

# 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: Wednesday 14<sup>th</sup> August, 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 viduals 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  $\exp(-x^2/2\sigma^2)$ .

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

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

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

171 An important step in creating any spatial model is dealing with range edges. Because local  
172 population density is used to model competition, edge or corner populations can be assigned  
173 artificially high fitness values because they lack neighbors within their interaction radius but  
174 outside the bounds of the simulation. We approximate a decline in habitat suitability near  
175 edges by decreasing the probability of survival proportional to the square root of distance to  
176 edges in units of  $\sigma$ . The final probability of survival for individual  $i$  is then

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

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

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

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

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

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

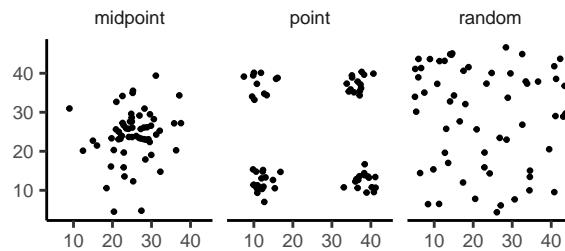
201    **Demographic Parameters**

202    Our demographic model includes parameters that control population density ( $K$ ), mean life  
203    span ( $L$ ), and dispersal distance ( $\sigma$ ). However, nonlinearity of local demographic stochasticity  
204    causes actual realized averages of these demographic quantities to deviate from the specified  
205    values in a way that depends on the neighborhood size. Therefore, to properly compare to  
206    theoretical expectations, we empirically calculated these demographic quantities in simula-  
207    tions. We recorded the census population size in all simulations. To estimate generation times,  
208    we stored ages of the parents of every new individual born across 200 timesteps, after a 100  
209    generation burn-in, and took the mean. To estimate variance in offspring number, we tracked  
210    the number of offspring for all individuals for 100 timesteps following a 100-timestep burn-in

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

215 **Sampling**

216 Our model records the genealogy and sequence variation of the complete population, but in  
217 real data, genotypes are only observed from a relatively small number of sampled individuals.  
218 We modeled three sampling strategies similar to common data collection methods in empirical  
219 genetic studies (Figure 1). “Random” sampling selects individuals at random from across  
220 the full landscape, “point” sampling selects individuals proportional to their distance from  
221 four equally spaced points on the landscape, and “midpoint” sampling selects individuals in  
222 proportion to their distance from the middle of the landscape. Downstream analyses were  
223 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.

224 **Summary Statistics**

225 We calculated the site frequency spectrum and a set of 18 summary statistics (Table S1) from  
226 60 diploid individuals sampled from the final generation of each simulation using the python  
227 package scikit-allel (Miles and Harding 2017). Statistics included common single-population  
228 summaries including mean pairwise divergence ( $\pi$ ), inbreeding coefficient ( $F_{IS}$ ), and Tajima’s  
229  $D$ , as well as an isolation-by-distance regression of genetic distance ( $D_{xy}$ ) against the logarithm  
230 of geographic distance analogous to Rousset (1997)’s approach, which we summarized as the  
231 correlation coefficient between the logarithm of the spatial distance and the proportion of  
232 identical base pairs across pairs of individuals.

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

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

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

## Demographic Modeling

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

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

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

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

279 **Association Studies**

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

289 In all simulations, the phenotype of each individual is determined by adding independent  
290 Gaussian noise with mean zero and standard deviation 10 to a mean that may depend on  
291 spatial position. We adjust the geographic pattern of mean phenotype to create spatially  
292 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 systematic 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).

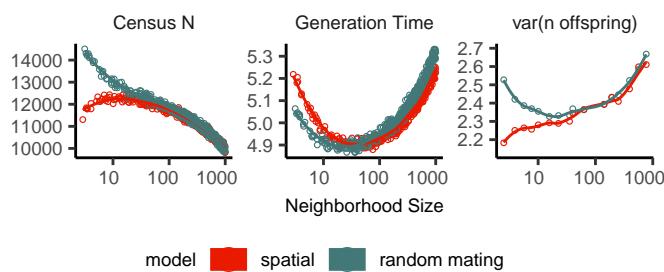
## 318 **Results**

### 319 **Demographic Parameters**

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

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

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



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

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

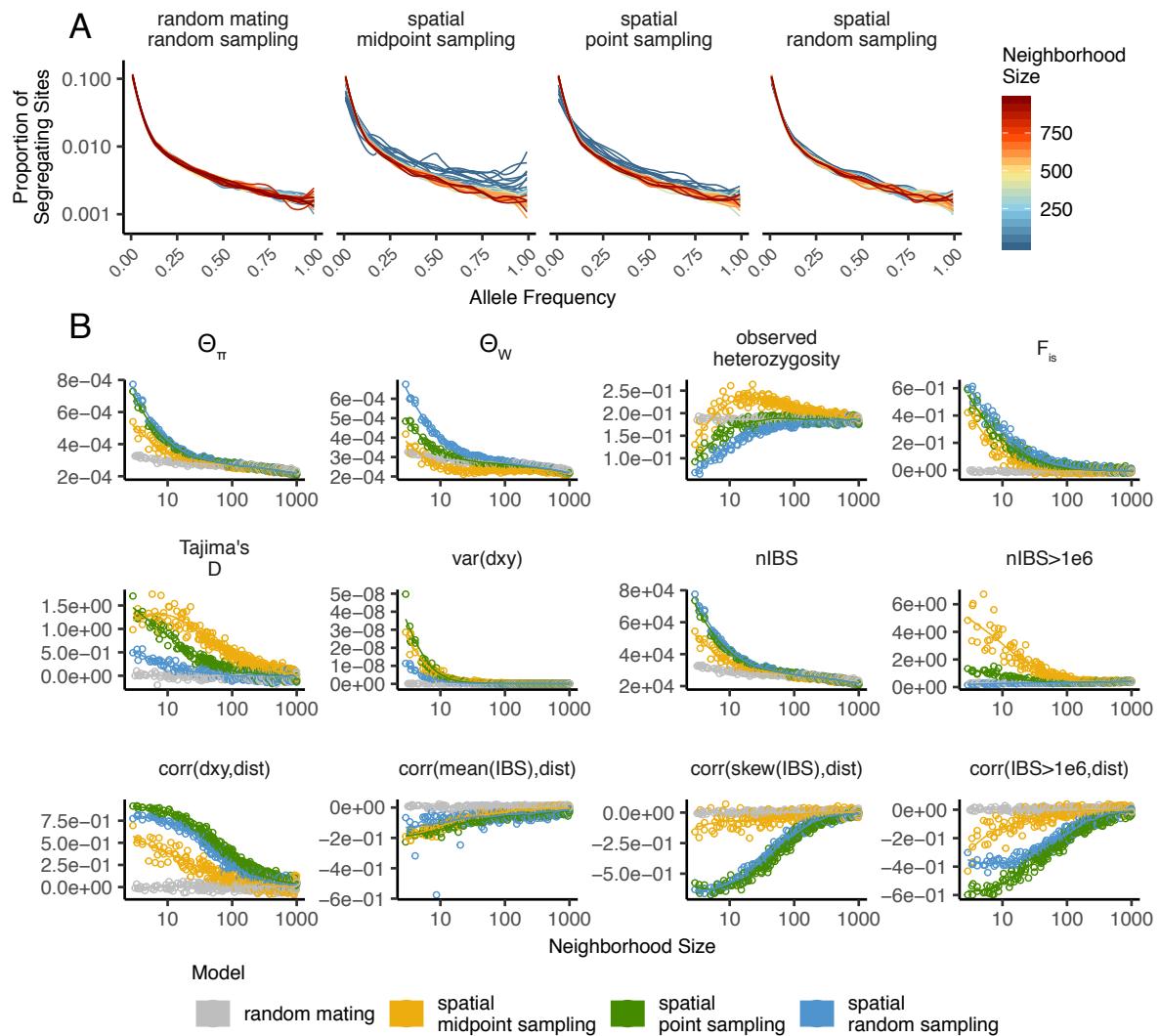
345 neighborhood size of around 50. Interestingly, under the range of neighborhood sizes that we  
346 examined, generation times between the random mating and spatial models are never quite  
347 equivalent – presumably this would cease to be the case at neighborhood sizes higher than  
348 we simulated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

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

### 355 ***Impacts of Continuous Space on Population Genetic Summary Statistics***

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

365 ***Site Frequency Spectra and Summaries of Diversity*** Figure 3 shows the effect of varying  
366 neighborhood size and sampling strategy on the site frequency spectrum (Figure 3A) and  
367 several standard population genetic summary statistics (Figure 3B). Consistent with findings  
368 in island and stepping stone simulations (Städler *et al.* 2009), the SFS shows a significant  
369 enrichment of intermediate frequency variants in comparison to the nonspatial expectation.  
370 This bias is most pronounced below neighborhood sizes  $\leq 100$  and is exacerbated by midpoint  
371 and point sampling of individuals (depicted in Figure 1). Reflecting this, Tajima's  $D$  is quite  
372 positive in the same situations (Figure 3B). Notably, the point at which Tajima's  $D$  approaches  
373 0 differs strongly across sampling strategies – varying from a neighborhood size of roughly 50  
374 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.

