size, and variance in offspring number. These are shown in Figure 2. This

309 occurs because, even though the “population density” ( $K$ ) and “mean lifetime” ( $L$ ) parameters  
310 were the same in all simulations, the strength of stochastic effects depends strongly on  $\sigma$ .  
311 For instance, the population density near to individual  $i$  (denoted  $n_i$  above) is computed by  
312 averaging over roughly  $N_W = 4\pi K\sigma^2$  individuals, and so has standard deviation proportional  
313 to  $1/\sqrt{N_W}$  – it is more variable at lower densities. (Recall that  $N_W$  is Wright’s neighborhood  
314 size.) Since the probability of survival is a nonlinear function of  $n_i$ , actual equilibrium  
315 densities and lifetimes differ from  $K$  and  $L$ . This is the reason that we included *random mating*  
316 simulations – where mate choice and offspring dispersal are both nonspatial – since this  
317 should preserve the random fluctuations in local population density while destroying any  
318 spatial genetic structure. We verified that random mating models retained no geographic  
319 signal by showing that summary statistics did not differ significantly between sampling  
320 regimes (Table S2), unlike in spatial models (discussed below).

321 There are a few additional things to note about Figure 2. First, all three quantities are  
322 non-monotone with neighborhood size. Census size largely declines as neighborhood size  
323 increases for both the spatial and random mating models. However, for spatial models this  
324 decline only begins for neighborhood size  $\geq 10$ . By a neighborhood sizes larger than 100, the  
325 spatial and random mating models are indistinguishable from one another, a sign that our  
326 simulations are performing as expected. Census sizes range from  $\approx 14,000$  at low  $\sigma$  in the  
327 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.

328 Generation time similarly shows complex behavior with respect to neighborhood sizes,  
329 and varies between 5.2 and 4.9 timesteps per generation across the parameter range explored.  
330 Under both the spatial and random mating models, generation time reaches a minimum at a

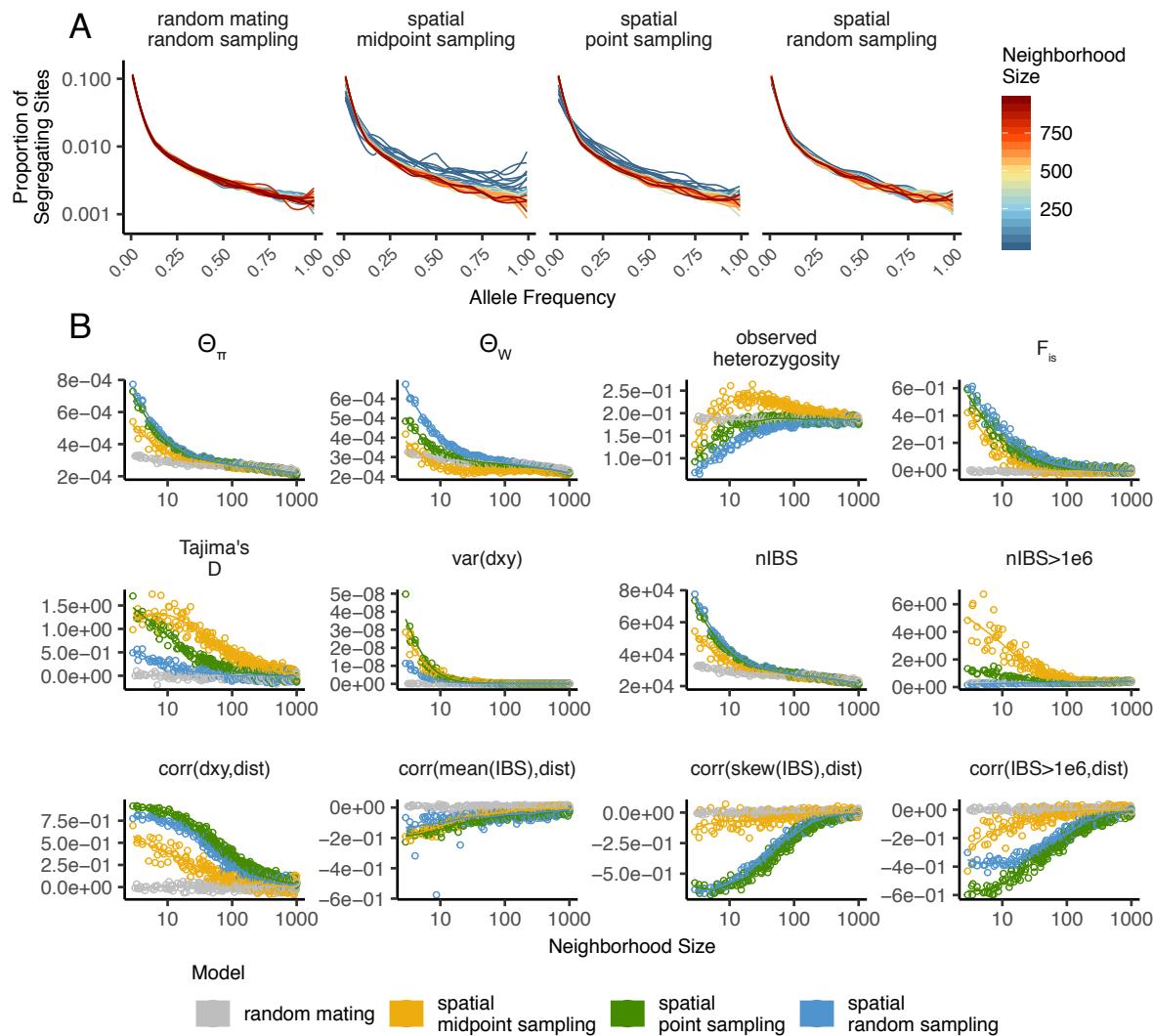
331 neighborhood size of around 50. Interestingly, under the range of neighborhood sizes that we  
332 examined, generation times between the random mating and spatial models are never quite  
333 equivalent – presumably this would cease to be the case at neighborhood sizes higher than  
334 we simulated here.

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

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

342 Even though certain aspects of population demography depend on the scale of spatial inter-  
343 actions, it still could be that population genetic variation is well-described by a well-mixed  
344 population model. Indeed, mathematical results suggest that genetic variation in some spatial  
345 models should be well-approximated by a Wright-Fisher population if neighborhood size is  
346 large and all samples are geographically widely separated (Wilkins 2004; Zähle *et al.* 2005).  
347 However, the behavior of most common population genetic summary statistics other than  
348 Tajima's  $D$  (Städler *et al.* 2009) has not yet been described in realistic geographic models.  
349 Moreover, as we will show, spatial sampling strategies can affect summaries of variation at  
350 least as strongly as the underlying population dynamics.

351 ***Site Frequency Spectra and Summaries of Diversity*** Figure 3 shows the effect of varying  
352 neighborhood size and sampling strategy on the site frequency spectrum (Figure 3A) and  
353 several standard population genetic summary statistics (Figure 3B). Consistent with findings  
354 in island and stepping stone simulations (Städler *et al.* 2009), the SFS shows a significant  
355 enrichment of intermediate frequency variants in comparison to the nonspatial expectation.  
356 This bias is most pronounced below neighborhood sizes  $\leq 100$  and is exacerbated by midpoint  
357 and point sampling of individuals (depicted in Figure 1). Reflecting this, Tajima's  $D$  is quite  
358 positive in the same situations (Figure 3B). Notably, the point at which Tajima's  $D$  approaches  
359 0 differs strongly across sampling strategies – varying from a neighborhood size of roughly 50  
360 for random sampling to at least 1000 for midpoint sampling.



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

361 One of the most commonly used summaries of variation is Tajima's summary of nucleotide  
362 divergence,  $\theta_\pi$ , calculated as the mean density of nucleotide differences averaged across pairs  
363 of samples. As can be seen in Figure 3B,  $\theta_\pi$  in the spatial model is inflated by up to three-fold  
364 relative to the random mating model. This pattern is opposite the expectation from census  
365 population size (Figure 2), because the spatial model has *lower* census size than the random  
366 mating model at neighborhood sizes less than 100. Differences between these models likely  
367 occur because  $\theta_\pi$  is a measure of mean time to most recent common ancestor between two  
368 samples, and at small values of  $\sigma$ , the time for dispersal to mix ancestry across the range  
369 exceeds the mean coalescent time under random mating. (For instance, at the smallest value  
370 of  $\sigma = 0.2$ , the range is 250 dispersal distances wide, and since the location of a diffusively  
371 moving lineage after  $k$  generations has variance  $k\sigma^2$ , it takes around  $250^2 = 62500$  generations  
372 to mix across the range, which is roughly ten times larger than the random mating effective  
373 population size).  $\theta_\pi$  using each sampling strategy approaches the random mating expectation  
374 at its own rate, but by a neighborhood size of around 100 all models are roughly equivalent.  
375 Interestingly, the effect of sampling strategy is reversed relative to that observed in Tajima's  
376 D – midpoint sampling reaches random mating expectations around neighborhood size 50,  
377 while random sampling is inflated until around neighborhood size 100.

