Spatial autocorrelation of the environment influences the patterns and genetics of local adaptation

Tom R. Booker[1,2]

1. Department of Zoology, Faculty of Science, University of British Columbia
2. Department of Forest and Conservation Science, Faculty of Forestry, University of British Columbia

Correspondence to: booker@zoology.ubc.ca

Abstract

Environmental heterogeneity can lead to spatially varying selection, which can, in turn, lead to local adaptation. Long-held views in evolutionary biology predict that environmental heterogeneity that is spatially autocorrelated will lead to stronger local adaptation than if it is not. Here I test this explicitly, combining analysis of a dataset from a large-scale provenance trial conducted on lodgepole pine with simulations modelling local adaptation. I find evidence that environmental heterogeneity that is spatially autocorrelated leads to stronger local adaptation in natural populations. Using simulations, I examine how the genetics of local adaptation are influenced by the spatial pattern of environmental heterogeneity. The simulations confirm that local adaptation increases with the degree of spatial autocorrelation in the selective environment, but also show that highly heterogeneous environments are more likely to exhibit high variation in local adaptation, a result not previously described. Spatial autocorrelation in the environment influences the genetic architecture of local adaptation, with different patterns of allele frequency and effect size arising under different patterns of environmental variation. These differences will likely impact the ability to study the genetic basis of local adaptation. Overall, this work emphasizes the profound importance that the spatial pattern of selection can have on the evolution and how spatial autocorrelation should be considered in ecological and genetic studies of local adaptation.

Introduction

Local adaptation is an important phenomenon in the natural world. Along with phenotypic plasticity, local adaptation can dictate the extent of environmental heterogeneity that a species can tolerate, shaping its geographic range [1], and help predict how it will respond to changing environments [2]. Local adaptation can be defined as a kind of genotype-by-environment interaction for fitness, where individuals have higher chances of survival and/or reproduction when they are reared at home as opposed to away, though several other definitions are used in the literature [3, 4]. Local adaptation is a property of a particular population at a particular point in time rather than a property of species as a whole. For example, populations at range edges are often expected to be maladapted to their conditions, while populations in the core of a range may be well adapted [5]. As such, locally adapted populations may harbour genetic variation that could help buffer susceptible populations against the detrimental effects of climate change [6], which are already wreaking havoc on important species around the world [7]. A deep understanding of local adaptation, the agents that have given rise to it and the genetics that underpin this phenomenon is thus important for our understanding of biodiversity and for species management and conservation in the Anthropocene [6, 8, 9].

The ultimate cause of local adaptation is variation in the environment. Whether it is biotic (e.g. disease/parasite prevalence or intraspecific competition) or abiotic (e.g. climate, geology or photoperiod), variation in the environment may lead to spatially varying selection pressures where phenotypic optima differ over a landscape. Such variation in selection across space has been well documented (e.g. [10]) and there are, of course, myriad aspects of the environment that could conceivably induce spatially varying selection, many of which would be highly inter-correlated. However, while there is an infinite number of ways to describe the environment, most of these may be functionally disconnected from a species' biology. A recent review by Wadgymer et al [8] highlighted a critical gap in our knowledge of local adaptation - that the aspects of environmental variation that have given rise to local adaptation (what they term the 'agents of selection') are unknown in most cases. For example, in the absence of experimental evidence many genetic studies have assumed that various climatic measures recorded in databases such as WorldClim correspond to agents of selection and search for the genetic basis of local adaptation using those data [11]. However, that a particular aspect environmental variation could conceivably induce spatially varying selection is not a guarantee that it will have led to local adaptation.

Population genetic studies have revealed numerous factors that can influence the evolution of local adaptation, the most important being the strength of selection and rates of gene flow among populations [12]. The strength of natural selection is of foremost importance because larger fitness consequences for deviating from the optimal phenotype in a particular location can potentially lead to greater evolutionary change [13]. The rate of gene flow is important because migration into a region experiencing idiosyncratic selection can overwhelm that selection, preventing regional trait differences (i.e. local adaptation) from accumulating [12, 14]. Indeed, the probability that local adaptation will arise in a particular location is related to the ratio of gene flow from dissimilar environments (m) to the strength of selection (s), m/s [12, 14, 15]. This general logic holds for models of single alleles and quantitative traits, both in discrete and continuous space, though in continuous space it is $\sqrt{s}$ that is important [16, 17]. A large proportion of species exhibit a specific pattern of population structure termed "isolation-by-distance" (IBD) where genetic distance is positively correlated with geographic distance or measures of resistance based on features of the landscape [18]. IBD can arise when rates of gene flow are highest among parts of a species' range that are in close proximity, though it can also reflect demographic histories such as past population expansion [19]. For a species experiencing

spatially varying selection, the relationship between the spatial pattern of population structure and the spatial pattern of selection will influence the *m/s* ratio across the landscape, influencing the probability that local adaptation will evolve in different parts of a species' range.

Since at least the 1960s, it has been recognised that the spatial pattern of environmental heterogeneity can influence the evolution of local adaptation [20–23]. The spatial pattern of environmental variation will influence how much gene flow a given population receives from dissimilar environments and how dissimilar those environments are. Of course, the scale over which the environment varies and the scale over which populations disperse are critically important parameters; if individuals disperse widely, the corresponding change in the environment observed over typical dispersal distances is what matters and this is captured in what is referred to as the "characteristic length" of a cline [24]. Previous theoretical studies have shown that the overall level of local adaptation a species exhibits, and potentially the probability of ecological speciation, can be strongly affected by the pattern of environmental variation [23, 24]. Several studies have framed this idea in terms of the spatial autocorrelation of the environment [23, 25]. When aspects of the environment that give rise to spatially varying selection exhibit high spatial autocorrelation (such as the right-hand map in Figure 2A) nearby populations will experience similar selection pressures. As a result, greater levels of local adaptation may evolve across a species' range if the spatial pattern of environmental variation aligns with the opportunity for migration. There is empirical evidence that spatial autocorrelation in patterns of biotic interactions can explain a substantial proportion of variation in trait differentiation among populations [25]. However, empirical evidence that patterns of local adaptation coincide with spatially autocorrelated features of the environment is lacking [25]. Despite strong theoretical predictions about the evolution of local adaptation when the agents of selection are autocorrelated in space, the genetics of local adaptation in such cases are not well characterised.

In this study, I examine how the patterns of local adaptation and the genetics underlying it are affected by the spatial pattern of environmental variation. I use empirical data from a large-scale experiment in lodgepole pine and find evidence that spatial autocorrelation in patterns of climatic/environmental variation are related to local adaptation in a natural system. I then build a simulation model of spatially varying selection to investigate how local adaptation and its genetic bases can be influenced by the pattern of environmental variation. I examine the genetic architecture of local adaptation and its evolution as it relates to varying degrees of spatial autocorrelation in the environment. Overall, I show that spatial autocorrelation in the environment can have a profound impact on the patterns and genetics of local adaptation and discuss implications from this work for the study of local adaptation in natural populations.

Results

Local adaptation and environmental structure in the Illingworth Trial

Local adaptation is widely observed in forest trees [26] and, in some regions, this is central to plans to adapt forestry practice in light of climate change [27, 28]. Provenance trials have been used for almost 200 years in forest trees and involve planting multiple populations of a species in numerous common gardens to assess how "transfer distance", the distance between home and the common garden, affects productivity (reviewed in [28]). The Illingworth trial is an exceptionally large provenance trial established by the Ministry of Forestry in British Columbia, Canada in the 1970s to establish seed-transfer guidelines for the lodgepole pine (*Pinus contorta*) [29]. Seeds were collected from 140 provenances from Northwestern North America and seedlings were planted in a set of 62 sites distributed across British Columbia (Figure 1A). Phenotypic data has been recorded for around 60,000 individual trees since the

Illingworth trial began and previous studies have used this data to demonstrate clear local adaptation in lodgepole pine [30, 31]. Given the geographic breadth of the Illingworth trial (Figure 1A), it represents a suitable dataset to test the prediction that the pattern of environmental variation influences the evolution of local adaptation.

Lodgepole pine exhibits the strongest signals of local adaptation when climatic/environmental variation is highly spatially autocorrelated (Figure 1). I analyzed data from the Illingworth trial using a mixed-modelling approach. Different aspects of climatic variation across the sites in the Illingworth trial are highly intercorrelated (Figure S1), so I used principal components analysis to identify independent axes of climatic/environmental variation in the dataset. Across all locations in the Illingworth trial, there were 8 principal components (PCs) that individually explained at least 1% of climatic/environmental variation (Figure S1A). I regressed phenotypes (either tree diameter at breast height or survival) on transfer distance between planting site and provenances in PC-space (Figure 1B-C), modelling provenance and planting sites as random effects (see Methods for details). Among the aspects of climatic variation that had significant slopes of diameter or survival on environmental distance, the steepest slopes was found for the PC with the greatest degree of spatial autocorrelation, as measured using Moran's I (Figure 1D-E). The PC in question (PC1) captures climatic/environmental differences between coastal and inland locations (Figure S2). Overall, there seemed to be a negative association between the strength of local adaptation and the degree of spatial autocorrelation in the environment (Figure 1D-E). However, only 4 or 5 PCs exhibited statistically significant evidence for local adaptation, so I did not conduct a formal statistical test of the relationship between spatial autocorrelation and local adaptation. Nevertheless, this result is suggestive that spatial autocorrelation in the abiotic environment predicts the strength of local adaptation in natural populations.

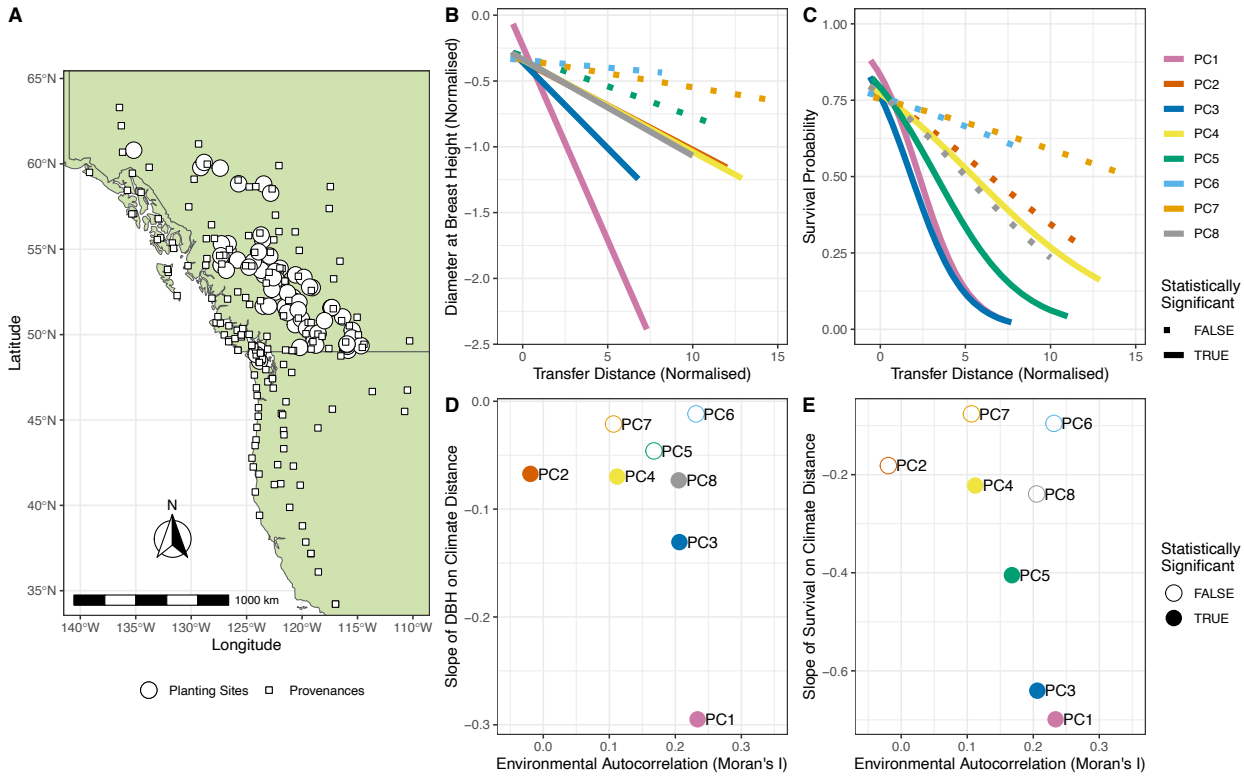


Figure 1 Analysis of local adaptation in lodgepole pine from the Illingworth provenance trial. Panel A) shows the map of provenances and planting sites in the Illingworth trials across the Northwest of North

America. B) Fitted relationships between tree diameter at breast height (DBH) and transfer distance for 8 principal components. C) Survival probability as a function of transfer function for 8 environmental principal components. D) Linear regression coefficients for the relationships shown in B compared to degree of autocorrelation in the environment. E) Logistic regression coefficients for the relationships shown in C compared to the degree of spatial autocorrelation in the environment. Statistical significance was assessed after correcting for multiple comparisons using the Dunn-Šidak method.