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

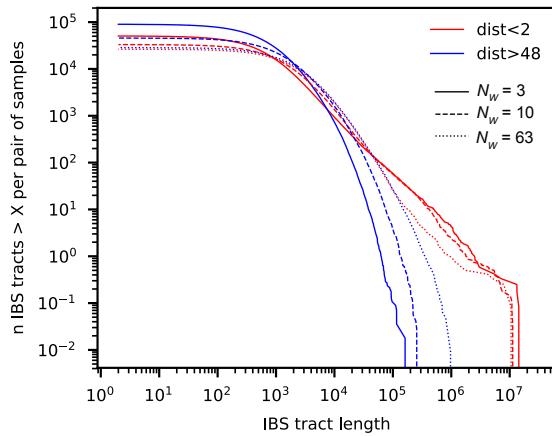
392 Values of observed heterozygosity and its derivative  $F_{IS}$  also depend heavily on neigh-  
393 borhood size under spatial models as well as the sampling scheme.  $F_{IS}$  is inflated above the  
394 expectation across most of the parameter space examined and across all sampling strategies.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

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

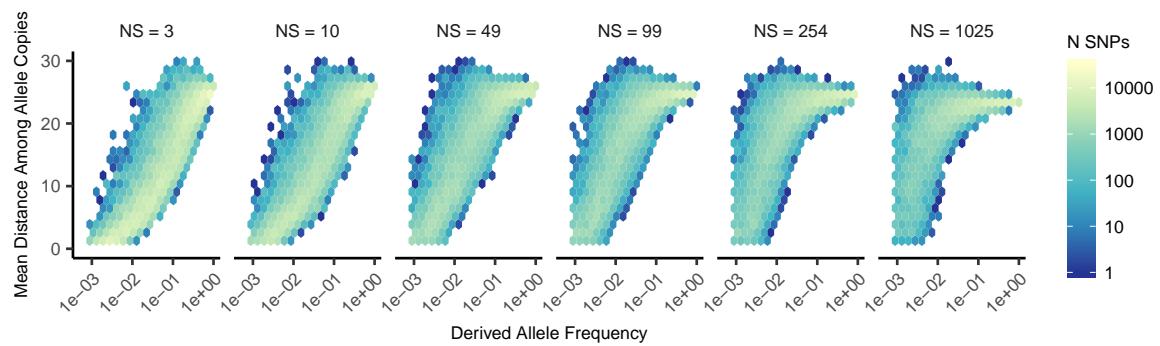
406 distant individuals in Figure 4. There are two main patterns to note. First, nearby individuals  
407 share more long IBS tracts than distant individuals (as expected because they are on average  
408 more closely related). Second, the difference in the number of long IBS tracts between nearby  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

415 The patterns observed for correlations of IBS tract lengths with geographic distance are  
416 similar to those observed in the more familiar regression of allele frequency measures such  
417 as  $D_{xy}$  (i.e., “genetic distance”) or  $F_{ST}$  against geographic distance (Rousset 1997).  $D_{xy}$  is  
418 positively correlated with the geographic distance between the individuals, and the strength  
419 of this correlation declines as dispersal increases (Figure 3B), as expected (Wright 1943; Rousset  
420 1997). This relationship is very similar across random and point sampling strategies, but is  
421 weaker for midpoint sampling, perhaps due to a dearth of long-distance comparisons. In  
422 much of empirical population genetics a regression of genetic differentiation against spatial  
423 distance is a de-facto metric of the significance of isolation by distance. The similar behavior of  
424 moments of the pairwise distribution of IBS tract lengths shows why haplotype block sharing  
425 has recently emerged as a promising source of information on spatial demography through  
426 methods described in Ringbauer *et al.* (2017) and Baharian *et al.* (2016).

427 **Spatial distribution of allele copies** Mutations occur in individuals and spread geographically  
428 over time. Because low frequency alleles generally represent recent mutations (Sawyer 1977;  
429 Griffiths *et al.* 1999), the geographic dispersion of an allele may covary along with its frequency  
430 in the population. To visualize this relationship we calculated the average distance among  
431 individuals carrying a focal derived allele across simulations with varying neighborhood sizes,  
432 shown in Figure 5. On average we find that low frequency alleles are the most geographically  
433 restricted, and that the extent to which geography and allele frequency are related depends on  
434 the amount of dispersal in the population. For populations with large neighborhood sizes we  
435 found that even very low frequency alleles can be found across the full landscape, whereas  
436 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.

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

441 ***Effects of Space on Demographic Inference***

442 One of the most important uses for population genetic data is inferring demographic history  
443 of populations. As demonstrated above, the site frequency spectrum and the distribution of  
444 IBS tracts varies across neighborhood sizes and sampling strategies. Does this variation lead to  
445 different inferences of past population sizes? To ask this we inferred population size histories  
446 from samples drawn from our simulated populations with two approaches: stairwayplot  
447 (Liu and Fu 2015), which uses a genome-wide estimate of the SFS, and SMC++ (Terhorst *et al.*  
448 2016), which incorporates information on both the SFS and linkage disequilibrium across the  
449 genome.

450 Figure 6A shows the median inferred population size histories from each method across all  
451 simulations, grouped by neighborhood size and sampling strategy. In general these methods  
452 tend to slightly overestimate ancient population sizes and infer recent population declines  
453 when neighborhood sizes are below 20 and sampling is spatially clustered (Figure 6A, Figure  
454 S4). The overestimation of ancient population sizes however is relatively minor, averaging  
455 around a two-fold inflation at 10,000 generations before present in the worst-affected bins.  
456 For stairwayplot we found that many runs infer dramatic population bottlenecks in the last  
457 1,000 generations when sampling is spatially concentrated, resulting in ten-fold or greater  
458 underestimates of recent population sizes. However SMC++ appeared more robust to this  
459 error, with runs on point- and midpoint-sampled simulations at the lowest neighborhood  
460 sizes underestimating recent population sizes by roughly half and those on randomly sampled  
461 simulations showing little error. Above neighborhood sizes of around 100, both methods  
462 performed relatively well when averaging across results from multiple simulations.

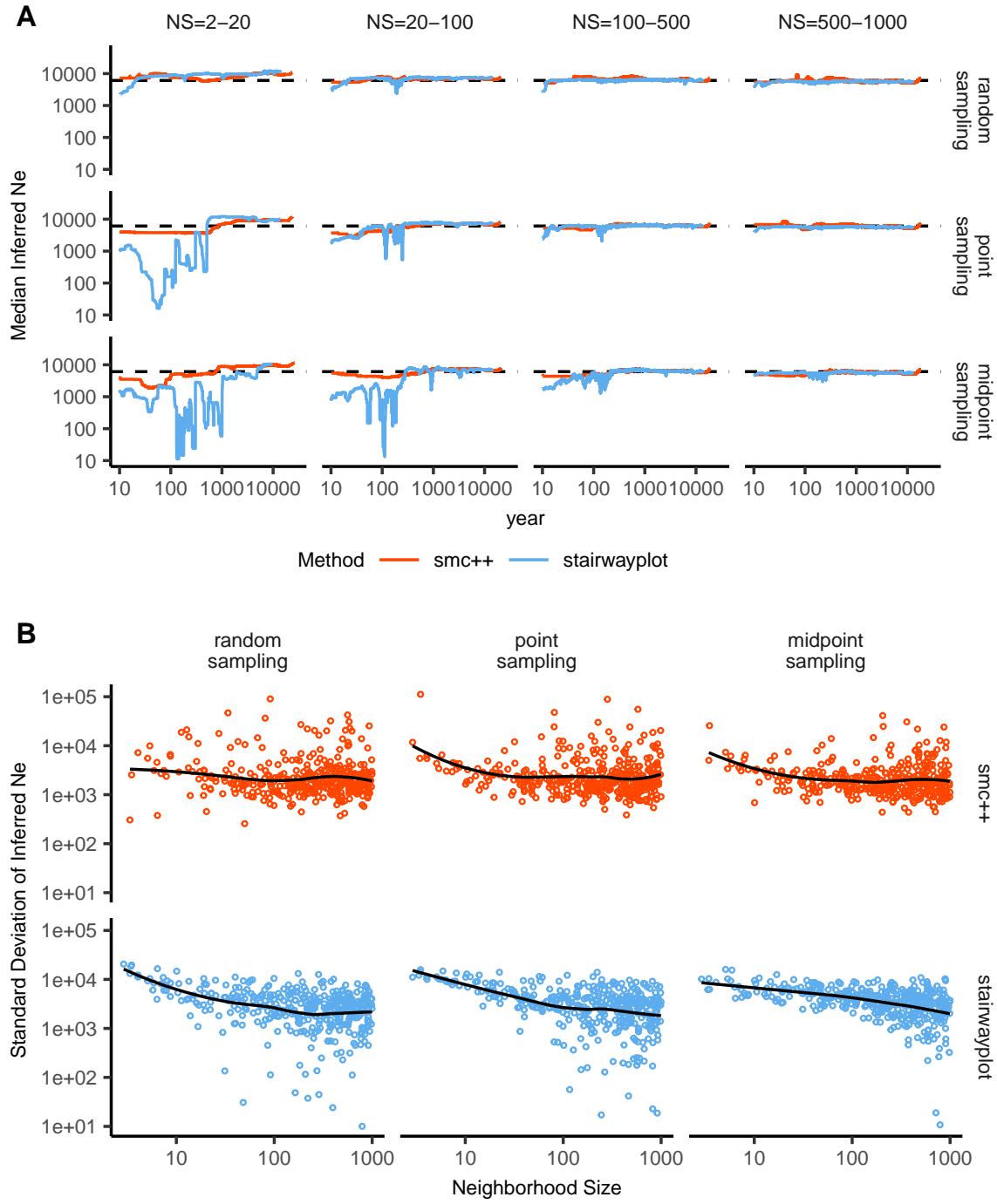
463 However, individual model fits from both methods frequently reflected turbulent demo-  
464 graphic histories (Figure S4), with the standard deviation of inferred  $N_e$  across time points  
465 often exceeding the expected  $N_e$  for both methods (Figure 6B). That is, despite the constant  
466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

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

471 **GWAS**

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

485 The confounding effects of geographic structure are well known, and it is common practice  
486 to control for this by including principal components (PCs) as covariates to control for these  
487 effects. This mostly works in our simulations – after incorporating the first ten PC axes as  
488 covariates, the vast majority of SNPs no longer surpass a significance threshold chosen to  
489 have a 5% false discovery rate (FDR). However, a substantial number of SNPs – up to 1.5% at  
490 the lowest dispersal distances – still surpass this threshold (and thus would be false positives  
491 in a GWAS), especially under “corner” and “patchy” environmental distributions (Figure 7C).  
492 At neighborhood sizes larger than 500, up to 0.31% of SNPs were significant for corner and  
493 clinal environments. Given an average of 132,000 SNPs across simulations after MAF filtering,  
494 this translates to up to 382 false-positive associations; for human-sized genomes, this number  
495 would be much larger. In most cases the  $p$  values for these associations were significant after  
496 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.

