

# 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>

<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. Stepping-stone models reproduce some of these effects, but discretizing the landscape introduces artifacts which in some cases are exacerbated at higher resolutions. 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 revealed surprisingly little systematic bias. 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

## Introduction

The inescapable reality that biological organisms live, move, and reproduce in continuous geography is usually omitted from population genetic models. However, mates tend to live near to one another and to their offspring, leading to a positive correlation between genetic differentiation and geographic distance. This pattern of “isolation by distance” (?) is one of the most widely replicated empirical findings in population genetics (??). Despite a long history of analytical work describing the genetics of populations distributed across continuous geography (e.g., ???????), much modern work still describes geographic structure as a set of discrete populations connected by migration (e.g., ?????) or as an average over such discrete models (??). For this reason, most population genetics statistics are interpreted with reference to discrete, well-mixed populations, and most empirical papers analyze variation within clusters of genetic variation inferred by programs like *STRUCTURE* (?) with methods that assume these are randomly mating units.

Manuscript compiled: Friday 20<sup>th</sup> March, 2020

<sup>†</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

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

50 Two areas in which spatial structure may be particularly important are demographic inference and  
51 genome-wide association studies (GWAS). Previous work has found that discrete population structure  
52 can create false signatures of population bottlenecks when attempting to infer demographic histories  
53 from microsatellite variation (?), statistics summarizing the site frequency spectrum (SFS) (??), or runs  
54 of homozygosity in a single individual (?). The increasing availability of whole-genome data has led  
55 to the development of many methods that attempt to infer detailed trajectories of population sizes  
56 through time based on a variety of summaries of genetic data (????). Because all of these methods  
57 assume that the populations being modeled are approximately randomly mating, they are likely  
58 affected by spatial biases in the genealogy of sampled individuals (?), which may lead to incorrect  
59 inference of population changes over time (?). However, previous investigations of these effects have  
60 focused on discrete rather than continuous space models, and the level of isolation by distance at which  
61 inference of population size trajectories become biased by structure is not well known. Here we test  
62 how two methods suitable for use with large samples of individuals – stairwayplot (?) and SMC++ (?)  
63 – perform when applied to populations evolving in continuous space with varying sampling strategies  
64 and levels of dispersal.

65 Spatial structure is also a major challenge for interpreting the results of genome-wide association  
66 studies (GWAS). This is because many phenotypes of interest have strong geographic differences  
67 due to the (nongenetic) influence of environmental or socioeconomic factors, which can therefore  
68 show spurious correlations with spatially patterned allele frequencies (??). Indeed, two recent studies  
69 found that previous evidence of polygenic selection on human height in Europe was confounded by  
70 subtle population structure (??), suggesting that existing methods to correct for population structure in  
71 GWAS are insufficient. However we have little quantitative idea of the population and environmental  
72 parameters that can be expected to lead to biases in GWAS.

73 Last, some of the most basic tools of population genetics are summary statistics like  $F_{IS}$  and Tajima’s  
74  $D$ , which are often interpreted as reflecting the influence of selection or demography on sampled  
75 populations (?). Statistics like Tajima’s  $D$  are essentially summaries of the site frequency spectrum,  
76 which itself reflects variation in branch lengths and tree structure of the underlying genealogies of  
77 sampled individuals. Geographically limited mate choice distorts the distribution of these genealogies  
78 (??), which can affect the value of Tajima’s  $D$  (?). Similarly, the distribution of tract lengths of identity by  
79 state among individuals contains information about not only historical demography (??) and selection  
80 (?), but also dispersal and mate choice (??). We are particularly keen to examine how such summaries  
81 will be affected by models that incorporate continuous space, both to evaluate the assumptions  
82 underlying existing methods and to identify where the most promising signals of geography lie.

83 To study this, we have implemented an individual-based model in continuous geography that  
84 incorporates overlapping generations, local dispersal of offspring, and density-dependent survival. We  
85 simulate chromosome-scale genomic data in tens of thousands of individuals from parameter regimes  
86 relevant to common subjects of population genetic investigation, and output the full genealogy and  
87 recombination history of all final-generation individuals. We use these simulations to test how sampling  
88 strategy interacts with geographic population structure to cause systematic variation in population  
89 genetic summary statistics typically analyzed assuming discrete population models. We then examine  
90 how the fine-scale spatial structures occurring under limited dispersal impact demographic inference  
91 from the site frequency spectrum. Last, we examine the impacts of continuous geography on genome-  
92 wide association studies (GWAS) and identify regions of parameter space under which the results from

<sup>93</sup> GWAS may be misleading.

## <sup>94</sup> Materials and Methods

### <sup>95</sup> ***Modeling Evolution in Continuous Space***

<sup>96</sup> The degree to which genetic relationships are geographically correlated depends on the chance that  
<sup>97</sup> two geographically nearby individuals are close relatives – in modern terms, by the tension between  
<sup>98</sup> migration (the chance that one is descended from a distant location) and coalescence (the chance that  
<sup>99</sup> they share a parent). A key early observation by ? is that this balance is often nicely summarized  
<sup>100</sup> by the “neighborhood size”, defined in two dimensions to be  $N_W = 4\pi\rho\sigma^2$ , where  $\sigma^2$  is one half of  
<sup>101</sup> the mean squared parent–offspring distance and  $\rho$  is population density (see ? for further discussion  
<sup>102</sup> of parameter definitions in one- and two-dimensional habitats) axialdist. This can be thought of as  
<sup>103</sup> proportional to the average number of potential mates for an individual (those within distance  $2\sigma$ ), or  
<sup>104</sup> the number of potential parents of a randomly chosen individual. Empirical estimates of neighborhood  
<sup>105</sup> size vary hugely across species – even in human populations, estimates range from 40 to over 5,000  
<sup>106</sup> depending on the population and method of estimation (Table ??).

<sup>107</sup> The first approach to modeling continuously distributed populations was to endow individuals  
<sup>108</sup> in a Wright-Fisher model with locations in continuous space. However, since the total size of the  
<sup>109</sup> population is constrained, this introduces interactions between arbitrarily distant individuals, which  
<sup>110</sup> (aside from being implausible) was shown by ? to eventually lead to unrealistic population clumping  
<sup>111</sup> if the range is sufficiently large. Another method for modeling spatial populations is to assume the  
<sup>112</sup> existence of a grid of discrete randomly mating populations connected by migration, thus enforcing  
<sup>113</sup> regular population density by edict. Among many other results drawn from this class of “lattice” or  
<sup>114</sup> “stepping stone” models (?), ? showed that the slope of the linear regression of genetic differentiation  
<sup>115</sup> ( $F_{ST}$ ) against the logarithm of spatial distance is an estimate of neighborhood size. Although these  
<sup>116</sup> grid models may be good approximations of continuous geography in many situations, they do not  
<sup>117</sup> model demographic fluctuations, and limit investigation of spatial structure below the level of the  
<sup>118</sup> deme, assumptions whose impacts are unknown. An alternative method for dealing with continuous  
<sup>119</sup> geography is a new class of coalescent models, the Spatial Lambda Fleming-Viot models (??).

<sup>120</sup> To avoid hard-to-evaluate approximations, we here used forward-time, individual-based simulations  
<sup>121</sup> across continuous geographical space. The question of what regulates real populations has a  
<sup>122</sup> long history and many answers (e.g., ???), ecolcites but it is clear that populations must at some point  
<sup>123</sup> have density-dependent feedback on population size, or else they would face eventual extinction or  
<sup>124</sup> explosion. In the absence of unrealistic global population regulation, this regulation must be local, and  
<sup>125</sup> there are many ways to achieve this (?). In our simulations, each individual’s probability of survival is a  
<sup>126</sup> decreasing function of local population density, which shifts reproductive output towards low-density  
<sup>127</sup> regions, and produces total census sizes that fluctuate around an equilibrium. This also prevents the  
<sup>128</sup> population clumping seen by ? (Supplemental ??)). Such models have been used extensively in ecological  
<sup>129</sup> modeling (?????) but rarely in population genetics, where to our knowledge implementations of  
<sup>130</sup> continuous space models before their availability through SLiM (?) have focused on a small number  
<sup>131</sup> of genetic loci (e.g., ?????), which limits the ability to investigate the impacts of continuous space  
<sup>132</sup> on genome-wide genetic variation as is now routinely sampled from real organisms. By simulating  
<sup>133</sup> chromosome-scale sequence alignments and complete population histories we are able to treat our  
<sup>134</sup> simulations as real populations and replicate the sampling designs and analyses commonly conducted  
<sup>135</sup> on real genomic data.

### <sup>136</sup> ***A Forward-Time Model of Evolution in Continuous Space***

<sup>137</sup> We simulated populations using the program SLiM v3.1 (?). Each time step consists of three stages:  
<sup>138</sup> reproduction, dispersal, and mortality. To reduce the number of parameters we use the same parameter,  
<sup>139</sup> denoted  $\sigma$ , to modulate the spatial scale of interactions at all three stages by adjusting the standard  
<sup>140</sup> deviation of the corresponding Gaussian functions. Informally, we think of  $\sigma$  as the “dispersal distance”,  
<sup>141</sup> although only one of those stages is dispersal.

At the beginning of the simulation individuals are distributed uniformly at random on a continuous, square landscape. Individuals are hermaphroditic, and each time step, each produces a Poisson number of offspring with mean  $1/L$ . 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  $d$  proportional to the Gaussian density with mean zero and standard deviation  $\sigma$ , which is  $g(d) = \exp(-d^2/2\sigma^2)/(2\sigma^2)$ .

To maintain a stable population, mortality increases with local population density. To do this we say that individuals at distance  $d$  have a competitive interaction with strength  $g(d)$ . Then, the sum of all competitive interactions with individual  $i$  is  $n_i = \sum_j g(d_{ij})$ , where  $d_{ij}$  is the distance between individuals  $i$  and  $j$  and the sum is over all neighbors within distance  $3\sigma$ . Since  $g$  is a probability density,  $n_i$  is an estimate of the number of nearby individuals per unit area. Then, given a per-unit carrying capacity  $K$ ,  $\text{Il:K}_d$  defines the probability of survival until the next timestep for individual  $i$  is  $p_i = \min\left(0.95, \frac{1}{1+n_i/(K(1+L))}\right)$ . (1) We chose this functional form so that the equilibrium population density per unit area is close to  $K$ ,  $\text{Il:K}_a$  again and the mean lifetime is around  $L$ ; for more description see the Appendix.

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

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

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

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

We stored the full genealogy and recombination history of final-generation individuals as tree sequences (?), as implemented in SLiM (?). Scripts for figures and analyses are available at <https://github.com/kern-lab/spaceness>.

We ran 400 simulations for the spatial and random-mating models on a square landscape of width  $W = 50$  with per-unit carrying capacity  $K = 5$  (census  $N \approx 10,000$ ), average lifetime  $L = 4$ , genome size  $10^8$  bp, recombination rate  $10^{-9}$  per bp per generation, and drawing  $\sigma$  values from a uniform distribution between 0.2 and 4. To speed up the simulations and limit memory overhead we used a mutation rate of 0 in SLiM and later applied mutations to the tree sequence with msprime’s `mutate` function (?). Because msprime applies mutations proportionally to elapsed time, we divided the mutation rate of  $10^{-8}$  mutations per site per generation by the average generation time estimated for each value of  $\sigma$  (see ‘Demographic Parameters’ below) to convert the rate to units of mutations per site per unit time. We verify that this procedure produces the same site frequency spectrum as applying mutations directly in SLiM in Figure ??, in agreement with theory (?). Simulations were run for 1.6 million timesteps (approximately  $30N$  generations).

We also compared our model’s output to a commonly-used approximation of continuous space, the stepping-stone model, which we simulated with msprime (?). These results are discussed in detail in the Appendix, but in general we find that the demographic structure of a stepping-stone model can depend strongly on the chosen discretization, and some artifacts of discretization seem to become stronger in the limit of a fine grid. For many summary statistics, finer discretizations (we used a  $50 \times 50$  grid) produced similar results to the continuous model, but this was not true for others (e.g.,  $F_{IS}$  and

177 Tajima's  $D$ ), which differed from the continuous model *more* at finer discretizations.

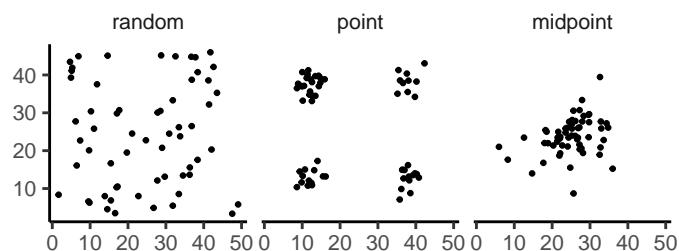
### 178 **Demographic Parameters**

Our demographic model includes parameters that control population density ( $K$ ), mean life span ( $L$ ), and dispersal distance ( $\sigma$ ). However, nonlinearity of local demographic stochasticity causes actual realized averages of these demographic quantities to deviate from the specified values in a way that depends on the neighborhood size. Therefore, to properly compare to theoretical expectations, we empirically calculated these demographic quantities in simulations. We recorded the census population size in all simulations, and used mean population density ( $\rho$ , census size divided by total area) to compute neighborhood size as  $N_W = 4\pi\rho\sigma^2$ . *ll:nw* defines *nTo estimate generation times, we stored edges of the parents of every new individual in, and took the mean. To estimate variance in offspring number, we tracked the lifetime total number of offspring for all individuals timestep burn – in period, and calculated the variance in number of offspring across all individuals in timesteps 50 – 100. All calculations were performed within information recorded in the tree sequence, using *pyslim*(?).*

179 Note that  $\sigma$  controls the dispersal of offspring away from only one parent *ll:sigma\_effective* (e.g., the seed parent for plants),  
180 – the distance between mates – has in our simulations a distribution that is off  $\sigma$  but that depends on  
181 the population's patchiness. If both between-mate distance and dispersal distance has variance  
182  $\sigma^2$  along each axis, then the mean squared distance to a randomly chosen parent along that axis would  
183 be  $(\sigma^2 + 2\sigma^2)/2 = 3\sigma^2/2$ . To match theory, neighborhood size should be defined in terms of *effective*  
184 dispersal distance, i.e., the mean squared displacement along an axis between parent-child pairs found  
185 moving back along a lineage (?), and *effective* population density (?). However, we use  $\sigma$  and  $\rho$  as  
186 defined here to compute  $N_W$  because these quantities are more easily observable in practice than their  
187 "effective" versions.