Simulating local adaptation to spatially heterogeneous environments

To further understand the effects that the spatial pattern of environmental variation can have on patterns of local adaptation, I constructed a simulation model of spatially varying selection. I used forward-in-time population genetic simulations in SLiM (4.1; [32]) modelling a 2-dimensional stepping-stone metapopulation of 196 demes (i.e. a 14x14 grid). Migration was restricted to adjacent demes with rates of gene flow that resulted in pronounced population structure with clear isolation-by-distance (Figure S3). Spatially varying selection was modelled as stabilising selection, where each deme (d) had a particular phenotypic optimum (θ_d), i.e. an individual in deme d had higher fitness if its phenotype was close to θ_d. The strength of stabilising selection was set such that an individual with the optimal phenotype for the deme with the most negative optimum translocated into the deme with the most positive optimum would experience a 50% reduction in fitness (strong selection) or a 25% reduction in fitness (moderate selection). I used a quantitative trait model to study local adaptation, because it is thought that the traits involved in local adaptation are generally polygenic [33]. Additionally, there is evidence that the genetic basis of local adaptation can involve alleles that have spatially antagonistic fitness effects as well as conditionally neutral effects [34] and both kinds of effects can arise in a quantitative trait model given variation in genetic backgrounds and environments. I constructed a set of 200 maps of normally distributed environmental variation that varied in the degree of spatial autocorrelation (three examples are shown in Figure 2A). I quantified spatial autocorrelation in the environment using Moran's I, which varied from 0.05 (weak autocorrelation) to 0.95 (strong autocorrelation) in the maps I constructed. The maps of environmental variation were used to specify phenotypic optima in individual simulations.

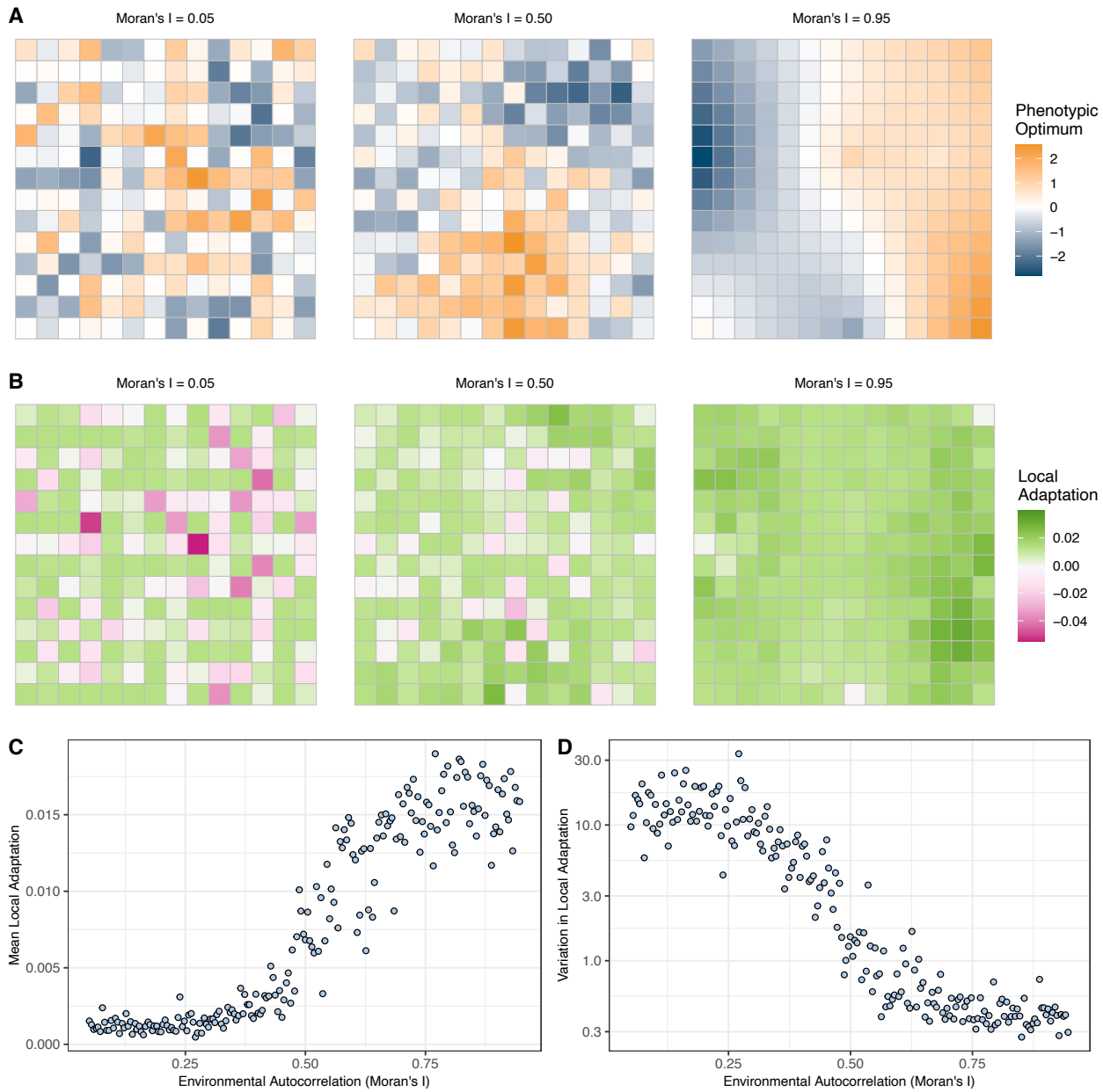


Figure 2 The pattern of environmental heterogeneity influences the outcomes of spatially varying selection. A) Three examples of environmental heterogeneity with similar distributions of phenotypic optima ranging from low autocorrelation on the left to high autocorrelation on the right. B) The pattern of local adaptation that evolved on the landscapes shown in panel A. C) The average local adaptation that arises as a function of Moran's I across 200 maps of environmental variation. C) The coefficient of variation for local adaptation across the 200 maps. The simulation results shown are for cases with mean F_{ST} ~ 2% and moderate stabilising selection.

The effects of environmental structure on patterns of local adaptation that evolved in simulations were profound. I measured local adaptation in simulated populations by comparing an individual's fitness at "home" versus "away" following the method outlined by [23]. Using this method the observed local adaptation for each deme in the metapopulation was computed (e.g. Figure 2B). As expected, the mean

local adaptation across populations increased with the degree of spatial autocorrelation in the environment (Figure 2C), consistent with [34]. This result held over various levels of gene flow and strengths of stabilising selection (Figure S4). Varying the rate of gene flow and/or the strength of selection had an effect on the level of local adaptation that arose in a given case, but a pattern of increasing mean local adaptation with Moran's I was always observed (Figure S4). In natural populations it is likely that distinct traits may be responding to separate selection pressures. In a model of two traits subject to spatially varying selection, local adaptation tended to be stronger for the trait with the pattern of selection that was more spatially autocorrelated (Figure S5), which agrees with the pattern found when analyzing data from the Illingworth trial (Figure 1).

The spatial pattern of environmental variation did not just affect the average level of local adaptation, though, it also had a large influence on the variation in local adaptation. The coefficient of variation in local adaptation across the landscape decreased rapidly with increasing autocorrelation (Figure 2D, S4). When environmental variation was weakly autocorrelated, the coefficient of variation in local adaptation was as much as 30x higher than for more highly autocorrelated environments (Figure 2D). In cases with weak environmental autocorrelation, the environmental variation in the immediate vicinity of a particular deme predicted levels of local adaptation (Figure S6). Demes with low local heterogeneity in the environment had greater local adaptation (Figure S6). Local heterogeneity in the environment also predicted levels of additive genetic variance within populations (Figure S6). Highly autocorrelated environments resulted in strong positive relationships between additive genetic variance (V_A) for phenotypes and local environmental heterogeneity, a pattern that has been documented for Lodgepole pine [35]. Environments that were weakly autocorrelated had weaker relationships between V_A and local environmental heterogeneity (Figure S6). Variation in the degree of local adaptation across a species range is understudied in the population genetics literature but has important implications (see Discussion).

Environmental structure and the genetic architecture of local adaptation

The results so far demonstrate that the spatial patterns of selection can have a clear impact on the patterns of local adaptation that evolve, but do they influence the genetic basis of that adaptation? Previous studies have examined the effects of landscape structure on the genetics of local adaptation in single locus models [21] or in models of population expansion [35–37], but it is unclear how the polygenic architecture of local adaptation will be influenced by the degree of autocorrelation in the environment. Under a model of spatially varying stabilising selection, each polymorphism that affects the phenotypes under selection will influence local adaptation to some extent. I quantified the contribution that a given polymorphism made to mean local adaptation ($\overline{LA}$) as follows. For each polymorphism in a simulation, I shuffled its distribution on the landscape then recalculated the average level of local adaptation across the species' range ($\overline{LA_\iota}$). Using this value, I quantify the relative contribution of that polymorphism to the average local adaptation across the landscape as $LA_{Rel} = (1 - \overline{LA_\iota}/\overline{LA})$. Shuffling the presence/absence of the polymorphism across the landscape effectively erases its pattern of local adaptation, while holding its contribution to overall additive genetic variance constant. Note that LA_{Rel} is not strictly a proportion, see Methods for details.