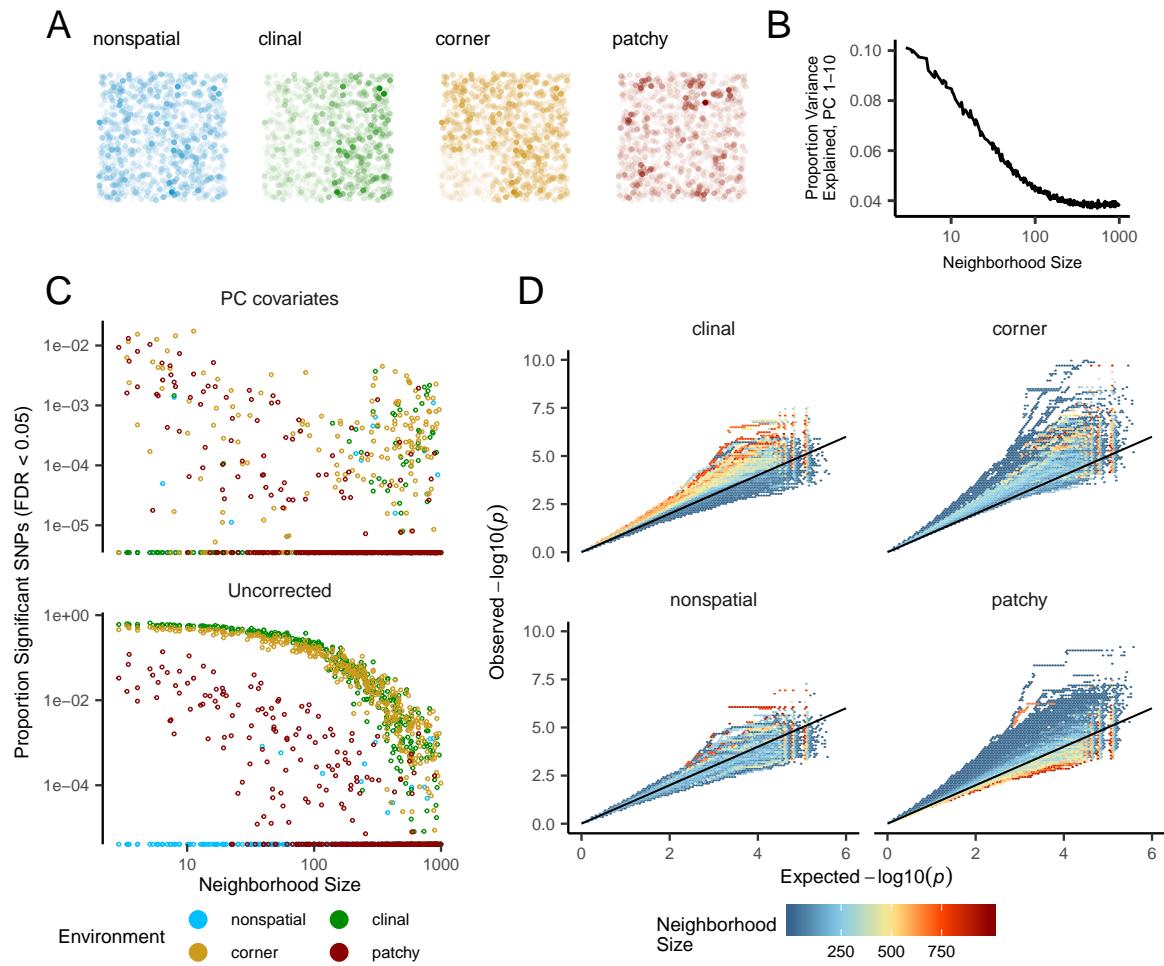
497 Bonferroni correction (see example Manhattan plots in figure S5).

498 Clinal environments cause an interesting pattern in false positives after PC correction:  
499 at low neighborhood sizes the correction removes nearly all significant associations, but at  
500 neighborhood sizes above roughly 250 the proportion of significant SNPs increases to up to  
501 0.4% (Figure 7). This may be due to a loss of descriptive power of the PCs – as neighborhood  
502 size increases, the total proportion of variance explained by the first 10 PC axes declines from  
503 roughly 10% to 4% (Figure 7B). Essentially, PCA seems unable to effectively summarize the  
504 weak population structure present in large-neighborhood simulations, but these populations  
505 continue to have enough spatial structure to create significant correlations between genotypes  
506 and the environment. A similar process can also be seen in the corner phenotype distribution,  
507 in which the count of significant SNPs initially declines as neighborhood size increases and  
508 then increases at approximately the point at which the proportion of variance explained by  
509 PCA approaches its minimum.

510 Figure 7D shows quantile-quantile plots that show the degree of genome-wide inflation of  
511 test statistics in PC-corrected GWAS across all simulations and environmental distributions.  
512 For clinal environments,  $-\log_{10}(p)$  values are most inflated when neighborhood sizes are  
513 large, consistent with the pattern observed in the count of significant associations after  
514 PC regression. In contrast corner and patchy environments cause the greatest inflation in  
515  $-\log_{10}(p)$  at neighborhood sizes less than 100, which likely reflects the inability of PCA to  
516 account for fine-scale structure caused by very limited dispersal. Finally, we observed that PC  
517 regression appears to overfit to some degree for all phenotype distributions, visible in Figure  
518 7D as points falling below the 1:1 line.

## 519 Discussion

520 In this study, we have used efficient forward time population genetic simulations to describe  
521 the myriad influence of continuous geography on genetic variation. In particular, we examine  
522 how three main types of downstream empirical inference are affected by unmodeled spatial  
523 population structure – 1) population genetic summary statistics, 2) inference of population  
524 size history, and 3) genome-wide association studies (GWAS). As discussed above, space often  
525 matters (and sometimes dramatically), both because of how samples are arranged in space,  
526 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.

527    **Effects of Dispersal**

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

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

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

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

563 **Effects of Sampling**

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

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

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

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

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

618    **Demography**

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

641    We found that SMC++ was more robust to the effects of space than stairwayplot, under-  
642    estimating recent populations by roughly half in the worst time periods rather than nearly  
643    10-fold as with stairwayplot. Though this degree of variation in population size is certainly  
644    meaningful in an ecological context, it is relatively minor in population genetic terms. A  
645    more worrying pattern was the high level of variance in inferred  $N_e$  trajectories for individual  
646    model fits using these methods, which was highest in simulations with the smallest neighbor-  
647    hood size (Figure 6, Figure S4). This suggests that, at a minimum, researchers working with  
648    empirical data should replicate analyses multiple times and take a rolling average if model

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

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

663 **GWAS**

664 Spatial structure is particularly challenging for genome-wide association studies, because  
665 the effects of dispersal on genetic variation are compounded by spatial variation in the  
666 environment (Mathieson and McVean 2012). Spatially restricted mate choice and dispersal  
667 causes variation in allele frequencies across the range of a species. If environmental factors  
668 affecting the phenotype of interest also vary over space, then groups of individuals in different  
669 regions will allele frequencies and environmental exposures will covary over space. In this  
670 scenario an uncorrected GWAS will infer genetic associations with a purely environmental  
671 phenotype at any site in the genome that is differentiated over space, and the relative degree  
672 of bias will be a function of the degree of covariation in allele frequencies and the environment  
673 (i.e., Figure 7C, bottom panel). This pattern has been demonstrated in a variety of simulation  
674 and empirical contexts (Price *et al.* 2006; Yu *et al.* 2005; Young *et al.* 2018; Mathieson and  
675 McVean 2012; Kang *et al.* 2008, 2010; Bulik-Sullivan *et al.* 2015; Berg *et al.* 2018; Sohail *et al.*  
676 2018).

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

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

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

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

710 interpretation of GWAS *p*-values, and correction using principal components is insufficient to  
711 fully separate these signals for polygenic traits under a variety of environmental and popu-  
712 lation parameter regimes. Other GWAS methods may be less sensitive to this confounding,  
713 but there is no obvious reason that this should be so. One approach to estimating the degree  
714 of bias in GWAS caused by population structure is LD score regression (Bulik-Sullivan *et al.*  
715 2015). Though this approach appears to work well in practice, its interpretation is not always  
716 straightforward and it is likely biased by the presence of linked selection (Berg *et al.* 2018).  
717 In addition, we observed that in many cases the false-positive SNPs we identified appeared  
718 to be concentrated in LD peaks similar to those expected from truly causal sites (Figure S5),  
719 which may confound LD score regression.

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

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

732 For how much of the tree of life do spatial patterns circumscribe genomic variation? In Table  
733 1 we gathered estimates of neighborhood size from a range of organisms to get an idea of  
734 how likely dispersal is to play an important role in patterns of variation. Though this sample  
735 is almost certainly biased towards small-neighborhood species (because few studies have  
736 quantified neighborhood size in species with very high dispersal or population density), we  
737 find that neighborhood sizes in the range we simulated are fairly common across a range of  
738 taxa. At the extreme low end of empirical neighborhood size estimates we see some flowering  
739 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)

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

#### 755 ***Future Directions and Limitations***

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

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

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

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

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

## 788 **Data Availability**

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

## 790 **Acknowledgements**

791 We thank Brandon Cooper, Matt Hahn, Doc Edge, and others for reading and thinking about  
792 this manuscript. CJB and ADK were supported by NIH award R01GM117241.

## 793 **Literature Cited**

794 Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, *et al.*, 2017 Deconstruct-  
795 ing isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics  
796 **13:** 1–27.