### 188 **Sampling**

189 Our model records the genealogy and sequence variation of the complete population, but in real data,  
190 genotypes are only observed from a relatively small number of sampled individuals. We modeled three  
191 sampling strategies similar to common data collection methods in empirical genetic studies (Figure ??).  
192 "Random" sampling selects individuals at random from across the full landscape, "point" sampling  
193 selects individuals proportional to their distance from four equally spaced points on the landscape,  
194 and "midpoint" sampling selects individuals in proportion to their distance from the middle of the  
195 landscape. Downstream analyses were 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.

### 196 **Summary Statistics**

197 We calculated the site frequency spectrum and a set of 18 summary statistics (Table ??) from 60 diploid  
198 individuals sampled from the final generation of each simulation using the python package scikit-allel  
199 (?). Statistics included common single-population summaries including mean pairwise divergence  
200 ( $\pi$ ), inbreeding coefficient ( $F_{IS}$ ), and Tajima's  $D$ , as well as (motivated by ?'s results) the correlation  
201 coefficient between the logarithm of the spatial distance and the proportion of identical base pairs  
202 across pairs of individuals.

Following recent studies that showed strong signals for dispersal and demography in the distribution of shared haplotype block lengths (e.g., ??), we also calculated various summaries of the distribution of pairwise identical-by-state (IBS) block lengths among sampled chromosomes, defined to be the set of distances between adjacent sites that differ between the two 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 calculated 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 (?). For a subset of simulations, we also calculated cumulative distributions for IBS tract lengths across pairs of distant (more than 40 map units) and nearby (less than 10 map units) individuals. Last, we examined the relationship between allele frequency and the spatial dispersion of an allele by calculating the average distance between individuals carrying each derived allele.

The effects of sampling on summary statistic estimates were summarized by testing for differences in mean (ANOVA, ?) and variance (Levene's test, ?) across sampling strategies for each summary statistic.

#### **Demographic Inference**

To assess the impacts of continuous spatial structure on demographic inference we inferred population size histories for all simulations using two approaches: stairwayplot (?) and SMC++ (?). 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.

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

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

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

#### **Association Studies**

To assess the degree to which spatial structure confounds GWAS we simulated four types of nongenetic phenotype variation for 1000 randomly sampled individuals in each spatial SLiM simulation and conducted a linear regression GWAS with principal components as covariates in PLINK (?). SNPs with a minor allele frequency less than 0.5% were excluded from this analysis. Phenotype values were set to vary by two standard deviations across the landscape in a rough approximation of the variation seen in height across Europe (???). Conceptually our approach is similar to that taken by ?, though here we model fully continuous spatial variation and compare GWAS output across a range of dispersal distances.

In all simulations, the phenotype of each individual is determined by drawing from a Gaussian distribution with standard deviation 10 and a mean that may depend on spatial position. In spatially varying models, the mean phenotype differs by two standard deviations across the landscape. We then adjust the geographic pattern of mean phenotype to create four types of spatially 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, the mean phenotype  $p$  for an individual at position  $(x, y)$  is  $p = 100 + 2x/5$ . Third, we simulated a more concentrated "corner" environmental effect by setting the mean phenotype to 120 for individuals with both  $x$  and  $y$  coordinates below 20 (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 (?) 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 ?.

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% (?). We also examined  $p$  values for systematic inflation by comparing to the values expected from a uniform distribution (because no SNPs were used when generating phenotypes, well-calibrated  $p$ -values should be uniform).

Results from all analyses were summarized and plotted with the "ggplot2" (?) and "cowplot" (?) packages in R (?).

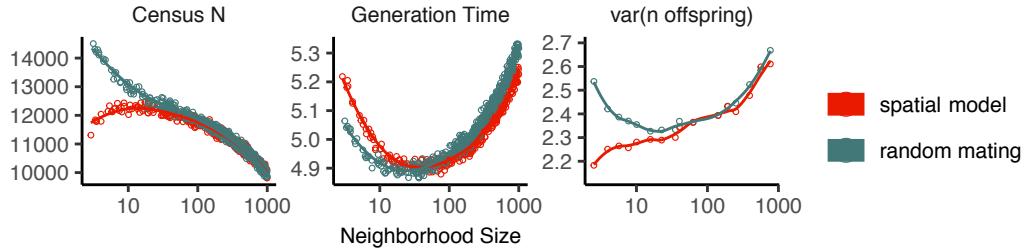
## Results

### Demographic Parameters and Run Times

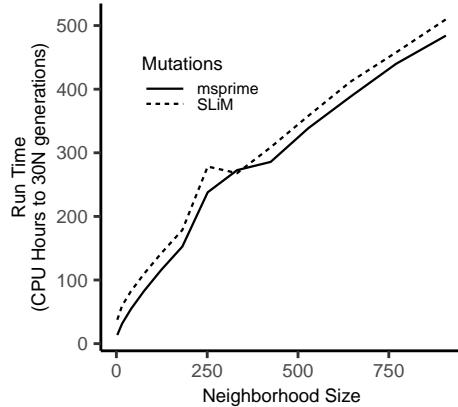
Adjusting the spatial dispersal and interaction distance,  $\sigma$ , has a surprisingly large effect on demographic quantities that are usually fixed in Wright-Fisher models – the generation time, census population size, and variance in offspring number, shown in Figure ???. Because our simulation is parameterized on an individual level, these population parameters emerge as a property of the interactions among individuals rather than being directly set. emergentparams Variation across runs occurs because, even though the parameters  $K$  and  $L$  that control population density and mean lifetime respectively ll:K<sub>m</sub>ore were the same in all simulations, the strength of stochastic effects depends strongly on the spatial interaction distance  $\sigma$ . 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 ??), unlike in spatial models (discussed below).

There are a few additional things to note about Figure ???. 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  $\geq 10$ . Spatial and random mating models are indistinguishable from one another for neighborhood sizes larger than 100. Census sizes range from around 14,000 at low  $\sigma$  in the random mating model to 10,000 for both models when neighborhood sizes approach 1,000. The scaling of census sizes in both random-mating and spatial models appears to be related to two consequences of the spatial

304 competition function: the decline of fitness at range edges, which effectively reduces the habitable area  
 305 by one  $\sigma$  around the edge of the map and so results in a smaller habitable area at high  $\sigma$  values; and  
 306 variation in the equilibrium population density given varying competition radii. Furthermore, census  
 307 size increases in spatial models as neighborhood size increases from 2 to 10. This may reflect an Allee  
 308 effect (?) in which some individuals are unable to find mates when the mate selection radius is very  
 309 small. censusscaling



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



**Figure 3** Run times of continuous space simulations with landscape width 50 and expected density 5 under varying neighborhood size. Times are shown for simulations run with mutations applied directly in SLiM (dashed lines) or later applied to tree sequences with msprime (solid lines). Times for simulations run with tree sequence recording disabled are shown in grey.

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

316 Last, we looked at the variance in number of offspring – a key parameter determining the effective  
 317 population size. Surprisingly, the spatial and random mating models behave quite differently: while  
 318 the variance in offspring number increases nearly monotonically under the spatial model, the random  
 319 mating model actually shows a decline in the variance in offspring number until a neighborhood size  
 320 of around 10 before it increases and eventually equals what we observe in the spatial case.

321 Run times for our model scale approximately linearly with neighborhood size (Figure ??), with the  
 322 lowest neighborhood sizes reaching 30N generations in around a day and those with neighborhood size

323 approaching 1,000 requiring up to three weeks of computation. As currently implemented, running  
324 simulations at neighborhood sizes more than 1,000 to coalescence is likely impractical, though running  
325 these models for more limited timescales and then “recapitulating” the simulation using reverse-time  
326 simulation from the resulting tree sequence in msprime is possible (?). runtimesxn

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

328 Even though certain aspects of population demography depend on the scale of spatial interactions, it  
329 still could be that population genetic variation is well-described by a well-mixed population model.  
330 Indeed, mathematical results suggest that genetic variation in some spatial models should be well-  
331 approximated by a Wright-Fisher population if neighborhood size is large and all samples are ge-  
332 ographically widely separated (?). However, the behavior of most common population genetic  
333 summary statistics other than Tajima’s  $D$  (?) has not yet been described in realistic geographic models.  
334 Moreover, as we will show, spatial sampling strategies can affect summaries of genetic variation at  
335 least as strongly as the underlying population dynamics.

336 **Site Frequency Spectra and Summaries of Diversity** Figure ?? shows the effect of varying neighbor-  
337 hood size and sampling strategy on the site frequency spectrum (Figure ??, Figure ??) and several  
338 standard population genetic summary statistics (Figure ??B; additional statistics are shown in Figure  
339 ??). Consistent with findings in island and stepping stone simulations (?), the SFS shows a significant  
340 enrichment of intermediate frequency variants in comparison to the nonspatial expectation. This bias  
341 is most pronounced below a neighborhood size of 100 and is exacerbated by midpoint and point  
342 sampling of individuals (depicted in Figure ??). Reflecting this, Tajima’s  $D$  is quite positive in the same  
343 situations (Figure ??B). Notably, the point at which Tajima’s  $D$  approaches 0 differs strongly across  
344 sampling strategies – varying from a neighborhood size of roughly 50 for random sampling to at least  
345 1000 for midpoint sampling.

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

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



**Figure 4** Site frequency spectrum (A; note axes are log-scaled) and summary statistic distributions (B) by sampling strategy and neighborhood size.

373 **IBS tracts and correlations with geographic distance** We next turn our attention to the effect of  
374 geographic distance on haplotype block length sharing, summarized for sets of nearby and distant  
375 individuals in Figure ???. There are two main patterns to note. First, nearby individuals share more  
376 long IBS tracts than distant individuals (as expected because they are on average more closely related).  
377 Second, the difference in the number of long IBS tracts between nearby and distant individuals  
378 decreases as neighborhood size increases. This reflects the faster spatial mixing of populations with  
379 higher dispersal, which breaks down the correlation between the IBS tract length distribution and  
380 geographic distance. This can also be seen in the bottom row of Figure ??B, where the correlation  
381 coefficients between the summaries of the IBS tract length distribution (the mean, skew, and count of  
382 tracts over  $10^6$ bp) and geographic distance approaches 0 as neighborhood size increases.

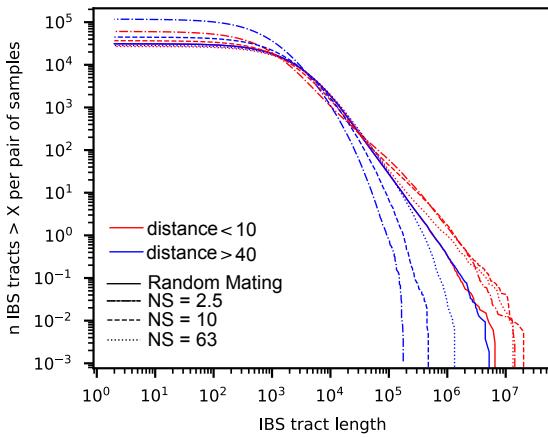
383 The patterns observed for correlations of IBS tract lengths with geographic distance are similar to  
384 those observed in the more familiar correlation of allele frequency measures such as  $D_{xy}$  (i.e., “genetic  
385 distance”) or  $F_{ST}$  against geographic distance (?).  $D_{xy}$  is positively correlated with the geographic  
386 distance between the individuals, and the strength of this correlation declines as dispersal increases  
387 (Figure ??B), as expected (?). This relationship is very similar across random and point sampling  
388 strategies, but is weaker for midpoint sampling, perhaps due to a dearth of long-distance comparisons.  
389 In much of empirical population genetics a regression of genetic differentiation against spatial distance  
390 is a de-facto metric of the significance of isolation by distance. The similar behavior of moments of the  
391 pairwise distribution of IBS tract lengths shows why haplotype block sharing has recently emerged as  
392 a promising source of information on spatial demography through methods described in ? and ?.

393 **Spatial distribution of allele copies** Mutations occur in individuals and spread geographically over  
394 time. Because low frequency alleles generally represent recent mutations (?), the geographic spread  
395 of an allele may covary along with its frequency in the population. To visualize this relationship we  
396 calculated the average distance among individuals carrying a focal derived allele across simulations  
397 with varying neighborhood sizes, shown in Figure ???. On average we find that low frequency alleles  
398 are the most geographically restricted, and that the extent to which geography and allele frequency are  
399 related depends on the amount of dispersal in the population. For populations with large neighborhood  
400 sizes we found that even very low frequency alleles can be found across the full landscape, whereas  
401 in populations with low neighborhood sizes the relationship between distance among allele copies  
402 and their frequency is quite strong. This is the basic process underlying ?’s (?) method for estimating  
403 dispersal distances based on the distribution of low frequency alleles, and also generates the greater  
404 degree of bias in GWAS effect sizes for low frequency alleles identified in ?.

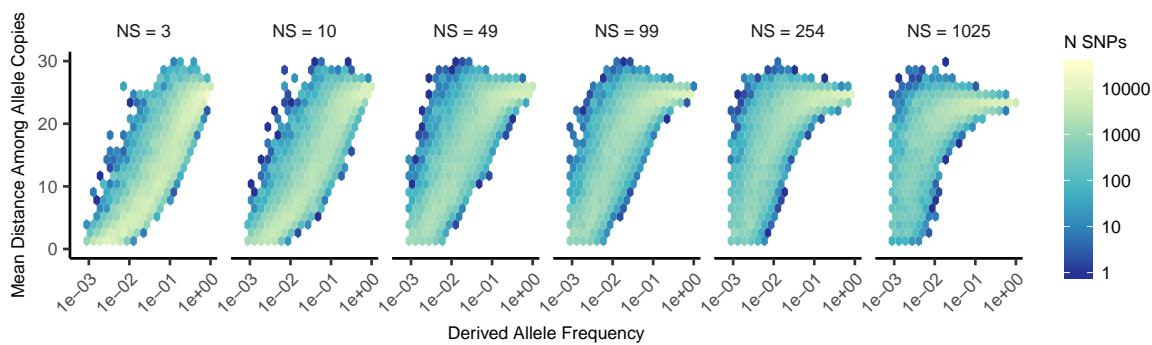
#### 405 **Effects of Space on Demographic Inference**

406 One of the most important uses for population genetic data is inferring demographic history of popu-  
407 lations. As demonstrated above, the site frequency spectrum and the distribution of IBS tracts varies  
408 across neighborhood sizes and sampling strategies. Does this variation lead to different inferences of  
409 past population sizes? To ask this we inferred population size histories from samples drawn from our  
410 simulated populations with two approaches: stairwayplot (?), which uses a genome-wide estimate of  
411 the SFS, and SMC++ (?), which incorporates information on both the SFS and linkage disequilibrium  
412 across the genome.

