A window-based approach for genotype-environment association studies

Tom R. Booker[1,2,$], **Samuel Yeaman**[3] **and Michael C. Whitlock**[1,2]

[1]Department of Zoology, University of British Columbia, Vancouver, Canada, [2]Biodiversity Research Centre, University of British Columbia, Vancouver, Canada,
[3]Department of Biological Sciences, University of Calgary, Calgary, Canada

---

1   **ABSTRACT** *I'm using the GENETICS template because it looks nice, but I would like to submit this there*

2   Genotype environment association (GEA) studies have the potential to elucidate the genetic basis of local adaptation in natural
3   populations. Specifically, GEA approaches look for a correlation between allele frequencies and putatively selective features of
4   the environment. Genetic markers with extreme evidence of correlation with the environment are presumed to be tagging
5   the location of alleles that contribute to local adaptation. In this study, we propose a new method for GEA studies called the
6   weighted-Z analysis (WZA) that combines information from closely linked sites into analysis windows in a way that was inspired
7   by methods for calculating F_{ST}. We analyse simulations modelling local adaptation to heterogeneous environments using GEA
8   methods that control for population structure or uncorrected approaches. In the majority of cases we tested, the WZA either
9   performs similarly to single-SNP based approaches or outperforms them. We found that correcting for population structure
10   using an estimated population covariance matrix resulted in a loss of statistical power, similar to results from a previous study.
11   The WZA outperforms individual SNP approaches when the measured environment is not perfectly correlated with the true
12   selection pressure, or the a small number of individuals or demes was sampled. Finally, we apply the WZA to empirical data
13   from lodgepole pine.

14

15   **KEYWORDS** Local adaptation, population genetics, landscape genomics

Introduction

With an understanding of the genes or genomic regions that contribute to local adaptation, we may develop an understanding of the genetic architecture of adaptation and inform models of the limits and constraints of evolvability (e.g. Yeaman *et al.* 2018). In addition, knowledge of the loci or alleles involved in local adaptation may inform conservation management programs for buffering against the consequences of anthropogenic climate change (Aitken and Whitlock 2013).

Alleles may vary in frequency across a species' range in response to changing environmental conditions that give rise to spatially varying selection pressures (Haldane 1948). For that reason, genetic variants that exhibit strong correlations with putatively selective features of the environment are often interpreted as a signature of local adaptation (Coop *et al.* 2010). Genotype-environment association (GEA) studies examine such correlations. Allele frequencies for many genetic markers, typically single nucleotide polymorphisms (hereafter SNPs), are estimated in numerous locations across a species' range. Correlations between allele frequency and environmental variables are calculated then contrasted for sites across the genome. It is assumed in GEA studies that current heterogeneity in the environment (whether biotic or abiotic) reflects the history of selection.

Numerous approaches for performing GEA analyses have been proposed. If individuals are sequenced,

GEA can be performed by regressing environments on genotypes as a form of genome-wide association study, for example using the *GEMMA* package (Zhou *et al.* 2013). However, to estimate SNP effects with reasonable statistical power, many individuals may need to be sequenced. A cost-effective alternative is pooled sequencing (hereafter pooled-seq), where allele frequencies for populations of individuals are estimated rather than individual genotypes (Schlötterer *et al.* 2014). In this study, we focus on analyses that can be performed on pooled-seq datasets given the wide adoption of that protocol in the GEA literature.

The most straightforward way to perform a GEA analysis is to simply examine the correlation between allele frequencies and environmental variables measured in multiple populations, for example using rank correlations such as Spearman's ρ or Kendall's τ. This simple approach may commonly lead to false positives, however, if there is environmental variation across the focal species's range that is correlated with patterns of gene flow or historical selection (Meirmans 2012; Novembre and Di Rienzo 2009). For example, consider a hypothetical species inhabiting a large latitudinal range. If the hypothetical species had restricted migration and exhibited isolation-by-distance, neutral alleles may be correlated with any environmental variable that happened to correlate with latitude simply due to population structure.