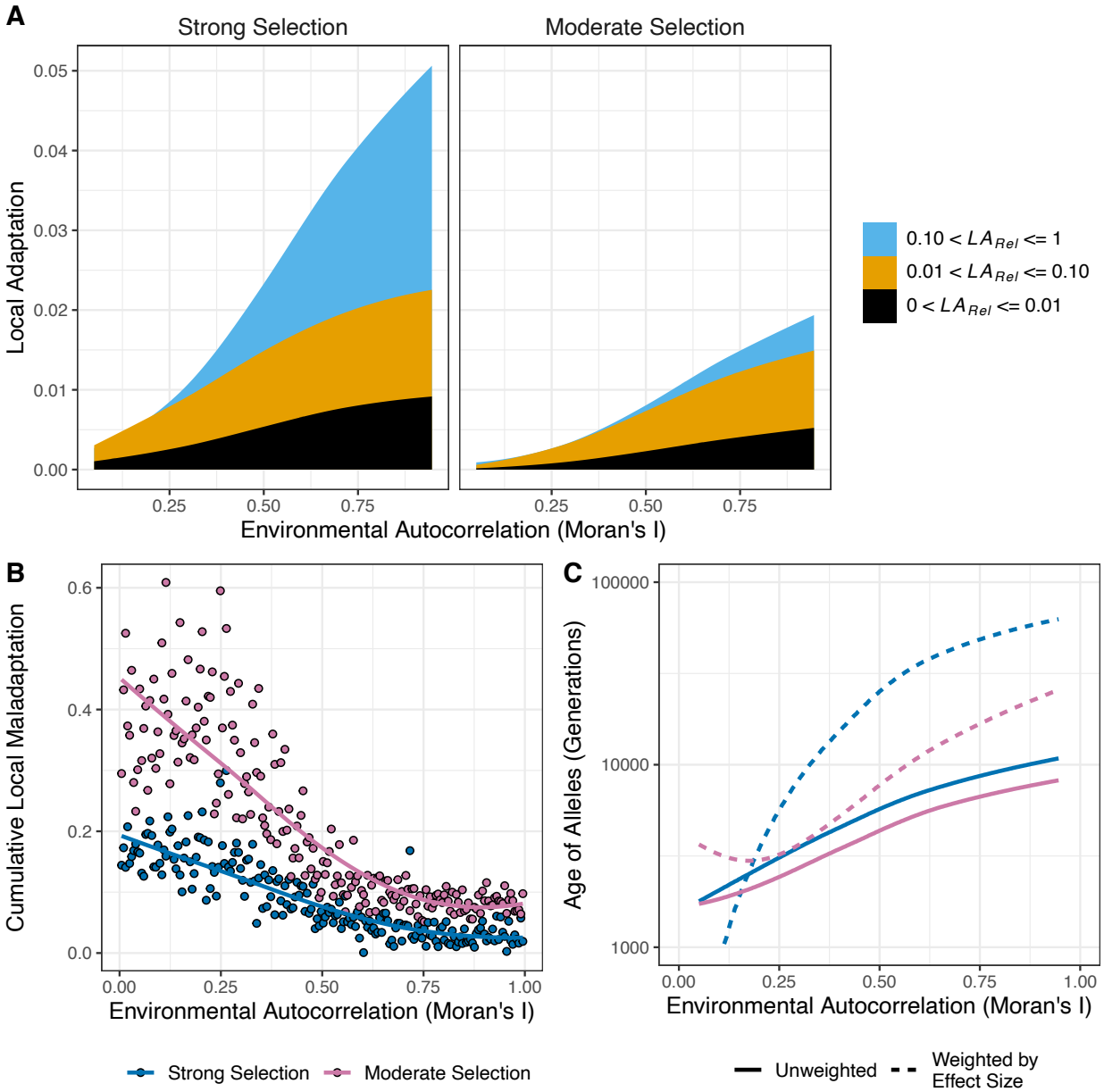


Figure 3 The genetic basis of local adaptation is influenced by the structure of the environment. A) The genetic architecture of local adaptation. The cumulative effects of locally maladaptive polymorphisms (B) and the age of alleles that contribute to local adaptation (C) varies as a function of environment. All plots show results for cases where mean F_{ST} = 2%. Lines represent LOESS curves fit with a span parameter of 1.5.

The degree of environmental autocorrelation had a clear influence on the genetic architecture of local adaptation (Figure 3). Figure 3A shows a discretized version of the distribution of LA_{Rel} as it relates to environmental autocorrelation. When the environment was weakly autocorrelated there was less local adaptation, but it was mostly explained by alleles with LA_{Rel} between 1% and 10% (Figure 3A). This was the case with both strong and moderate stabilising selection and over varying levels of gene flow (Figure 3A, S7). Alleles that made very large contributions to local adaptation ($LA_{Rel} > 10\%$) were largely

absent when autocorrelation was weak (Figure 3A, S7). Increasing the degree of spatial autocorrelation in the environment increased the fraction of local adaptation that was explained by polymorphisms that made large contributions (Figure 3A). Indeed, with high spatial autocorrelation, substantial fractions of local adaptation were explained by alleles across all levels of LA_{Rel} (Figure 3A). Locally maladaptive alleles were found at all levels of autocorrelation in the environment, but the cumulative effects of maladaptive alleles decreased sharply with increasing spatial autocorrelation (Figure 3B, S8). Finally, there was a positive relationship between the age of locally adaptive alleles and the extent of spatial autocorrelation in the environment, particularly for alleles that made large contributions to local adaptation (Figure 3C). Given that levels of local adaptation had reached a steady state in the simulations (Figure S9), this suggests that there is a higher turnover of locally adaptive alleles when space is weakly autocorrelated.

While the spatial pattern of environmental variation influences the relative contribution of individual polymorphisms to overall local adaptation, the contribution that a given polymorphism makes to local adaptation will depend on its phenotypic effect size and allele frequency across the landscape. These properties differ for locally adaptive polymorphisms depending on the spatial pattern of environmental variation (Figure 4). When the environment was weakly autocorrelated, locally adaptive alleles tended to be present at low frequencies and have large phenotypic effects (Figure 4). In weakly autocorrelated environments, a given deme likely receives migrants from dissimilar environments, so locally adaptive alleles will require large phenotypic effect sizes to resist the swamping effects of gene flow (see Introduction). In such cases, locally adaptive alleles may be spatially restricted, explaining their low allele frequencies (Figure 4).

In highly autocorrelated environments, locally adaptive alleles may readily spread among the demes where they are beneficial because neighbouring populations experience similar selection pressures. Indeed, locally adaptive alleles tended to exhibit high allele frequencies in high autocorrelation environments (Figure 4). However, the phenotypic effect sizes of locally adaptive alleles tended to be small in high autocorrelation environments (Figure 4). Alleles with large phenotypic effects in a high autocorrelation environment may often cause individuals to overshoot the optimum in neighbouring locations with slightly different optima. In highly autocorrelated landscapes, alleles with very large effects (LA_{Rel}> 10%) sometimes had both large phenotypic effects and large allele frequencies (Figure S10).

These general patterns were observed with both strong and moderate selection (Figure 4) as well as over varying levels of gene flow (Figure S10). The patterns of allele frequency versus phenotypic effect still held when looking at absolute effects on local adaptation, though they were much less pronounced (Figure S10). There are numerous factors that may interact with the pattern of spatially varying selection to shape the genetics of local adaptation that I did not explore here. The degree of genetic redundancy in relevant traits, distribution of phenotypic effect sizes, mutation rates and patterns of dispersal can all influence the genetics of local adaptation [36–38]. Follow up studies looking at how such factors influence the genetic basis of local adaptation in differently structured environments are needed.

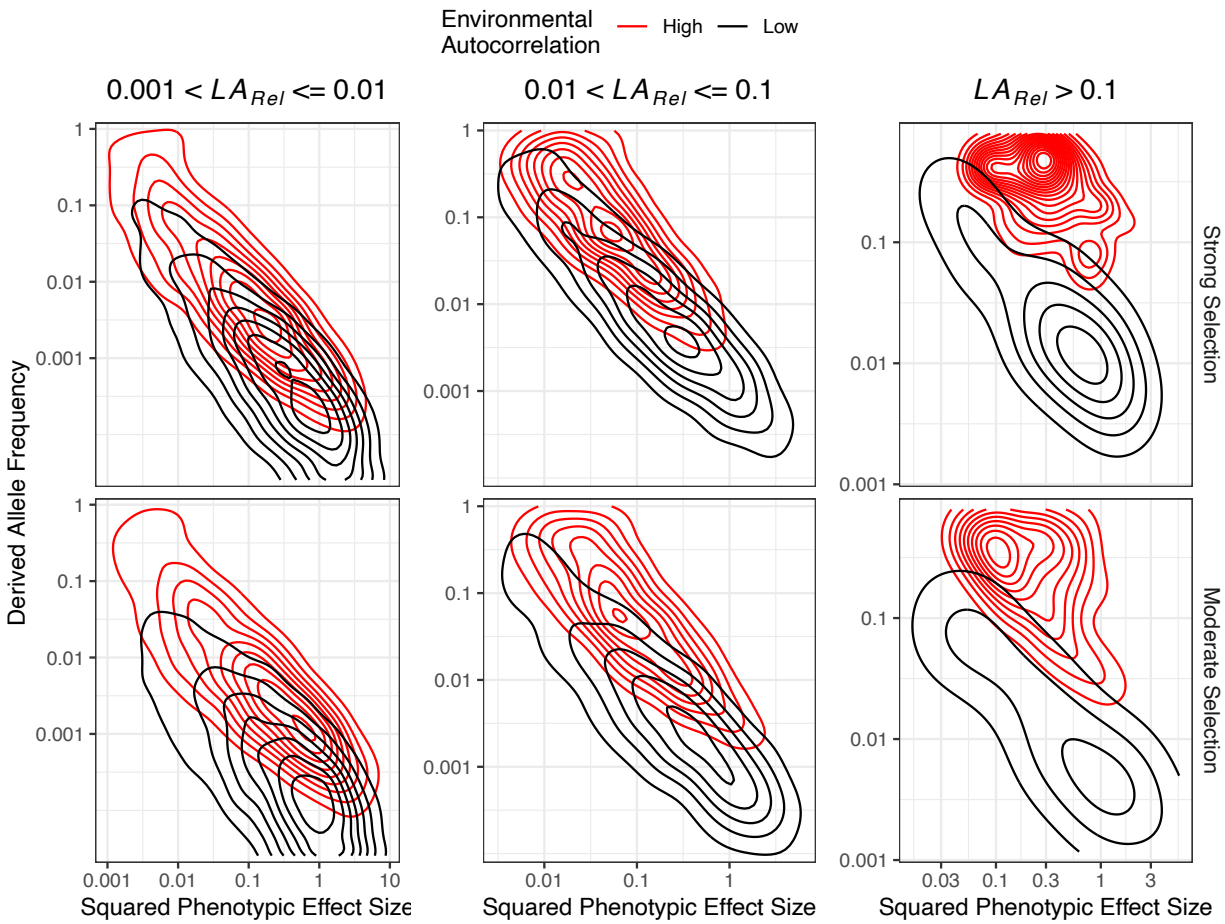


Figure 4 The properties of locally adaptive polymorphisms vary depending on the structure of the environment. The contour lines indicate regions with high densities of points. High autocorrelation refers to data from maps with the 50 highest values of Moran's I. Low autocorrelation refers to data from maps with the 50 lowest values of Moran's I. Results come from simulations with F_{ST} = 2%. Note that both horizontal and vertical axes are on log scales.

Discussion

It is important to stress that the only difference between the simulations shown in Figure 2 is the pattern of environmental variation. The average level of local adaptation in these simulations increased by a factor of roughly 5x and the coefficient of variation in local adaptation decreased by 30x when comparing the cases with the highest and lowest amount of spatial autocorrelation in the environment (Figure 2). These differences, along with the analysis of the Illingworth trial (Figure 1), should serve to emphasise the importance of spatial autocorrelation in the climate/environment as an important predictor of local adaptation in natural populations. While the effect of spatial autocorrelation on the mean local adaptation is a long-standing expectation of evolutionary biology [e.g. 20, 23], the effect of spatial autocorrelation on the variation in local adaptation has not previously been described. Patterns of spatial autocorrelation in the environment can lead to testable hypotheses about the strength of local adaptation that would be hard to formulate based on unrealistic population genetic models such as two-patch models or the island model. Unlike migration rates or the strength of selection, which are hard to quantify, spatial autocorrelation in the environment is readily measurable.

Knowing the genetic basis of local adaptation in natural systems would give us a better understanding of evolution, but may also be informative for conservation and management [39]. Many studies have used methods that associate environmental/phenotypic variation with genotypes or allele frequencies to characterise the genetic basis of local adaptation [39]. Such association methods often treat population structure as a nuisance variable and various approaches are taken to correct for it. This is done for the statistical necessity of establishing a suitable null model [40]. However, characterising the genetic basis of local adaptation when the agents of selection are co-linear or co-autocorrelated with patterns of gene flow, which could alternatively be described as "isolation by environment", is particularly challenging [41].

The structure of the environment likely impacts our ability to study the genetic basis of local adaptation. The differences shown in Figure 4 are relevant for studies that seek to characterise the genetic basis of local adaptation. To demonstrate this point, I performed a genome wide association study on phenotypes for 1,000 randomly chosen individuals from my simulations and corrected for population structure using the kinship matrix. Figure S11 shows that the $-\log_{10}(p\text{-values})$ for alleles that contribute similar levels of local adaptation tend to be smaller (i.e. there is less power) for high versus low autocorrelation environments. Much of this difference will be due to the differences in frequency and phenotypic effect size for the alleles involved in local adaptation previously described (Figure 4, S9), but some may be due to the effects of population structure correction itself. For example, rather than examining phenotypes, studies may examine the relationship between genotypes and environmental variation, i.e. genotype-environment association (GEA) analyses. Latent factor mixed models (LFMMs) are widely used to conduct GEA analyses while correcting for population structure [42, 43]. Lotterhos [44] recently found that the sensitivity of the LFMM method declined with increasing correlation between the environment and major axes of population structure. This all suggests that characterising the genetic basis of local adaptation is particularly difficult in the exact cases where local adaptation is expected to be strongest, namely when selection pressures and population structure are highly co-autocorrelated over space. More careful sampling strategies may alter the power of association methods [41, 45, 46], but such planning requires *a priori* hypotheses about locally adapted traits, the agents of selection (e.g. [47]) and/or the genes involved [48].