413 Figure ??A shows rolling medians of inferred population size histories from each method across all  
414 simulations, grouped by neighborhood size and sampling strategy. In general these methods tend to  
415 slightly overestimate ancient population sizes and infer recent population declines when neighborhood  
416 sizes are below 20 and sampling is spatially clustered. The overestimation of ancient population sizes  
417 however is relatively minor, averaging around a two-fold inflation at 10,000 generations before present  
418 in the worst-affected bins. For stairwayplot we found that many runs infer dramatic population  
419 bottlenecks in the last 1,000 generations when sampling is spatially concentrated, resulting in ten-fold  
420 or greater underestimates of recent population sizes. However SMC++ appeared more robust to  
421 this error, with runs on point- and midpoint-sampled simulations at the lowest neighborhood sizes  
422 underestimating recent population sizes by roughly half and those on randomly sampled simulations  
423 showing little error. Above neighborhood sizes of around 100, both methods performed relatively well



**Figure 5** Cumulative distributions for IBS tract lengths per pair of individuals at different geographic distances, across three neighborhood sizes (NS). Nearby pairs (red curves) share many more long IBS tracts than do distant pairs (blue curves), except in the random mating model. The distribution of long IBS tracts between nearby individuals are very similar across neighborhood sizes, but distant individuals are much more likely to share long IBS tracts at high neighborhood size than at low neighborhood size.



**Figure 6** Spatial spread of rare alleles by neighborhood size (NS): Each plot shows the distribution (across derived alleles and simulations) of average pairwise distance between individuals carrying a focal derived allele and derived allele frequency.

424 when averaging across results from multiple simulations.

425 However, individual simulations were often inferred to have turbulent demographic histories, as  
426 shown by the individually inferred histories (shown in Figure ??). Indeed, the standard deviation of  
427 inferred  $N_e$  across time points (shown in Figure ??B) often exceeds the expected  $N_e$  for both methods.  
428 That is, despite the nearly constant population sizes in our simulations, both methods tended to infer  
429 large fluctuations in population size over time, which could potentially result in incorrect biological  
430 interpretations. On average the variance of inferred population sizes was elevated at the lowest  
431 neighborhood sizes and declines as dispersal increases, with the strongest effects seen in stairwayplot  
432 results with clustered sampling and neighborhood sizes less than 20 (Figure ??B).

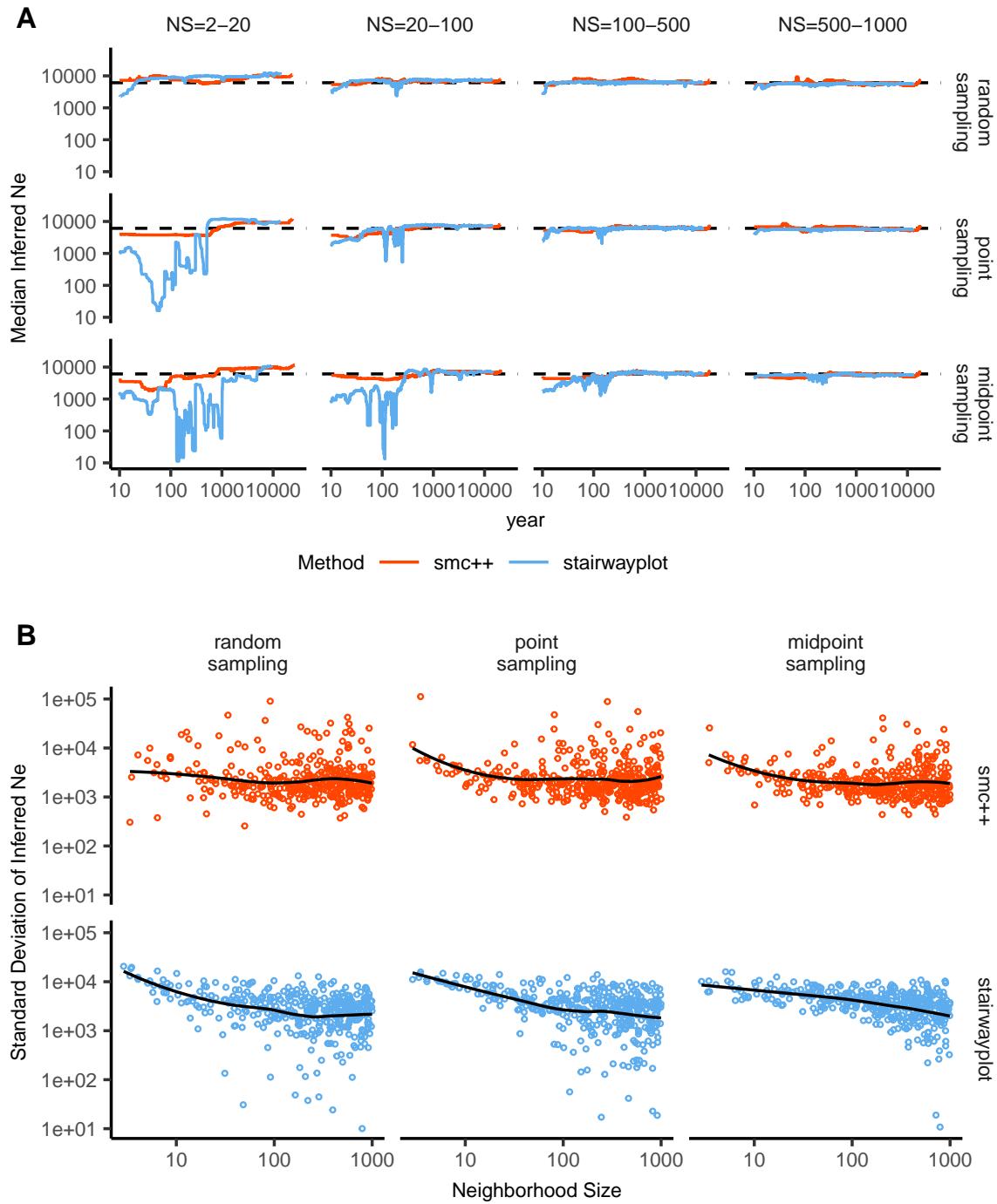
### 433 GWAS

434 To ask what confounding effects spatial genetic variation might have on genome-wide association  
435 studies we performed GWAS on our simulations using phenotypes that were determined solely by  
436 the environment – so, any SNP showing statistically significant correlation with phenotype is a false  
437 positive. As expected, spatial autocorrelation in the environment causes spurious associations across  
438 much of the genome if no correction for genetic relatedness among samples is performed (Figures  
439 ?? and ??). This effect is particularly strong for clinal and corner environments, for which the lowest  
440 dispersal levels cause over 60% of SNPs in the sample to return significant associations. Patchy  
441 environmental distributions, which are less strongly spatially correlated (Figure ??A), cause fewer  
442 false positives overall but still produce spurious associations at roughly 10% of sites at the lowest  
443 neighborhood sizes. Interestingly we also observed a small number of false positives in roughly 3% of  
444 analyses on simulations with nonspatial environments, both with and without PC covariates included  
445 in the regression.

446 The confounding effects of geographic structure are well known, and it is common practice to  
447 control for this by including principal components (PCs) as covariates to control for these effects. This  
448 mostly works in our simulations – after incorporating the first ten PC axes as covariates, the vast  
449 majority of SNPs no longer surpass a significance threshold chosen to have a 5% false discovery rate  
450 (FDR). However, a substantial number of SNPs – up to 1.5% at the lowest dispersal distances – still  
451 surpass this threshold (and thus would be false positives in a GWAS), especially under “corner” and  
452 “patchy” environmental distributions (Figure ??C). At neighborhood sizes larger than 500, up to 0.31%  
453 of SNPs were significant for corner and clinal environments. Given an average of 132,000 SNPs across  
454 simulations after MAF filtering, this translates to up to 382 false-positive associations; for human-sized  
455 genomes, this number would be much larger. In most cases the  $p$  values for these associations were  
456 significant after FDR correction but would not pass the threshold for significance under the more  
457 conservative Bonferroni correction (see example Manhattan plots in figure ??).

458 Clinal environments cause an interesting pattern in false positives after PC correction: at low  
459 neighborhood sizes the correction removes nearly all significant associations, but at neighborhood  
460 sizes above roughly 250 the proportion of significant SNPs increases to up to 0.4% (Figure ??). This  
461 may be due to a loss of descriptive power of the PCs – as neighborhood size increases, the total  
462 proportion of variance explained by the first 10 PC axes declines from roughly 10% to 4% (Figure ??B).  
463 Essentially, PCA seems unable to effectively summarize the weak population structure present in large-  
464 neighborhood simulations given the sample sizes we tested, 22 but these populations continue to have  
465 enough spatial structure to create significant correlations between genotypes and the environment. A  
466 similar process can also be seen in the corner phenotype distribution, in which the count of significant  
467 SNPs initially declines as neighborhood size increases and then increases at approximately the point at  
468 which the proportion of variance explained by PCA approaches its minimum.

469 Figure ??D shows quantile-quantile plots for a subset of simulations that show the degree of  
470 genome-wide inflation of test statistics in PC-corrected GWAS across all simulations and environ-  
471 mental distributions. An alternate visualization is also included in figure ???. For clinal environments,  
472  $-\log_{10}(p)$  values are most inflated when neighborhood sizes are large, consistent with the pattern  
473 observed in the count of significant associations after PC regression. In contrast corner and patchy en-  
474 vironments cause the greatest inflation in  $-\log_{10}(p)$  at neighborhood sizes less than 100, which likely  
475 reflects the inability of PCA to account for fine-scale structure caused by very limited dispersal. Finally,



**Figure 7** 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 ??.

476 we observed that PC regression appears to overfit to some degree for all phenotype distributions,  
477 visible in Figure ??D as points falling below the 1:1 line.

## 478 Discussion

479 In this study, we have used efficient forward time population genetic simulations to describe the  
480 myriad influence of continuous geography on genetic variation. In particular, we examine how three  
481 main types of downstream empirical inference are affected by unmodeled spatial population structure  
482 – population genetic summary statistics, inference of population size history, and genome-wide associa-  
483 tion studies (GWAS). As discussed above, space often matters (and sometimes dramatically), both  
484 because of how samples are arranged in space, and because of the inherent patterns of relatedness  
485 established by geography.

### 486 Effects of Dispersal

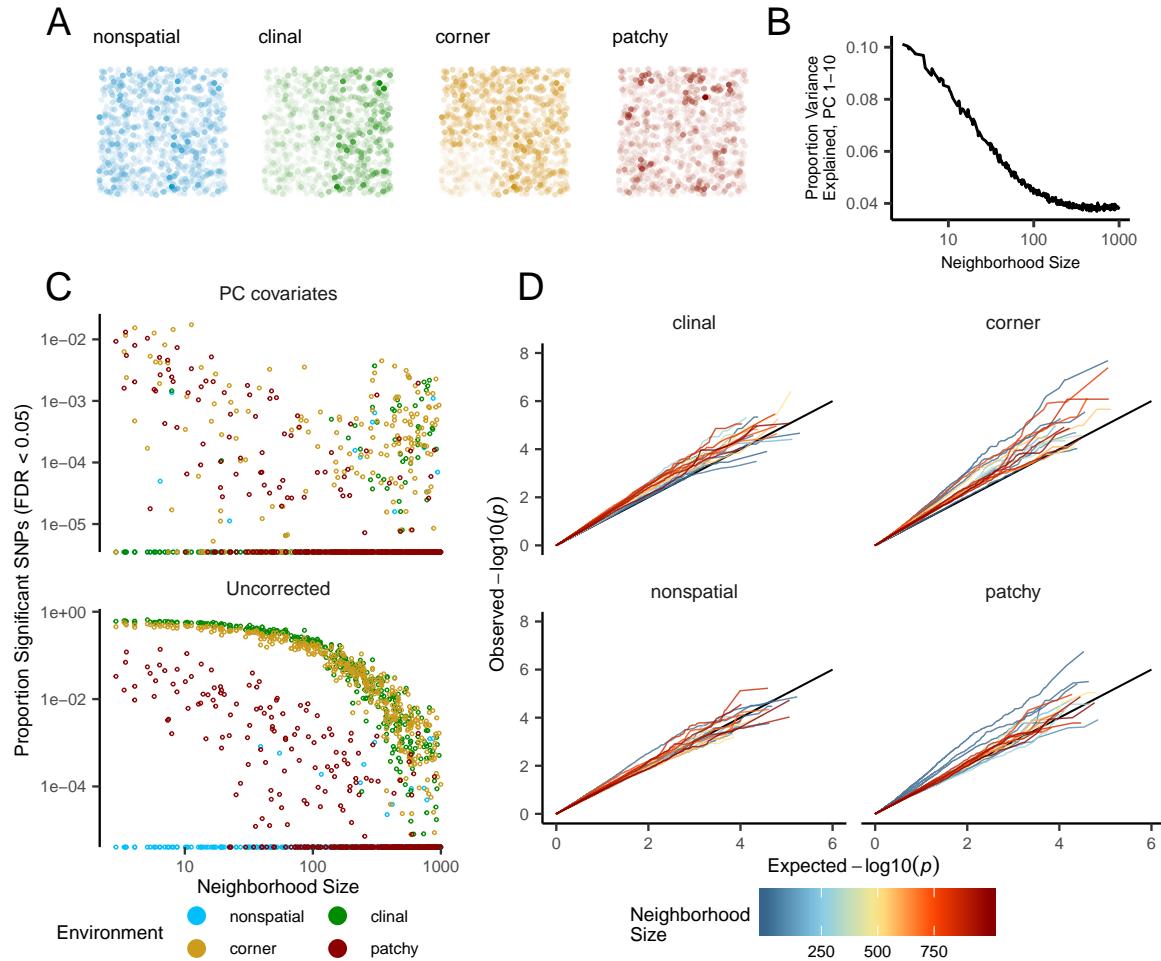
487 Limited dispersal inflates effective population size, creates correlations between genetic and spatial  
488 distances, and introduces strong distortions in the site frequency spectrum that are reflected in a  
489 positive Tajima's  $D$  (Figure ??). At the lowest dispersal distances, this can increase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

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

506 The genealogical patterns introduced by limited dispersal are particularly apparent in the distribution  
507 of haplotype block lengths (Figure ??). This is because identical-by-state tract lengths reflect the  
508 impacts of two processes acting along the branches of the underlying genealogy – both mutation and  
509 recombination – rather than just mutation as is the case when looking at the site frequency spectrum or  
510 related summaries. This means that the pairwise distribution of haplotype block lengths carries with  
511 it important information about genealogical variation in the population, and correlation coefficients  
512 between moments of the this distribution and geographic location contain signal similar to the corre-  
513 lations between  $F_{ST}$  or  $D_{xy}$  and geographic distance (?). Indeed this basic logic underlies two recent  
514 studies explicitly estimating dispersal from the distribution of shared haplotype block lengths (??).  
515 Conversely, because haplotype-based measures of demography are particularly sensitive to variation  
516 in the underlying genealogy, inference approaches that assume random mating when analyzing the  
517 distribution of shared haplotype block lengths are likely to be strongly affected by spatial processes.