Several approaches have been proposed to identify genotype-environment correlations above and beyond what is expected given underlying population structure. The commonly used BayPass (Gautier 2015) package, which is an extension of BayEnv by Coop *et al.* (2010), estimates correlations between alleles and environmental variables in a two-step process. First, a population covariance matrix (Ω) is estimated from SNP data. Second, correlations between the frequencies of individual SNPs and environmental variation is estimated treating Ω in a manner similar to a random effect in a generalised mixed model. Latent-factor mixed models (LFMM) are an alternative approach for performing GEA (Frichot *et al.* 2013), and are implemented in the LEA package (Frichot and François 2015). In LFMMs, an allele frequency matrix is regressed on environments with population structure modelled as a series of 'latent factors', which are major axes of variation in the allele frequency matrix. For clarification, if one ran an LFMM on SNP data without environmental data and set the number of latent factors equal to the number of loci in the allele frequency matrix, the result would be similar to calculating the loadings in a principle components analysis. Finally, redundancy analysis (RDA) has been proposed as a way to perform GEA studies incorporating multiple environmental variables (Forester *et al.* 2016, 2018). In RDA, a multiple regression of allele frequencies at each SNP on environmental variables is performed. This leads to a matrix of fitted values with the same dimensions as the allele frequency matrix. A principle components analysis is then performed on the matrix of fitted values and the loadings on the first few axes of variation are used as test statistics (Legendre and Legendre 2012). Note that RDA does not explicitly control for population structure. A recent study used simulated data to compare the performance of the GEA methods we have outlined (Lotterhos 2019). Overall, Lotterhos (2019) found that a comparatively simple analysis calculating Spearman's ρ between allele frequency and environment performed similarly to RDA, as measured by precision-recall curves, and both uncorrected approaches outperformed LFMMs and *BayPass*. In all of the GEA analysis methods outlined above, each SNP is treated independently, though the "AUX" mode in *BayPass* can account for linkage disequilibrium by modelling the spatial dependency among SNPs. However, the "AUX" mode is recommended for zeroing in on particular regions rather than genome-wide scans (Gautier 2015).

Closely linked SNPs are not independently inherited and may have highly correlated evolutionary histories. Individual SNPs may provide very noisy estimates of summary statistics, but by combining information across closely linked sites noise can be reduced. It is for that reason that genome scan studies often aggregate data across adjacent markers into analysis windows based on a fixed physical or genetic distance or number of SNPs (Hoban *et al.* 2016). In the case of F_{ST}, the standard measure of population differentiation, there are numerous methods for combining estimates across sites (*See* Bhatia *et al.* 2013). In Weir and Cockerham's (1984) method, for example, estimates of F_{ST} for individual loci are combined into a single value with each marker's contribution weighted by its allele frequency.

In the context of GEA studies, each marker or SNP provides a test of whether a particular genealogy is correlated with environmental heterogeneity. In the extreme case of a non-recombining region, each SNP would be present on the exact same genealogy and thus provide a test of identical hypotheses. However, the SNPs that are the most informative in this context are those at high frequencies as they contain the most information about the shape of the underlying genealogy. For recombining portions of the genome, however, linked sites will not have the exact same genealogy, but genealogies may be highly correlated. Similar to combining estimates of F_{ST} to decrease statistical noise, combining GEA tests performed on individual markers into analysis windows may increase the power of GEA studies to identify genomic regions that contribute to local adaptation.

In this study, we propose a general method for combining the results of single SNP GEA scores into analysis

windows that we call the weighted-Z analysis (WZA). Using simulations, we generate datasets modelling a pooled-sequencing experiment where estimates of allele frequency are obtained for numerous populations across a species' range. We analyse our simulated data and compare the performance of the WZA to current GEA approaches. We show that the WZA is particularly useful when GEA analyses on individual SNPs are noisy. We re-analyse previously published lodgepole pine data using the WZA and find several candidate loci that were not identified using the methods of the original study.

The Weighted-Z Analysis

In this study, we propose the Weighted-Z Analysis (hereafter, the WZA) for combining information across linked sites in the context of GEA studies. The WZA uses the weighted-Z test, a test from the meta-analysis literature that combines p-values from multiple independent hypothesis tests into a single score. In the weighted-Z test, each of the n independent tests is given a weight that is proportional to the inverse of its error variance (Whitlock 2005). Inspired by Weir and Cockerham's (1984) method for combining estimates of F_{ST} across sites, in the WZA we use a marker's allele frequency to determine weights when performing the weighted-Z test on GEA data. At a given polymorphic site, we denote the average frequency of the minor allele across populations as $\overline{p}$ ($\overline{q}$ corresponds to the major allele). The sampling variance of frequency for the minor allele from one generation to the next is $\frac{\overline{pq}}{2N}$, where N is the population size (Charlesworth and Charlesworth 2010). In the WZA, we $\overline{pq}$ in place of the error variance in the weighted-Z test.