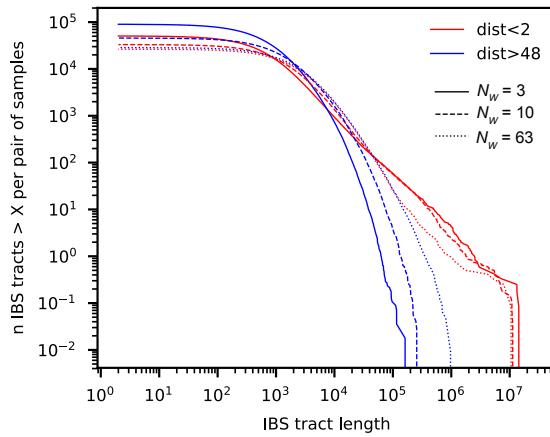
378 Values of observed heterozygosity and its derivative  $F_{IS}$  also depend heavily on neighbor-  
379 hood size under spatial models as well as the sampling scheme.  $F_{IS}$  is inflated above the  
380 expectation across most of the parameter space examined and across all sampling strategies.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

390 **IBS tracts and correlations with geographic distance** We next turn our attention to the effect  
391 of geographic distance on haplotype block length sharing, summarized for sets of nearby and

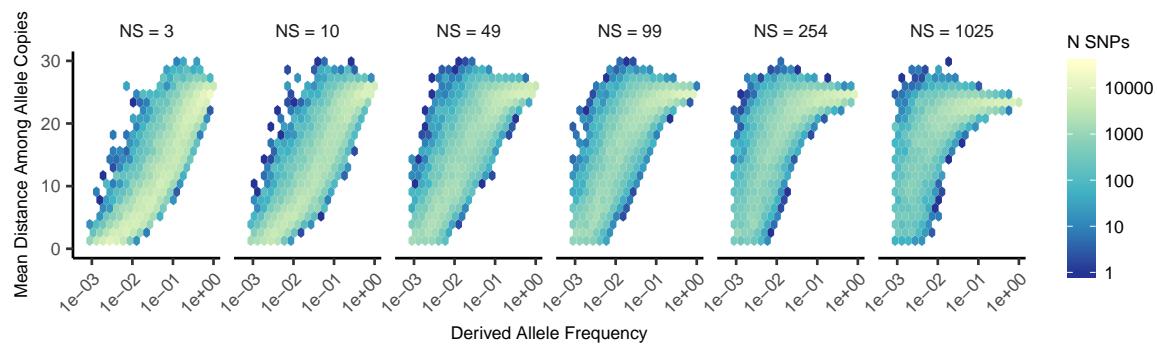
392 distant individuals in Figure 4. There are two main patterns to note. First, nearby individuals  
393 share more long IBS tracts than distant individuals (as expected because they are on average  
394 more closely related). Second, the difference in the number of long IBS tracts between nearby  
395 and distant individuals decreases as neighborhood size increases. This reflects the faster  
396 spatial mixing of populations with higher dispersal, which breaks down the correlation  
397 between the IBS tract length distribution and geographic distance. This can also be seen in  
398 the bottom row of Figure 3B, where the correlation coefficients between the summaries of the  
399 IBS tract length distribution (the mean, skew, and count of tracts over  $10^6$ bp) and geographic  
400 distance approaches 0 as neighborhood size increases.

401 The patterns observed for correlations of IBS tract lengths with geographic distance are  
402 similar to those observed in the more familiar regression of allele frequency measures such  
403 as  $D_{xy}$  (i.e., “genetic distance”) or  $F_{ST}$  against geographic distance (Rousset 1997).  $D_{xy}$  is  
404 positively correlated with the geographic distance between the individuals, and the strength  
405 of this correlation declines as dispersal increases (Figure 3B), as expected (Wright 1943; Rousset  
406 1997). This relationship is very similar across random and point sampling strategies, but is  
407 weaker for midpoint sampling, perhaps due to a dearth of long-distance comparisons. In  
408 much of empirical population genetics a regression of genetic differentiation against spatial  
409 distance is a de-facto metric of the significance of isolation by distance. The similar behavior of  
410 moments of the pairwise distribution of IBS tract lengths shows why haplotype block sharing  
411 has recently emerged as a promising source of information on spatial demography through  
412 methods described in Ringbauer *et al.* (2017) and Baharian *et al.* (2016).

413 **Spatial distribution of allele copies** Mutations occur in individuals and spread geographically  
414 over time. Because low frequency alleles generally represent recent mutations (Sawyer 1977;  
415 Griffiths *et al.* 1999), the geographic dispersion of an allele may covary along with its frequency  
416 in the population. To visualize this relationship we calculated the average distance among  
417 individuals carrying a focal derived allele across simulations with varying neighborhood sizes,  
418 shown in Figure 5. On average we find that low frequency alleles are the most geographically  
419 restricted, and that the extent to which geography and allele frequency are related depends on  
420 the amount of dispersal in the population. For populations with large neighborhood sizes we  
421 found that even very low frequency alleles can be found across the full landscape, whereas  
422 in populations with low neighborhood sizes the relationship between distance among allele



**Figure 4** Cumulative distributions for IBS tract lengths per pair of individuals at different geographic distances, across three neighborhood sizes ( $N_W$ ).



**Figure 5** Trends in the distance among allele copies at varying derived allele frequencies and neighborhood sizes.

423 copies and their frequency is quite strong. This is the basic process underlying Novembre and  
424 Slatkin's (2009) method for estimating dispersal distances based on the distribution of low  
425 frequency alleles, and also generates the greater degree of bias in GWAS effect sizes for low  
426 frequency alleles identified in Mathieson and McVean (2012).

427 ***Effects of Space on Demographic Inference***

428 One of the most important uses for population genetic data is inferring demographic history  
429 of populations. As demonstrated above, the site frequency spectrum and the distribution of  
430 IBS tracts varies across neighborhood sizes and sampling strategies. Does this variation lead to  
431 different inferences of past population sizes? To ask this we inferred population size histories  
432 from samples drawn from our simulated populations with two approaches: stairwayplot  
433 (Liu and Fu 2015), which uses a genome-wide estimate of the SFS, and SMC++ (Terhorst *et al.*  
434 2016), which incorporates information on both the SFS and linkage disequilibrium across the  
435 genome.

436 Figure 6A shows the median inferred population size histories from each method across all  
437 simulations, grouped by neighborhood size and sampling strategy. In general these methods  
438 tend to slightly overestimate ancient population sizes and infer recent population declines  
439 when neighborhood sizes are below 20 and sampling is spatially clustered (Figure 6A, Figure  
440 S4). The overestimation of ancient population sizes however is relatively minor, averaging  
441 around a two-fold inflation at 10,000 generations before present in the worst-affected bins.  
442 For stairwayplot we found that many runs infer dramatic population bottlenecks in the last  
443 1,000 generations when sampling is spatially concentrated, resulting in ten-fold or greater  
444 underestimates of recent population sizes. However SMC++ appeared more robust to this  
445 error, with runs on point- and midpoint-sampled simulations at the lowest neighborhood  
446 sizes underestimating recent population sizes by roughly half and those on randomly sampled  
447 simulations showing little error. Above neighborhood sizes of around 100, both methods  
448 performed relatively well when averaging across results from multiple simulations.

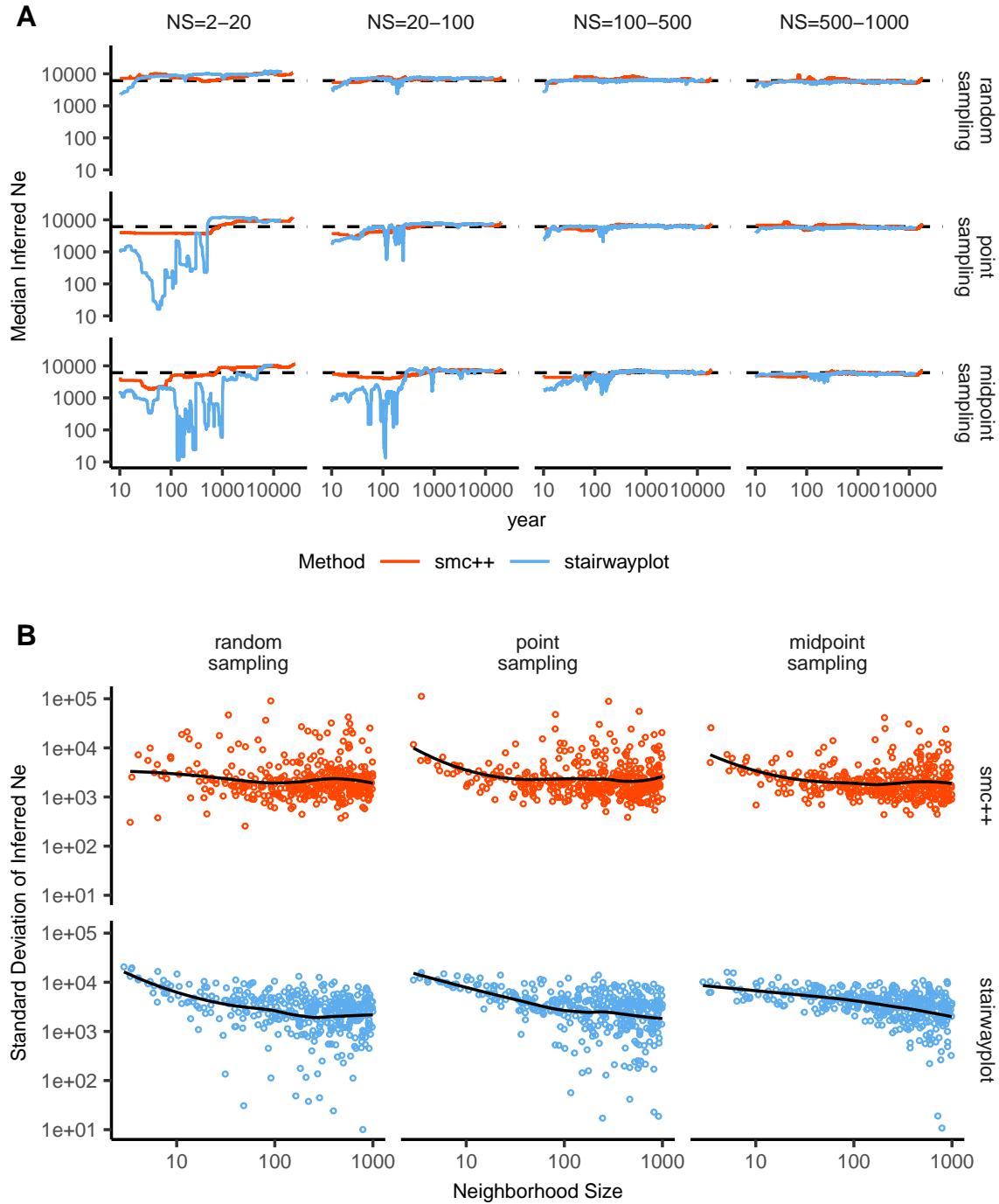
449 However, individual model fits from both methods frequently reflected turbulent demo-  
450 graphic histories (Figure S4), with the standard deviation of inferred  $N_e$  across time points  
451 often exceeding the expected  $N_e$  for both methods (Figure 6B). That is, despite the constant  
452 population sizes in our simulations, both methods tended to infer large fluctuations in popu-

453 lation size over time, which could potentially result in incorrect biological interpretations. On  
454 average the variance of inferred population sizes was elevated at the lowest neighborhood  
455 sizes and declines as dispersal increases, with the strongest effects seen in stairwayplot model  
456 fits with for clustered sampling and neighborhood sizes less than 20 (Figure 6B).