### 518 Effects of Sampling

519 One of the most important differences between random mating and spatial models is the effect of  
520 sampling: in a randomly mating population the spatial distribution of sampling effort has no effect on  
521 estimates of genetic variation (Table ??), but when dispersal is limited sampling strategy can compound  
522 spatial patterns in the underlying genealogy and create pervasive impacts on all downstream genetic  
523 analyses (see also ?). In most species, the difficulty of traveling through all parts of a species range and  
524 the inefficiency of collecting single individuals at each sampling site means that most studies follow



**Figure 8** 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 for simulations with spatially structured environments, with line colors showing the neighborhood size of each simulation.

525 something closest to the “point” sampling strategy we simulated, in which multiple individuals are  
526 sampled from nearby points on the landscape. For example, in ornithology a sample of 10 individuals  
527 per species per locality is a common target when collecting for natural history museums. In classical  
528 studies of *Drosophila* variation the situation is considerably worse, in which a single orchard might be  
529 extensively sampled.

530 When sampling is clustered at points on a landscape and dispersal is limited, the sampled indi-  
531 viduals will be more closely related than a random set of individuals. Average coalescence times of  
532 individuals collected at a locality will then be more recent and branch lengths shorter than expected by  
533 analyses assuming random mating. This leads to fewer mutations and recombination events occurring  
534 since their last common ancestor, causing a random set of individuals to share longer average IBS tracts  
535 and have fewer nucleotide differences. For some data summaries, such as Tajima’s  $D$ , Watterson’s  
536  $\theta$ , or the correlation coefficient between spatial distance and the count of long haplotype blocks, this  
537 can result in large differences in estimates between random and point sampling (Figure ??). Inferring  
538 underlying demographic parameters from these summary statistics – unless the spatial locations of the  
539 sampled individuals are somehow taken into account – will likely be subject to bias.

540 We observed the largest sampling effects using “midpoint” sampling. This model is meant to reflect  
541 a bias in sampling effort towards the middle of a species’ range. In empirical studies this sampling  
542 strategy could arise if, for example, researchers choose to sample the center of the range and avoid range  
543 edges to maximize probability of locating individuals during a short field season. Because midpoint  
544 sampling provides limited spatial resolution it dramatically reduces the magnitude of observed  
545 correlations between spatial and genetic distances. More surprisingly, midpoint sampling also leads  
546 to strongly positive Tajima’s  $D$  and an inflation in the proportion of heterozygous individuals in the  
547 sample – similar to the effect of sampling a single deme in an island model as reported in ?. This increase  
548 in observed heterozygosity appears to reflect the effects of range edges, which are a fundamental  
549 facet of spatial genetic variation. If individuals move randomly in a finite two-dimensional landscape  
550 then regions in the middle of the landscape receive migrants from all directions while those on the  
551 edge receive no migrants from at least one direction. The average number of new mutations moving  
552 into the middle of the landscape is then higher than the number moving into regions near the range  
553 edge, leading to higher heterozygosity and lower inbreeding coefficients ( $F_{IS}$ ) away from range edges.  
554 Though here we used only a single parameterization of fitness decline at range edges we believe  
555 this is a general property of non-infinite landscapes as it has also been observed in previous studies  
556 simulating under lattice models (?).

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

### 568 **Demography**

569 Previous studies have found that population structure and nonrandom sampling can create spurious  
570 signals of population bottlenecks when attempting to infer demographic history with microsatellite  
571 variation, summary statistics, or runs of homozygosity (?????). Here we found that methods that infer  
572 detailed population trajectories through time based on the SFS and patterns of LD across the genome  
573 are also subject to this bias, with some combinations of dispersal and sampling strategy systematically  
574 inferring deep recent population bottlenecks and overestimating ancient  $N_e$  by around a factor of 2.  
575 We were surprised to see that both stairwayplot and SMC++ can tolerate relatively strong isolation  
576 by distance – i.e., neighborhood sizes of 20 – and still perform well when averaging results across

577 multiple simulations. (However, note the high amount of between-simulation variance seen in ??).  
578 Il:variance<sub>note</sub>Inferenceinpopulationswithneighborhoodsizesover20wasrelativelyunbiasedunlesssampleswereconcentratedin  
579 than isolation by distance.

580 We found that SMC++ was more robust to the effects of space than stairwayplot, underestimating  
581 recent populations by roughly half in the worst time periods rather than nearly 10-fold as with  
582 stairwayplot. Though this degree of variation in population size is certainly meaningful in an ecological  
583 context, it is relatively minor in population genetic terms. ibdnesxn Methods directly assessing  
584 haplotype structure in phased data example, (e.g., ?) are thought to provide increased resolution for  
585 recent demographic events, but in this case the error we observed was essentially an accurate reflection  
586 of underlying genealogies in which terminal branches are anomalously short. Combined with our  
587 analysis of IBS tract length variation (Figure ??) this suggests that haplotype-based methods are likely  
588 to be affected by similar biases.

589 A more worrying pattern was the high level of variance in inferred  $N_e$  trajectories for individual  
590 model fits using these methods, which was highest in simulations with the smallest neighborhood  
591 size (Figure ??, Figure ??). This suggests that, at a minimum, researchers working with empirical data  
592 should replicate analyses multiple times and take a rolling average if model fits are inconsistent across  
593 runs. Splitting samples and running replicates on separate subsets – the closest an empirical study can  
594 come to our design of averaging the results from multiple simulations – may also alleviate this issue.

595 Our analysis suggests that many empirical analyses of population size history using methods like  
596 SMC++ are robust to error caused by spatial structure within continuous landscapes. Inferences drawn  
597 from static SFS-based methods like stairwayplot should be treated with caution when there are signs  
598 of isolation by distance in the underlying data (for example, if a regression of  $F_{ST}$  against the logarithm  
599 of geographic distance has a significantly positive slope), and in particular an inference of population  
600 bottlenecks in the last 1000 years should be discounted if sampling is clustered, but estimates of deeper  
601 time patterns are likely to be fairly accurate. The biases in the SFS and haplotype structure identified  
602 above (see also ???) are apparently small enough that they fall within the range of variability regularly  
603 inferred by these approaches, at least on datasets of the size we simulated.

## 604 **GWAS**

605 Spatial structure is particularly challenging for genome-wide association studies, because the effects of  
606 dispersal on genetic variation are compounded by spatial variation in the environment (?). Spatially  
607 restricted mate choice and dispersal causes variation in allele frequencies across the range of a species.  
608 If environmental factors affecting the phenotype of interest also vary over space, then allele frequencies  
609 and environmental exposures will covary over space. Il:1:11:2 In this scenario an uncorrected GWAS  
610 will infer genetic associations with a purely environmental phenotype at any site in the genome that is  
611 differentiated over space, and the relative degree of bias will be a function of the degree of covariation  
612 in allele frequencies and the environment (i.e., Figure ??C, bottom panel). This pattern has been  
613 demonstrated in a variety of simulation and empirical contexts (??????????).

614 Incorporating PC positions as covariates in a linear-regression GWAS (?) is designed to address this  
615 challenge by regressing out a baseline level of “average” differentiation. In essence, a PC-corrected  
616 GWAS asks “what regions of the genome are more associated with this phenotype than the average  
617 genome-wide association observed across populations?” In our simulations, we observed that this  
618 procedure can fail under a variety of circumstances. If dispersal is limited and environmental variation  
619 is clustered in space (i.e., corner or patchy distributions in our simulations), PC positions fail to capture  
620 the fine-scale spatial structure required to remove all signals of association. Conversely, as dispersal  
621 increases, PCA loses power to describe population structure before spatial mixing breaks down the  
622 relationship between genotype and the environment. These effects were observed with all spatially  
623 correlated environmental patterns, but were particularly pronounced if environmental effects are  
624 concentrated in one region, as was also found by ?. Though increasing the number of PC axes used in  
625 the analysis may reduce the false-positive rate, this may also decrease the power of the test to detect  
626 truly causal alleles (?).

627 In this work we simulated a single chromosome with size roughly comparable to one human  
628 chromosome. If we scale the number of false-positive associations identified in our analyses to a

GWAS conducted on whole-genome data from humans, we would expect to see several thousand weak false-positive associations after PC corrections in a population with neighborhood sizes up to at least 1000 (which should include values appropriate for many human populations). Notably, very few of the spurious associations we identified would be significant at a conservative Bonferroni-adjusted *p*-value cutoff (see Figure ??). This suggests that GWAS focused on finding strongly associated alleles for traits controlled by a limited number of variants in the genome are likely robust to the impacts of continuous spatial structure. However, methods that analyze the combined effects of thousands or millions of weakly associated variants such as polygenic risk scores (??) are likely to be affected by subtle population structure. 23 Indeed as recently identified in studies of genotype associations for human height in Europe (??), 24 PC regression GWAS in modern human populations do include residual signal of population structure in large-scale analyses of polygenic traits. In addition to error associated with varying patterns of linkage disequilibrium and allele frequency among populations, the confounding of environmental and genetic effects on phenotypes introduced by population structure is expected to lead to low predictive power when polygenic scores are generated for populations outside the original GWAS cohort, ldcaveats as was shown in a recent study finding lower polygenic score predictive power outside European populations (?).

In summary, spatial covariation in population structure and the environment confounds the interpretation of GWAS *p*-values, and correction using principal components is insufficient to fully separate these signals for polygenic traits under a variety of environmental and population parameter regimes. Other GWAS methods such as mixed models (?) mixedmodelcallout may be less sensitive to this confounding, but there is no obvious reason that this should be so. One approach to estimating the degree of bias in GWAS caused by population structure is LD score regression (?). Though this approach appears to work well in practice, its interpretation is not always straightforward and it is likely biased by the presence of linked selection (?). In addition, we observed that in many cases the false-positive SNPs we identified appeared to be concentrated in LD peaks similar to those expected from truly causal sites (Figure ??), which may confound LD score regression.

We suggest a straightforward alternative for species in which the primary axes of population differentiation are associated with geography (note this is likely not the case for some modern human populations): run a GWAS with spatial coordinates as phenotypes and check for *p*-value inflation or significant associations. If significant associations with sample locality are observed after correcting for population structure, the method is sensitive to false positives induced by spatial structure. This is essentially the approach taken in our "clinal" model (though we add normally distributed noise to our phenotypes). This approach has recently been taken with polygenic scores for UK Biobank samples in ?, finding that scores are correlated with birth location even in this relatively homogenous sample. Of course, it is possible that genotypes indirectly affect individual locations by adjusting organismal fitness and thus habitat selection across spatially varying environments, but we believe that this hypothesis should be tested against a null of stratification bias inflation rather than accepted as true based on GWAS results.

### Where are natural populations on this spectrum?

For how much of the tree of life do spatial patterns circumscribe genomic variation? In Table ?? we gathered estimates of neighborhood size from a range of organisms to get an idea of how strongly local geographic dispersal affects patterns of variation. This is an imperfect measure: some aspects of genetic variation are most strongly determined by neighborhood size (?), others (e.g., number of segregating sites) by global  $N_e$ , or the ratio of the two. popsizecaveat In addition, definitions of "population density" in genetic versus ecological studies may lead to varying estimates of neighborhood size for a given species rhocaveat, and these empirical examples may be biased towards small-neighborhood species because few studies have quantified neighborhood size in species with very high dispersal or population density.

However, from the available data we find that neighborhood sizes in the range we simulated are fairly common across a range of taxa. At the extreme low end of empirical neighborhood size estimates we see some flowering plants, large mammals, and colonial insects like ants with neighborhood sizes less than roughly 100. Species such as this have neighborhood size estimates small enough that spatial

**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	(?)
<i>Borreria frutescens</i>	salt marsh plant	20 - 30	Genetic+Survey	(?)
<i>Oreamnos americanus</i>	mountain goat	36 - 100	Genetic	(?)
<i>Homo sapiens</i>	Gainj- and Kalam-speaking people, Papua New Guinea	40 - 213	Genetic	(?)
<i>Formica sp.</i>	colonial ants	50 - 100	Genetic	(?)
<i>Astrocaryum mexicanum</i>	palm tree	102 - 895	Genetic+survey	(?)
<i>Spermophilus mollis</i>	ground squirrel	204 - 480	Genetic+Survey	(?)
<i>Sceloporus olivaceus</i>	lizard	225 - 270	Survey	(?)
<i>Dieffenbachia longispatha</i>	beetle-pollinated colonial herb	227 - 611	Survey	(?)
<i>Aedes aegypti</i>	Yellow-fever mosquito	268	Genetic	(?)
<i>Homo sapiens</i>	Gainj- and Kalam-speaking people, Papua New Guinea	410	Survey	(?)
<i>Quercus laevis</i>	Oak tree	> 440	Genetic	(?)
<i>Drosophila pseudoobscura</i>	fruit fly	500 - 1,000	Survey+Crosses	(?)
<i>Homo sapiens</i>	POPRES data NE Europe	1,342 - 5,425	Genetic	(?)