We combine information from the markers (typically SNPs) present in a focal genomic region into a single weighted-Z score (Z_W). The genomic region in question could be a gene or genomic analysis window. We calculate $Z_{W,k}$ for genomic region k, which contains n SNPs as

$$Z_{W,k} = \frac{\sum\limits_{i=1}^{n} \overline{p_i q_i} z_i}{\sqrt{\sum\limits_{i=1}^{n} (\overline{p_i q_i})^2}}, \quad (1)$$

where z_i is the standard normal deviate calculated from the one-sided p-value for SNP i.

A feature of the WZA is that many statistics can potentially be used as input as long as individual SNPs provide a measure for the strength of evidence against a null hypothesis.

Under the null hypothesis that there is no correlation between allele frequency and environment, the expected distribution of correlation coefficients in a GEA would be normal about 0, with a uniform distribution of p-values. However, as will often be the case in nature, there may be an underlying correlation between population structure and environmental heterogeneity that will cause these genome-wide distributions to deviate from this null expectation. The average effect of population structure on individual SNP scores can be incorporated into an analysis by converting an individual SNP's correlation coefficient or parametric p-value into empirical p-values based on the genome-wide distribution. With this procedure, we lose some ability to ascribe significance to particular regions, but aggregating the empirical p-values using the WZA may identify genomic regions with a pattern of GEA statistics that deviate from the average genome-wide. In empirical studies, it may be preferable to use the parametric p-values rather than the correlation coefficients themselves as there may be varying power to calculate correlations across the genome, e.g. due to varying levels of missing data.

When we apply the WZA in this study, we used two different statistics as input. The first was empirical p-values calculated from the genome-wide distribution of parametric p-values from Kendall's τ. The second is empirical p-values calculated from the genome-wide distribution of Bayes factors as obtained using the *BayPass* program (see below).

Materials and Methods

Simulating local adaptation

We performed forward-in-time population genetic simulations of local adaptation to test the WZA and compare its performance to the commonly used *BayPass* package. GEA studies are often performed on large spatially extended populations that may be comprised of hundreds of thousands of individuals. However, it is computationally infeasible to model selection and linkage in long chromosomal segments (>1Mbp) for such large populations. For that reason, we simulated relatively small populations containing 19,600 diploid individuals in total and scaled population genetic parameters so as to model a large population. We based our choice of population genetic parameters on estimates for conifer species and in the Appendix we give a breakdown and justification of the parameters we simulated. A representative set of parameters is given in Table **??**. All simulations were performed in SLiM v3.4 **?**.

The simulated organism had a genome containing 1,000 genes uniformly distributed onto 5 chromosomes. We simulated a chromossome structure in SLiM by including nucleotides that recombined at $r = 0.5$ at the hypothetical chromosome boundaries. Each chromosome contained 200 10,000bp long segments. We refer to these segments as genes for brevity, although we did not model an explicit exon/intron or codon structure. Recombination within genes was uniform and occurred at a rate of $r = 10^{-7}$, giving a population-scaled recombination

rate ($4N_d r$) of 0.0004. The final base-pair in each gene recombined at a rate of 0.005, effectively modelling a stretch of 50,000bp of intergenic sequence.

Individuals inhabited a 2-dimensional stepping-stone population made up of 196 demes (i.e. a 14×14 grid). Each deme consisted of $N_d = 100$ diploid individuals. We assumed a Wright-Fisher model so demes did not fluctuate in size over time. Migration was limited to neighbouring demes in the cardinal directions and the migration rate (m) was set to $m = \frac{7.5}{2N_d}$ in each possible direction to achieve an overall F_{ST} for the metapopulation of around 0.04 (Figure S1). Additionally, we simulated metapopulations with no spatial structure (i.e. finite deme island models). In these simulations, we used formula $m = \frac{\frac{1}{F_{ST}}-1}{4N_d 196}$ (Charlesworth and Charlesworth 2010 pp319) to achieve a target F_{ST} of 0.03.