- 797 Antlfinger, A. E., 1982 Genetic neighborhood structure of the salt marsh composite, *Borrichia*  
798 *frutescens*. *Journal of Heredity* **73**: 128–132.
- 799 Antolin, M. F., B. V. Horne, M. D. Berger, Jr., A. K. Holloway, J. L. Roach, *et al.*, 2001 Effective  
800 population size and genetic structure of a piute ground squirrel (*Spermophilus mollis*)  
801 population. *Canadian Journal of Zoology* **79**: 26–34.
- 802 Ashander, J., P. Ralph, E. McCartney-Melstad, and H. B. Shaffer, 2018 Demographic inference  
803 in a spatially-explicit ecological model from genomic data: a proof of concept for the mojave  
804 desert tortoise. *bioRxiv* .
- 805 Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, *et al.*, 2016 The great  
806 migration and african-american genomic diversity. *PLOS Genetics* **12**: 1–27.
- 807 Barton, N. H., F. Depaulis, and A. M. Etheridge, 2002 Neutral evolution in spatially continuous  
808 populations. *Theoretical Population Biology* **61**: 31–48.
- 809 Barton, N. H., J. Kelleher, and A. M. Etheridge, 2010 A new model for extinction and recolo-  
810 nization in two dimensions: Quantifying phylogeography. *Evolution* **64**: 2701–2715.
- 811 Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing  
812 under dependency. *The Annals of Statistics* **29**: 1165–1188.
- 813 Berg, E. E. and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. *Evolution*  
814 **49**: 110–120.
- 815 Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, *et al.*, 2018  
816 Reduced signal for polygenic adaptation of height in uk biobank. *bioRxiv* .
- 817 Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 Ld score regression  
818 distinguishes confounding from polygenicity in genome-wide association studies. *Nature*  
819 *Genetics* **47**: 291 EP –.
- 820 Campbell, D. R. and J. L. Dooley, 1992 The spatial scale of genetic differentiation in a  
821 hummingbird-pollinated plant: Comparison with models of isolation by distance. *The*  
822 *American Naturalist* **139**: 735–748.
- 823 Chapman, N. H. and E. A. Thompson, 2002 The effect of population history on the lengths of  
824 ancestral chromosome segments. *Genetics* **162**: 449–458.
- 825 Chikhi, L., V. C. Sousa, P. Luisi, B. Goossens, and M. A. Beaumont, 2010 The confounding  
826 effects of population structure, genetic diversity and the sampling scheme on the detection  
827 and quantification of population size changes. *Genetics* **186**: 983–995.

- 828 Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, *et al.*, 1993 Direct  
829 and indirect estimates of neighborhood and effective population size in a tropical palm,  
830 *astrocaryum mexicanum*. *Evolution* **47**: 75–87.
- 831 Epperson, B., 2003 *Geographical Genetics*. Monographs in Population Biology, Princeton Uni-  
832 versity Press.
- 833 Felsenstein, J., 1975 A pain in the torus: Some difficulties with models of isolation by distance.  
834 *The American Naturalist* **109**: 359–368.
- 835 Fox, J. and S. Weisberg, 2011 *An R Companion to Applied Regression*. Sage, Thousand Oaks CA,  
836 second edition.
- 837 Garcia, J. and C. Quintana-Domeque, 2006 The evolution of adult height in europe: A brief  
838 note. Working Paper .
- 839 Garcia, J. and C. Quintana-Domeque, 2007 The evolution of adult height in europe: A brief  
840 note. *Economics & Human Biology* **5**: 340 – 349.
- 841 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in  
842 north american *drosophila melanogaster* show signatures of soft sweeps. *PLOS Genetics* **11**:  
843 1–32.
- 844 Griffiths, R., S. Tavaré, *et al.*, 1999 The ages of mutations in gene trees. *The Annals of Applied  
845 Probability* **9**: 567–590.
- 846 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence  
847 recording in SLiM opens new horizons for forward-time simulation of whole genomes.  
848 *Molecular Ecology Resources* **19**: 552–566.
- 849 Haller, B. C. and P. W. Messer, 2019 Slim 3: Forward genetic simulations beyond the wright–  
850 fisher model. *Molecular biology and evolution* **36**: 632–637.
- 851 Harris, K. and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared  
852 haplotype lengths. *PLOS Genetics* **9**: 1–20.
- 853 Huillet, T. and M. Möhle, 2011 On the extended Moran model and its relation to coalescents  
854 with multiple collisions. *Theoretical Population Biology* pp. –.
- 855 Jasper, M., T. Schmidt, N. Ahmad, S. Sinkins, and A. Hoffmann, 2019 A genomic approach to  
856 inferring kinship reveals limited intergenerational dispersal in the yellow fever mosquito.  
857 *bioRxiv* .
- 858 Jay, F., P. Sjödin, M. Jakobsson, and M. G. Blum, 2012 Anisotropic Isolation by Distance: The

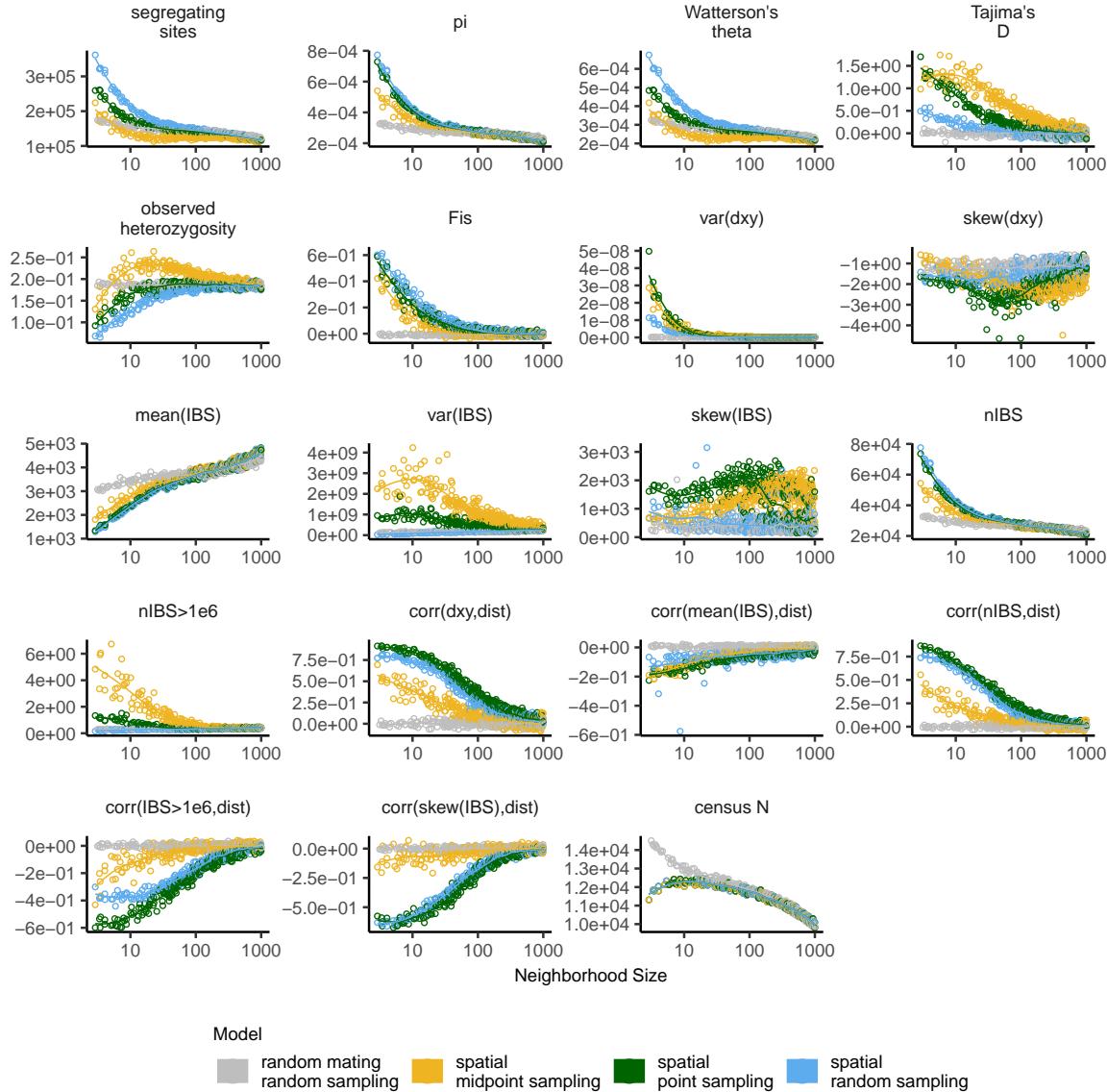
- 859 Main Orientations of Human Genetic Differentiation. *Molecular Biology and Evolution* **30**:  
860 513–525.
- 861 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component  
862 model to account for sample structure in genome-wide association studies. *Nature Genetics*  
863 **42**: 348 EP –.
- 864 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, *et al.*, 2008 Efficient control  
865 of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- 866 Kelleher, J., A. Etheridge, and N. Barton, 2014 Coalescent simulation in continuous space:  
867 Algorithms for large neighbourhood size. *Theoretical Population Biology* **95**: 13 – 23.
- 868 Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and ge-  
869 nealogical analysis for large sample sizes. *PLoS Comput Biol* **12**: 1–22.
- 870 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for  
871 fast population genetics simulation. *PLOS Computational Biology* **14**: 1–21.
- 872 Kerster, H. W., 1964 Neighborhood size in the rusty lizard, *sceloporus olivaceus*. *Evolution* **18**:  
873 445–457.
- 874 Khera, A. V., M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, *et al.*, 2018 Genome-wide poly-  
875 genic scores for common diseases identify individuals with risk equivalent to monogenic  
876 mutations. *Nature Genetics* **50**: 1219–1224.
- 877 Kingman, J., 1982 The coalescent. *Stochastic Processes and their Applications* **13**: 235 – 248.
- 878 Lawson, D. J., N. M. Davies, S. Haworth, B. Ashraf, L. Howe, *et al.*, 2019 Is population structure  
879 in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* .
- 880 Liu, X. and Y.-X. Fu, 2015 Exploring population size changes using snp frequency spectra.  
881 *Nature Genetics* **47**: 555 EP –.
- 882 Lundgren, E. and P. L. Ralph, 2018 Are populations like a circuit? The relationship between  
883 isolation by distance and isolation by resistance. *bioRxiv* .
- 884 Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, *et al.*, 2019 Clinical use of current  
885 polygenic risk scores may exacerbate health disparities. *Nature Genetics* **51**: 584–591.
- 886 Maruyama, T., 1972 Rate of decrease of genetic variability in a two-dimensional continuous  
887 population of finite size. *Genetics* **70**: 639–651.
- 888 Mathieson, I. and G. McVean, 2012 Differential confounding of rare and common variants in  
889 spatially structured populations. *Nature Genetics* **44**: 243 EP –.