457 **GWAS**

458 To ask what confounding effects spatial genetic variation might have on genome-wide associa-  
459 tion studies we performed GWAS on our simulations using phenotypes that were determined  
460 solely by the environment – so, any SNP showing statistically significant correlation with  
461 phenotype is a false positive. As expected, spatial autocorrelation in the environment causes  
462 spurious associations across much of the genome if no correction for genetic relatedness  
463 among samples is performed (Figures 7 and S5). This effect is particularly strong for clinal  
464 and corner environments, for which the lowest dispersal levels cause over 60% of SNPs in the  
465 sample to return significant associations. Patchy environmental distributions, which are less  
466 strongly spatially correlated (Figure 7A), cause fewer false positives overall but still produce  
467 spurious associations at roughly 10% of sites at the lowest neighborhood sizes. Interestingly  
468 we also observed a small number of false positives in roughly 3% of analyses on simula-  
469 tions with nonspatial environments, both with and without PC covariates included in the  
470 regression.

471 The confounding effects of geographic structure are well known, and it is common practice  
472 to control for this by including principal components (PCs) as covariates to control for these  
473 effects. This mostly works in our simulations – after incorporating the first ten PC axes as  
474 covariates, the vast majority of SNPs no longer surpass a significance threshold chosen to  
475 have a 5% false discovery rate (FDR). However, a substantial number of SNPs – up to 1.5% at  
476 the lowest dispersal distances – still surpass this threshold (and thus would be false positives  
477 in a GWAS), especially under “corner” and “patchy” environmental distributions (Figure 7C).  
478 At neighborhood sizes larger than 500, up to 0.31% of SNPs were significant for corner and  
479 clinal environments. Given an average of 132,000 SNPs across simulations after MAF filtering,  
480 this translates to up to 382 false-positive associations; for human-sized genomes, this number  
481 would be much larger. In most cases the *p* values for these associations were significant after  
482 FDR correction but would not pass the threshold for significance under the more conservative



**Figure 6** A: Rolling median inferred  $N_e$  trajectories for stairwayplot and smc++ across sampling strategies and neighborhood size bins. The dotted line shows the mean  $N_e$  of random-mating simulations. B: Standard deviation of individual inferred  $N_e$  trajectories, by neighborhood size and sampling strategy. Black lines are loess curves. Plots including individual model fits are shown in Figure S4.

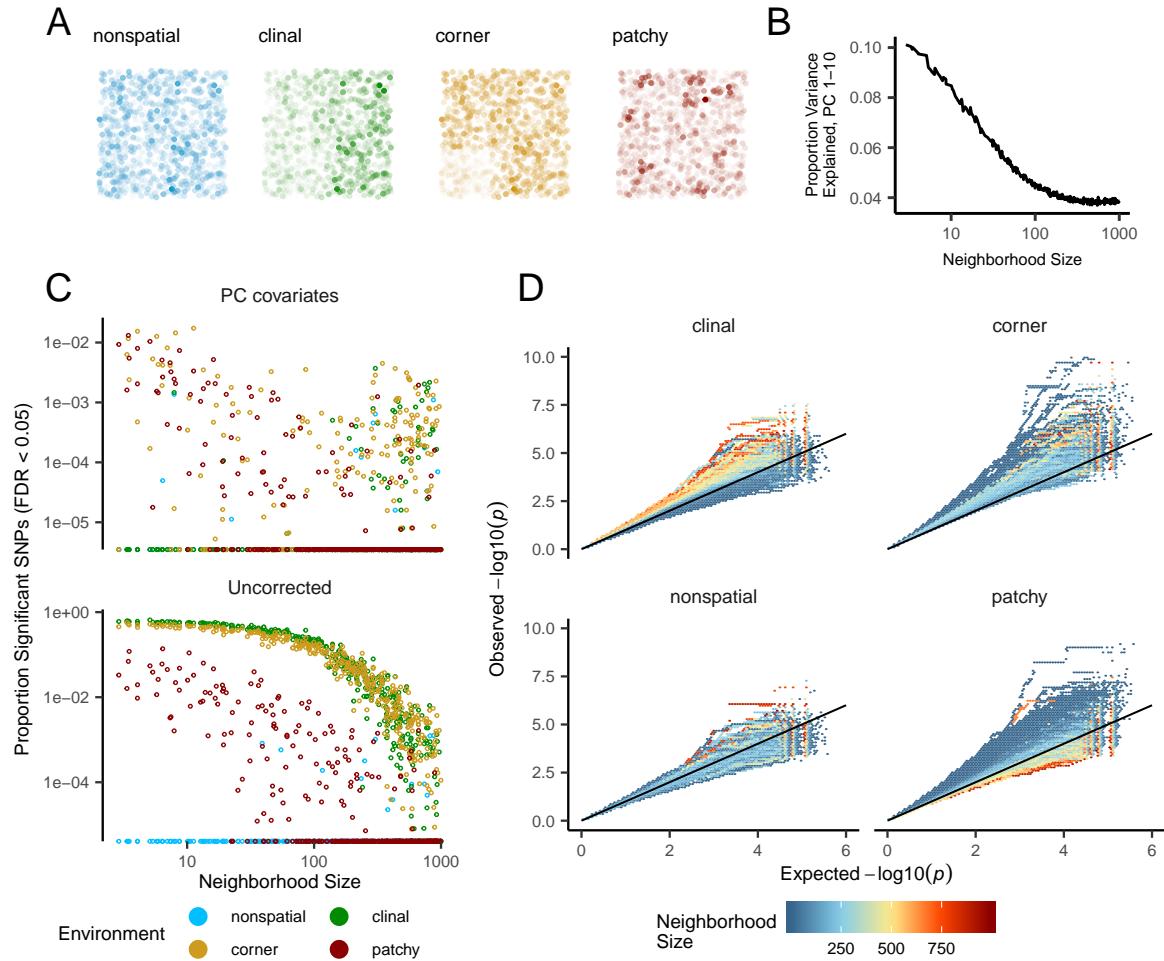
483 Bonferroni correction (see example Manhattan plots in figure S5).

484 Clinal environments cause an interesting pattern in false positives after PC correction:  
485 at low neighborhood sizes the correction removes nearly all significant associations, but at  
486 neighborhood sizes above roughly 250 the proportion of significant SNPs increases to up to  
487 0.4% (Figure 7). This may be due to a loss of descriptive power of the PCs – as neighborhood  
488 size increases, the total proportion of variance explained by the first 10 PC axes declines from  
489 roughly 10% to 4% (Figure 7B). Essentially, PCA seems unable to effectively summarize the  
490 weak population structure present in large-neighborhood simulations, but these populations  
491 continue to have enough spatial structure to create significant correlations between genotypes  
492 and the environment. A similar process can also be seen in the corner phenotype distribution,  
493 in which the count of significant SNPs initially declines as neighborhood size increases and  
494 then increases at approximately the point at which the proportion of variance explained by  
495 PCA approaches its minimum.

496 Figure 7D shows quantile-quantile plots that show the degree of genome-wide inflation of  
497 test statistics in PC-corrected GWAS across all simulations and environmental distributions.  
498 For clinal environments,  $-\log_{10}(p)$  values are most inflated when neighborhood sizes are  
499 large, consistent with the pattern observed in the count of significant associations after  
500 PC regression. In contrast corner and patchy environments cause the greatest inflation in  
501  $-\log_{10}(p)$  at neighborhood sizes less than 100, which likely reflects the inability of PCA to  
502 account for fine-scale structure caused by very limited dispersal. Finally, we observed that PC  
503 regression appears to overfit to some degree for all phenotype distributions, visible in Figure  
504 7D as points falling below the 1:1 line.

## 505 Discussion

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



**Figure 7** 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 the first 10 PC coordinates 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.

513 **Effects of Dispersal**

514 Limited dispersal inflates effective population size, creates correlations between genetic and  
515 spatial distances, and introduces strong distortions in the site frequency spectrum that are  
516 reflected in a positive Tajima's  $D$  (Figure 3). At the lowest dispersal distances, this can  
517 increase genetic diversity threefold relative to random-mating expectations. These effects  
518 are strongest when neighborhood sizes are below 100, but in combination with the effects of  
519 nonrandom sampling they can persist up to neighborhood sizes of at least 1000 (e.g., inflation  
520 in Tajima's  $D$  and observed heterozygosity under midpoint sampling). If samples are chosen  
521 uniformly from across space, the general pattern is similar to expectations of the original  
522 analytic model of Wright (1943), which predicts that populations with neighborhood sizes  
523 under 100 will differ substantially from random mating, while those above 10,000 will be  
524 nearly indistinguishable from panmixia.

525 The patterns observed in sequence data reflect the effects of space on the underlying  
526 genealogy. Nearby individuals coalesce rapidly under limited dispersal and so are connected  
527 by short branch lengths, while distant individuals take much longer to coalesce than they  
528 would under random mating. Mutation and recombination events in our simulation both  
529 occur at a constant rate along branches of the genealogy, so the genetic distance and number  
530 of recombination events separating sampled individuals simply gives a noisy picture of the  
531 genealogies connecting them. Tip branches (i.e., branches subtending only one individual)  
532 are then relatively short, and branches in the middle of the genealogy connecting local groups  
533 of individuals relatively long, leading to the biases in the site frequency spectrum shown in  
534 Figure 3.

535 The genealogical patterns introduced by limited dispersal are particularly apparent in the  
536 distribution of haplotype block lengths (Figure 3). This is because identical-by-state tract  
537 lengths reflect the impacts of two processes acting along the branches of the underlying  
538 genealogy – both mutation and recombination – rather than just mutation as is the case  
539 when looking at the site frequency spectrum or related summaries. This means that the  
540 pairwise distribution of haplotype block lengths carries with it important information about  
541 genealogical variation in the population, and correlation coefficients between moments of the  
542 this distribution and geographic location contain signal similar to the correlations between  
543  $F_{ST}$  or  $D_{xy}$  and geographic distance (Rousset 1997). Indeed this basic logic underlies two

544 recent studies explicitly estimating dispersal from the distribution of shared haplotype block  
545 lengths (Ringbauer *et al.* 2017; Baharian *et al.* 2016). Conversely, because haplotype-based  
546 measures of demography are particularly sensitive to variation in the underlying genealogy,  
547 inference approaches that assume random mating when analyzing the distribution of shared  
548 haplotype block lengths are likely to be strongly affected by spatial processes.