In nature, species may inhabit large spatial ranges and environmental variation may shape selection pressures. Obviously environmental variation is autocorrelated in space, so when modelling local adaptation to variable environments incorporating realistic spatial autocorrelation may be important.

We incorporated spatial heterogeneity in the environment into our simulations using a discretised map of degree days below 0 (DD0) across British Columbia (BC). We generated the discretised DD0 map as follows. We downloaded the map of DD0 for BC from ClimateBC (http://climatebc.ca/; Wang *et al.* (2016)). From the DD0 map (Figure 1A). Using Dog Mountain, BC as the reference point in the South-West corner (Latitude = 48.37, Longitude = -122.97), we extracted data in a rectangular grid with edges 3.6 degrees long in terms of both latitude and longitude, an area of approximately $266 \times 400 km^2$ (Figure 1A). We divided this map into a 14×14 grid, calculated the the mean DD0 scores in each grid cell, converted them into standard normal deviates (i.e. Z-scores) and rounded up to the nearest third. We used the number of thirds of a Z-score as phenotypic optima in our simulations. We refer to this map of phenotypic optima as the *BC* map.

We used data from the *BC* map to generate two additional maps of environmental heterogeneity. First, we ordered the data from the *BC* map along one axis of the 14×14 grid and randomised optima along the non-ordered axis. We refer to this re-ordered map as the *Gradient* map (Figure 1C). Both the BC map and the gradient map have a normally distributed range of phenotypic optima. For some species, selection may only apply beyond a certain environmental threshold, leading to a non-normal distribution of phenotypic optima. To model such a situation, we set the phenotypic optimum of 20 demes in the top-right corner of the meta-population to +3 and set the optimum for all other populations to -1. We chose 20 demes as it represented 10.2% of the total population. We refer to this map as the *Truncated* map (Figure 1D).

We used either directional or stabilising selection to model local adaptation. In both cases, there were 12 genes distributed evenly across four of simulated chromosomes that potentially contributed to local adaptation. When assuming directional selection, fitness affecting mutations could only occur at a single nucleotide position in the centre of the 12 potentially selected genes. Directionally selected mutations had a spatially antagonistic effect on fitness. In deme d with phenotypic optimum θ_d, the fitness of a selected allele was calculated as $1 + s_a \theta_d$ for an individual homozygous for a locally beneficial allele. The directional selection simulations had a mutation rate of 10^{-7}.

When assuming stabilising selection, the phenotypic effects of mutations that occurred in the 12 genes had a normal distribution of phenotypic effects, with variance $\sigma_a^2 = 0.5$. Phenotype affecting mutations occurred at a rate of 10^{-10} per base-pair in the 12 genes. An individual's phenotype was calculated as the the sum of the phenotypic effects of all phenotype affecting mutations. We calculated an individual's fitness using the standard expression for Gaussian stabilising selection,

$$W[z_{i,j} = exp\left[\frac{-(f_{i,j} - \theta_d)^2}{2V_s}\right],$$

where f_i is the phenotype of the ith individual in environment j and V_s is the variance of the Gaussian fitness function (Walsh and Lynch 2018). We set $V_s = 196$ so that there was a 50% fitness difference between individuals perfectly adapted to the two extremes of the distribution of phenotypic optima. This was motivated by empirical studies of local adaptation that have demonstrated such fitness differences in numerous species (Hereford 2009; Bontrager *et al.* 2020).

We ran simulations for a total of 200,102 generations. The 19,600 individuals initially inhabited a panmictic population that evolved neutrally. After 100 generations, the panmictic population divided into the 14×14 grid and evolved strictly neutrally when modelling directional selection), or with a phenotypic optimum of 0 for all demes (when modelling stabilising selection). After 180,000 generations, we imposed the various maps of phenotypic optima and simulated for further 20,000 generations. For selected mutations, we used the "f" option for SLiM's mutation stack policy, so only the first mutational change was retained. Using the tree-sequence option in SLiM (Haller *et al.* 2019) we tracked the coalescent history of each individual in the population. At the end of each simulations, neutral mutations were added at a rate of 10^{-8} usipysling

PySLiM (https://pyslim.readthedocs.io/en/latest/). For each combination of map and mode of selection, we performed 20 replicate simulations.