- 890 Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi, 2015 On the importance of being  
891 structured: instantaneous coalescence rates and human evolution—lessons for ancestral  
892 population size inference? *Heredity* **116**: 362 EP –.
- 893 Miles, A. and N. Harding, 2017 *cgh/scikit-allel*: v1.1.8.
- 894 Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, *et al.*, 2013 Estimation of effec-  
895 tive population size in continuously distributed populations: there goes the neighborhood.  
896 *Heredity* **111**: 189 EP –.
- 897 Novembre, J. and M. Slatkin, 2009 Likelihood-based inference in isolation-by-distance models  
898 using the spatial distribution of low-frequency alleles. *Evolution* **63**: 2914–2925.
- 899 Pamilo, P., 1983 Genetic differentiation within subdivided populations of *formica* ants. *Evolu-  
900 tion* **37**: 1010–1022.
- 901 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLOS  
902 Genetics* **2**: 1–20.
- 903 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, *et al.*, 2006 Principal  
904 components analysis corrects for stratification in genome-wide association studies. *Nature  
905 Genetics* **38**: 904 EP –.
- 906 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using  
907 multilocus genotype data. *Genetics* **155**: 945–959.
- 908 Ptak, S. E. and M. Przeworski, 2002 Evidence for population growth in humans is confounded  
909 by fine-scale population structure. *Trends in Genetics* **18**: 559–563.
- 910 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.*, 2007 Plink: A tool set for  
911 whole-genome association and population-based linkage analyses. *The American Journal  
912 of Human Genetics* **81**: 559 – 575.
- 913 R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for  
914 Statistical Computing, Vienna, Austria.
- 915 Ralph, P. and G. Coop, 2013 The geography of recent genetic ancestry across Europe. *PLoS  
916 Biol* **11**: e1001555.
- 917 Ringbauer, H., G. Coop, and N. H. Barton, 2017 Inferring recent demography from isolation  
918 by distance of long shared sequence blocks. *Genetics* **205**: 1335–1351.
- 919 Robledo-Arnuncio, J. J. and F. Rousset, 2010 Isolation by distance in a continuous population  
920 under stochastic demographic fluctuations. *Journal of Evolutionary Biology* **23**: 53–71.

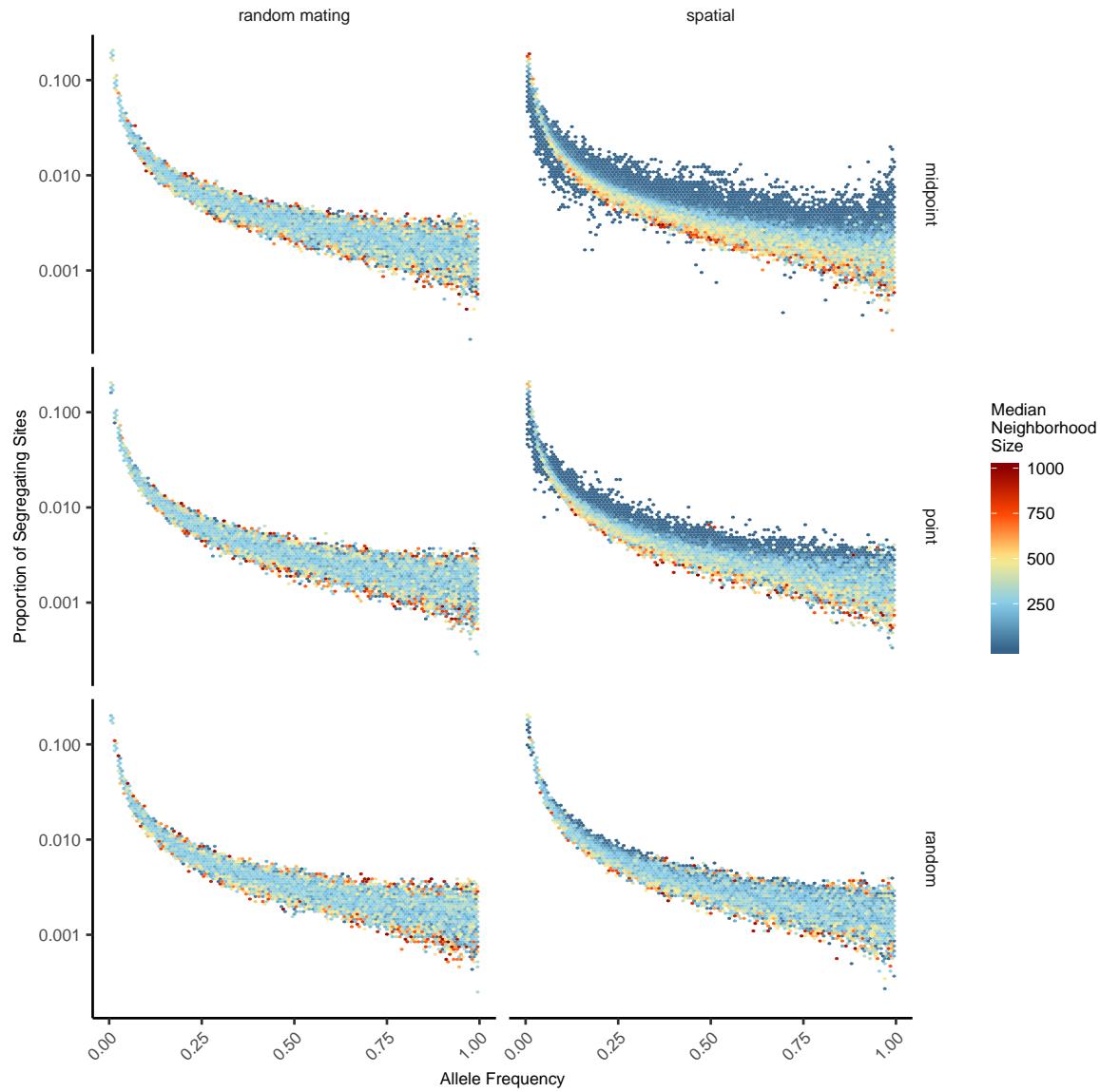
- 921 Rousset, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under  
922 isolation by distance. *Genetics* **145**: 1219–1228.
- 923 Rousset, F. and R. Leblois, 2011 Likelihood-based inferences under isolation by distance:  
924 Two-dimensional habitats and confidence intervals. *Molecular Biology and Evolution* **29**:  
925 957–973.
- 926 Sawyer, S., 1977 On the past history of an allele now known to have frequency p. *Journal of  
927 Applied Probability* **14**: 439–450.
- 928 Schiffels, S. and R. Durbin, 2014 Inferring human population size and separation history from  
929 multiple genome sequences. *Nature Genetics* **46**: 919 EP –.
- 930 Schrider, D. R. and A. D. Kern, 2018 Supervised machine learning for population genetics: A  
931 new paradigm. *Trends in Genetics* **34**: 301 – 312.
- 932 Sharbel, T. F., B. Haubold, and T. Mitchell-Olds, 2000 Genetic isolation by distance in ara-  
933 bidopsis thaliana: biogeography and postglacial colonization of europe. *Molecular Ecology*  
934 **9**: 2109–2118.
- 935 Sheehan, S., K. Harris, and Y. S. Song, 2013 Estimating variable effective population sizes from  
936 multiple genomes: A sequentially markov conditional sampling distribution approach.  
937 *Genetics* **194**: 647–662.
- 938 Shirk, A. J. and S. A. Cushman, 2014 Spatially-explicit estimation of wright's neighborhood  
939 size in continuous populations. *Frontiers in Ecology and Evolution* **2**: 62.
- 940 Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, *et al.*, 2018 Signals of polygenic  
941 adaptation on height have been overestimated due to uncorrected population structure in  
942 genome-wide association studies. *bioRxiv* .
- 943 St. Onge, K. R., A. E. Palmé, S. I. Wright, and M. Lascoux, 2012 Impact of sampling schemes on  
944 demographic inference: An empirical study in two species with different mating systems  
945 and demographic histories. *G3: Genes, Genomes, Genetics* **2**: 803–814.
- 946 Städler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber, 2009 The impact of sam-  
947 pling schemes on the site frequency spectrum in nonequilibrium subdivided populations.  
948 *Genetics* **182**: 205–216.
- 949 Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA poly-  
950 morphism. *Genetics* **123**: 585–595.
- 951 Terhorst, J., J. A. Kamm, and Y. S. Song, 2016 Robust and scalable inference of population

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

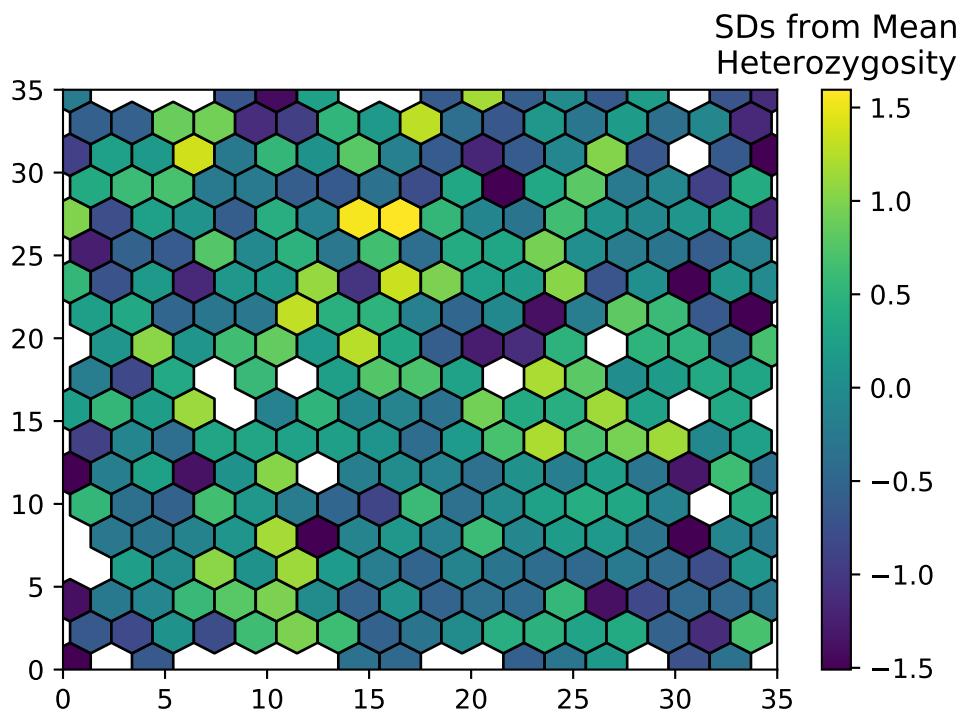
981 **Supplementary Figures and Tables**