549 **Effects of Sampling**

550 One of the most important differences between random mating and spatial models is the effect  
551 of sampling: in a randomly mating population the spatial distribution of sampling effort has  
552 no effect on estimates of genetic variation (Table S1), but when dispersal is limited sampling  
553 strategy can compound spatial patterns in the underlying genealogy and create pervasive  
554 impacts on all downstream genetic analyses (see also Städler *et al.* (2009)). In most species, the  
555 difficulty of traveling through all parts of a species range and the inefficiency of collecting  
556 single individuals at each sampling site means that most studies follow something closest  
557 to the “point” sampling strategy we simulated, in which multiple individuals are sampled  
558 from nearby points on the landscape. For example, in ornithology a sample of 10 individuals  
559 per species per locality is a common target when collecting for natural history museums. In  
560 classical studies of *Drosophila* variation the situation is considerably worse, in which a single  
561 orchard might be extensively sampled.

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

573 However, we observed the largest sampling effects using “midpoint” sampling. This model

574 is meant to reflect a bias in sampling effort towards the middle of a species' range. In empirical  
575 studies this sampling strategy could arise if, for example, researchers choose to sample the  
576 center of the range and avoid range edges to maximize probability of locating individuals  
577 during a short field season. Because midpoint sampling provides limited spatial resolution  
578 it dramatically reduces the magnitude of observed correlations between spatial and genetic  
579 distances. More surprisingly, midpoint sampling also leads to strongly positive Tajima's  $D$   
580 and an inflation in the proportion of heterozygous individuals in the sample – similar to the  
581 effect of sampling a single deme in an island model as reported in (Städler *et al.* 2009). This  
582 increase in observed heterozygosity appears to reflect the effects of range edges, which are  
583 a fundamental facet of spatial genetic variation. If individuals move randomly in a finite  
584 two-dimensional landscape then regions in the middle of the landscape receive migrants from  
585 all directions while those on the edge receive no migrants from at least one direction. The  
586 average number of new mutations moving into the middle of the landscape is then higher  
587 than the number moving into regions near the range edge, leading to higher heterozygosity  
588 and lower inbreeding coefficients ( $F_{IS}$ ) away from range edges. Though here we used only a  
589 single parameterization of fitness decline at range edges we believe this is a general property  
590 of non-infinite landscapes as it has also been observed in previous studies simulating under  
591 lattice models (Neel *et al.* 2013; Shirk and Cushman 2014).

592 In summary, we recommend that empirical researchers collect individuals from across as  
593 much of the species' range as practical, choosing samples separated by a range of spatial scales.  
594 Many summary statistics are designed for well-mixed populations, and so provide different  
595 insights into genetic variation when applied to different subsets of the population. Applied  
596 to a cluster of samples, summary statistics based on segregating sites (e.g., Watterson's  $\Theta$   
597 and Tajima's  $D$ ), heterozygosity, or the distribution of long haplotype blocks, can be expected  
598 to depart significantly from what would be obtained from a wider distribution of samples.  
599 Comparing the results of analyses conducted on all individuals versus those limited to single  
600 individuals per locality can provide an informative contrast. Finally we wish to point out  
601 that the bias towards intermediate allele frequencies that we observe may mean that the  
602 importance of linked selection, at least as is gleaned from the site frequency spectrum, may be  
603 systematically underestimated currently.

604 **Demography**

605 Previous studies have found that population structure and nonrandom sampling can create  
606 spurious signals of population bottlenecks when attempting to infer demographic history with  
607 microsatellite variation, summary statistics, or runs of homozygosity (Chikhi *et al.* 2010; Städler  
608 *et al.* 2009; Ptak and Przeworski 2002; Mazet *et al.* 2015). Here we found that methods that  
609 infer detailed population trajectories through time based on the SFS and patterns of LD across  
610 the genome are also subject to this bias, with some combinations of dispersal and sampling  
611 strategy systematically inferring deep recent population bottlenecks and overestimating  
612 ancient  $N_e$  by a around a factor of 2. We were surprised to see that both stairwayplot and  
613 SMC++ can tolerate relatively strong isolation by distance – i.e., neighborhood sizes of 20  
614 – and still perform well when averaging results across multiple simulations. Inference in  
615 populations with neighborhood sizes over 20 was relatively unbiased unless samples were  
616 concentrated in the middle of the range (Figure 6). Although median demography estimates  
617 across many independent simulations were fairly accurate, empirical work has only a single  
618 estimate to work with, and individual model fits (Figure S4) suggest that spuriously inferred  
619 population size changes and bottlenecks are common, especially at small neighborhood  
620 sizes. As we will discuss below, most empirical estimates of neighborhood size, including all  
621 estimates for human populations, are large enough that population size trajectories inferred  
622 by these approaches should not be strongly affected by spatial biases created by dispersal  
623 in continuous landscapes. In contrast, Mazet *et al.* (2015) found that varying migration rates  
624 through time could create strong biases in inferred population trajectories from an  $n$ -island  
625 model with parameters relevant for human history, suggesting that changes in migration rates  
626 through time are more likely to drive variation in inferred  $N_e$  than isolation by distance.

627 We found that SMC++ was more robust to the effects of space than stairwayplot, under-  
628 estimating recent populations by roughly half in the worst time periods rather than nearly  
629 10-fold as with stairwayplot. Though this degree of variation in population size is certainly  
630 meaningful in an ecological context, it is relatively minor in population genetic terms. A  
631 more worrying pattern was the high level of variance in inferred  $N_e$  trajectories for individual  
632 model fits using these methods, which was highest in simulations with the smallest neighbor-  
633 hood size (Figure 6, Figure S4). This suggests that, at a minimum, researchers working with  
634 empirical data should replicate analyses multiple times and take a rolling average if model

635 fits are inconsistent across runs. Splitting samples and running replicates on separate subsets –  
636 the closest an empirical study can come to our design of averaging the results from multiple  
637 simulations – may also alleviate this issue.

638 Our analysis suggests that many empirical analyses of population size history using meth-  
639 ods like SMC++ are robust to error caused by spatial structure within continuous landscapes.  
640 Inferences drawn from static SFS-based methods like stairwayplot should be treated with  
641 caution when there are signs of isolation by distance in the underlying data (for example, if  
642 a regression of  $F_{ST}$  against the logarithm of geographic distance has a significantly positive  
643 slope), and in particular an inference of population bottlenecks in the last 1000 years should  
644 be discounted if sampling is clustered, but estimates of deeper time patterns are likely to  
645 be fairly accurate. The biases in the SFS and haplotype structure identified above (see also  
646 Wakeley 1999; Chikhi *et al.* 2010; Städler *et al.* 2009) are apparently small enough that they fall  
647 within the range of variability regularly inferred by these approaches, at least on datasets of  
648 the size we simulated.

649 **GWAS**

650 Spatial structure is particularly challenging for genome-wide association studies, because  
651 the effects of dispersal on genetic variation are compounded by spatial variation in the  
652 environment (Mathieson and McVean 2012). Spatially restricted mate choice and dispersal  
653 causes variation in allele frequencies across the range of a species. If environmental factors  
654 affecting the phenotype of interest also vary over space, then groups of individuals in different  
655 regions will allele frequencies and environmental exposures will covary over space. In this  
656 scenario an uncorrected GWAS will infer genetic associations with a purely environmental  
657 phenotype at any site in the genome that is differentiated over space, and the relative degree  
658 of bias will be a function of the degree of covariation in allele frequencies and the environment  
659 (i.e., Figure 7C, bottom panel). This pattern has been demonstrated in a variety of simulation  
660 and empirical contexts (Price *et al.* 2006; Yu *et al.* 2005; Young *et al.* 2018; Mathieson and  
661 McVean 2012; Kang *et al.* 2008, 2010; Bulik-Sullivan *et al.* 2015; Berg *et al.* 2018; Sohail *et al.*  
662 2018).

663 Incorporating PC positions as covariates in a linear-regression GWAS (Price *et al.* 2006) is  
664 designed to address this challenge by regressing out a baseline level of “average” differentia-

665 tion. In essence, a PC-corrected GWAS asks “what regions of the genome are more associated  
666 with this phenotype than the average genome-wide association observed across populations?”  
667 In our simulations, we observed that this procedure can fail under a variety of circumstances.  
668 If dispersal is limited and environmental variation is clustered in space (i.e., corner or patchy  
669 distributions in our simulations), PCA positions fail to capture the fine-scale spatial structure  
670 required to remove all signals of association. Conversely, as dispersal increases, PCA loses  
671 power to describe population structure before spatial mixing breaks down the relationship  
672 between genotype and the environment. These effects were observed with all spatially cor-  
673 related environmental patterns, but were particularly pronounced if environmental effects  
674 are concentrated in one region, as was also found by Mathieson and McVean (2012). Though  
675 increasing the number of PC axes used in the analysis may reduce the false-positive rate, this  
676 may also decrease the power of the test to detect truly causal alleles (Lawson *et al.* 2019).

677 In this work we simulated a single chromosome with size roughly comparable to one  
678 human chromosome. If we scale the number of false-positive associations identified in our  
679 analyses to a GWAS conducted on whole-genome data from humans, we would expect to  
680 see several thousand weak false-positive associations after PC corrections in a population  
681 with neighborhood sizes up to at least 1000 (which should include values appropriate for  
682 many human populations). Notably, very few of the spurious associations we identified  
683 would be significant at a conservative Bonferroni-adjusted *p*-value cutoff (see Figure S5). This  
684 suggests that GWAS focused on finding strongly associated alleles for traits controlled by  
685 a limited number of variants in the genome are likely robust to the impacts of continuous  
686 spatial structure. However, methods that analyze the combined effects of thousands or  
687 millions of weakly associated variants such as polygenic risk scores (Khera *et al.* 2018) are  
688 likely to be affected by subtle population structure. Indeed as recently identified in studies  
689 of genotype associations for human height in Europe (Berg *et al.* 2018; Sohail *et al.* 2018), PC  
690 regression GWAS in modern human populations do include residual signal of population  
691 structure in large-scale analyses of polygenic traits. When attempting to make predictions  
692 across populations with different environmental exposures, polygenic risk scores affected by  
693 population structure can be expected to offer low predictive power, as was shown in a recent  
694 study finding lower performance outside European populations (Martin *et al.* 2019).