Classifying genes as locally adapted

To evaluate the performance of different GEA methods, we needed to know which genes contribute to local adaptation and which do not in our simulated data. As described above, our simulations incorporated a stochastic mutation model so from replicate to replicate the genes that contributed to local adaptation varied and, in the case of stabilising selection, so did the effect size of the alleles in those genes. The condition we set to determine whether a gene was involved in local adaptation was by measuring the covariance of fitness contributed by alleles present in a particular gene in a particular environment and environmental heterogeneity. For a genetic variant to be considered locally adaptive, there needs to be a positive covariance between the fitness advantage it confers and phenotypic optima across space.

We calculated the covariance between environmental heterogeneity and the fitness or phenotypic contribution of a gene for directional and stabilising selection simulations, respectively. For each gene that could potentially contribute to local adaptation there are k causal SNPs. In deme d, there are n diploid individuals so we have $M_{g,d}$, an $n \times k$ matrix in which the genotype of each individual in d at each causal SNP is coded as 0, 1 or 2 corresponding to aa, aA and AA genotypes, respectively. Under directional selection, each gene that could contribute to phenotypic variation had at most a single causal SNP ($k = 1$) with a selective effect of $\theta_d s_a$. For each gene we calculated the covariance between $M_{g,d}\theta_d s_a$ and environmental heterogeneity. Under stabilising selection, we calculated the proportion of covariance between the phenotypic contribution of a gene in a particular deme and the environment. We use ν_g to refer to the vector of phenotypic effects for each of the k causal SNPs in gene g. The mean contribution that a gene makes to the phenotype in a particular deme is calculated as $P_{g,d} = \frac{1}{k} \sum M_{g,d} \nu_{g,d}$. We took the covariance between $PB_{g,d}$ and the vector of phenotypic optima as our measure of local adaptation for stabilising selection simulations.

Analysis of simulation data

We performed GEA on our simulated data using either Kendall's τ, a rank correlation that does not model population structure, or *BayPass*, a commonly used package for estimating correlations between allele frequencies and environments while correcting for a population covariance matrix (Gautier 2015). For all analyses, except where specified, we analysed data for set of 40 randomly selected demes and sampled 50 individuals from each to estimate allele frequencies. We sampled individuals from the same set of 40 demes for all analyses (Figure S8). From a given replicate, we calculated Kendall's τ between population-specific allele frequencies and the local environment for each SNP. Any SNP with an average minor allele frequency less than 0.05 across demes was filtered out. We ran *BayPass* (v2.1; Gautier 2015) on simulated data following the "worked example" in section 5.1.2 of the manual provided with the software.

For each gene in a given simulation replicate, we applied the WZA to either the p-values from Kendall's τ, converted into empirical p-values or Bayes factors from *BayPass*, converted into empirical p-values.

We used three different methods to summarise the GEA results for each gene, the WZA, the top-candidate method developed by Yeaman *et al.* (2016) and a single SNP-based approach.

Yeaman *et al.* (2016) proposed the top-candidate method for combining information across sites in GEA studies that they referred to as the top-candidate method. The top-candidate method attempts to identify regions of the genome involved in local adaptation under the assumption that such regions may contain multiple sites that exhibit strong correlation with environmental variables. The top-candidate method asks whether there is a significant excess of "outlier" SNPs in a region compared to what one would expect given the genome wide distribution. To apply the top-candidate method Yeaman *et al.* (2016) classified p-values below the 1st percentile genome-wide as outliers. The number of outliers in a given genomic region is tested against the average number of outliers seen in a gene. A binomial test is then used to determine whether a given gene has an excess of outliers relative to the genome-wide expectation. Analysis windows or genes with a p-value from the binomial test less than 0.0001 were taken as "top-candidates" for local adaptation. In this study, we do not use the p-value from the binomial test to categorise windows as being significant or not, we use it as a continuous index.

Single SNP-based approach

We used a single SNP-based approach to identify genes as outliers in the analysis of GEA data. For each gene, the SNPs with the most extreme test statistic (i.e. smallest p-value or largest Bayes factor) for each gene were recorded.

was recorded, when using the uncorrected approach or *BayPass*, respectively, all other SNPs were ignored. The most extreme

Across a simulated genome, the gene containing the SNP with the most extreme test statistic (e.g. the smallest p-value) was scored as a hit and other SNPs in the identified gene were subsequently ignored. The gene with the second-most extreme test statistic was then examined, then the third, and so on.