Rather than defining local adaptation in terms of fitness in home-like versus away-like conditions (as I have done so far), one could examine phenotypic clines (Figure S9). In their recent paper, Lotterhos [44] discussed such phenotypic clines in the context of classic theory that predicts parallel clines in allele frequency. In their model, Lotterhos [44] examined local adaptation for multiple traits each responding to variation in two different environments. Lotterhos [44] reported that a monotonic relationship between allele frequency and environment was quite rare, contradicting the expectations from classic theory, which they attributed to selection acting simultaneously on multiple environments. However, I found a similar pattern in my simulations of just a single environmental variable. Specifically, I find clear correlations between phenotypes and environments (Figure S9), but alleles with a monotonic genotype-environment association (GEA) make up only a small fraction of local adaptation even when the environment is highly autocorrelated (Figure S12). Only by aggressively smoothing GEA relationships for individual polymorphisms do monotonic GEAs making a substantial contribution to local adaptation. Overall, this suggests that the phenotypic clines can arise in the absence of monotonic genetic clines when only a single heterogeneously distributed variable induces spatially varying selection.

Understanding the scale of environmental variation that is relevant to patterns of local adaptation requires an understanding of a species' life history and physiology. That local adaptation is predicted to be stronger with increasing autocorrelation in the environment does not imply that strong local adaptation cannot arise in highly heterogeneous environments or with little spatial autocorrelation.

There are numerous examples of local adaptation to environmental heterogeneity that is not smoothly distributed in space. For example, local adaptation to heavy metal concentrations in mine tailings and serpentine soils have been thoroughly studied, and in some cases it seems as if selection is so strong that it overwhelms the effects of gene flow [49]. The patchy distributions of heavy-metal rich soils may be more or less autocorrelated from the perspective of a given species depending on its dispersal behaviour. In my simulations, I matched the granularity of dispersal with that of the environment (Figure 2A). Understanding the scale over which species disperse and how their environments vary will be important for predicting the effects of environmental structure on local adaptation in natural populations. Recent population genetic advances may make estimating dispersal much less time-intensive than previously [50, 51]. Finally, an aspect of the environment may be highly autocorrelated in space, but if the variation it exhibits does not correspond to varying selection pressures, then it will not relate to local adaptation.

In recent years, population genetic analyses have been developed to identify parts of a species range that are particularly vulnerable to climate change (i.e. genomic offset; [51–53]). Such methods analyse present-day relationships between allele frequency and the environment to predict how species will fare given predicted patterns of change. If the agents of local adaptation are highly spatially autocorrelated, neutral population structure may be partially aligned with gradients of selection, which could explain why the use of “adaptive” genetic markers and randomly chosen markers seem to perform equally well in such offset analyses [52–54]. Analyses such as genomic offset or conservation interventions that assume homogeneity in levels of local adaptation across a species range may be negatively impacted if there is high heterogeneity in local adaptation. Violating the homogeneity assumption in offset analyses, for example, would likely introduce noise into predictions and potentially lead to spurious results.

Conclusions

Across a species’ range, local adaptation may arise due to spatially varying selection, but the outcomes will depend on the spatial pattern of that selection. Analysis of a large provenance trial in lodgepole pine indicates that spatially autocorrelated aspects of the environment have shaped local adaptation in that species. Through a simulation model I showed that varying the degree of spatial autocorrelation in environmental variation can lead to dramatically different genetic bases of local adaptation. The differences that arise may make it particularly difficult to characterise the genetic basis of local adaptation in exactly the cases that are expected to lead to the strongest local adaptation.

Materials and Methods

Analysis of data from the Illingworth trial

ClimateNA [55] was used to extract climatic data for each location in the Illingworth Trial. Across the locations in the Illingworth trial, many aspects of climatic/environmental variation are highly inter-correlated (Figure S1B). Because such inter-correlation would make it difficult to tease apart the effects of individual aspects of climatic/environmental variation on local adaptation, I conducted a principal components analysis to separate the variation onto independent axes. I only included the 8 principal components that individually explained at least 1% of variance in the data. Diameter at breast height (DBM) and tree height exhibit a strong positive correlation, I chose to restrict analyses solely to DBM.

Phenotype and survival data after 20 years for individual trees from the Illingworth trial were analyzed using mixed models. Mean normalised DBH was modelled as a normally distributed variable using the *lme4* package and survival using a generalised linear mixed model with a "logit" link function using the *glmer* package. The normalised Euclidean distance between each individual's provenance and planting site (i.e. transfer distance) was used as a predictor in the model. Provenance, planting site and planting block within sites were included as having random effects on the slope and intercept of the relationship between phenotype and transfer distance.

Simulating spatially varying selection

To explore the effects of landscape structure on the outcomes of spatially varying selection, I constructed set of maps that exhibited varying degrees of spatial autocorrelation. Maps of normally distributed environmental heterogeneity were constructed using the midpoint displacement algorithm as implemented in the *NLMpy* package [56]. I simulated a 14x14 cell grid (i.e. landscape), specifying the desired level of autocorrelation to achieve a set of 200 maps, spanning the range of Moran's I values from 0.05 to 0.95 (i.e. Moran's I varied in increments of 0.0045). Simulated maps were rejected if the mean value across the landscape was less than 0.4 or greater than 0.6. This ensured that the distribution of environmental values across the landscape were roughly equivalent across maps.

Using SLiM v4.1 [32], I modelled a 2-dimensional stepping-stone meta-populations with 196 demes (i.e. a 14x14 grid). Each deme contained 100 diploid individuals for total meta-population size of 19,600. Migration occurred between adjacent demes in the four cardinal directions (except for populations at the range edge) at rate of 0.07, 0.035 or 0.0175, leading to population-wide neutral F_{ST} of 2%, 5%, 10%, respectively (Figure 3A). Each individual's simulated genome was 10Mbp long, and recombined at constant rate of r=1 x 10^{-7}. When modelling a single trait, phenotype-affecting mutations were distributed as $N(0,1)$ and occurred at random along the sequence at a rate of $\mu = 10^{-10}$, corresponding to a mutational variance of 0.001 for the trait subject to stabilising selection. When modelling two traits, the mutation rate was the same, but effects were modelled as multivariate normal with means of 0, variances of 1 and covariances of 0. Note that mutations are pleiotropic in this model, but not deterministically so. An individual's phenotype for a given trait was the additive combination of the effects of the alleles it possessed (with semi-dominance).

Spatially varying stabilising selection was modelled using the maps of environmental heterogeneity to specify the distribution of phenotypic optima across the landscape. An individual's relative fitness W_i was calculated using the standard expression for Gaussian stabilising selection [57]:

$$W_i = \exp\left[-\left(\frac{\left(\alpha_{i,d} - \theta_d\right)^2}{2V_s}\right)\right]$$

where V_S is the variance of the Gaussian fitness function, α_i is the phenotype of the ith individual in deme d, and θ_d is the phenotypic optimum of deme d. When modelling stabilising selection in cases with two traits, an individual's fitness was calculated as follows:

$$W_i = \exp\left[-\left(\frac{\left(\alpha_{i,1,d}-\theta_{d,1}\right)^2+\left(\alpha_{i,2,d}-\theta_{d,2}\right)^2}{2V_s}\right)/2\right],$$

where $\alpha_{i,1,d}$ and $\alpha_{i,2,d}$ are the values for traits 1 and 2 for individual i in deme d, respectively, and $\theta_{d,1}$ and $\theta_{d,2}$ are the phenotypic optima for traits 1 and 2, respectively. In effect, an individual's relative fitness in this 2-trait model is the average of the marginal finesses for each trait.

To achieve an equilibrium of migration, selection and drift, meta-populations were allowed to evolve for 100,401 generations. Initially, meta-populations evolved under stabilising selection with an optimum of 0 in all demes. After 400 generations, the landscape was altered to one of the 200 maps of environmental heterogeneity (see below). At the end of the simulation, the tree-sequence of the simulated meta-population was recorded. *PySlim*, *tskit* and *msprime* packages ([58, 59]) were used to work with the output files.

Analyzing simulated data

Local adaptation was quantified for each deme using the "home-versus-away" (HA) method outlined by Blanquart [59]. Specifically, each individual's fitness was quantified in its home deme and every other possible location on the landscape. The mean local adaptation was calculated in each deme as the mean difference in fitness between home and away conditions across all individuals. For each deme d not on the edge of the simulated landscape, also I quantified local heterogeneity in the landscape as the mean sum of squares between a focal deme's environment and that of the four adjacent demes (in the cardinal directions). For each deme, across the n polymorphisms that affected the phenotype additive genetic variance for the trait was calculated as $V_{A,d} = \sum_{i=1}^{n} p_{i,d}(1 - p_{i,d})\gamma_i$, where $p_{i,d}$ is the allele frequency in for SNP i in deme d and γ_i is the phenotypic effect of SNP i.

The contribution of individual SNPs to local adaptation was quantified as follows. For each polymorphism that affected the trait under selection, I shuffled its pattern of allele frequency over the landscape. I then recomputed all phenotypes and quantified local adaptation as $\overline{LA_\iota}$. I quantified the relative contribution of a polymorphism to local adaptation as:

$$LA_{Rel,i} = (1 - \overline{LA_\iota}/\overline{LA}).$$

Note that LA_{Rel} is not strictly a proportion, as epistasis for fitness that arises in models of stabilising selection means that the $\sum_{i=1}^{n} LA_{Rel} \neq 1$ for the n SNPs that affect phenotypes. Local maladaptation was the additive combination of all polymorphisms with negative LA_{Rel}.

I combined results across the 50 simulations with the lowest levels of spatial autocorrelation (weak autocorrelation) and the 50 simulations with the highest autocorrelation (high autocorrelation) and examined the relationship between allele frequency and phenotypic effect sizes for the alleles underlying local adaptation.

For each simulation, randomly sampled 1,000 individuals from the landscape and I recording their trait values as well as neutral and phenotype affecting polymorphisms. I applied LD-pruning (with a threshold of $r^2 < 0.2$) to the neutral SNPs and used these data to infer the kinship matrix using PLINK (v2; [60]). I performed genome-wide association studies on results from individual simulations with GEMMA (v0.98.5; [61]), using the inferred kinship matrix as a random effect to model population structure.

Data accessibility

All the code used to perform, analyse and plot the results of simulations is available at https://github.com/TBooker/LocalAdaptationArchitechture. R scripts to analyse and plot the results of the Illingworth trial data are available at https://github.com/TBooker/LocalAdaptationArchitechture, but the raw data files were used by permission of the BC Ministry of Forestry.

Acknowledgements

Thanks to the Aitken lab, Katie Lotterhos, Ailene Macpherson, Anna Bazzicalupo, Greg O'Neil and Wouter van der Wijl for lots of discussions. Thanks to Mike Whitlock, Sam Yeaman, Julia Kreiner, Asad Hasan and Tyler Kent for comments on the manuscript. The data from the Illingworth trial was compliments of Nick Ukrainetz, the lodgepole pine breeder with the Forest Improvement and Research Management branch of the BC Ministry of Forests. TRB is supported by a Bioinformatics Postdoctoral Fellowship from the Biodiversity Research Centre at the University of British Columbia.