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 ? 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 the recent ancestors of modern humans in northeastern Europe (?). 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.

#### 694 **Other demographic models**

Any simulation of a population of reproducing organisms requires some kind of control on population sizes, or else the population will either die out or grow very large after a sufficiently long period of time. demog<sub>d</sub>iscoThe usual choice of population regulation for population genetics -- a constant size, as in the Wright -- Fisher model -- implies biologically unrealistic interactions between geographically distant parts of the species range. Our choice dependent control on mortality is only one of many possible ways to do this. We could have instead regulated fecundity, or recruitment, -- Pacala model" (?). It is not currently clear how much different choices of demographic parameters, or of functional forms for the Volterra model), or extrinsically specified by local resource availability (e.g., by food or nest site availability). Indeed, our model could as local density increases, good habitat is increasingly occupied, pushing individuals into more marginal habitat and increasing their

Population genetic simulations often use grids of discrete demes, which are assumed to approximate continuous space. However, there are theoretical reasons to expect that increasingly fine grids of discrete demes do not approach the continuous model (?). If continuous space can be approximated by a limit of discrete models, this should be true regardless of the precise details of the discrete model. Although we carefully chose parameters to match our continuous models, we found that some aspects of genetic variation diverged from the continuous case as the discretization got finer. This suggests that these models do not converge in the limit. However, many populations may indeed be well-modeled as a series of discrete, randomly-mating demes if, for example, suitable habitats are patchily distributed across the landscape. There is a clear need for greater exploration of the consequences for population genetics of ecologically realistic population models.

#### 705 **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., ?) for the accurate estimation of dispersal from genetic data. Indeed, ? 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 ? and ? have shown that these parameters are identifiable under some continuous models. Similarly, though the scaling effects of dispersal we show in Figure ?? should occur in populations of any total size, other aspects such as the number of segregating sites are also likely affected by the total landscape size (and so total census size). landscapesizecaveat Indeed, our finding that stepping-stone and continuous-space models match in only certain aspects of genetic variation (Figure ??) shows that qualitatively similar models can produce different results dependent on the specific parameterizations used. While we believe our continuous model is a more appropriate depiction of many species' demographies than lattice models, it is likely that some populations and

breeding systems do more closely resemble a series of interconnected random-mating populations. As with all population models, the best approximation for any empirical system will depend on the natural history of the species in question. bestmodelcaveat Much additional work remains to be done to better understand how life history, range size, and habitat geometry interact to shape genetic variation in continuous space, which we leave to future studies.

Though our simulation allows incorporation of realistic demographic and spatial processes, it is inevitably limited by the computational burden of tracking tens or hundreds of thousands of individuals in every generation. In particular, computations required for mate selection and spatial competition scale approximately with the product of the total census size and the neighborhood size and so increase rapidly for large populations and dispersal distances. The reverse-time spatial Lambda-Fleming-Viot model described by ? and implemented by ? allows exploration of larger population and landscape sizes, but the precise connection of these models to forward-time demography is not yet clear. Alternatively, implementation of parallelized calculations may allow progress with forward-time simulations.

Finally, we believe that the difficulties in correcting for population structure in continuous populations using principal components analysis or similar decompositions is a difficult issue, well worth considering on its own. How can we best avoid spurious correlations while correlating genetic and phenotypic variation without underpowering the methods? Perhaps optimistically, we posit that process-driven descriptions of ancestry 27 and/or more generalized unsupervised methods may be able to better account for carry out this task.

## 744 Data Availability

745 Scripts used for all analyses and figures are available at <https://github.com/kern-lab/spaceness>.

## 746 Acknowledgements

747 We thank Brandon Cooper, Matt Hahn, Doc Edge, and others for reading and thinking about this  
748 manuscript. Three anonymous reviewers and the suggestions from editors of GENETICS significantly  
749 improved the final manuscript. CJB and ADK were supported by NIH award R01GM117241.

## 750 Literature Cited

- 751 Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, *et al.*, 2017 Deconstructing  
752 isolation-by-distance: The genomic consequences of limited dispersal. PLOS Genetics 13: 1–27.  
753 Al-Asadi, H., D. Petkova, M. Stephens, and J. Novembre, 2019 Estimating recent migration and  
754 population-size surfaces. PLoS genetics 15: e1007908.  
755 Allee, W. C., O. Park, A. E. Emerson, T. Park, K. P. Schmidt, *et al.*, 1949 Principles of animal ecology.  
756 Technical report, Saunders Company Philadelphia, Pennsylvania, USA.  
757 Antlfinger, A. E., 1982 Genetic neighborhood structure of the salt marsh composite, *Borrichia frutescens*.  
758 Journal of Heredity 73: 128–132.  
759 Antolin, M. F., B. V. Horne, M. D. Berger, Jr., A. K. Holloway, J. L. Roach, *et al.*, 2001 Effective population  
760 size and genetic structure of a piute ground squirrel (*Spermophilus mollis*) population. Canadian  
761 Journal of Zoology 79: 26–34.  
762 Antonovics, J. and D. A. Levin, 1980 The ecological and genetic consequences of density-dependent  
763 regulation in plants. Annual Review of Ecology and Systematics 11: 411–452.  
764 Ashander, J., P. Ralph, E. McCartney-Melstad, and H. B. Shaffer, 2018 Demographic inference in a  
765 spatially-explicit ecological model from genomic data: a proof of concept for the mojave desert  
766 tortoise. bioRxiv .  
767 Baharian, S., M. Barakatt, C. R. Gignoux, S. Shringarpure, J. Errington, *et al.*, 2016 The great migration  
768 and African-American genomic diversity. PLOS Genetics 12: 1–27.  
769 Barton, N. H., F. Depaulis, and A. M. Etheridge, 2002 Neutral evolution in spatially continuous  
770 populations. Theoretical Population Biology 61: 31–48.

- 771 Barton, N. H., J. Kelleher, and A. M. Etheridge, 2010 A new model for extinction and recolonization in  
 772 two dimensions: Quantifying phylogeography. *Evolution* **64**: 2701–2715.
- 773 Benjamini, Y. and D. Yekutieli, 2001 The control of the false discovery rate in multiple testing under  
 774 dependency. *The Annals of Statistics* **29**: 1165–1188.
- 775 Berg, E. E. and J. L. Hamrick, 1995 Fine-scale genetic structure of a turkey oak forest. *Evolution* **49**:  
 776 110–120.
- 777 Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, *et al.*, 2018 Reduced  
 778 signal for polygenic adaptation of height in UK Biobank. *bioRxiv*.
- 779 Bolker, B. and S. W. Pacala, 1997 Using moment equations to understand stochastically driven spatial  
 780 pattern formation in ecological systems. *Theoretical Population Biology* **52**: 179 – 197.
- 781 Bolker, B. M., S. W. Pacala, and C. Neuhauser, 2003 Spatial dynamics in model plant communities:  
 782 What do we really know? *The American Naturalist* **162**: 135–148, PMID: 12858259.
- 783 Browning, S. R. and B. L. Browning, 2015 Accurate non-parametric estimation of recent effective  
 784 population size from segments of identity by descent. *The American Journal of Human Genetics* **97**:  
 785 404–418.
- 786 Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang, *et al.*, 2015 LD score regression  
 787 distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*  
 788 **47**: 291 EP –.
- 789 Campbell, D. R. and J. L. Dooley, 1992 The spatial scale of genetic differentiation in a hummingbird-  
 790 pollinated plant: Comparison with models of isolation by distance. *The American Naturalist* **139**:  
 791 735–748.
- 792 Champer, J., I. Kim, S. E. Champer, A. G. Clark, and P. W. Messer, 2019 Suppression gene drive in  
 793 continuous space can result in unstable persistence of both drive and wild-type alleles. *bioRxiv*.
- 794 Chapman, N. H. and E. A. Thompson, 2002 The effect of population history on the lengths of ancestral  
 795 chromosome segments. *Genetics* **162**: 449–458.
- 796 Chikhi, L., V. C. Sousa, P. Luisi, B. Goossens, and M. A. Beaumont, 2010 The confounding effects of  
 797 population structure, genetic diversity and the sampling scheme on the detection and quantification  
 798 of population size changes. *Genetics* **186**: 983–995.
- 799 Crawley, M. J., 1990 The population dynamics of plants. *Philosophical Transactions of the Royal Society*  
 800 of London. Series B: Biological Sciences **330**: 125–140.
- 801 Durrett, R. and S. Levin, 1994 The importance of being discrete (and spatial). *Theoretical Population  
 802 Biology* **46**: 363–394.
- 803 Eguiarte, L. E., A. Búrquez, J. Rodríguez, M. Martínez-Ramos, J. Sarukhán, *et al.*, 1993 Direct and  
 804 indirect estimates of neighborhood and effective population size in a tropical palm, *Astrocaryum  
 805 mexicanum*. *Evolution* **47**: 75–87.
- 806 Epperson, B., 2003 *Geographical Genetics*. Monographs in Population Biology, Princeton University  
 807 Press.
- 808 Felsenstein, J., 1975 A pain in the torus: Some difficulties with models of isolation by distance. *The  
 809 American Naturalist* **109**: 359–368.
- 810 Fournier, N. and S. Méléard, 2004 A microscopic probabilistic description of a locally regulated  
 811 population and macroscopic approximations. *The Annals of Applied Probability* **14**: 1880–1919.
- 812 Fox, J. and S. Weisberg, 2011 *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second  
 813 edition.
- 814 Garcia, J. and C. Quintana-Domeque, 2006 The evolution of adult height in Europe: A brief note.  
 815 Working Paper .
- 816 Garcia, J. and C. Quintana-Domeque, 2007 The evolution of adult height in Europe: A brief note.  
 817 Economics & Human Biology **5**: 340 – 349.
- 818 Garud, N. R., P. W. Messer, E. O. Buzbas, and D. A. Petrov, 2015 Recent selective sweeps in North  
 819 American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genetics* **11**: 1–32.
- 820 Griffiths, R., S. Tavaré, *et al.*, 1999 The ages of mutations in gene trees. *The Annals of Applied Probability*  
 821 **9**: 567–590.
- 822 Guindon, S., H. Guo, and D. Welch, 2016 Demographic inference under the coalescent in a spatial  
 823 continuum. *Theoretical population biology* **111**: 43–50.

- 824 Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence recording  
825 in SLiM opens new horizons for forward-time simulation of whole genomes. *Molecular Ecology*  
826 Resources **19**: 552–566.
- 827 Haller, B. C. and P. W. Messer, 2019 SLiM 3: Forward genetic simulations beyond the Wright–Fisher  
828 model. *Molecular biology and evolution* **36**: 632–637.
- 829 Harris, K. and R. Nielsen, 2013 Inferring demographic history from a spectrum of shared haplotype  
830 lengths. *PLOS Genetics* **9**: 1–20.
- 831 Haworth, S., R. Mitchell, L. Corbin, K. H. Wade, T. Dudding, *et al.*, 2019 Apparent latent structure within  
832 the UK Biobank sample has implications for epidemiological analysis. *Nature communications* **10**:  
833 333.
- 834 Huillet, T. and M. Möhle, 2011 On the extended Moran model and its relation to coalescents with  
835 multiple collisions. *Theoretical Population Biology* pp. –.
- 836 International Schizophrenia Consortium *et al.*, 2009 Common polygenic variation contributes to risk of  
837 schizophrenia and bipolar disorder. *Nature* **460**: 748.
- 838 Jackson, N. D. and L. Fahrig, 2014 Landscape context affects genetic diversity at a much larger spatial  
839 extent than population abundance. *Ecology* **95**: 871–881.
- 840 Jasper, M., T. Schmidt, N. Ahmad, S. Sinkins, and A. Hoffmann, 2019 A genomic approach to inferring  
841 kinship reveals limited intergenerational dispersal in the yellow fever mosquito. *bioRxiv* .
- 842 Jay, F., P. Sjödin, M. Jakobsson, and M. G. Blum, 2012 Anisotropic isolation by distance: The main  
843 orientations of human genetic differentiation. *Molecular Biology and Evolution* **30**: 513–525.
- 844 Kang, H. M., J. H. Sul, S. K. Service, N. A. Zaitlen, S.-y. Kong, *et al.*, 2010 Variance component model to  
845 account for sample structure in genome-wide association studies. *Nature Genetics* **42**: 348 EP –.
- 846 Kang, H. M., N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, *et al.*, 2008 Efficient control of  
847 population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- 848 Kelleher, J., A. Etheridge, and N. Barton, 2014 Coalescent simulation in continuous space: Algorithms  
849 for large neighbourhood size. *Theoretical Population Biology* **95**: 13 – 23.
- 850 Kelleher, J., A. M. Etheridge, and G. McVean, 2016 Efficient coalescent simulation and genealogical  
851 analysis for large sample sizes. *PLoS Comput Biol* **12**: 1–22.
- 852 Kelleher, J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for fast  
853 population genetics simulation. *PLOS Computational Biology* **14**: 1–21.
- 854 Kerster, H. W., 1964 Neighborhood size in the rusty lizard, *Sceloporus olivaceus*. *Evolution* **18**: 445–457.
- 855 Kingman, J., 1982 The coalescent. *Stochastic Processes and their Applications* **13**: 235 – 248.
- 856 Law, R., D. J. Murrell, and U. Dieckmann, 2003 Population growth in space and time: Spatial logistic  
857 equations. *Ecology* **84**: 252–262.
- 858 Lawson, D. J., N. M. Davies, S. Haworth, B. Ashraf, L. Howe, *et al.*, 2019 Is population structure in the  
859 genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* .
- 860 Leblois, R., A. Estoup, and R. Streiff, 2006 Genetics of recent habitat contraction and reduction in  
861 population size: does isolation by distance matter? *Molecular Ecology* **15**: 3601–3615.
- 862 Liu, X. and Y.-X. Fu, 2015 Exploring population size changes using SNP frequency spectra. *Nature*  
863 Genetics **47**: 555 EP –.
- 864 Lloyd, M., 1967 ‘Mean crowding’. *Journal of Animal Ecology* **36**: 1–30.
- 865 Lundgren, E. and P. L. Ralph, 2019 Are populations like a circuit? Comparing isolation by resistance to  
866 a new coalescent-based method. *Molecular Ecology Resources* **19**: 1388–1406.
- 867 Martin, A. R., M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, *et al.*, 2019 Clinical use of current polygenic  
868 risk scores may exacerbate health disparities. *Nature Genetics* **51**: 584–591.
- 869 Maruyama, T., 1972 Rate of decrease of genetic variability in a two-dimensional continuous population  
870 of finite size. *Genetics* **70**: 639–651.
- 871 Mathieson, I. and G. McVean, 2012 Differential confounding of rare and common variants in spatially  
872 structured populations. *Nature Genetics* **44**: 243 EP –.
- 873 Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi, 2015 On the importance of being  
874 structured: instantaneous coalescence rates and human evolution—lessons for ancestral population  
875 size inference? *Heredity* **116**: 362 EP –.
- 876 Miles, A. and N. Harding, 2017 *cgh/scikit-allel*: v1.1.8.