695 In summary, spatial covariation in population structure and the environment confounds the

696 interpretation of GWAS *p*-values, and correction using principal components is insufficient to  
697 fully separate these signals for polygenic traits under a variety of environmental and popu-  
698 lation parameter regimes. Other GWAS methods may be less sensitive to this confounding,  
699 but there is no obvious reason that this should be so. One approach to estimating the degree  
700 of bias in GWAS caused by population structure is LD score regression (Bulik-Sullivan *et al.*  
701 2015). Though this approach appears to work well in practice, its interpretation is not always  
702 straightforward and it is likely biased by the presence of linked selection (Berg *et al.* 2018).  
703 In addition, we observed that in many cases the false-positive SNPs we identified appeared  
704 to be concentrated in LD peaks similar to those expected from truly causal sites (Figure S5),  
705 which may confound LD score regression.

706 We suggest a straightforward alternative for species in which the primary axes of population  
707 differentiation is space (note this is likely not the case for some modern human populations):  
708 run a GWAS with spatial coordinates as phenotypes and check for *p*-value inflation or  
709 significant associations. If significant associations with sample locality are observed after  
710 correcting for population structure, the method is sensitive to false positives induced by  
711 spatial structure. This is essentially the approach taken in our “clinal” model (though we  
712 add normally distributed noise to our phenotypes). Of course, it is possible that genotypes  
713 indirectly affect individual locations by adjusting organismal fitness and thus habitat selection  
714 across spatially varying environments, but we believe that this hypothesis should be tested  
715 against a null of stratification bias inflation rather than accepted as true based on GWAS  
716 results.

### 717 **Where are natural populations on this spectrum?**

718 For how much of the tree of life do spatial patterns circumscribe genomic variation? In Table  
719 1 we gathered estimates of neighborhood size from a range of organisms to get an idea of  
720 how likely dispersal is to play an important role in patterns of variation. Though this sample  
721 is almost certainly biased towards small-neighborhood species (because few studies have  
722 quantified neighborhood size in species with very high dispersal or population density), we  
723 find that neighborhood sizes in the range we simulated are fairly common across a range of  
724 taxa. At the extreme low end of empirical neighborhood size estimates we see some flowering  
725 plants, large mammals, and colonial insects like ants. Species such as this have neighborhood

**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>Aedes aegypti</i>	Yellow-fever mosquito	268	Genetic	(Jasper <i>et al.</i> 2019)
<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)

726 size estimates small enough that spatial processes are likely to strongly influence inference.  
727 These include some human populations such as the Gainj- and Kalam-speaking people of  
728 Papua New Guinea, in which the estimated neighborhood sizes in (Rousset 1997) range from  
729 40 to 410 depending on the method of estimation. Many more species occur in a middle range  
730 of neighborhood sizes between 100 and 1000 – a range in which spatial processes play a minor  
731 role in our analyses under random spatial sampling but are important when sampling of  
732 individuals in space is clustered. Surprisingly, even some flying insects with huge census  
733 population sizes fall in this group, including fruit flies (*D. melanogaster*) and mosquitoes (*A.*  
734 *aegypti*). Last, many species likely have neighborhood sizes much larger than we simulated,  
735 including modern humans in northeastern Europe (Ringbauer *et al.* 2017). For these species  
736 demographic inference and summary statistics are likely to reflect minimal bias from spatial  
737 effects as long as dispersal is truly continuous across the landscape. While that is so we  
738 caution that association studies in which the effects of population structure are confounded  
739 with spatial variation in the environment are still sensitive to dispersal even at these large  
740 neighborhood sizes.

#### 741 ***Future Directions and Limitations***

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

751 One facet of spatial variation that we did not address in this study is the confounding of  
752 dispersal and population density implicit in the definition of Wright's neighborhood size.  
753 Our simulations were run under constant densities, but Ringbauer *et al.* (2017)'s approach  
754 to demographic inference in space suggests that density and dispersal can in some cases be  
755 estimated separately from genetic data. Much additional work remains to be done to better

756 understand how these parameters interact to shape genetic variation in continuous space,  
757 which we leave to future studies.

758 Though our simulation allows incorporation of realistic demographic and spatial processes,  
759 it is inevitably limited by the computational burden of tracking tens or hundreds of thousands  
760 of individuals in every generation. In particular, computations required for mate selection  
761 and spatial competition scale approximately with the product of the total census size and  
762 the neighborhood size and so increase rapidly for large populations and dispersal distances.  
763 The reverse-time model of continuous space evolution described by Barton *et al.* (2010)  
764 and implemented by Kelleher *et al.* (2014) allows exploration of parameter regimes with  
765 population and landscape sizes more directly comparable to empirical cases like humans.  
766 Alternatively, implementation of parallelized calculations may allow progress with forward-  
767 time simulations.

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

## 774 Data Availability

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

## 776 Acknowledgements

777 We thank Brandon Cooper, Matt Hahn, Doc Edge, and others for reading and thinking about  
778 this manuscript. We thank the Hearth for having such good, nearby coffee, and Falling Sky  
779 Brewing for creative support during manuscript drafting. CJB and ADK were supported by  
780 NIH award R01GM117241.

## 781 Literature Cited

782 Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, *et al.*, 2017 Deconstruct-  
783 ing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics

- 784      **13:** 1–27.
- 785      Antlfinger, A. E., 1982 Genetic neighborhood structure of the salt marsh composite, *Borrichia*  
786      *frutescens*. *Journal of Heredity* **73**: 128–132.
- 787      Antolin, M. F., B. V. Horne, M. D. Berger, Jr., A. K. Holloway, J. L. Roach, *et al.*, 2001 Effective  
788      population size and genetic structure of a piute ground squirrel (*Spermophilus mollis*)  
789      population. *Canadian Journal of Zoology* **79**: 26–34.
- 790      Ashander, J., P. Ralph, E. McCartney-Melstad, and H. B. Shaffer, 2018 Demographic inference  
791      in a spatially-explicit ecological model from genomic data: a proof of concept for the mojave  
792      desert tortoise. *bioRxiv* .
- 793      Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, *et al.*, 2016 The great  
794      migration and african-american genomic diversity. *PLOS Genetics* **12**: 1–27.
- 795      Barton, N. H., F. Depaulis, and A. M. Etheridge, 2002 Neutral evolution in spatially continuous  
796      populations. *Theoretical Population Biology* **61**: 31–48.
- 797      Barton, N. H., J. Kelleher, and A. M. Etheridge, 2010 A new model for extinction and recolo-  
798      nization in two dimensions: Quantifying phylogeography. *Evolution* **64**: 2701–2715.
- 799      Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing  
800      under dependency. *The Annals of Statistics* **29**: 1165–1188.
- 801      Berg, E. E. and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. *Evolution*  
802      **49**: 110–120.
- 803      Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, *et al.*, 2018  
804      Reduced signal for polygenic adaptation of height in uk biobank. *bioRxiv* .
- 805      Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 Ld score regression  
806      distinguishes confounding from polygenicity in genome-wide association studies. *Nature*  
807      *Genetics* **47**: 291 EP –.
- 808      Campbell, D. R. and J. L. Dooley, 1992 The spatial scale of genetic differentiation in a  
809      hummingbird-pollinated plant: Comparison with models of isolation by distance. *The*  
810      *American Naturalist* **139**: 735–748.
- 811      Chapman, N. H. and E. A. Thompson, 2002 The effect of population history on the lengths of  
812      ancestral chromosome segments. *Genetics* **162**: 449–458.
- 813      Chikhi, L., V. C. Sousa, P. Luisi, B. Goossens, and M. A. Beaumont, 2010 The confounding  
814      effects of population structure, genetic diversity and the sampling scheme on the detection

- 815 and quantification of population size changes. *Genetics* **186**: 983–995.
- 816 Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, *et al.*, 1993 Direct  
817 and indirect estimates of neighborhood and effective population size in a tropical palm,  
818 *astrocaryum mexicanum*. *Evolution* **47**: 75–87.
- 819 Epperson, B., 2003 *Geographical Genetics*. Monographs in Population Biology, Princeton Uni-  
820 versity Press.
- 821 Felsenstein, J., 1975 A pain in the torus: Some difficulties with models of isolation by distance.  
822 *The American Naturalist* **109**: 359–368.
- 823 Fox, J. and S. Weisberg, 2011 *An R Companion to Applied Regression*. Sage, Thousand Oaks CA,  
824 second edition.
- 825 Garcia, J. and C. Quintana-Domeque, 2006 The evolution of adult height in europe: A brief  
826 note. Working Paper .
- 827 Garcia, J. and C. Quintana-Domeque, 2007 The evolution of adult height in europe: A brief  
828 note. *Economics Human Biology* **5**: 340 – 349.
- 829 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in  
830 north american *drosophila melanogaster* show signatures of soft sweeps. *PLOS Genetics* **11**:  
831 1–32.
- 832 Griffiths, R., S. Tavaré, *et al.*, 1999 The ages of mutations in gene trees. *The Annals of Applied  
833 Probability* **9**: 567–590.
- 834 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence  
835 recording in SLiM opens new horizons for forward-time simulation of whole genomes.  
836 *Molecular Ecology Resources* **19**: 552–566.
- 837 Haller, B. C. and P. W. Messer, 2019 Slim 3: Forward genetic simulations beyond the wright–  
838 fisher model. *Molecular biology and evolution* **36**: 632–637.
- 839 Harris, K. and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared  
840 haplotype lengths. *PLOS Genetics* **9**: 1–20.
- 841 Huillet, T. and M. Möhle, 2011 On the extended Moran model and its relation to coalescents  
842 with multiple collisions. *Theoretical Population Biology* pp. –.
- 843 Jasper, M., T. Schmidt, N. Ahmad, S. Sinkins, and A. Hoffmann, 2019 A genomic approach to  
844 inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito.  
845 bioRxiv .