Literature cited

1. Kirkpatrick M, Barton NH (1997) Evolution of a species' range. Am Nat 150:1–23

2. Rellstab C, Dauphin B, Exposito-Alonso M (2021) Prospects and limitations of genomic offset in conservation management. Evol Appl 14:1202–1212

3. Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7:1225–1241

4. Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. Ecol Lett 16:1195–1205

5. Angert AL, Bontrager MG, Ågren J (2020) What Do We Really Know About Adaptation at Range Edges? Annu Rev Ecol Evol Syst 51:341–361

6. Aitken SN, Whitlock MC (2013) Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu Rev Ecol Evol Syst 44:367–388

7. Hartmann H, Bastos A, Das AJ, et al (2022) Climate Change Risks to Global Forest Health: Emergence of Unexpected Events of Elevated Tree Mortality Worldwide. Annu Rev Plant Biol 73:673–702

8. Wadgymar SM, DeMarche ML, Josephs EB, et al (2022) Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence. Annu Rev Ecol Evol Syst 53:87–111

9. Exposito-Alonso M (2023) Understanding local plant extinctions before it is too late: bridging evolutionary genomics with global ecology. New Phytol 237:2005–2011

10. Siepielski AM, Gotanda KM, Morrissey MB, et al (2013) The spatial patterns of directional phenotypic selection. Ecol Lett 16:1382–1392

11. Lasky JR, Josephs EB, Morris GP (2023) Genotype-environment associations to reveal the molecular basis of environmental adaptation. Plant Cell 35:125–138

12. Yeaman S (2015) Local Adaptation by Alleles of Small Effect. Am Nat 186 Suppl 1:S74-89

13. Falconer DS, MacKay TFC (1995) Introduction to Quantitative Genetics, 4th ed. Longman, London, England

14. Wright S (1931) Evolution in Mendelian populations. Genetics 16:290–290

15. Slatkin M (1978) Spatial patterns in the distributions of polygenic characters. J Theor Biol 70:213–228

16. Felsenstein J (1976) The theoretical population genetics of variable selection and migration. Annu Rev Genet 10:253–280

17. Lenormand T (2002) Gene flow and the limits to natural selection. Trends Ecol Evol 17:183–189

18. Jenkins DG, Carey M, Czerniewska J, et al (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? Ecography 33:315–320

19. Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264

20. Levins R (1966) THE STRATEGY OF MODEL BUILDING IN POPULATION BIOLOGY. Am Sci 54:421–431

21. Schiffers K, Schurr FM, Travis JMJ, et al (2014) Landscape structure and genetic architecture jointly impact rates of niche evolution. Ecography 37:1218–1229

22. Forester BR, Jones MR, Joost S, et al (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Mol Ecol 25:104–120

23. Hadfield JD (2016) The spatial scale of local adaptation in a stochastic environment. Ecol Lett 19:780–788

24. Slatkin M (1973) Gene flow and selection in a cline. Genetics 75:733–756

25. Urban MC (2011) The evolution of species interactions across natural landscapes. Ecol Lett 14:723–732

26. Leites L, Benito Garzón M (2023) Forest tree species adaptation to climate across biomes: Building on the legacy of ecological genetics to anticipate responses to climate change. Glob Chang Biol. https://doi.org/10.1111/gcb.16711

27. O'Neill GA, Gómez-Pineda E (2021) Local*was*best: sourcing tree seed for future climates. Can J For Res 51:1432–1439

28. Ying CC, Yanchuk AD (2006) The development of British Columbia's tree seed transfer guidelines: Purpose, concept, methodology, and implementation. For Ecol Manage 227:1–13

29. Illingworth K (1978) Study of lodgepole pine genotype–environment interaction in B.C. In: Proceedings International Union of Forestry Research Organizations (IUFRO) joint meeting of working parties: Douglas-fir provenances, lodgepole pine provenances, Sitka spruce provenances, and Abies provenances. Vancouver, British Columbia, Canada. pp 151–158

30. Mahony CR, MacLachlan IR, Lind BM, et al (2020) Evaluating genomic data for management of local adaptation in a changing climate: A lodgepole pine case study. Evol Appl 13:116–131

31. Wang T, Hamann A, Yanchuk A, et al (2006) Use of response functions in selecting lodgepole pine populations for future climates. Glob Chang Biol 12:2404–2416

32. Haller BC, Messer PW (2023) SLiM 4: Multispecies Eco-Evolutionary Modeling. Am Nat 201:E127–E139

33. Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nat Rev Genet 14:807–820

34. Anderson JT, Lee C-R, Rushworth CA, et al (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. Mol Ecol 22:699–708

35. Yeaman S, Jarvis A (2006) Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. Proc Biol Sci 273:1587–1593

36. Láruson ÁJ, Yeaman S, Lotterhos KE (2020) The importance of genetic redundancy in evolution. Trends Ecol Evol 35:809–822

37. Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci U S A 110:E1743-51

38. Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. Evolution 65:1897–1911

39. Grummer JA, Booker TR, Matthey-Doret R, et al (2022) The immediate costs and long-term benefits of assisted gene flow in large populations. Conserv Biol 36:e13911

40. Meirmans PG (2012) The trouble with isolation by distance. Mol Ecol 21:2839–2846

41. Wang IJ, Bradburd GS (2014) Isolation by environment. Mol Ecol 23:5649–5662

42. Caye K, Jumentier B, Lepeule J, François O (2019) LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. Mol Biol Evol 36:852–860

43. Frichot E, Schoville SD, Bouchard G, François O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol 30:1687–1699

44. Lotterhos KE (2023) The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. Proc Natl Acad Sci U S A 120:e2220313120

45. Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol 24:1031–1046

46. Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. Mol Ecol 24:3223–3231

47. Kreiner JM, Caballero A, Wright SI, Stinchcombe JR (2022) Selective ancestral sorting and de novo evolution in the agricultural invasion of Amaranthus tuberculatus. Evolution 76:70–85

48. Fournier-Level A, Korte A, Cooper MD, et al (2011) A map of local adaptation in Arabidopsis thaliana. Science 334:86–89

49. Jain SK, Bradshaw AD (1966) Evolutionary divergence among adjacent plant populations I. The evidence and its theoretical analysis. Heredity 21:407–441

50. Smith CCR, Tittes S, Ralph PL, Kern AD (2023) Dispersal inference from population genetic variation using a convolutional neural network. bioRxiv 2022.08.25.505329

51. Bradburd GS, Ralph PL (2019) Spatial Population Genetics: It's About Time. Annu Rev Ecol Evol Syst 50:427–449

52. Láruson ÁJ, Fitzpatrick MC, Keller SR, et al (2022) Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. Evol Appl 15:403–416

53. Fitzpatrick MC, Chhatre VE, Soolanayakanahally RY, Keller SR (2021) Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. Mol Ecol Resour 21:2749–2765

54. Lind BM, Candido-Ribeiro R, Singh P, et al (2023) How useful is genomic data for predicting maladaptation to future climate? bioRxiv 2023.02.10.528022

55. Wang T, Hamann A, Spittlehouse D, Carroll C (2016) Locally downscaled and spatially customizable climate data for historical and future periods for north America. PLoS One 11:e0156720

56. Etherington TR, Holland EP, O'Sullivan D (2015) NLM py: a python software package for the creation of neutral landscape models within a general numerical framework. Methods Ecol Evol 6:164–168

57. Walsh B, Lynch M (2018) Evolution and selection of quantitative traits. Oxford University Press, London, England

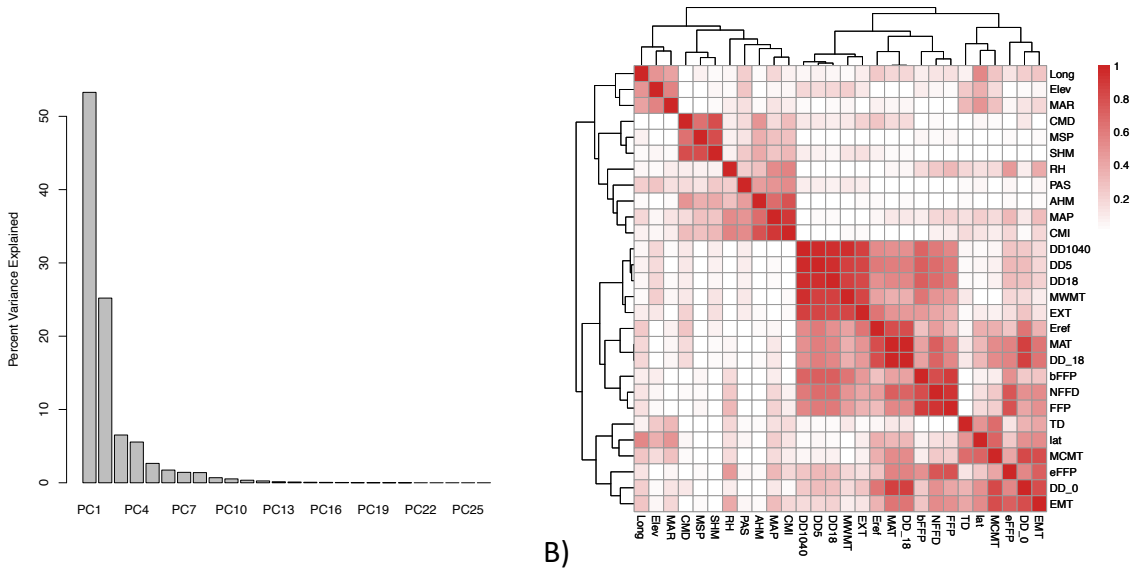
58. Baumdicker F, Bisschop G, Goldstein D, et al (2022) Efficient ancestry and mutation simulation with msprime 1.0. Genetics 220:. https://doi.org/10.1093/genetics/iyab229

59. Haller BC, Galloway J, Kelleher J, et al (2019) Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol Ecol Resour 19:552–566

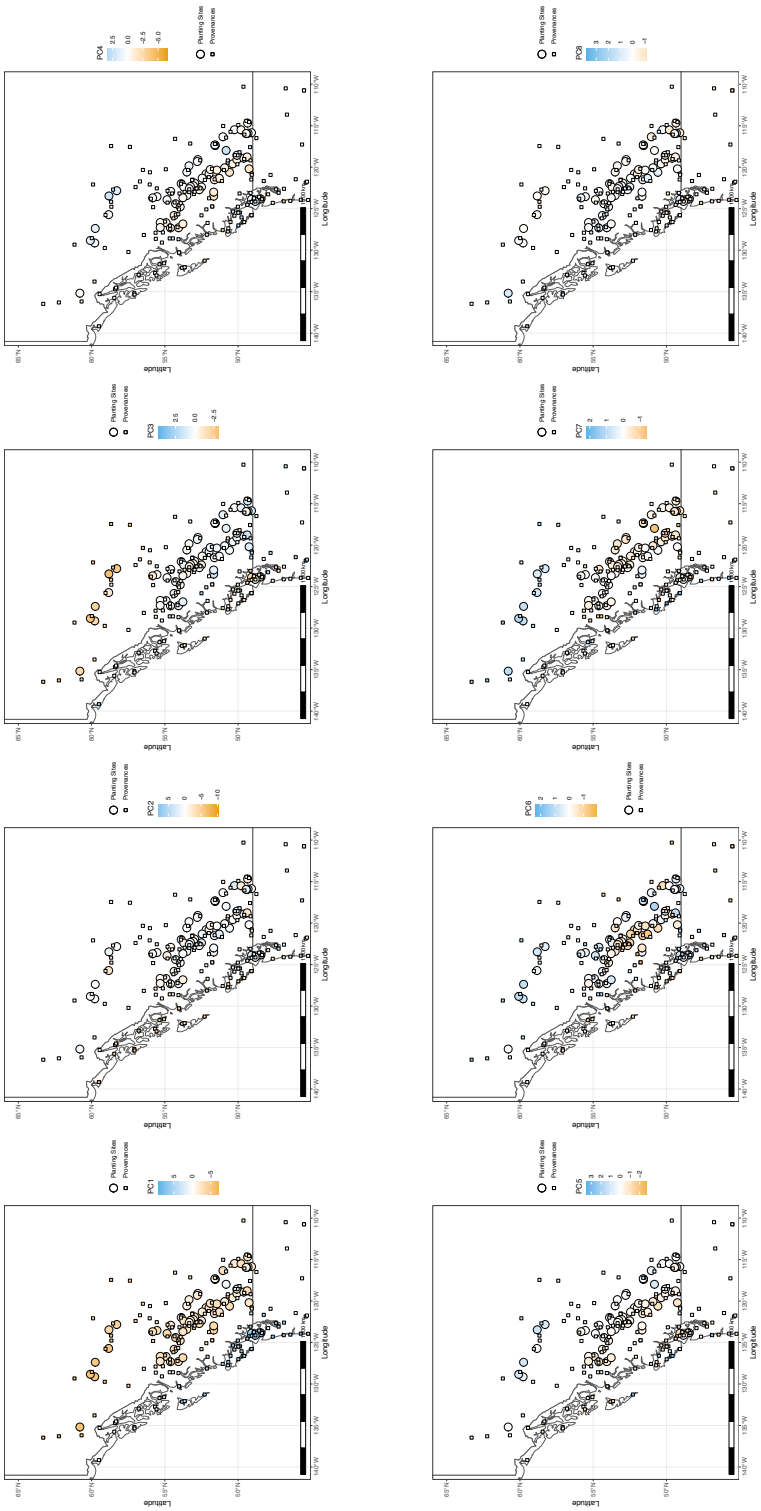
60. Chang CC, Chow CC, Tellier LC, et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7

61. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44:821–824

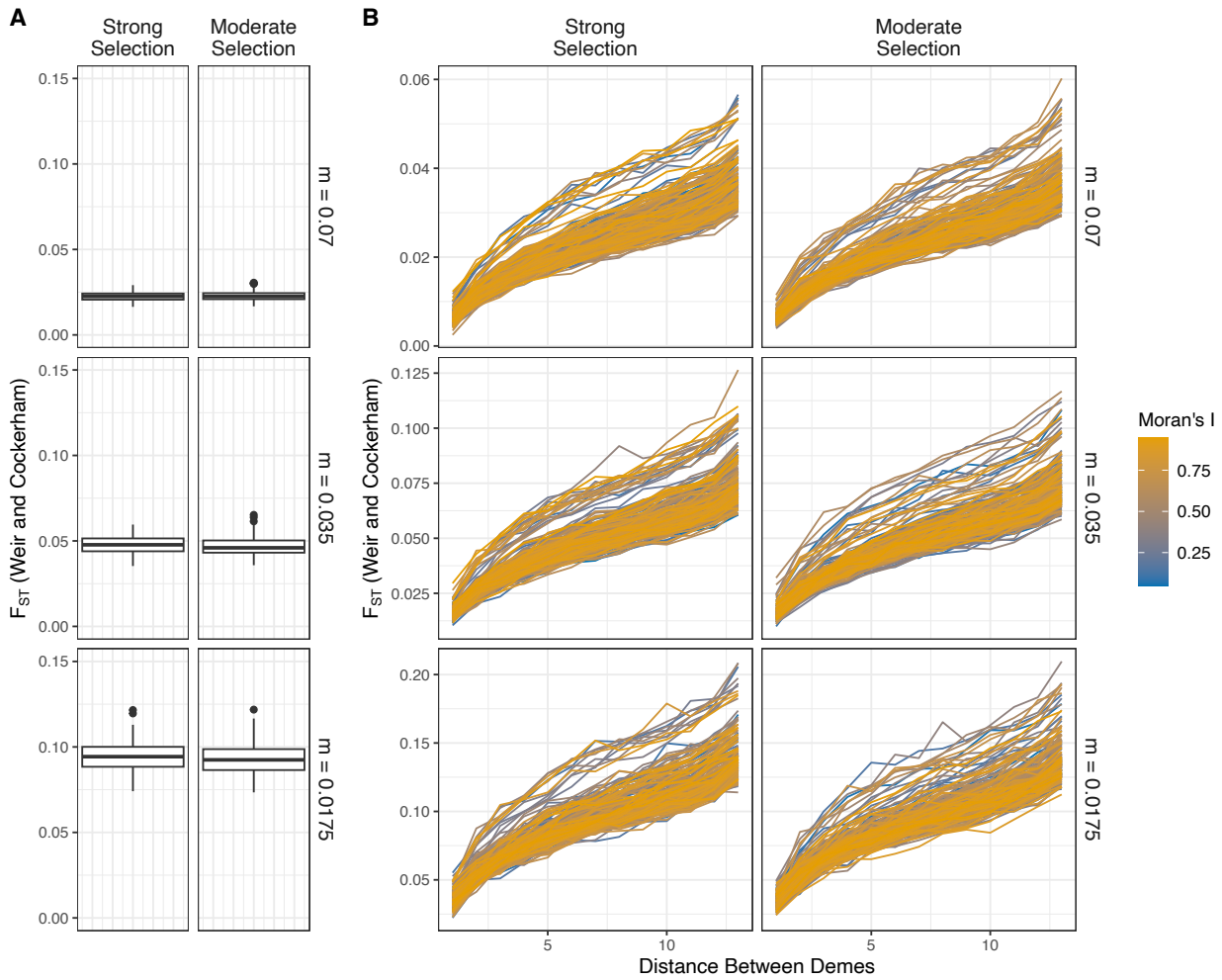
Supplementary Material



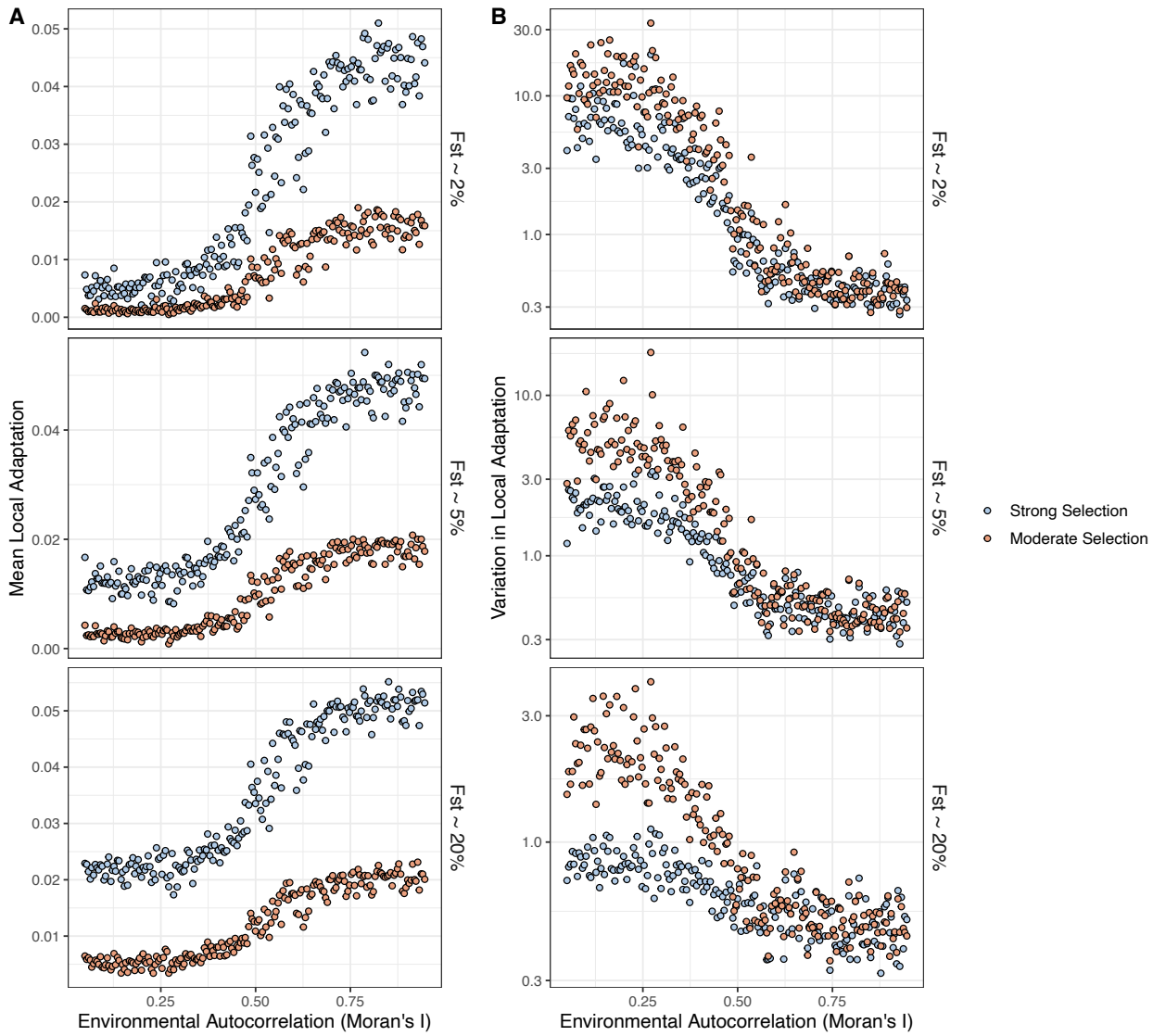
Supplementary Figure 1 A) Percent variance explained by the principal component analysis conducted on climatic/environmental variation in the Illingworth trial data. B) The correlation matrix for the 28 climatic/environmental variables for planting sites and provenances in the Illingworth trial. A key to the abbreviations for the 25 annual climatic variables from ClimateBC along can be obtained from https://climatebc.ca/Help2. Additionally, latitude (lat), longitude (Long) and elevation (Elev) are included.



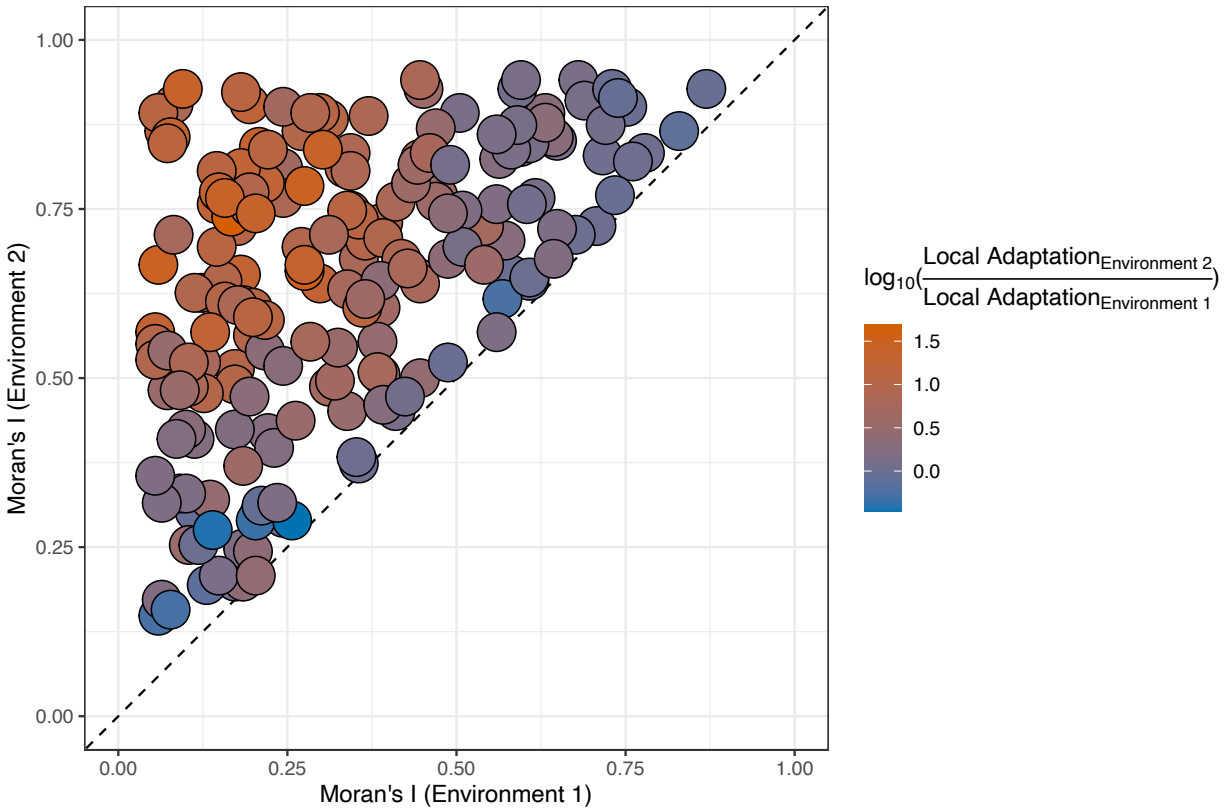
Supplementary Figure 2 The spatial pattern of loadings onto the first 8 principle components of environmental/climatic variation across provenances and planting sites in the Illingworth Trial. The first 8 principal components each explained at least 1% of variation in the data.