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

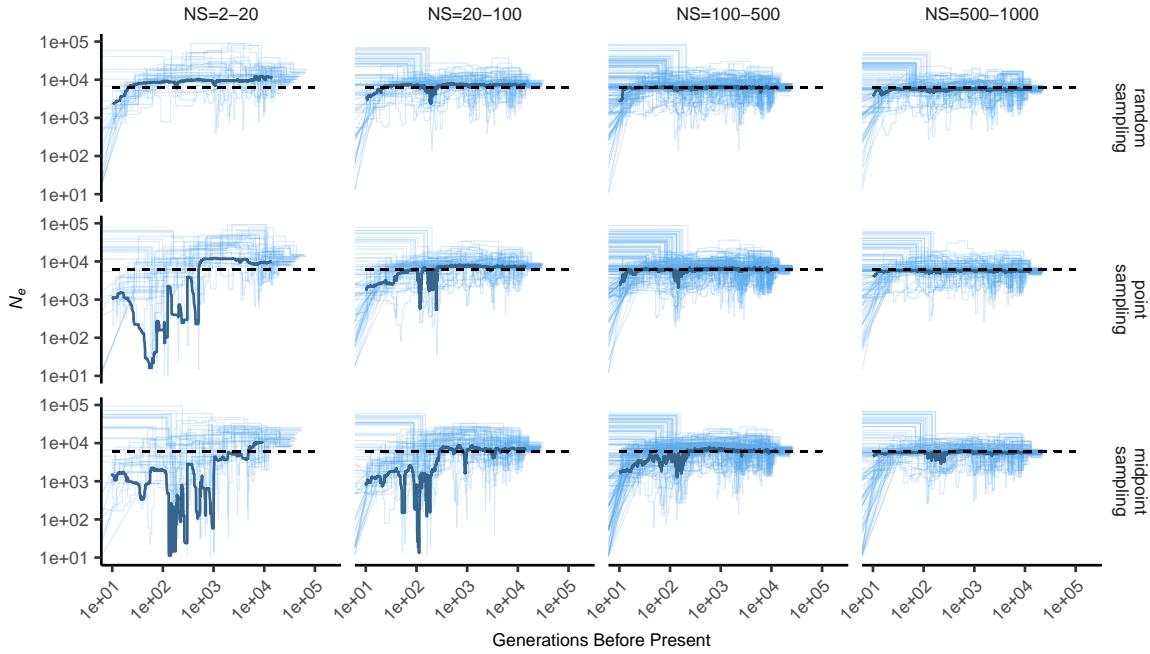


**Figure S2** Site frequency spectra for random mating and spatial SLiM models under all sampling schemes.

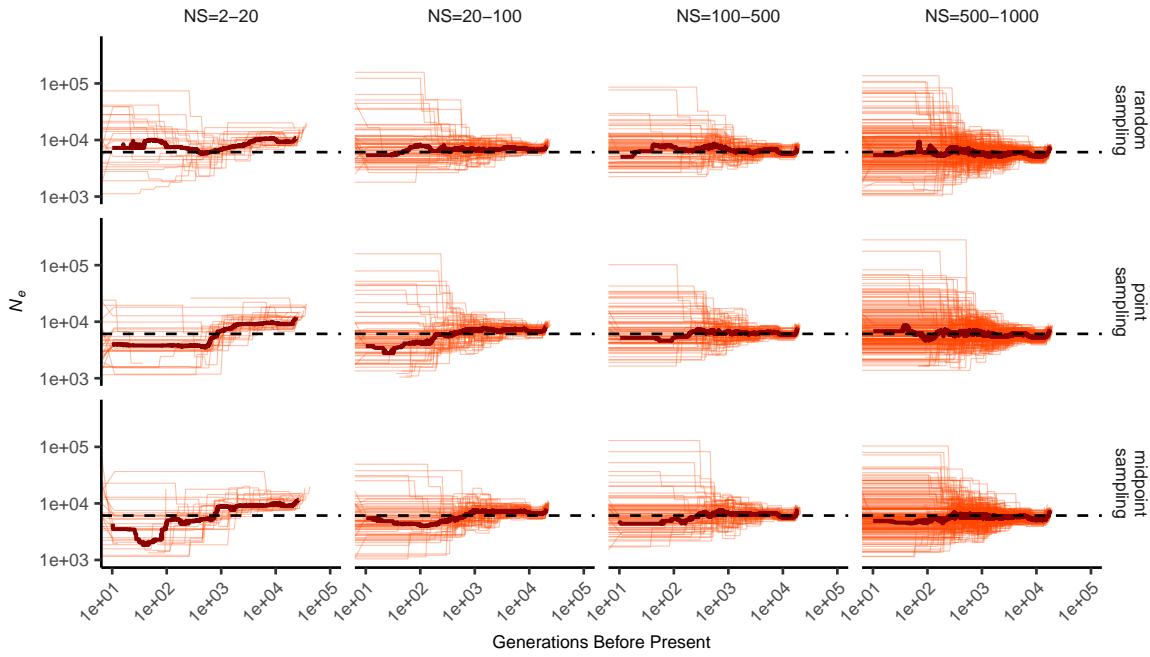


**Figure S3** Normalized mean observed heterozygosity by location across 200 randomly sampled individuals

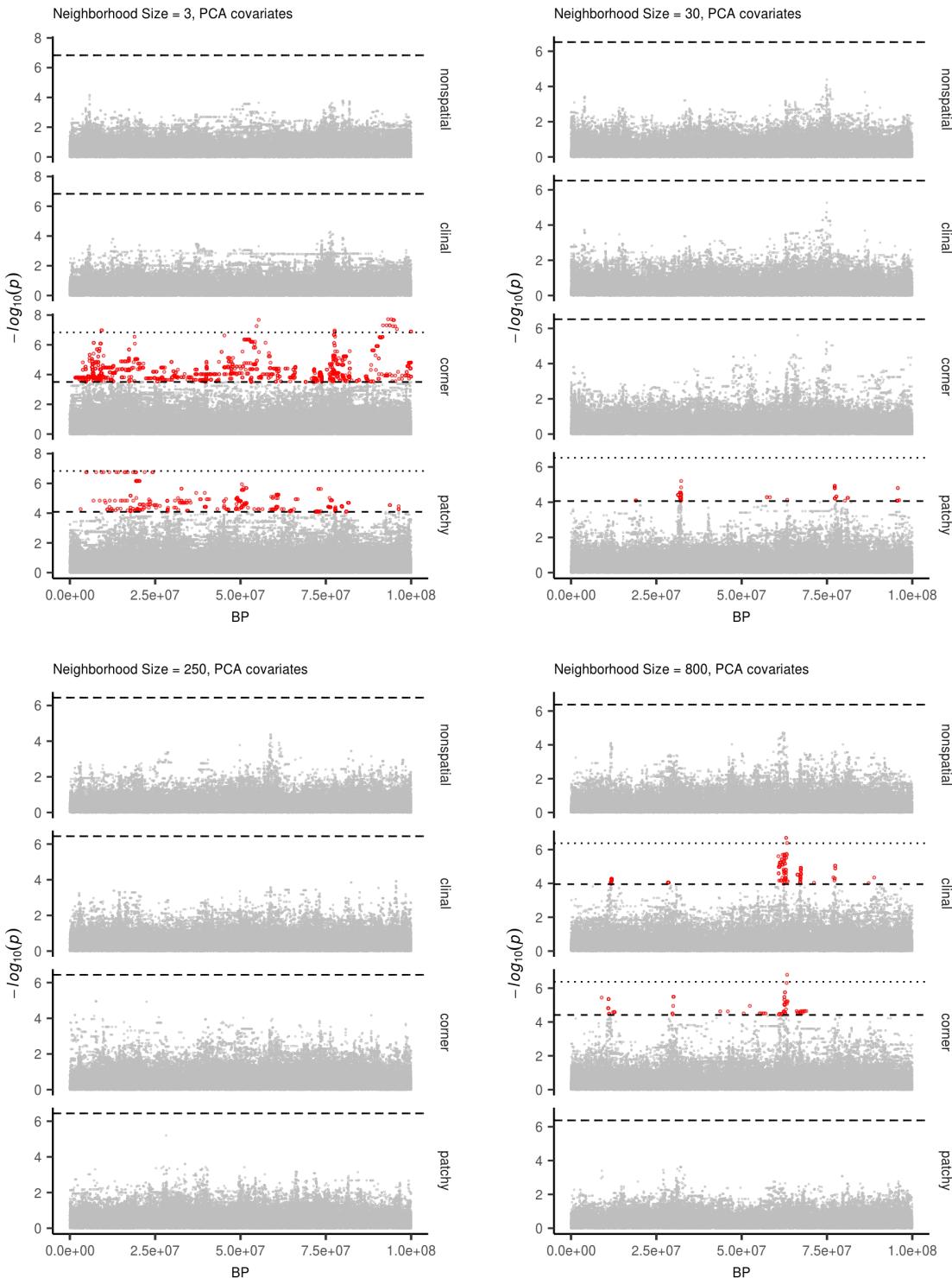
Stairwayplot



SMC++



**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
$mean(IBS)$	Mean of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$var(IBS)$	Variance of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$skew(IBS)$	Skew of the distribution of pairwise identical-by-state (IBS) tract lengths taken over all pairs.
$nIBS$	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 $\log_{10}(spatial\ distance)$
$corr(mean(IBS), dist)$	Pearson correlation between the mean of the IBS tract distribution for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(nIBS, dist)$	Pearson correlation between the number of IBS tracts for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(IBS > 1e6, dist)$	Pearson correlation between the number of IBS tracts > $1 \times 10^6$ bp for each pair of samples and $\log_{10}(spatial\ distance)$
$corr(skew(IBS), dist)$	Pearson correlation between the skew of the distribution of pairwise haplotype block lengths for each pair of samples and $\log_{10}(spatial\ distance)$

**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>

Resubmission Cover Letter  
*Genetics*

C. J. Battey,  
Peter Ralph,  
*and* Andrew Kern  
Wednesday 14<sup>th</sup> August, 2019

**To the Editor(s) –**

We are writing to submit a revised version of our manuscript, “Space is the Place: Effects of Continuous Spatial Structure on Analysis of Population Genetic Data”.