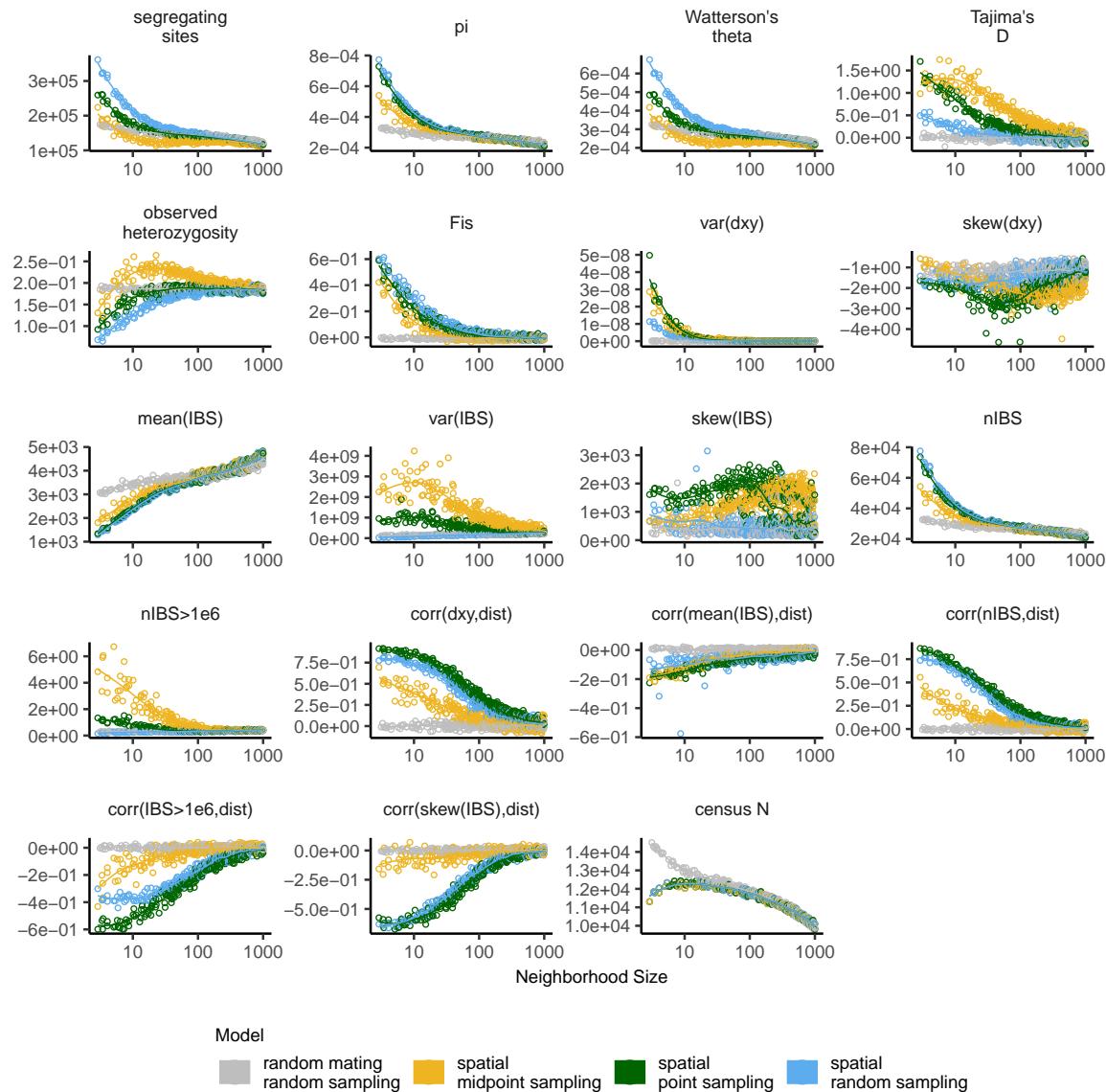
- 846 Jay, F., P. Sjödin, M. Jakobsson, and M. G. Blum, 2012 Anisotropic Isolation by Distance: The  
847 Main Orientations of Human Genetic Differentiation. *Molecular Biology and Evolution* **30**:  
848 513–525.
- 849 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component  
850 model to account for sample structure in genome-wide association studies. *Nature Genetics*  
851 **42**: 348 EP –.
- 852 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, *et al.*, 2008 Efficient control  
853 of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- 854 Kelleher, J., A. Etheridge, and N. Barton, 2014 Coalescent simulation in continuous space:  
855 Algorithms for large neighbourhood size. *Theoretical Population Biology* **95**: 13 – 23.
- 856 Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and ge-  
857 nealogical analysis for large sample sizes. *PLoS Comput Biol* **12**: 1–22.
- 858 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for  
859 fast population genetics simulation. *PLOS Computational Biology* **14**: 1–21.
- 860 Kerster, H. W., 1964 Neighborhood size in the rusty lizard, *sceloporus olivaceus*. *Evolution* **18**:  
861 445–457.
- 862 Khera, A. V., M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, *et al.*, 2018 Genome-wide poly-  
863 genic scores for common diseases identify individuals with risk equivalent to monogenic  
864 mutations. *Nature Genetics* **50**: 1219–1224.
- 865 Kingman, J., 1982 The coalescent. *Stochastic Processes and their Applications* **13**: 235 – 248.
- 866 Lawson, D. J., N. M. Davies, S. Haworth, B. Ashraf, L. Howe, *et al.*, 2019 Is population structure  
867 in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* .
- 868 Liu, X. and Y.-X. Fu, 2015 Exploring population size changes using snp frequency spectra.  
869 *Nature Genetics* **47**: 555 EP –.
- 870 Lundgren, E. and P. L. Ralph, 2018 Are populations like a circuit? The relationship between  
871 isolation by distance and isolation by resistance. *bioRxiv* .
- 872 Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, *et al.*, 2019 Clinical use of current  
873 polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**: 584–591.
- 874 Maruyama, T., 1972 Rate of decrease of genetic variability in a two-dimensional continuous  
875 population of finite size. *Genetics* **70**: 639–651.
- 876 Mathieson, I. and G. McVean, 2012 Differential confounding of rare and common variants in

- 877 spatially structured populations. *Nature Genetics* **44**: 243 EP –.
- 878 Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi, 2015 On the importance of being  
879 structured: instantaneous coalescence rates and human evolution—lessons for ancestral  
880 population size inference? *Heredity* **116**: 362 EP –.
- 881 Miles, A. and N. Harding, 2017 *cghg/scikit-allel*: v1.1.8.
- 882 Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, *et al.*, 2013 Estimation of effec-  
883 tive population size in continuously distributed populations: there goes the neighborhood.  
884 *Heredity* **111**: 189 EP –.
- 885 Novembre, J. and M. Slatkin, 2009 Likelihood-based inference in isolation-by-distance models  
886 using the spatial distribution of low-frequency alleles. *Evolution* **63**: 2914–2925.
- 887 Pamilo, P., 1983 Genetic differentiation within subdivided populations of formica ants. *Evolu-  
888 tion* **37**: 1010–1022.
- 889 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLOS  
890 Genetics* **2**: 1–20.
- 891 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, *et al.*, 2006 Principal  
892 components analysis corrects for stratification in genome-wide association studies. *Nature  
893 Genetics* **38**: 904 EP –.
- 894 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using  
895 multilocus genotype data. *Genetics* **155**: 945–959.
- 896 Ptak, S. E. and M. Przeworski, 2002 Evidence for population growth in humans is confounded  
897 by fine-scale population structure. *Trends in Genetics* **18**: 559–563.
- 898 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.*, 2007 Plink: A tool set for  
899 whole-genome association and population-based linkage analyses. *The American Journal  
900 of Human Genetics* **81**: 559 – 575.
- 901 R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for  
902 Statistical Computing, Vienna, Austria.
- 903 Ralph, P. and G. Coop, 2013 The geography of recent genetic ancestry across Europe. *PLoS  
904 Biol* **11**: e1001555.
- 905 Ringbauer, H., G. Coop, and N. H. Barton, 2017 Inferring recent demography from isolation  
906 by distance of long shared sequence blocks. *Genetics* **205**: 1335–1351.
- 907 Robledo-Arnuncio, J. J. and F. Rousset, 2010 Isolation by distance in a continuous population