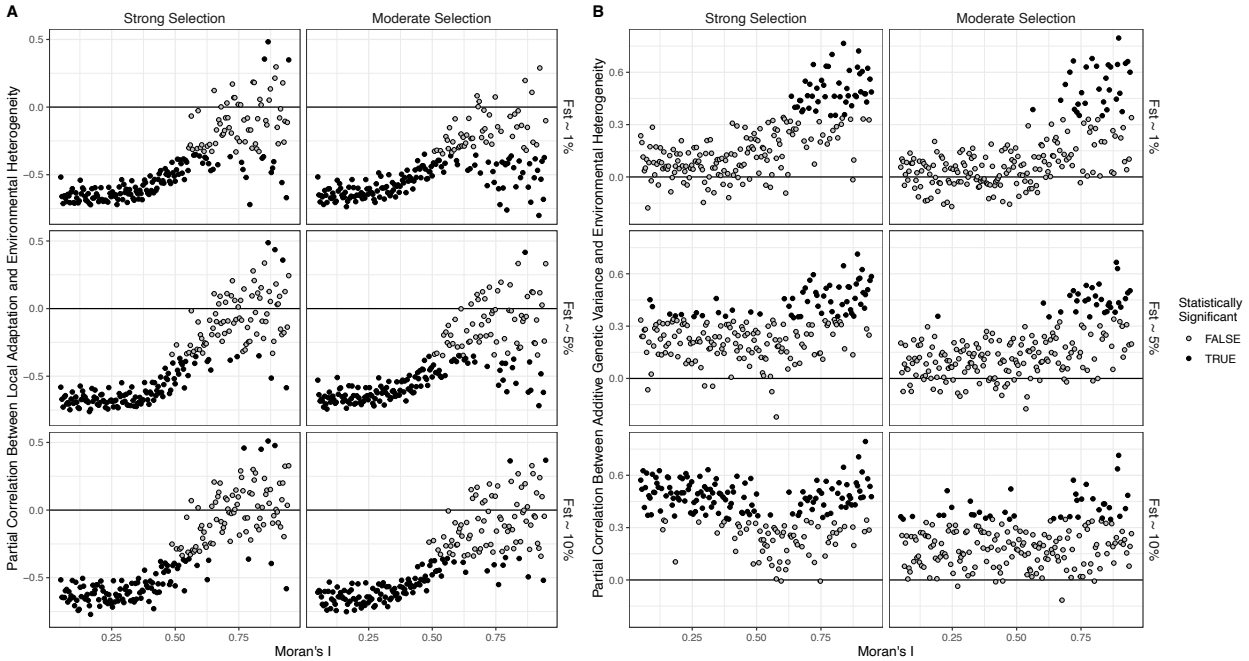
Supplementary Figure 3 Overall F_{ST} (panel A) and isolation by distance (panel B) in simulated populations. Note the varying y-axes in panel B. In the main text, F_{ST} is used to refer to the panels of individual graphs. In panel A values from 200 independent simulations were used to construct the boxplot and in panel B individual simulations are shown as lines.



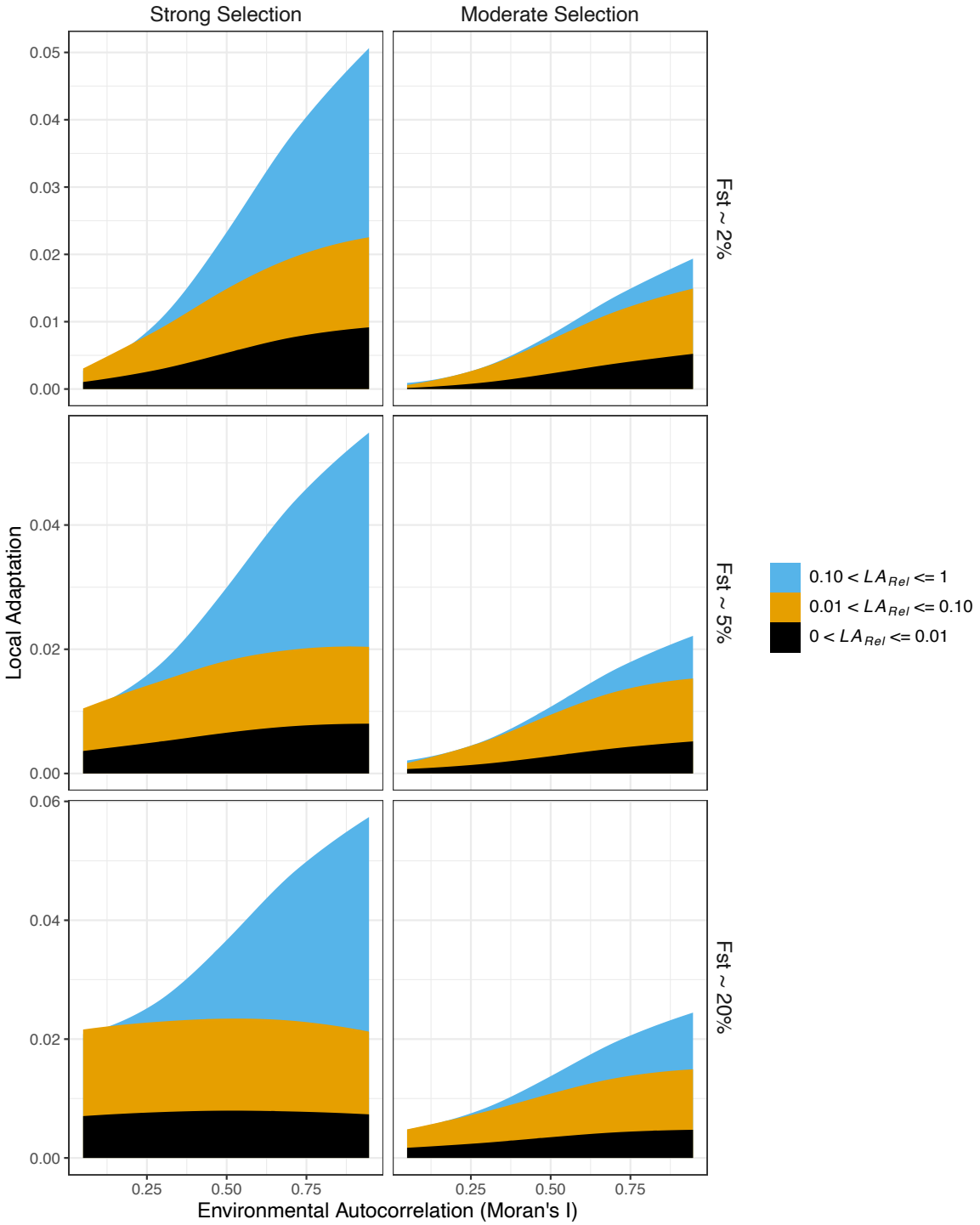
Supplementary Figure 4 The average extent of local adaptation (panel A) and coefficient of variation in local adaptation (panel B) as a function of spatial autocorrelation in the environment from simulated datasets. The upper cell of each column is included in Figure 1 of the main text.



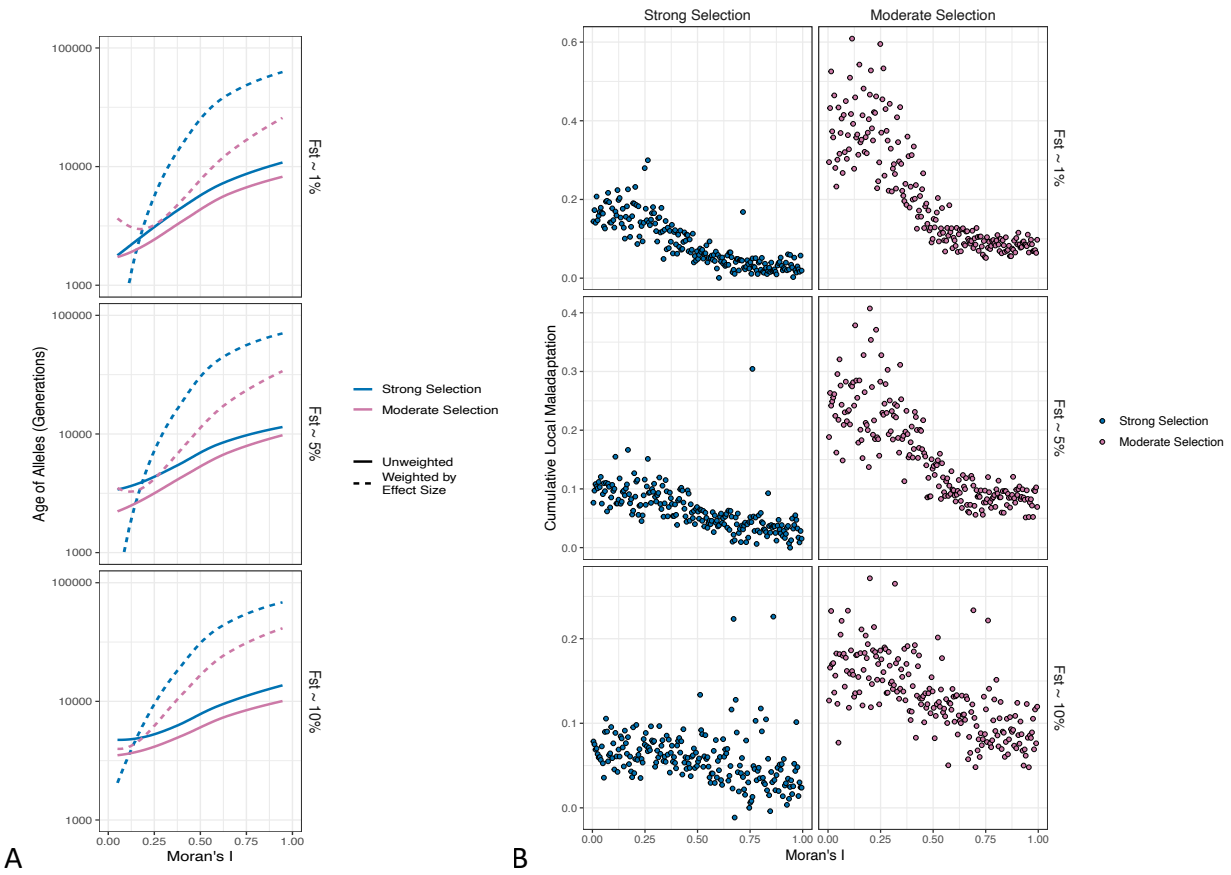
Supplementary Figure 5 Comparison of local adaptation that evolves for two traits subject to spatially varying selection. Selection on each trait was dictated by distinct maps of environmental variation/phenotypic optima. The environment that exhibited the greater degree of spatial autocorrelation (as measured by Moran's I) was designated "Environment 2". The 1:1 line is shown for reference.



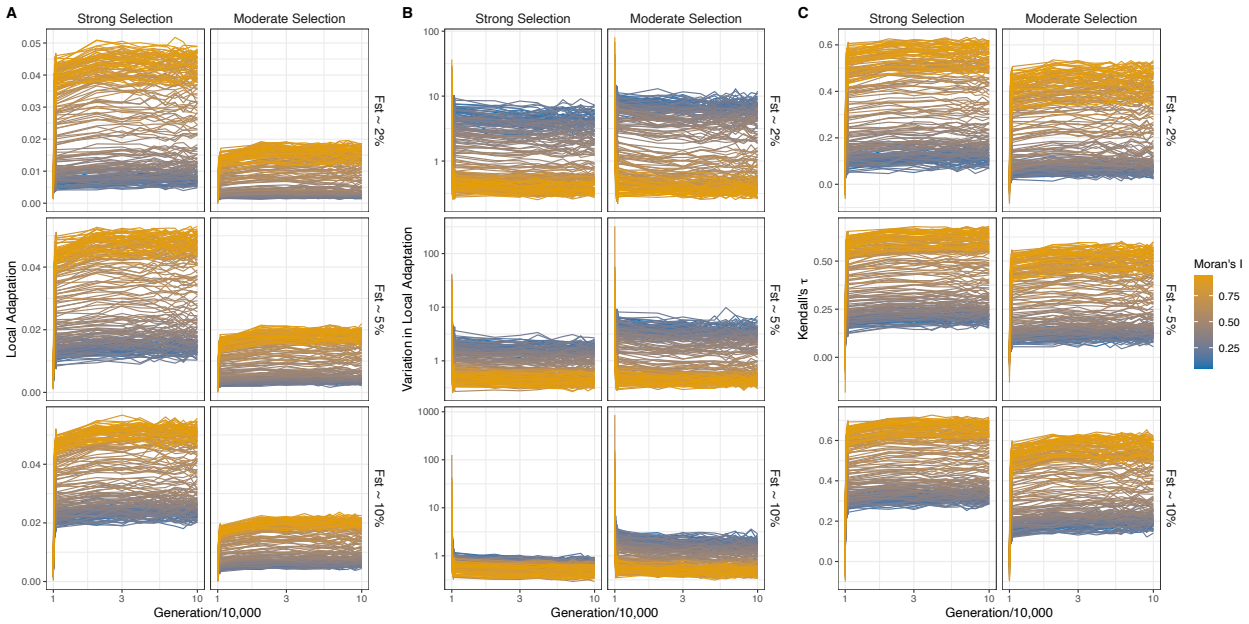
Supplementary Figure 6 The effect of spatial autocorrelation in the environment on the correlations of local environmental heterogeneity with local adaptation and additive genetic variance. A) The partial correlation between local adaptation and local environmental heterogeneity, controlling for additive genetic variance. B) The partial correlation between additive genetic variance and environmental heterogeneity, controlling for local adaptation. Statistical significance was assessed after correcting for multiple comparisons. The solid black line indicates the statistical null expectation of 0.