- 877 Neel, M. C., K. McKelvey, N. Ryman, M. W. Lloyd, R. Short Bull, *et al.*, 2013 Estimation of effective  
 878 population size in continuously distributed populations: there goes the neighborhood. *Heredity* **111**:  
 879 189 EP –.
- 880 Novembre, J. and M. Slatkin, 2009 Likelihood-based inference in isolation-by-distance models using  
 881 the spatial distribution of low-frequency alleles. *Evolution* **63**: 2914–2925.
- 882 Pamilo, P., 1983 Genetic differentiation within subdivided populations of formica ants. *Evolution* **37**:  
 883 1010–1022.
- 884 Patterson, N., A. L. Price, and D. Reich, 2006 Population structure and eigenanalysis. *PLOS Genetics* **2**:  
 885 1–20.
- 886 Petkova, D., J. Novembre, and M. Stephens, 2015 Visualizing spatial population structure with esti-  
 887 mated effective migration surfaces. *Nature Genetics* **48**: 94 EP –.
- 888 Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, *et al.*, 2006 Principal  
 889 components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*  
 890 **38**: 904 EP –.
- 891 Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus  
 892 genotype data. *Genetics* **155**: 945–959.
- 893 Ptak, S. E. and M. Przeworski, 2002 Evidence for population growth in humans is confounded by  
 894 fine-scale population structure. *Trends in Genetics* **18**: 559–563.
- 895 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, *et al.*, 2007 Plink: A tool set for  
 896 whole-genome association and population-based linkage analyses. *The American Journal of Human  
 897 Genetics* **81**: 559 – 575.
- 898 R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical  
 899 Computing, Vienna, Austria.
- 900 Ralph, P. and G. Coop, 2013 The geography of recent genetic ancestry across Europe. *PLoS Biol* **11**:  
 901 e1001555.
- 902 Ralph, P., J. Kelleher, J. Galloway, , and J. Ashander, 2019a pyslim.
- 903 Ralph, P., K. Thornton, and J. Kelleher, 2019b Efficiently summarizing relationships in large samples: a  
 904 general duality between statistics of genealogies and genomes. *bioRxiv* .
- 905 Ringbauer, H., G. Coop, and N. H. Barton, 2017 Inferring recent demography from isolation by distance  
 906 of long shared sequence blocks. *Genetics* **205**: 1335–1351.
- 907 Robledo-Arnuncio, J. J. and F. Rousset, 2010 Isolation by distance in a continuous population under  
 908 stochastic demographic fluctuations. *Journal of Evolutionary Biology* **23**: 53–71.
- 909 Rossine, F. W. S., 2014 *Espaço e diversificação: uma perspectiva teórica*. Master's dissertation in ecologia:  
 910 Ecossistemas terrestres e aquáticos, University of São Paulo, São Paulo : Instituto de Biociências.
- 911 Rousset, F., 1997 Genetic differentiation and estimation of gene flow from F-statistics under isolation  
 912 by distance. *Genetics* **145**: 1219–1228.
- 913 Rousset, F. and R. Leblois, 2011 Likelihood-based inferences under isolation by distance: Two-  
 914 dimensional habitats and confidence intervals. *Molecular Biology and Evolution* **29**: 957–973.
- 915 Sawyer, S., 1977 On the past history of an allele now known to have frequency p. *Journal of Applied  
 916 Probability* **14**: 439–450.
- 917 Schiffels, S. and R. Durbin, 2014 Inferring human population size and separation history from multiple  
 918 genome sequences. *Nature Genetics* **46**: 919 EP –.
- 919 Schrider, D. R. and A. D. Kern, 2018 Supervised machine learning for population genetics: A new  
 920 paradigm. *Trends in Genetics* **34**: 301 – 312.
- 921 Sharbel, T. F., B. Haubold, and T. Mitchell-Olds, 2000 Genetic isolation by distance in *Arabidopsis  
 922 thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology* **9**: 2109–2118.
- 923 Sheehan, S., K. Harris, and Y. S. Song, 2013 Estimating variable effective population sizes from multiple  
 924 genomes: A sequentially Markov conditional sampling distribution approach. *Genetics* **194**: 647–662.
- 925 Shirk, A. J. and S. A. Cushman, 2014 Spatially-explicit estimation of Wright's neighborhood size in  
 926 continuous populations. *Frontiers in Ecology and Evolution* **2**: 62.
- 927 Slatkin, M. and N. H. Barton, 1989 A comparison of three indirect methods for estimating average  
 928 levels of gene flow. *Evolution* **43**: 1349–1368.
- 929 Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, *et al.*, 2018 Signals of polygenic

- adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. *bioRxiv*.
- St. Onge, K. R., A. E. Palmé, S. I. Wright, and M. Lascoux, 2012 Impact of sampling schemes on demographic inference: An empirical study in two species with different mating systems and demographic histories. *G3: Genes, Genomes, Genetics* **2**: 803–814.
- Städler, T., B. Haubold, C. Merino, W. Stephan, and P. Pfaffelhuber, 2009 The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**: 205–216.
- Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Terhorst, J., J. A. Kamm, and Y. S. Song, 2016 Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* **49**: 303 EP –.
- Turchin, M. C., C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, *et al.*, 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nature Genetics* **44**: 1015 EP –.
- Wahlund, S., 1928 Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. *Hereditas* **11**: 65–106.
- Wakeley, J., 1999 Nonequilibrium migration in human history. *Genetics* **153**: 1863–1871.
- Wakeley, J., 2009 *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village, CO.
- Wakeley, J. and T. Takahashi, 2003 Gene genealogies when the sample size exceeds the effective size of the population. *Mol Biol Evol* **20**: 208–213.
- Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wilke, C. O., 2019 *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. R package version 0.9.4.
- Wilkins, J. F., 2004 A separation-of-timescales approach to the coalescent in a continuous population. *Genetics* **168**: 2227–2244.
- Wilkins, J. F. and J. Wakeley, 2002 The coalescent in a continuous, finite, linear population. *Genetics* **161**: 873–888.
- Wray, N. R., M. E. Goddard, and P. M. Visscher, 2007 Prediction of individual genetic risk to disease from genome-wide association studies. *Genome research* **17**: 1520–1528.
- Wright, S., 1931 Evolution in Mendelian populations. *Genetics* **16**: 97.
- Wright, S., 1943 Isolation by distance. *Genetics* **28**: 114–138.
- Wright, S., 1946 Isolation by distance under diverse systems of mating. *Genetics* **31**: 336.
- Young, A. I., M. L. Frigge, D. F. Gudbjartsson, G. Thorleifsson, G. Bjornsdottir, *et al.*, 2018 Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics* **50**: 1304–1310.
- Young, H. J., 1988 Neighborhood size in a beetle pollinated tropical aroid: effects of low density and asynchronous flowering. *Oecologia* **76**: 461–466.
- Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki, *et al.*, 2005 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203 EP –.
- Zähle, I., J. T. Cox, and R. Durrett, 2005 The stepping stone model. II. Genealogies and the infinite sites model. *Ann. Appl. Probab.* **15**: 671–699.

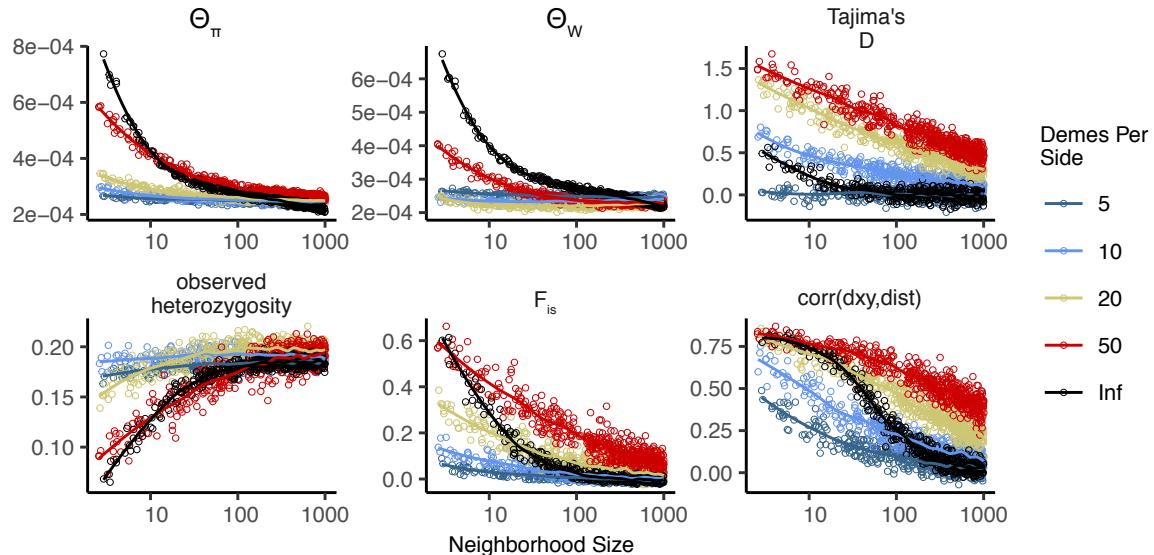
## 972 Comparisons with Stepping-Stone Models

973 apx:stepping,tone  
 We also compared our model results to a regular grid of discrete populations, which is commonly  
 974 used as an approximation of continuous geography. An important reason that this approximation  
 975 is often made is that it allows more efficient, coalescent simulations; we implemented these using  
 976 msprime (?). In this class of models we imagine an  $n \times n$  grid of populations exchanging migrants  
 977 with neighboring populations at rate  $m$ . If these models are good approximations of the continuous  
 978 case we expect that results will converge as  $n \rightarrow \infty$  (while scaling  $m$  appropriately and keeping total

population size fixed), so we ran simulations while varying  $n$  from 5 to 50 (Table ??). To compare with continuous models we first distributed the same “effective” number of individuals across the landscape as in our continuous-space simulations ( $\approx 6100$ , estimated from  $\theta_\pi$  of random-mating continuous-space simulations). We then approximate the mean per-generation dispersal distance  $\sigma$  given a total landscape width  $W$  as the product of the probability of an individual being a migrant and the distance traveled by migrants:  $\sigma = 4m(W/n)$ . This means that  $m$  in different simulations with the same  $\sigma$  scales with  $\sqrt{n}$ . We ran 500 simulations for each value of  $n$  while sampling  $\sigma$  from  $U(0.2, 4)$ . We then randomly selected 60 diploid individuals from each simulation (approximating diploidy by combining pairs of chromosomes with contiguous indices within demes) and calculated a set of six summary statistics using the scripts described in the summary statistics portion of the main text.

demes per side ( $n$ )	$N_e$ per deme	samples per deme
5	244	20
10	61	10
20	15.25	2
50	2.44	1

**Table A1** stepping-stone simulation parameters



**Figure A1** Summary statistics for 2-dimensional coalescent stepping-stone models with fixed total  $N_e$  and varying numbers of demes per side. The black “infinite” points are from our forward-time continuous space model. Inter-deme migration rates are related to  $\sigma$  as described above.

In general we find many of the qualitative trends are similar among continuous and stepping-stone models and that, at low neighborhood sizes, many (but not all) statistics from stepping-stone models approach the continuous model as the resolution of the grid increases. For example,  $\theta_\pi$  is lower in stepping-stone models at low neighborhood sizes (i.e., low  $m$ ), but increases to approach the continuous case as the resolution of the landscape increases. Similar patterns are observed for observed heterozygosity. However,  $\theta_W$  behaves differently, showing a non-monotonic relationship with grid resolution. This results in an increasingly positive Tajima’s  $D$  in grid simulations at small neighborhood

sizes, to a much greater extent than seen in a continuous model. In contrast to  $\theta_\pi$ , increasing the resolution of the grid causes Tajima's  $D$  to deviate *more* from what is seen in the continuous model. Similarly, although  $F_{IS}$  approaches the continuous case as the resolution of the grid increases at very small neighborhood sizes, at intermediate neighborhood sizes the continuous case best matches intermediate grid resolutions.

These differences relative to our continuous model mainly reflect two shortcomings of the reverse-time stepping stone model. If we simulate a coarse grid with relatively large populations in each deme, we cannot accurately capture the dynamics of small neighborhood sizes because mating within each deme remains random regardless of the migration rate connecting demes. This likely explains the trends in  $\theta_\pi$ , observed heterozygosity, and  $F_{IS}$ . However increasing the number of demes while holding the total number of individuals constant results in small within-deme populations for which even the minimum sample size of 1 approaches the local  $N_e$  (Table ??). This results in an excess of short terminal branches in the coalescent tree, which decreases the total branch length and leads to fewer segregating sites, deflated  $\theta_W$ , and inflated Tajima's  $D$ . Overall, the stepping-stone model reproduces important features of spatial structure in our continuous space model, such as a decline in  $\theta_\pi$  and correlations between spatial and genetic distance with increasing migration, but introduces artifacts caused by binning the landscape into discrete demes.