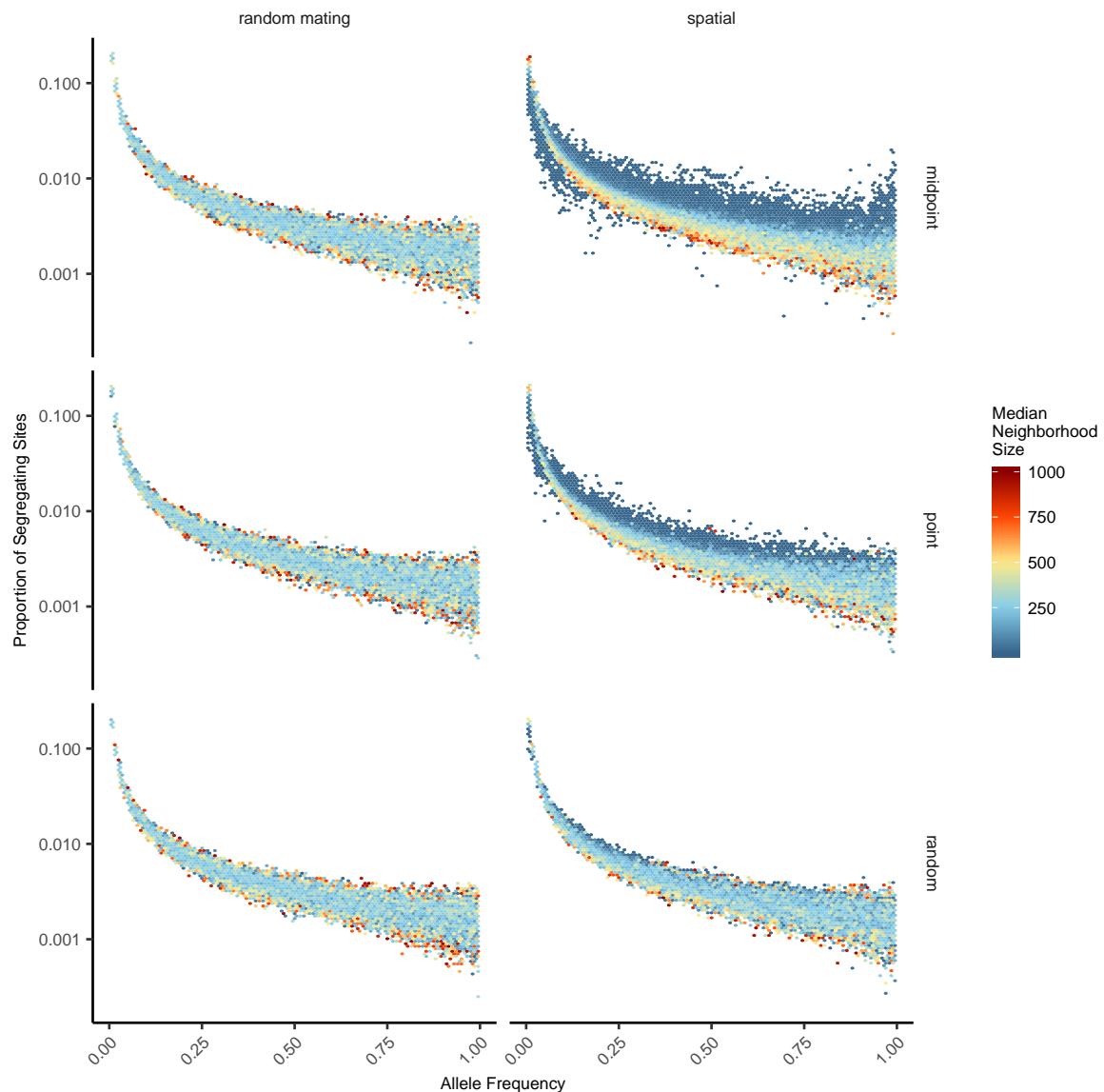
- 908 under stochastic demographic fluctuations. *Journal of Evolutionary Biology* **23**: 53–71.
- 909 Rousset, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under  
910 isolation by distance. *Genetics* **145**: 1219–1228.
- 911 Rousset, F. and R. Leblois, 2011 Likelihood-based inferences under isolation by distance:  
912 Two-dimensional habitats and confidence intervals. *Molecular Biology and Evolution* **29**:  
913 957–973.
- 914 Sawyer, S., 1977 On the past history of an allele now known to have frequency p. *Journal of*  
915 *Applied Probability* **14**: 439–450.
- 916 Schiffels, S. and R. Durbin, 2014 Inferring human population size and separation history from  
917 multiple genome sequences. *Nature Genetics* **46**: 919 EP –.
- 918 Schrider, D. R. and A. D. Kern, 2018 Supervised machine learning for population genetics: A  
919 new paradigm. *Trends in Genetics* **34**: 301 – 312.
- 920 Sharbel, T. F., B. Haubold, and T. Mitchell-Olds, 2000 Genetic isolation by distance in arabi-  
921 dopsis thaliana: biogeography and postglacial colonization of europe. *Molecular Ecology*  
922 **9**: 2109–2118.
- 923 Sheehan, S., K. Harris, and Y. S. Song, 2013 Estimating variable effective population sizes from  
924 multiple genomes: A sequentially markov conditional sampling distribution approach.  
925 *Genetics* **194**: 647–662.
- 926 Shirk, A. J. and S. A. Cushman, 2014 Spatially-explicit estimation of wright's neighborhood  
927 size in continuous populations. *Frontiers in Ecology and Evolution* **2**: 62.
- 928 Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, *et al.*, 2018 Signals of polygenic  
929 adaptation on height have been overestimated due to uncorrected population structure in  
930 genome-wide association studies. *bioRxiv* .
- 931 St. Onge, K. R., A. E. Palmé, S. I. Wright, and M. Lascoux, 2012 Impact of sampling schemes on  
932 demographic inference: An empirical study in two species with different mating systems  
933 and demographic histories. *G3: Genes, Genomes, Genetics* **2**: 803–814.
- 934 Städler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber, 2009 The impact of sam-  
935 pling schemes on the site frequency spectrum in nonequilibrium subdivided populations.  
936 *Genetics* **182**: 205–216.
- 937 Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA poly-  
938 morphism. *Genetics* **123**: 585–595.

- 939 Terhorst, J., J. A. Kamm, and Y. S. Song, 2016 Robust and scalable inference of population  
940 history from hundreds of unphased whole genomes. *Nature Genetics* **49**: 303 EP –.
- 941 Turchin, M. C., C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, *et al.*, 2012 Evidence  
942 of widespread selection on standing variation in europe at height-associated snps. *Nature*  
943 *Genetics* **44**: 1015 EP –.
- 944 Wahlund, S., 1928 Zusammensetzung von populationen und korrelationserscheinungen vom  
945 standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**: 65–106.
- 946 Wakeley, J., 1999 Nonequilibrium migration in human history. *Genetics* **153**: 1863–1871.
- 947 Wakeley, J., 2009 *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village,  
948 CO.
- 949 Wakeley, J. and T. Takahashi, 2003 Gene genealogies when the sample size exceeds the effective  
950 size of the population. *Mol Biol Evol* **20**: 208–213.
- 951 Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- 952 Wilke, C. O., 2019 *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package  
953 version 0.9.4.
- 954 Wilkins, J. F., 2004 A separation-of-timescales approach to the coalescent in a continuous  
955 population. *Genetics* **168**: 2227–2244.
- 956 Wright, S., 1931 Evolution in mendelian populations. *Genetics* **16**: 97.
- 957 Wright, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- 958 Wright, S., 1946 Isolation by distance under diverse systems of mating. *Genetics* **31**: 336.
- 959 Young, A. I., M. L. Frigge, D. F. Gudbjartsson, G. Thorleifsson, G. Bjornsdottir, *et al.*, 2018  
960 Relatedness disequilibrium regression estimates heritability without environmental bias.  
961 *Nature Genetics* **50**: 1304–1310.
- 962 Young, H. J., 1988 Neighborhood size in a beetle pollinated tropical aroid: effects of low  
963 density and asynchronous flowering. *Oecologia* **76**: 461–466.
- 964 Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, *et al.*, 2005 A unified mixed-model  
965 method for association mapping that accounts for multiple levels of relatedness. *Nature*  
966 *Genetics* **38**: 203 EP –.
- 967 Zähle, I., J. T. Cox, and R. Durrett, 2005 The stepping stone model. II. Genealogies and the  
968 infinite sites model. *Ann. Appl. Probab.* **15**: 671–699.

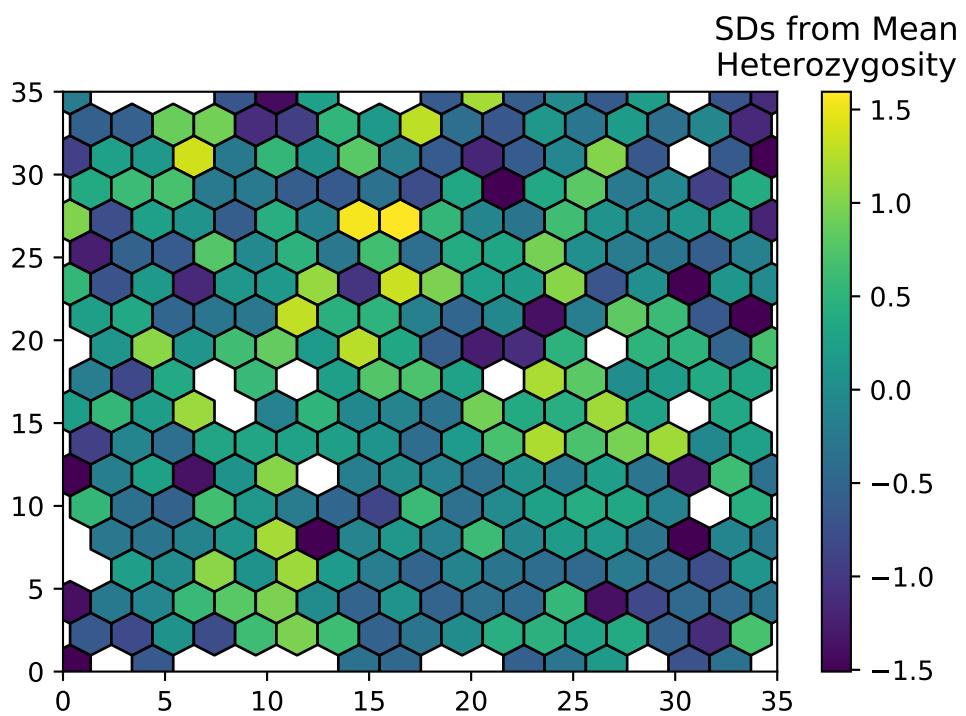
969 **Supplementary Figures and Tables**



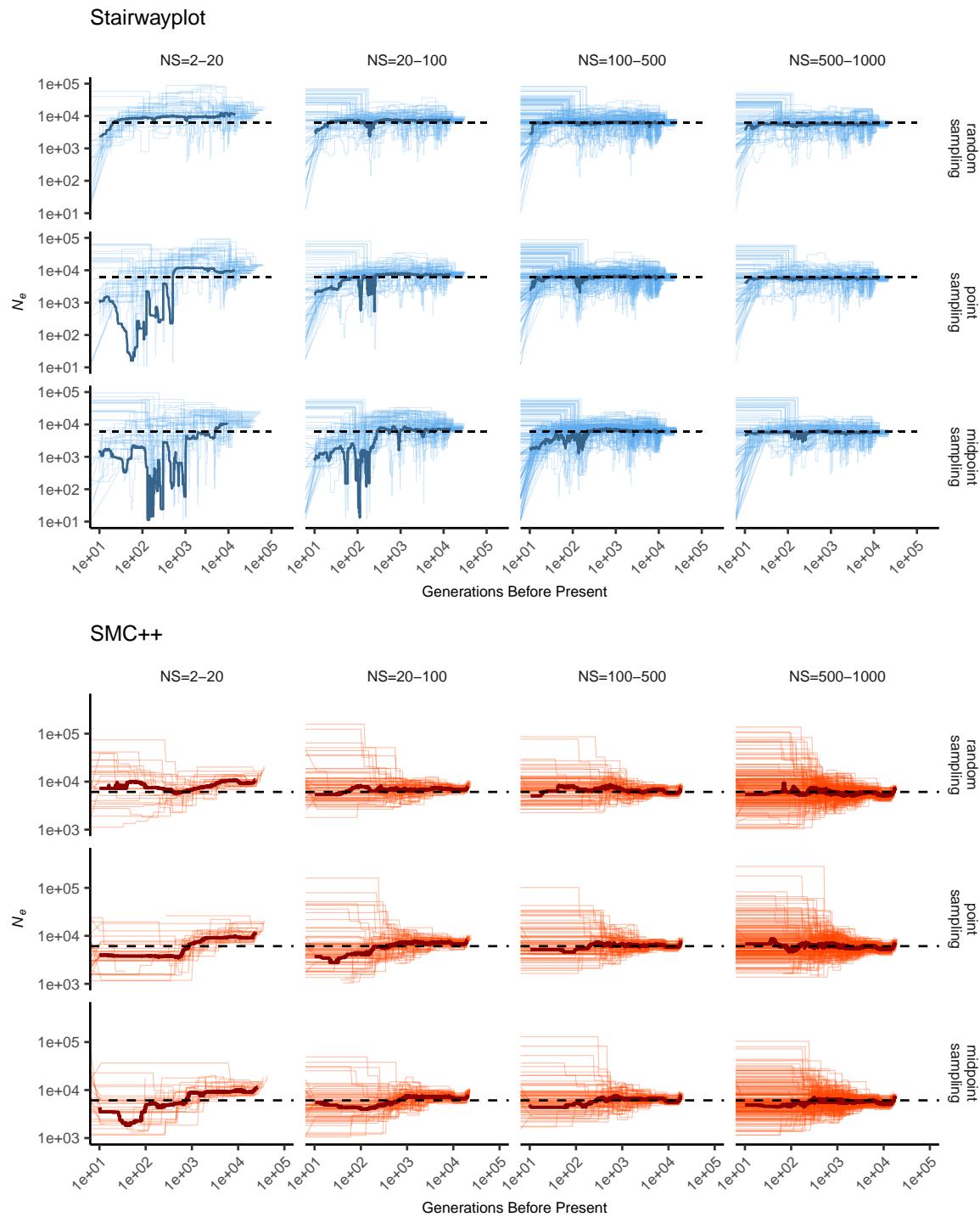
**Figure S1** Change in summary statistics by neighborhood size and sampling scheme calculated from simulated sequence data of 60 individuals.