**Sincerely,**

**C. J. Battey, Peter Ralph, and Andrew Kern**

**Reviewer AE:**

The manuscript admirably explores a lot of consequences of isolation-by-distance in the context of a novel model that is easily amenable to forward simulation; however, given that this model may be used in a lot of future studies based on the precedent set here, there is some concern about the model and its support. Reviewers 2 and 3 highlight this in particular (it underlies the main 2 points of reviewer 2's review, and the core of Reviewer 3's comment), and I agree. Whatever can be done to strengthen the standing of this model, and/or connect it to more thoroughly studied models, will be helpful for the manuscript. The concern would be that there are peculiarities of this model that do not generalize well. A new supplemental section or opener to the results section establishing the model more thoroughly would make the strongest response.

*(I would generally cut down the quoted bits like the above to only what's essential, but haven't done that yet.)*

*(IMPORTANT: don't reorder or delete "points" below - it messes up the automatic numbering!)*

---

**(AE.1) Line 35:** Also cite Wilkins and Wakeley, *Genetics* 2002; Wilkins 2004

**Reply:**

---

**(AE.2) (p. 5, l. 144)** "Such models have been used extensively in ecological modeling but rarely in population genetics" Detailing these previous uses via citations and elaboration may help alleviate the major concern about the provenance of this model and its unique behaviors (see general comments above and R2 and R3 comments).

**Reply:**

---

**(AE.3) (p. 7, l. 190)** Please describe computation time needed per replicate

**Reply:**

---

**(AE.4) (p. 32, l. 792)** I read the acknowledgement to the Hearth and Creative Sky Brewing with a sense of familiarity in feeling of gratitude to my own favorite cafes and breweries, but I it's not a great precedent for Acknowledgements to be filled this way. Please cut.

**Reply:** Good point; we have done this.

---

**(AE.5) Figure 4:** Show random-mating expectation

**Reply:**

---

**(AE.6) Figure 3A, S2:** Perhaps more revealing to show on log-log scale?

**Reply:**

---

**(AE.7) Figure S3:** Caption seems to be missing detail

**Reply:**

**Reviewer 1:**

This study explores biases arising in population-based inference when 1) real population samples are coming from spatial habitat with various degree of structuring while inference is made assuming random mating population; 2) imperfect sampling in practice that fails to represent full diversity across entire population habitat; 3) phenotypes that vary across geography and create spurious associations with genotypes. While earlier studies explored the effect of strong structure on population genetic inference and GWAS, this work focuses on less extreme scenarios of structuring that arises in populations evolving in continuous habitat. By using non-Wright-Fisher model, authors simulated chromosome-scale samples from populations that evolved in continuous space, and that can model environmental factors to create phenotypes varying over space. As a result, this study identified spatial structuring scenarios (small neighbourhood size 10-100) that coupled with imperfect sampling strategies lead to a biased inference of widely used population genetic statistics (altogether 18 statistics) such as pi (average pairwise sequence differences), heterozygosity (and inbreeding coefficient), and IBS tract sharing. Accordingly, inference of the effective population size history was also strongly affected under these parameter ranges. Finally, the authors use their spatial modelling to demonstrate that typical GWAS with PC-based correction cannot entirely remove spurious signals of genotype-phenotype associations arising from purely environmental factors. Overall, the authors explore an important but

often neglected source of bias that can affect inference in many population-based studies (in medical genetics, evolutionary biology and ecology). This study can be of interest to a broader audience of readership, and I have only minor comments to improve clarity and increase accessibility for readers:

---

(1.1) When neighbourhood size is small (10-100), the mean number of IBS tracts  $> 2bp$  ( $nIBS$  as in Table S1) is elevated similar to Wright's inbreeding coefficient, but mean of the distribution of pairwise IBS (mean(IBS)) is decreased. What could be the source of this discrepancy? How exactly mean(IBS) was calculated?

Reply:

---

(1.2) The authors use  $K$  to denote both carrying capacity (p. 6, l. 167) and population density (p. 6, l. 170). It might be better to use a different notation for these quantities since carrying capacity is fixed while density is an emergent quantity in the non-Wright-Fisher model. Use of  $K$  to denote carrying capacity and density is a bit confusing. For example, on Page 12, line 309, it is said that 'the "population density" ( $K$ ) and "mean lifetime" ( $L$ ) parameters were the same in all simulations'. Here  $K$  seems to indicate carrying capacity rather than density? The latter is an emergent quantity and varies across simulation runs?

Reply: We agree that this distinction is worth emphasizing! We've adjusted our language to hopefully remind the reader that  $K$  is a parameter that controls population density, rather than being equal to it, at (p. 6, l. 170) and (p. 7, l. 202) and (p. 12, l. 324).

---

(1.3) Concerning the non-Wright-Fisher model used, it would be helpful to emphasize that some of the parameters are emergent in contrast to Wright-Fisher model. For example, on Page 11, lines 306-308, the author's goal was to look at census size variation and variation in other quantities. This would be better understood if to emphasize that these parameters are emergent properties in the non-Wright-Fisher model used.

Reply:

---

(1.4) Page 9, line 242, Perhaps 'Demographic Inference' might better reflect the content of this section.

Reply:

**(1.5) (p. 10, l. 290)** This sentence with 'Gaussian noise with mean zero and standard deviation 10' is confusing since it was mentioned earlier that the modelled phenotype must vary as human height across Europe, and human height varies 2 standard deviations. Only after reading the whole paragraph it becomes clear that 'standard deviation 10' here refers to unit of height. Please consider rephrasing this sentence.

**Reply:**

---

**(1.6) (p. 11, l. 313)** In the sentence, 'We also examined p values for systemic inflation' I think the authors meant 'systematic inflation'.

**Reply:** Whoops; thanks. Fixed.

---

**(1.7)** Page 11, Please correct the legend in Figure 2: must be 'spatial model' and 'random mating' model.

**Reply:**

---

**(1.8)** Optional: a dashed line in Figure 2 that shows the total carrying capacity of  $50*50*5=12500$  would be helpful.

**Reply:**

---

**(1.9)** Page 13, line 349, The phrase 'affect summaries of variation' is better to replace with 'summaries of genetic variation'.

**Reply:**

---

**(1.10)** Please add or correct references to supplementary figures: For example, Figure S2 was probably meant to accompany Figure 3A, while Figure S1 Figure 3B, but references in the text are absent. In fact, the first reference is made to Figure S3 on page 15.

**Reply:**

---

**(1.11)** There are also several typos and errors in the text. For example, on Page 12, lines 309; Page 27, line 655.

**Reply:**

**Reviewer 2:**

Battey et al. use spatially explicit population genetic simulations to analyze the effects of spatial structure on (i) the estimation of key population genetic parameters, in turn used to (ii) make inferences about population history, and on (iii) confounding in genome-wide association studies (GWAS). I Liked the paper a lot. It's interesting, well-written and addresses an important question - the effect of spatial population structure on population genetic statistics and inference-and I enjoyed reading it. The most positive aspects were:

1. It nice to actually see spatially explicit simulations and I'm happy that forward simulation is now fast enough that you can do this sort of thing.
2. The paper is very clear and well-written, easy to understand the motivation and most of the details. That's not always the case for this sort of paper.
3. I felt that the section about the effect on GWAS was the most interesting and novel part of the paper and gave me some intuition that I hadn't had before.

I don't have any major criticisms. There were a few aspects that I thought might warrant some additional discussion, and a few specific questions below. The general questions I had after reading it were:

---

**(2.1)** *To what extent are any of the results dependent on the exact method of simulation. There are a number of choices about the exact details of the simulations (e.g. the way the overlapping generations are handled, the edge effects and, particularly, the form of Equation 1 - see below). It's not so much that these are non-standard (since I don't think there is a standard) and they all sort of make sense heuristically, and I was left wondering whether these sorts of choices actually make a difference. Do the authors have some thoughts/intuition/results about that? Given that the results in Fig. 3 seem quite consistent with expectations, I suspect that on some level it doesn't make much difference but then there are intermediate results like Fig. 2 which seem a bit counter-intuitive and I wonder if those aspects depend on the simulation scheme.*

**Reply:**

---

**(2.2)** Related to the first point, to what extent are the results qualitatively different to those that would be obtained in a stepping-stone model? My interpretation is that they are actually very similar, but I didn't see whether that was explicitly discussed. In some sense, it's still easier to do large simulations in a stepping-stone model so it would be nice to be reassured that that's still ok.

**Reply:**

---

**(2.3)** The source of equation (1) is not obvious to me. I sort of see how it makes sense, but a little bit more intuition or a brief derivation or an illuminating either in the main text or the supplement, would be helpful.

**Reply:**

---

**(2.4)** The authors use a scaling factor in equation (2) to counteract the increase in fitness of individuals at the edges. Can they provide a figure showing that this is the case. What does "roughly" mean on line 164. Perhaps a heatmap of the fitness of individuals across the grid with and without the scaling factor?

**Reply:**

---

**(2.5)** It would be helpful provide the figure showing that generating mutations during the forward simulations in SLIM is equivalent to applying mutations using msprime on pre-generated trees (line 185)? It sounds like this procedure would underestimate the variance in the number of mutations, since you remove the effect of random generation time. Is this effect small?