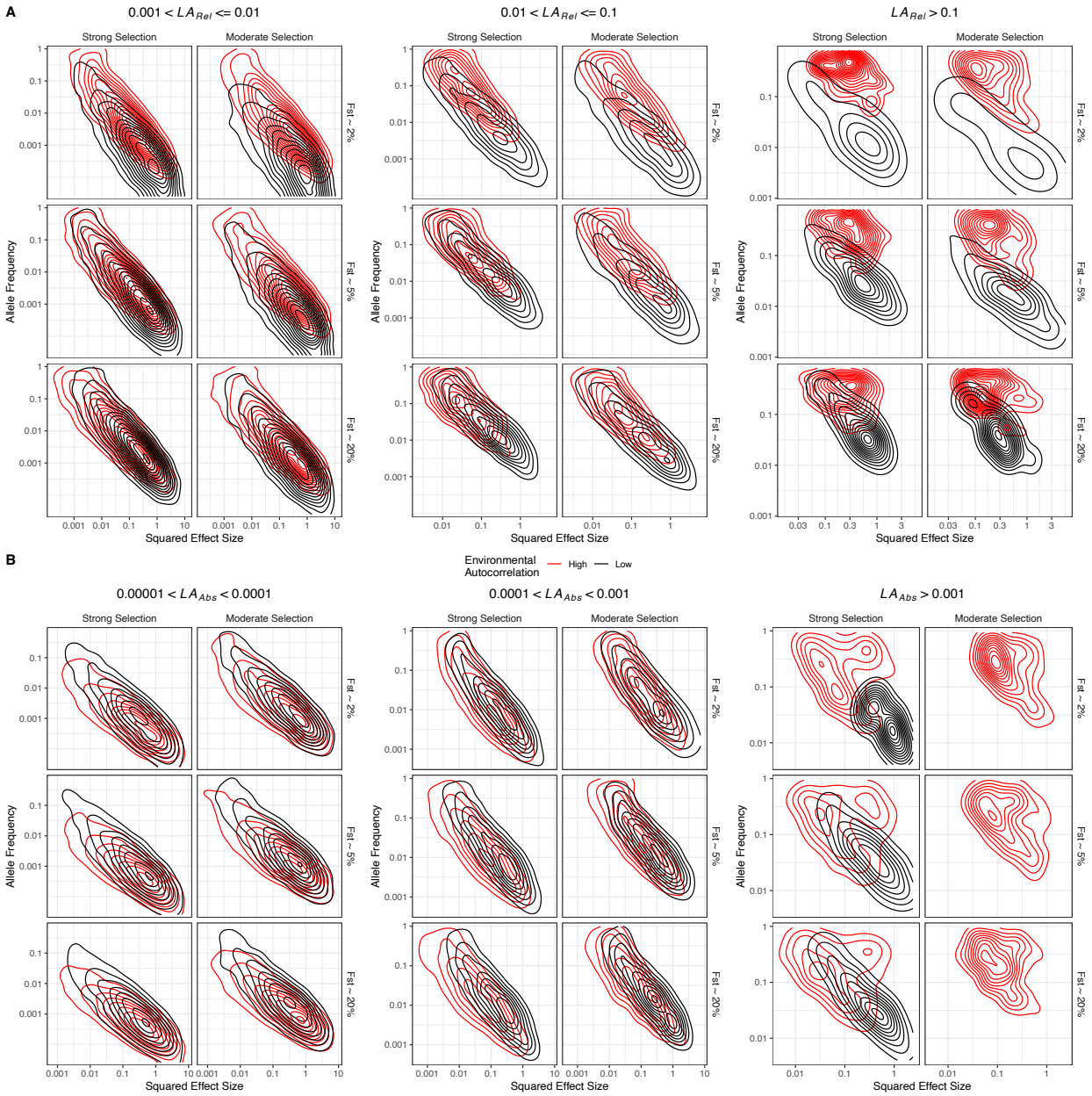
Supplementary Figure 7 The distribution of locally adaptive effects as a function of spatial autocorrelation in the environment. The area shown was calculated across 200 independent simulations and smoothed using a LOESS regression with span 1.5. The upper cells of each column are included in Figure 2 of the main text.



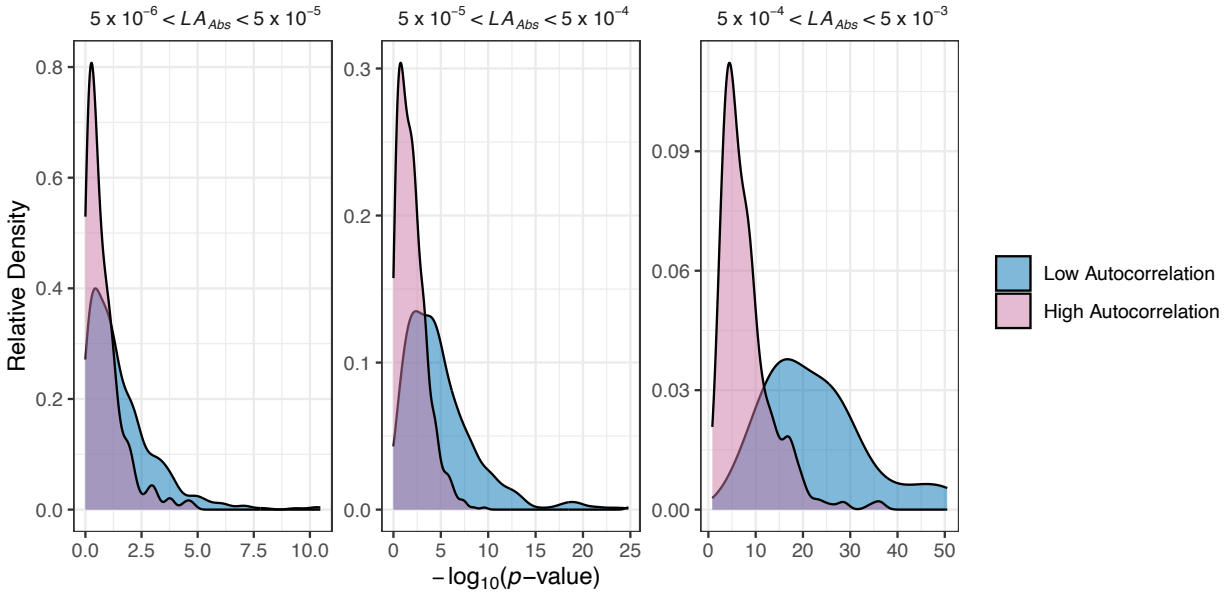
Supplementary Figure 8 A) Cumulative local maladaptation as a function of spatial autocorrelation in the environment across all parameter combinations. B) The average age of locally adaptive alleles in meta-populations subject to spatially varying selection. The lines represent LOESS regression curves calculated with span parameters of 1.5.



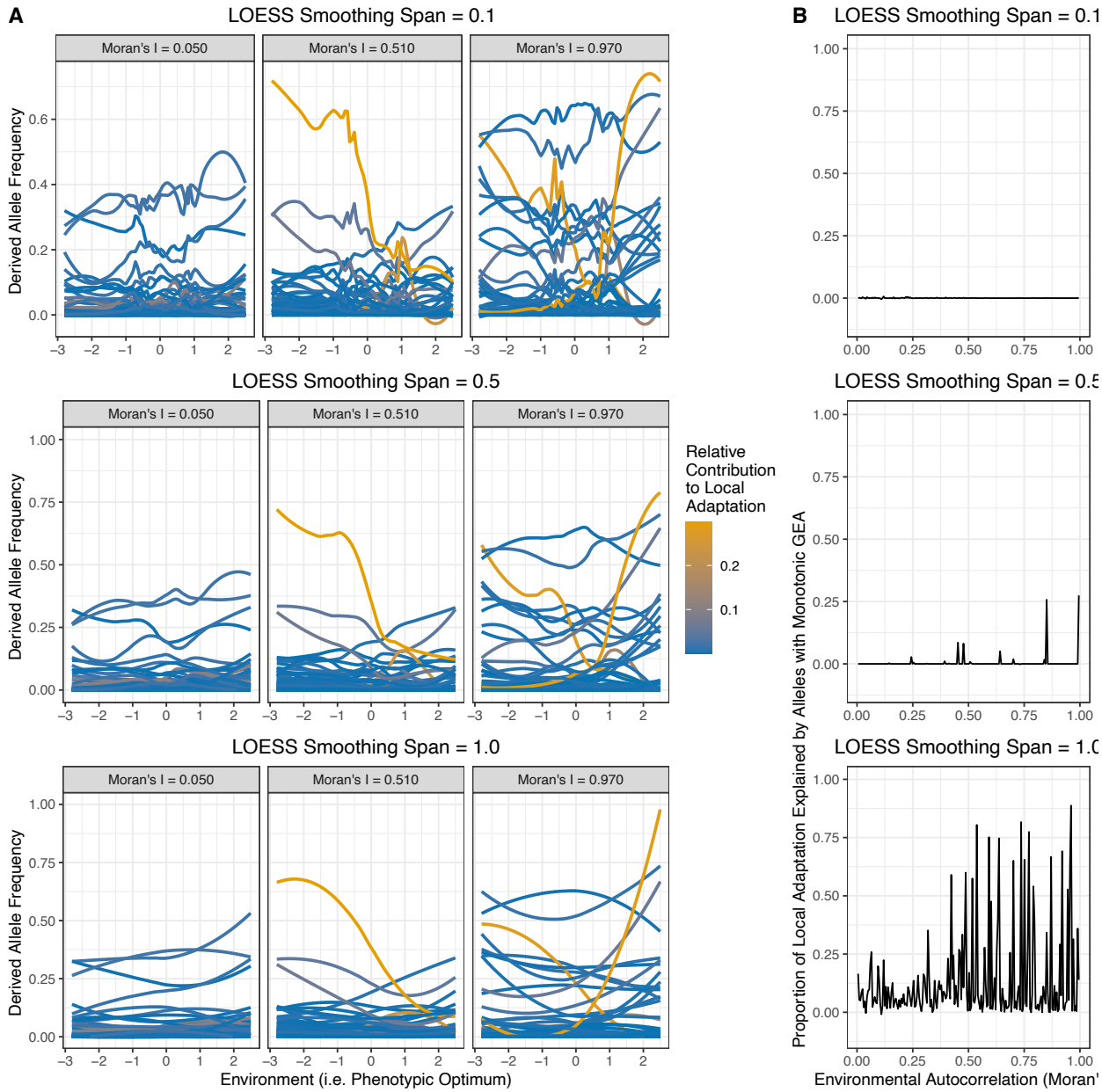
Supplementary Figure 9 Establishment of local adaptation in the simulations. Panel A) shows the average level of local adaptation across all demes. Panel B) shows the coefficient of variation in local adaptation across demes. Panel C) shows the Kendall's tau rank correlation between phenotypes and the environment.



Supplementary Figure 10 The relationship between allele frequency and the squared phenotypic effect size for polymorphisms that contribute varying degrees of local adaptation in either relative (panel A) or absolute terms (panel B).



Supplementary Figure 11 Results from a GWAS on 1,000 randomly chosen individuals from either high or low autocorrelation environments. Each panel compares the relative density of -log10(*p*-values) from a GWAS conducted on data from the 50 maps with the highest or lowest levels of spatial autocorrelation.



Supplementary Figure 12 The relationship between allele frequency and the environment for 3 representative simulations (panel A) and the proportion of local adaptation explained by alleles with a monotonic relationship between allele frequency and environmental variation (panel B). Results for simulations modelling strong selection (Vs = 15) and a high migration rate that led to mean F_{ST} ~ 2% are shown.