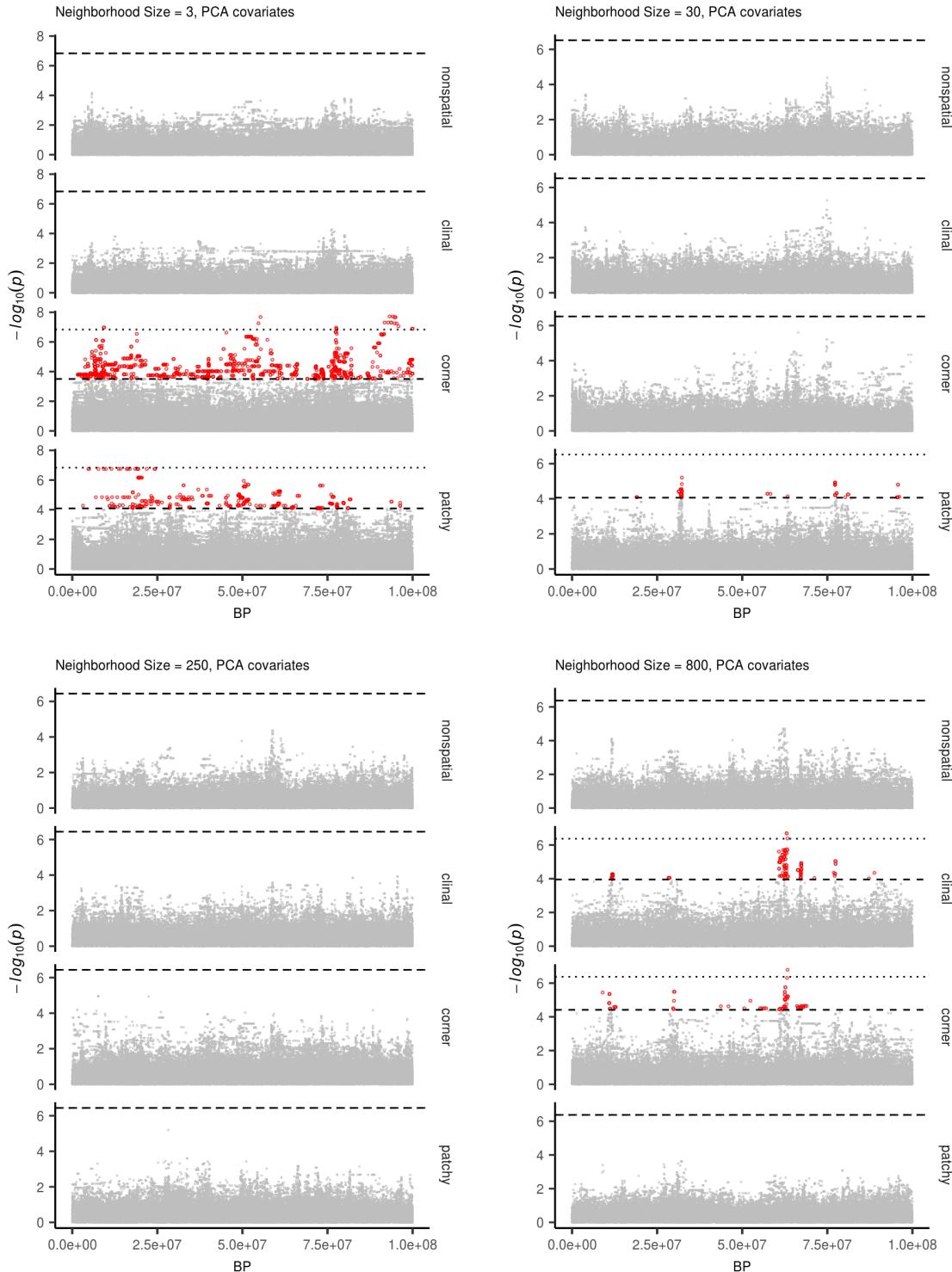
**Figure S2** Site frequency spectra for random mating and spatial SLiM models under all sampling schemes.



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



**Figure S4** Inferred demographic histories for spatial SLiM simulations, by sampling scheme and neighborhood size (NS) range. Thick lines are rolling medians across all simulations in a bin and thin lines are best fit models for each simulation. Dashed horizontal lines are the average  $N_e$  across random-mating SLiM models estimated from  $\theta_\pi$ .



**Figure S5** Manhattan plots for a sample of simulations at varying neighborhood sizes. Labels on the right of each plot describes the spatial distribution of environmental factors (described in the methods section of the main text). Points in red are significantly associated with a nongenetic phenotype using a 5% FDR threshold (dashed line). For runs with significant associations the dotted line is a Bonferroni-adjusted cutoff for  $p = 0.05$ .

**Table S1 Summary statistics calculated on simulated genotypes.**

Statistic	Description
$\Theta_{pi}$	Mean of the distribution of pairwise genetic differences
$\Theta_{W}$	Effective population size based on segregating sites
Segregating Sites	Total number of segregating sites in the sample
Tajima's $D$	Difference in $\Theta_{pi}$ and $\Theta_{W}$ 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
$mean(IVS)$	Mean of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$var(IVS)$	Variance of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$skew(IVS)$	Skew of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$nIBS$	Mean number of IBS tracts with length > 2bp across all pairs in the sample.
$nIBS > 1e6$	Mean number of IBS tracts over $1 \times 10^6$ bp per pair across all pairs in the sample.
$corr(D_{xy}, dist)$	Pearson correlation between genetic distance and $\log_{10}(spatial\ distance)$
$corr(mean(IVS), dist)$	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 $\log_{10}(spatial\ distance)$
$corr(IVS > 1e6, dist)$	Pearson correlation between the number of IBS tracts > $1 \times 10^6$ bp for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(skew(IVS), dist)$	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	$p(\text{equal means})$	$p(\text{equal variance})$
segsites	random mating	0.998190	0.980730
$\Theta_\pi$	random mating	0.997750	0.996450
$\Theta_W$	random mating	0.998190	0.980730
Tajima's $D$	random mating	0.879690	0.188770
observed heterozygosity	random mating	0.531540	0.433230
$F_{IS}$	random mating	0.474790	0.785730
$\text{mean}(D_{xy})$	random mating	0.997770	0.996510
$\text{var}(D_{xy})$	random mating	0.283630	0.647240
$\text{skew}(D_{xy})$	random mating	0.958320	0.260750
$\text{corr}(D_{xy}, \text{dist})$	random mating	0.601980	<b>0.000000</b>
$\text{mean}(IBS)$	random mating	0.997960	0.997730
$\text{var}(IBS)$	random mating	0.486450	0.399490
$\text{skew}(IBS)$	random mating	0.117980	0.069770
$nIBS$	random mating	0.997680	0.996570
$nIBS > 1e6$	random mating	0.834870	0.888730
$\text{corr}(\text{mean}(IBS), \text{dist})$	random mating	0.073270	0.308420
$\text{corr}(IBS > 1e6, \text{dist})$	random mating	0.268440	<b>0.002100</b>
$\text{corr}(\text{skew}(IBS), \text{dist})$	random mating	0.396920	<b>0.000620</b>
$\text{corr}(nIBS, \text{dist})$	random mating	0.581090	<b>0.000000</b>
segsites	spatial	<b>0.000000</b>	<b>0.000000</b>
$\Theta_\pi$	spatial	<b>0.026510</b>	<b>0.013440</b>
$\Theta_W$	spatial	<b>0.000000</b>	<b>0.000000</b>
Tajima's $D$	spatial	<b>0.000000</b>	<b>0.000000</b>
observed heterozygosity	spatial	<b>0.000000</b>	<b>0.000000</b>
$F_{IS}$	spatial	<b>0.000000</b>	<b>0.000120</b>
$\text{mean}(D_{xy})$	spatial	<b>0.025390</b>	<b>0.012910</b>
$\text{var}(D_{xy})$	spatial	<b>0.004970</b>	<b>0.006230</b>
$\text{skew}(D_{xy})$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{corr}(D_{xy}, \text{dist})$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{mean}(IBS)$	spatial	0.272400	0.114250
$\text{var}(IBS)$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{skew}(IBS)$	spatial	<b>0.000000</b>	<b>0.000000</b>
$nIBS$	spatial	<b>0.033920</b>	<b>0.016640</b>
$nIBS > 1e6$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{corr}(\text{mean}(IBS), \text{dist})$	spatial	<b>0.000000</b>	0.590540
$\text{corr}(IBS > 1e6, \text{dist})$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{corr}(\text{skew}(IBS), \text{dist})$	spatial	<b>0.000000</b>	<b>0.000000</b>
$\text{corr}(nIBS, \text{dist})$	spatial	<b>0.000000</b>	<b>0.000000</b>