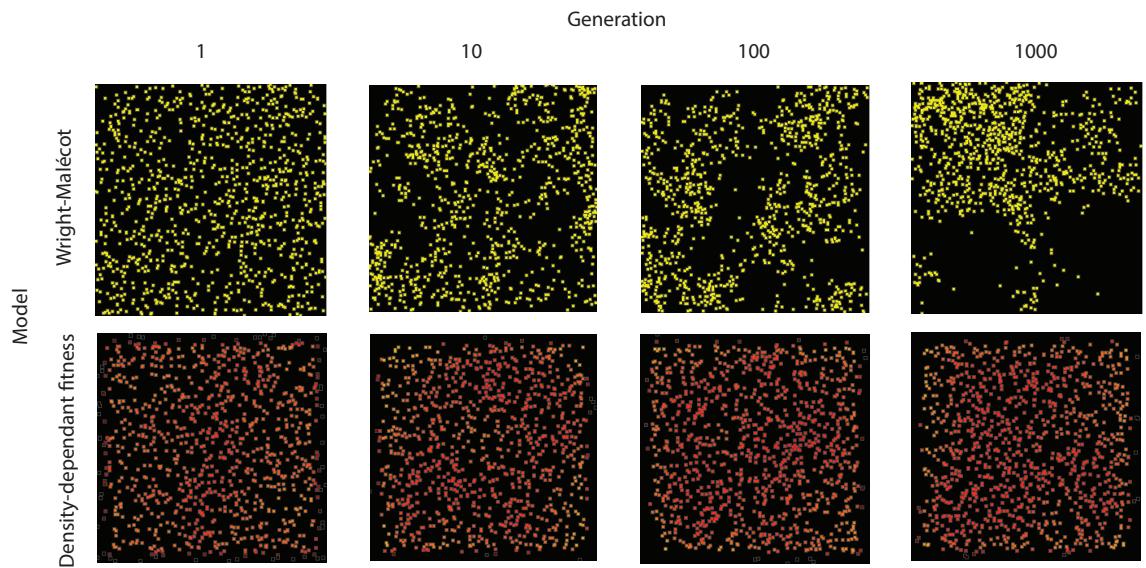
## Demographic model

apx:demographic<sub>model</sub> Local population regulation is controlled by two parameters,  $L$ , and  $K$ . Here, we show that these should be close to  $\min\left(0.95, \frac{1}{1+n/(K(1+L))}\right)$ . (3) We capped survival at 0.95 so that we would not have exceptionally long-lived individuals in sparsely populated areas – otherwise, an isolated individual might live for a very long time. Since  $1 - p \approx n/(K(1 + L))$ , mortality goes up roughly linearly with the number of neighbors (on a scale given by  $K$ ), as would be obtained if, for instance, mortality is due to agonistic interactions. Ignoring migration, a region is at demographic equilibrium if the per-capita probability of death is equal to the birth rate, i.e., if  $1 - p = 1/L$ . (Note that there is no effect of age in the model, which would make the analysis more complicated.) Solving this for  $n$ , we get that in a well-mixed population, the equilibrium density should be around

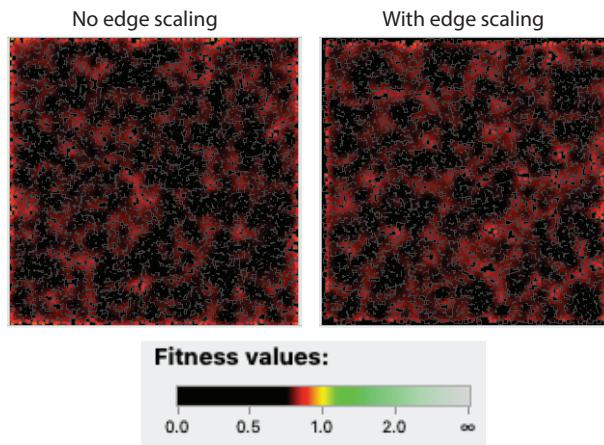
$$n = K \frac{L + 1}{L - 1} \quad (4)$$

individuals per unit area. At this density, the per-capita death rate is  $1/L$ , so the mean lifetime is  $L$ . This equilibrium density is *not*  $K$ , but (since  $L = 4$ ) is two-thirds larger. However, in practice this model leads to a total population size which is around  $K$  multiplied by total geographic area (but which depends on  $\sigma$ , as discussed above). The main reason for this is that since offspring tend to be near their parents, individuals tend to be “clumped”, and so experience a higher average density than the “density” one would compute by dividing census size by geographic area (?). To maintain a constant expected total population size would require making (say)  $K$  depend on  $\sigma$ ; however, typical local population densities might then be more dissimilar.

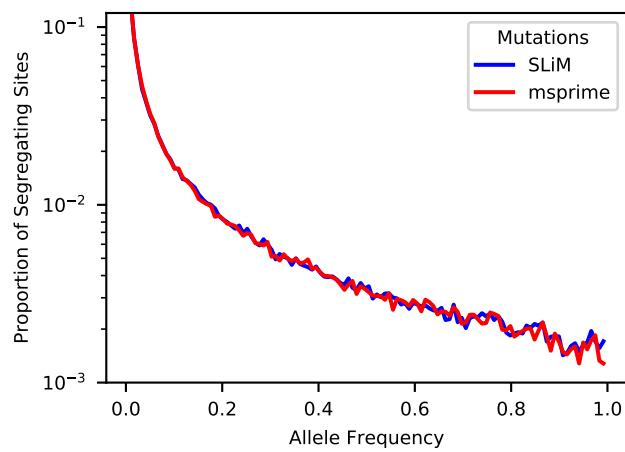
<sup>1022</sup> **Supplementary Figures and Tables**



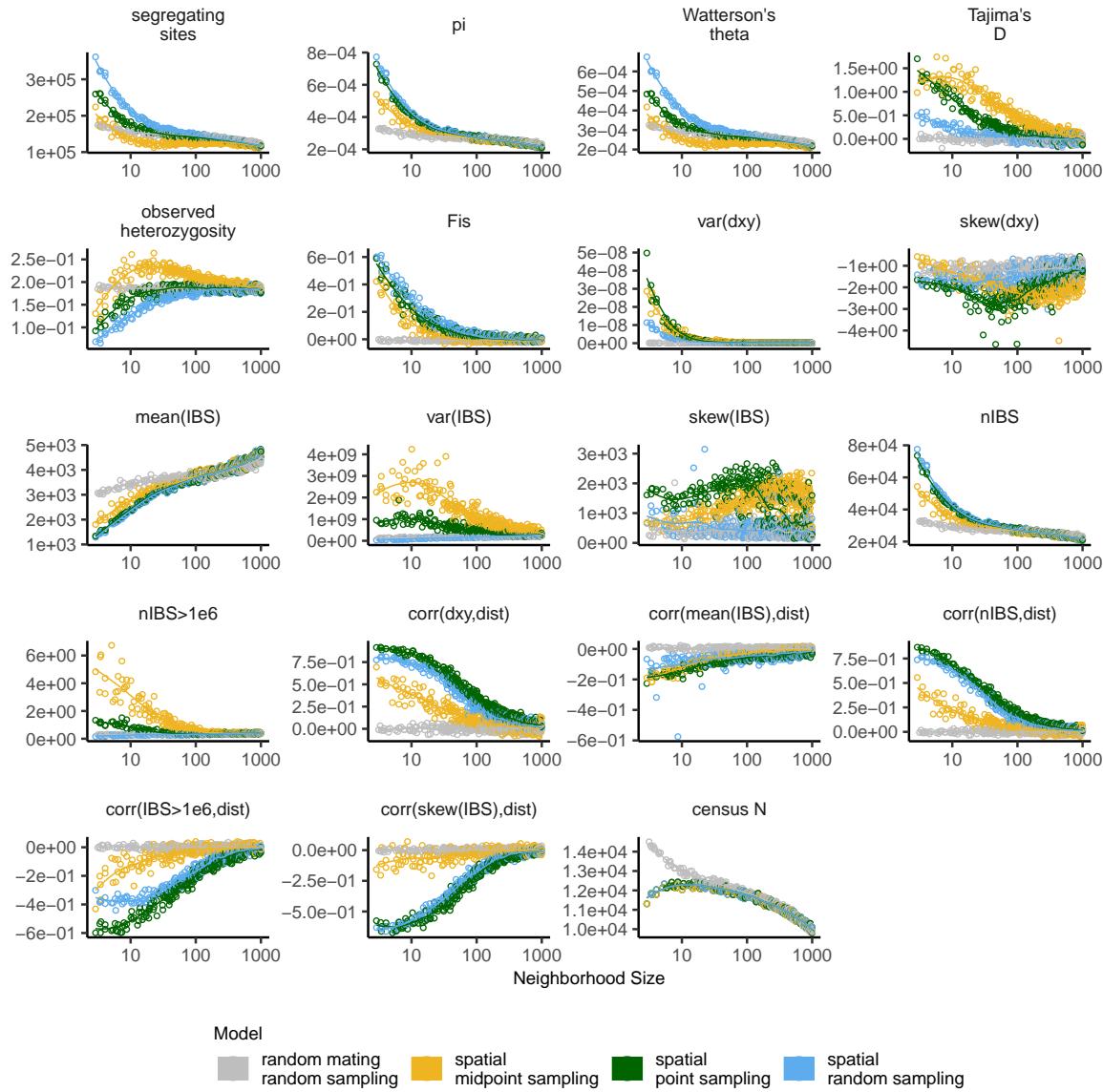
**Figure S1** Maps of individual locations in a continuous-space Wright-Malécot model with independent dispersal of all individuals (top) and under our continuous space model incorporating density-dependant fitness (bottom). The clustering seen in the top row is the “Pain in the Torus” described by ?.



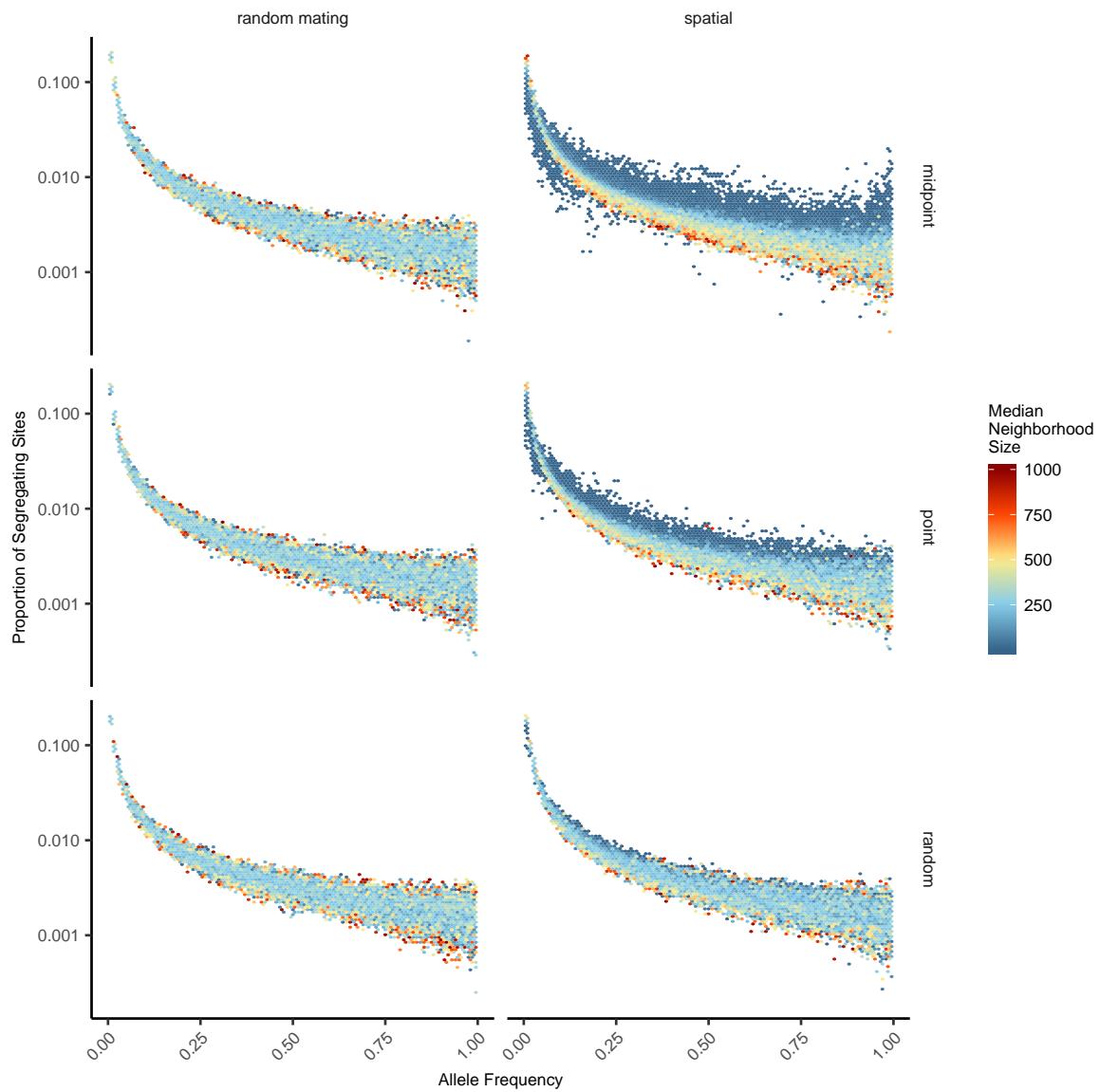
**Figure S2** Comparison of individual fitness across the landscape in simulations with (right) and without (left) a decline in fitness approaching range edges. Note the slight excess of high-fitness individuals at edges on the left, which is (partially) counteracted by the scaling procedure.