**Reply:**

---

**(2.6)** Can the authors provide a bit more intuition behind the patterns of variation seen in generation time, census population size, and variance in the number of offspring with respect to neighborhood size seen in Figure 2? For example, it is not obvious to me why the census population size, for example, should decline systematically with respect to neighborhood size. Presumably this isn't just due to the local demographic stochasticity. Could the authors briefly interpret the observed patterns or cite appropriate literature?

**Reply:**

**(2.7)** Fig. 7D: I am surprised by the extent to which the observed values of  $-\log_{10}(p)$  fall below the  $y=x$  line. Particularly in the lower right panel for large neighbourhood sizes. I would expect that to be close to panmictic - why are the P-values underdispersed? That seems like a potential bug, or else something weird is going on.

**Reply:**

---

**(2.8)** Lines 706-716, It might be worth citing Haworth et al *Nature Communications* 2019 (<https://doi.org/10.1038/s41467-018-08219-1>) who do the proposed test (GWAS for birth location) in UK Biobank to illustrate the population structure.

**Reply:**

---

**(2.9)** The analysis and discussion around the effect of GWAS is focused on PCA correction. Do mixed models help at all?

**Reply:**

---

**(2.10)** The github link to the code didn't work for me. I assume it will be made public later, but at this point I can't tell whether the code is available/useable.

**Reply:**

### **Reviewer 3:**

The present study deals with a "hot topic" in spatial population genetics. Most inferential and descriptive methods in statistical spatial population genetic rely on a discrete approximation of space and it is not clear what impact this approximation may have when individuals migrate along a continuum instead. Spatial patterns in sampling is also another major issue which is often simply dismissed, mainly because of the paucity of statistical methods to deal with it. This work touches on these important issues in a timely manner.

Although I was enthusiastic about the topic, I was quite disappointed with the core of the study, i.e., the forward-in-time simulation of populations in continuous space. The field has been struggling with this issue for decades – examples of spectacular failures like the Wright-Malecot model (see Felsenstein's "pain

in the torus' article, 1975) or, more recently, the "mugration" or "discrete trait analysis" model in phylodynamics (see De Maio et al. 2015) have probably mostly harmed our research field – that one cannot make the economy of using a sound probabilistic model for generating geo-referenced genetic data. It does not seem to be the case here unfortunately.

---

**(3.1)** *First, the simulation starts with individuals distributed uniformly at random in space. Is there any indication that the three-step algorithm used here maintains this distribution during the course of evolution? If it does not, then is there any stationary regime and how many generations does one need to wait before reaching it? I do appreciate that the competitive interaction term was introduced in order to avoid seeing the "clumping" of individuals that hampers the Wright-Malecot model. Yet, just because there are no such clusters does not mean that the spatial distribution of individuals reaches a stable regime and that the distribution reached, if any, is reasonable from a biological perspective.*

**Reply:**

Second, the demographic process used here involves birth and death of individuals. Does the population survive asymptotically or, like any birth-death process, eventually dies with probability one? In fact, one needs to know a little about the dynamics of the population size to decide whether the corresponding process is reasonable from a biological standpoint.

---

**(3.2)**

**Reply:**

---

**(3.3)** *Third, it is not clear what the relationship between the expected lifespan and the probability of survival is. The expected lifespan,  $L$ , is first defined as the inverse of the expected number of offspring produced by a parent. The authors also define the probability of survival of a given individual at a given point in space,  $p_i$ . Hence, the expected lifespan at a point in space (and time) is the mean of a geometric distribution with parameter  $p_i$ , i.e.,  $1/p_i$ . Now, it is far from being obvious what the relationship between these two approaches for defining the expected lifespan actually is.*

**Reply:**

---

(3.4) Also, the web page <https://github.com/petrelharp/spaceness> does not seem to exist so that I was not able to experiment with the forward-in-time generator used here unfortunately.

Reply:

---

(3.5) All in all, more efforts need to be made here in my opinion to show that the forward-in-time simulations generate sensible outcomes. Sensible in terms of the behavior of the population demography at equilibrium (provided such equilibrium indeed exists) along with that of the spatial distribution of individuals. The authors could provide some guarantee of the good behavior of their model as evidenced from simulations using a broad range of parameter values for generating data. Alternatively, they could elect to use the spatial-Lambda-Fleming-Viot model for their simulations, which, in my opinion would seem the most sensible option given that (1) it is possible to run backward-in-time simulations under this model, thereby saving a lot of computation time and (2) it is a well-studied model with good mathematical and biological properties and (3) it is implemented in a publicly available software program (<https://github.com/jeromekelleher/discsim>)

Reply:

---

(3.6) Figure 2: I do not understand why the neighborhood size varies to the same extent in the random mating model as it does for the spatial model. For the random mating model, I would have expected the neighborhood size to be equal to the census size since all individuals have the same probability of being a parent of any given offspring. From lines 166->171, it is clear that the spatial model would converge to the random mating model when the mean parent-offspring distance tends to infinity only if we were to ignore the impact of range edges. I am thus wondering whether the variation of neighborhood size one observes in Fig 2 for the random mating model is just a consequence of border effects. If that is the case, then the authors should state it clearly and try to justify it from a biological perspective.

Reply:

---

(3.7) Line 729-731: "Many more species occur in a middle range of neighborhood sizes between 100 and 1000 - a range in which spatial processes play a minor role in our analyses [...]" Do the authors think that the spatial processes would still play a minor role when neighborhood sizes exceed 100-1000 if the habitat was larger than that taken in the present

*simulations? It would also probably be useful to mention that neighborhood sizes given in Table 1 should be compared with extreme caution since the size of the corresponding habitats vary across species. More generally, I suspect that the size of the habitat has a substantial impact on the vast majority of statistics examined in this study. Indeed, the mean parent-offspring distance, which is at the core of the definition of Wright's neighborhood size, is only small or large relative to the size of the habitat.*

**Reply:**

---

**(3.8)** Line 753-757: please add a reference to Guindon, Guo and Welch (2016). This study clearly shows that population density and dispersal parameters are identifiable and can indeed be estimated in practice under the spatial Lambda-Fleming-Viot model.

**Reply:**

**Reviewer 4:**

The manuscript by Battey et al explores the consequence of a well-known violation to population genetic models: the fact that populations are spatially structured and mate along a geographical cline, rather than randomly. This topic is important, particularly in light of recent working describing how spatially correlated genetic and environmental impacts can confound some population genetic insights, such as positive selection for height in Europe. The analyses and investigations presented here are thorough and sensible, and my comments are primarily intended to broaden accessibility for this interesting topic.

---

**(4.1) Introduction.** The discussion is very clear, articulating the three primary goals of the project: the impact of failing to model spatial population structure on 1) population genetic summary statistics, 2) inference on demographic history from population genetic data, and 3) impacts on GWAS summary statistics. I found the discussion a bit easier to follow than the introduction and would suggest streamlining and introducing the topic a bit more. Since the paper follows the flow described in the discussion, it might help orient readers by introducing these topics in the same order.

**Reply:**

---

**(4.2)** I agree that most modern work describes structure as discrete populations connected by migration. However, some methods/studies have explicitly modeled spatial structure, e.g. especially in ecology or using methods like dadi (diffusion approximations). Highlighting some examples of previously identified structure not possible to infer without modeling geography would be helpful to contextualize this work.

**Reply:**

---

**(4.3)** There is some reference to spatial models using grids (e.g. Rousset 1997). Some additional discussion contextualizing more recent methods like EEMS that also construct demes and model migration through divergence between neighboring demes would be helpful and interesting.

**Reply:**

---

**(4.4)** Demographic modeling. Both approaches tested, stairwayplot and SMC++, are most sensitive to older demographic events, and consequently are very noisy and underestimate effect population sizes, especially in smaller neighborhood sizes. Models that consider haplotype structure are much better suited to this time period. It would be helpful to either 1) discuss the varying time sensitivities of different classes of demographic inference methods and how spatial patterns of genetic variation would influence these inferences, or 2) apply a method of this class (many options, e.g. DoRIS, IBDNe, Tracts, Globetrotter, etc) and show how it performs.

**Reply:**

---

**(4.5)** GWAS mixed models. To what extent can spatial signals (e.g. corner, patchy) be corrected with mixed models, e.g. with PCs and PC-adjusted GRM as in Conomos et al, 2016 using PC-AiR and PC-Relate? Is patchiness related to dispersal? I'm curious how this relates to the predictive ability of GWAS phenotypes with some spatial association that may or may not be associated with environmental effects.

**Reply:**

---

**(4.6)** Code availability. This github link doesn't work, but is important to be able to evaluate for review: <https://github.com/petrelharp/spaceness>

**Reply:**

---

(4.7) Definitions and interpretations. There are quite a large number of metrics discussed in Figure 3B, and it's a lot to take in. It might be helpful to have a table with a reminder of what the metric is, its interpretation, and how it is computed.

**Reply:**

---

(4.8) Notation: "Offspring disperse a Gaussian-distributed distance away from the parent with mean zero and standard deviation  $\sigma$  in both the x and y coordinates. Each offspring is produced with a mate selected randomly from those within distance  $3\sigma$ , with probability of choosing a neighbor at distance  $x$  proportional to  $\exp(-x^2/2\sigma^2)$ ." I think x may be overloaded here, or I'm confused. Clarify?

**Reply:**

---

(4.9) When introducing the "spatial model" as opposed to this "random model," the more concrete illustration in Figure 1 hasn't yet been referenced, which makes it harder to follow. It would be helpful to introduce this figure with the model. Additionally, when Figure 1 is introduced, the order is from right to left (random, then point, then midpoint). It would be helpful to rearrange the figure to mirror what's in the text.

**Reply:**

---

(4.10) Not sure I follow this example: "Concretely, an individual at position  $(x, y)$  in a  $50 \times 50$  landscape has mean phenotype  $100 + 2x/5$ ."

**Reply:**

---

(4.11) Minor typo (through vs though): "This occurs because, even through the "population density" (K) and "mean lifetime" (L) parameters..."

**Reply:**

---

(4.12) Define NS abbreviation in Figure 5.

**Reply:**