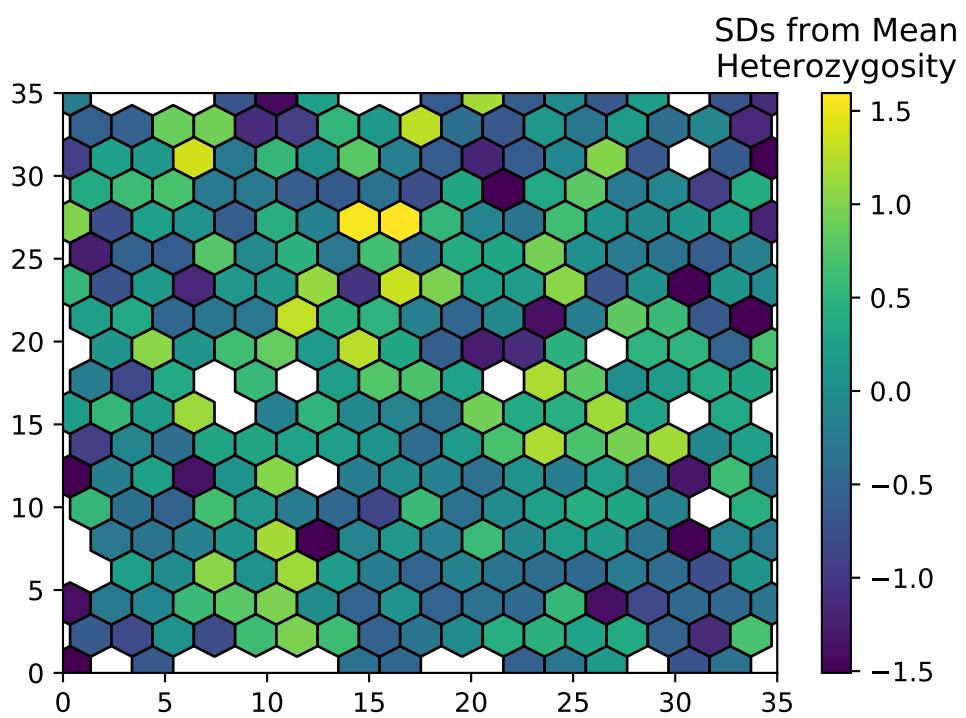
**Figure S3** Site frequency spectra from a simulation with neighborhood size = 12.5 when mutations are recorded directly in SLiM (blue line) or applied later in msprime (red line).



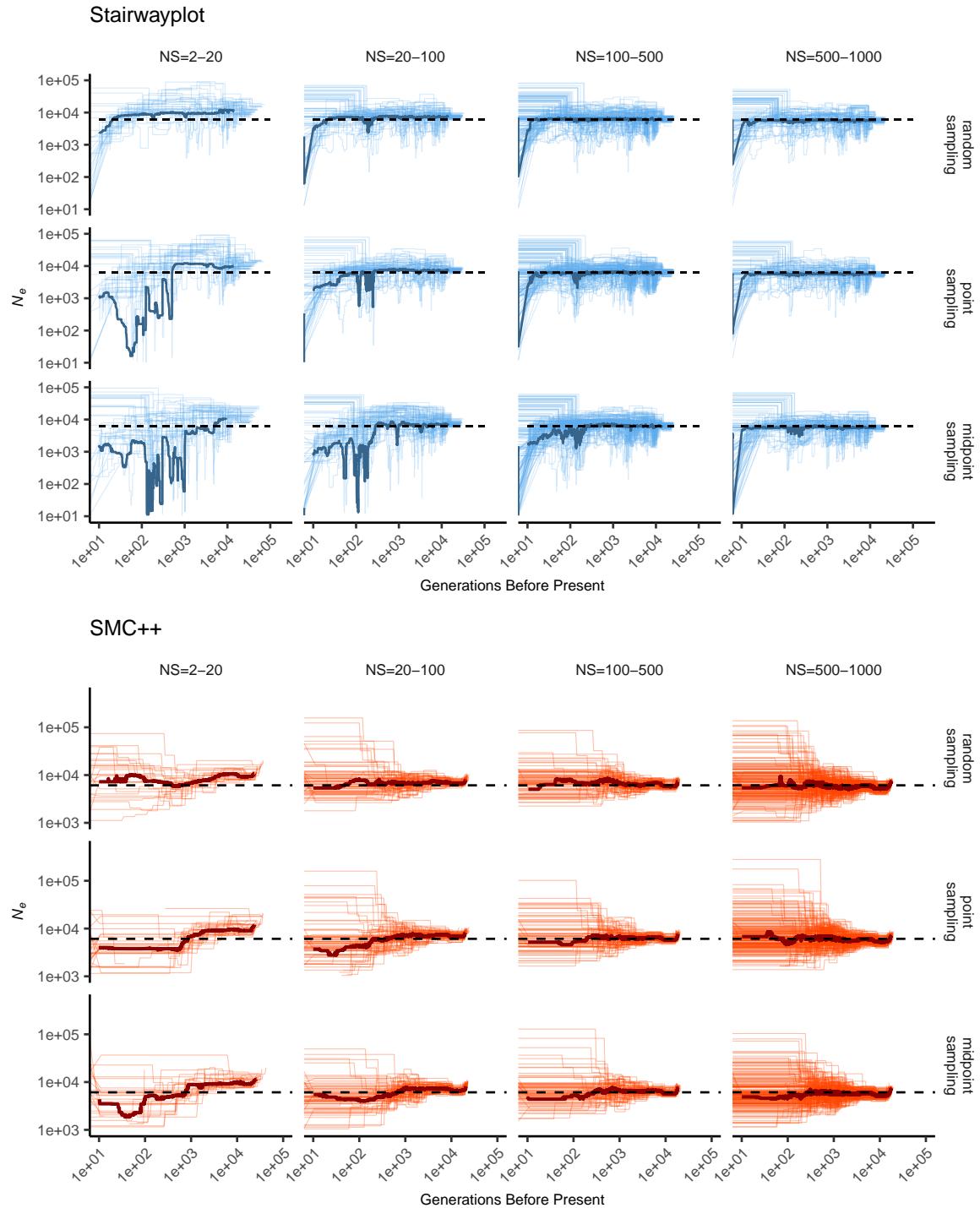
**Figure S4** Change in summary statistics by neighborhood size and sampling scheme calculated from simulated sequence data of 60 individuals.



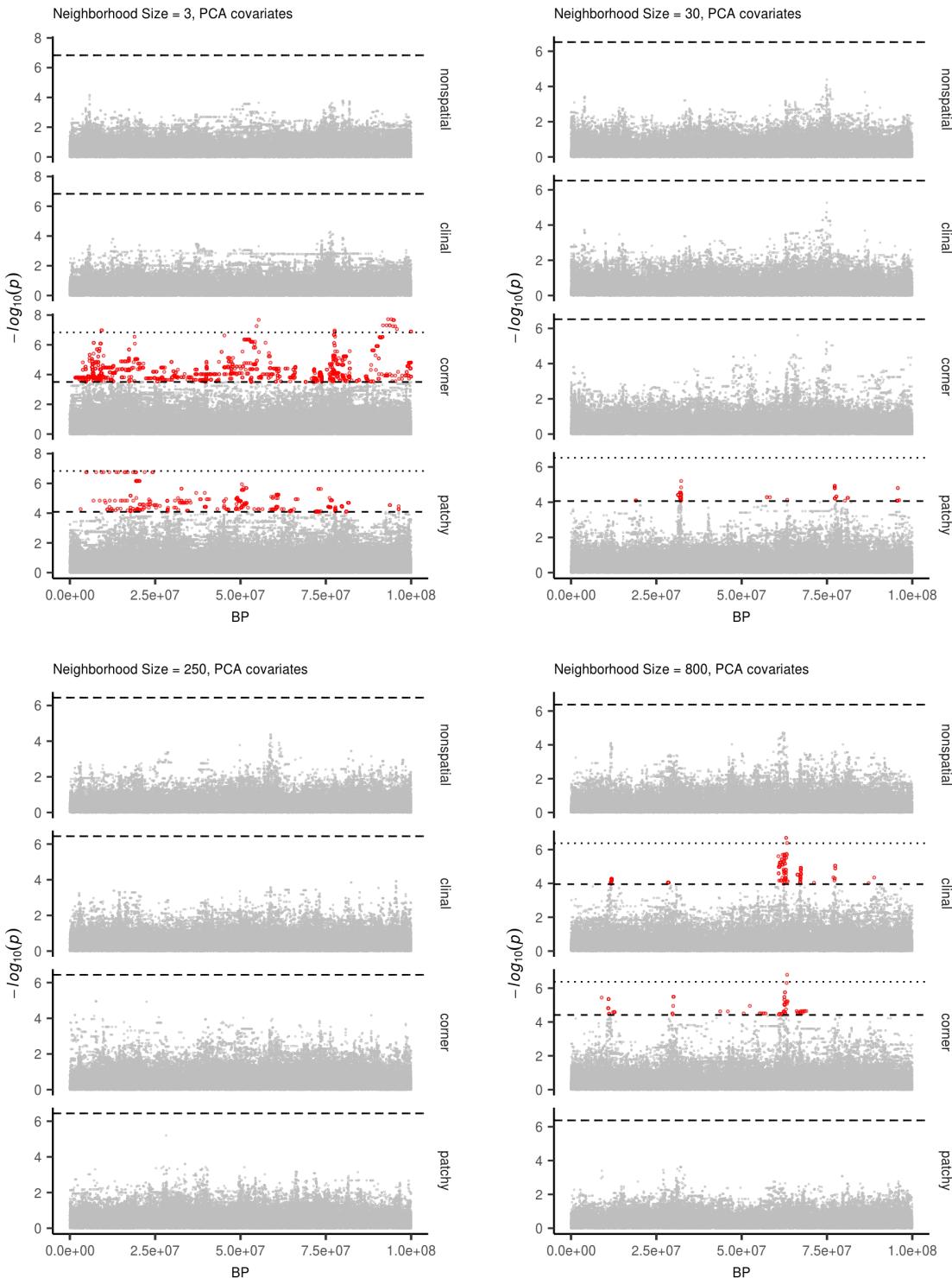
**Figure S5** Site frequency spectra for random mating and spatial SLiM models under all sampling schemes.



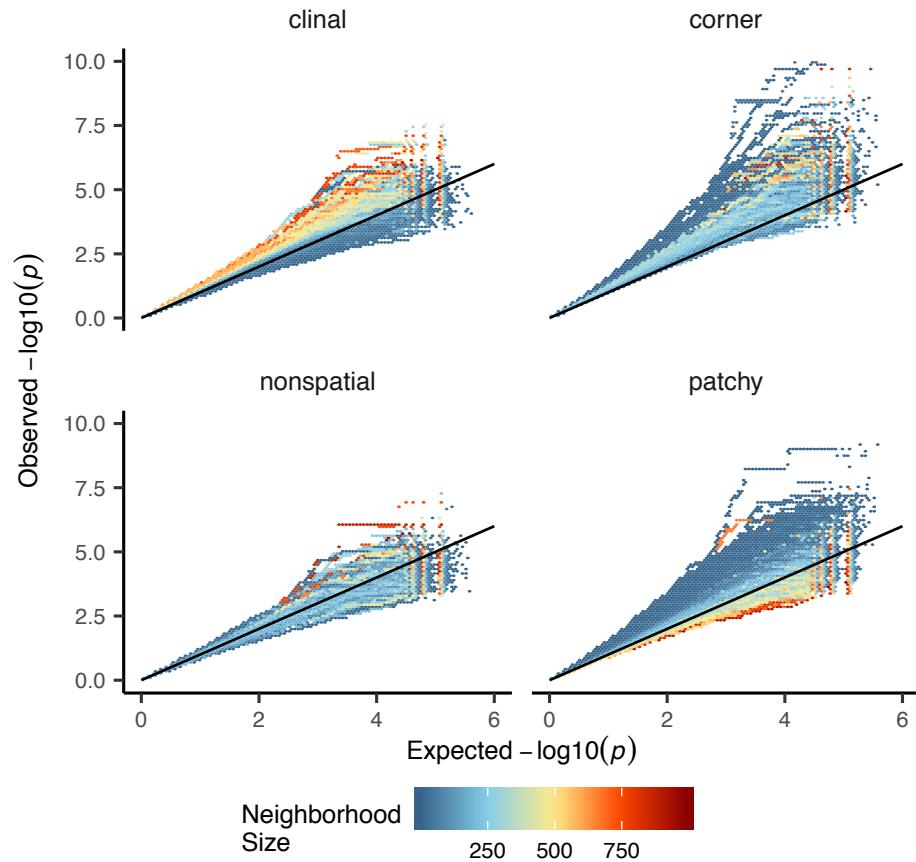
**Figure S6** Variation in observed heterozygosity (i.e. proportion of heterozygous individuals) in hexagonal bins across the landscape, estimated from a random sample of 200 individuals from the final generation of a simulation with neighborhood size  $\approx 25$ . Values were Z-normalized for plotting.



**Figure S7** 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 S8** 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  $p = 0.05$ .



**Figure S9** Quantile-quantile plots showing observed  $-\log_{10}(p)$  for PC-corrected GWAS run on simulations with varying neighborhood sizes and environmental distributions. Hexagonal bins are colored by the average neighborhood size of simulations with points falling in a given region of quantile-quantile space. Qqplots for a subset of these simulations are shown as lines in Figure ??D.

**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(equal means)	p(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
$mean(D_{xy})$	random mating	0.997770	0.996510
$var(D_{xy})$	random mating	0.283630	0.647240
$skew(D_{xy})$	random mating	0.958320	0.260750
$corr(D_{xy}, dist)$	random mating	0.601980	0.000000
$mean(IBS)$	random mating	0.997960	0.997730
$var(IBS)$	random mating	0.486450	0.399490
$skew(IBS)$	random mating	0.117980	0.069770
$nIBS$	random mating	0.997680	0.996570
$nIBS > 1e6$	random mating	0.834870	0.888730
$corr(mean(IBS), dist)$	random mating	0.073270	0.308420
$corr(IBS > 1e6, dist)$	random mating	0.268440	0.002100
$corr(skew(IBS), dist)$	random mating	0.396920	0.000620
$corr(nIBS, dist)$	random mating	0.581090	0.000000
segsites	spatial	0.000000	0.000000
$\Theta\pi$	spatial	0.026510	0.013440
$\Theta_W$	spatial	0.000000	0.000000
Tajima's $D$	spatial	0.000000	0.000000
observed heterozygosity	spatial	0.000000	0.000000
$F_{IS}$	spatial	0.000000	0.000120
$mean(D_{xy})$	spatial	0.025390	0.012910
$var(D_{xy})$	spatial	0.004970	0.006230
$skew(D_{xy})$	spatial	0.000000	0.000000
$corr(D_{xy}, dist)$	spatial	0.000000	0.000000
$mean(IBS)$	spatial	0.272400	0.114250
$var(IBS)$	spatial	0.000000	0.000000
$skew(IBS)$	spatial	0.000000	0.000000
$nIBS$	spatial	0.033920	0.016640
$nIBS > 1e6$	spatial	0.000000	0.000000
$corr(mean(IBS), dist)$	spatial	0.000000	0.590540
$corr(IBS > 1e6, dist)$	spatial	0.000000	0.000000
$corr(skew(IBS), dist)$	spatial	0.000000	0.000000
$corr(nIBS, dist)$	spatial	0.000000	0.000000