To assess the performance of the different tests, we examined the genes with the most extreme test statistics and asked whether they contributed to local adaptation. We examined the top 1, 2, 3,... 50 genes in terms of Z_W scores, $-log[10](p$-values$)$ from the top-candidate method, or the
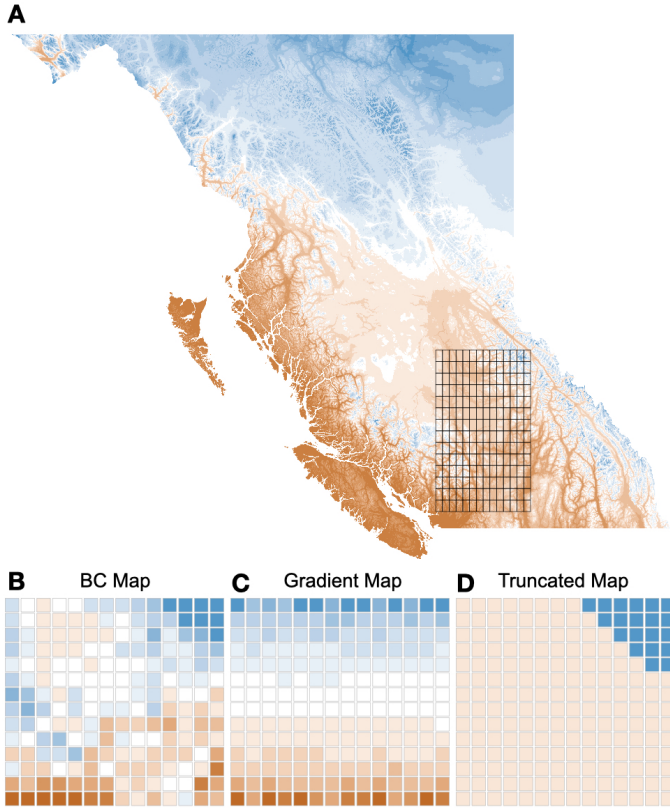
Figure 1 A) Degree days below zero across British Columbia, the overlain grid in A shows the locations we used to construct phenotypes for our simulated populations. B) A discretized map of DD0 in Southern British Columbia, we refer to the map in B as the BC map. C) A 1-dimensional gradient in phenotypic optimum, we refer to this as the 1D gradient map. D) A heterogenous distribution of phenotyic optima.

SNP-based approach scores. For simulations modelling directional selection, we assessed statistical power by examining the total number of genes that contribute to local adaptation among the test set. Genes that had had a $Cov(PB_g, env) > 0.005$ were classified as true positives. For the stabilising selection simulations we calculated the proportion of $Cov(PB, env)$ explained by a given set of genes. We calculated the false discovery rate using the total number of genes with that do not contribute to local adaptation in a given set of genes.

We examined the properties of the WZA in regions of low recombination by manipulating the tree-sequences we recorded in *SLiM*. In our simulations, genes were 9,999bp long, so to model genomic regions of low recombination rate, we extracted the coalescent trees that corresponded to the central 1,000bp or 100bp of each gene. For the 1,000bp and 100bp intervals, we added mutations at $10\times$ and $100\times$ the standard mutation rate.

Tree sequences were manipulated using the tskit package. Mutations were added to trees using msprime (REF) through the PySLiM package (version). F_{ST} and r^2 (a measure of linkage disequilibrium) were calculated using custom Python scripts that invoked the scikit-allel package (REF).

Analysis of data from Lodgepole pine

We re-analysed a previously published population genomic dataset for lodgepole pine, *Pinus contorta*, a conifer that is widely distributed across the North West of North America. Briefly, Yeaman *et al.* (2016) collected samples from 254 populations across British Columbia and Alberta, Canada and Northern Washington, USA. The lodgepole pine genome is very large (20Gbp), so Yeaman *et al.* (2016) used a sequence capture technique based on the *P. contorta* transcriptome. Allele frequencies were estimated for many markers across the captured portion of the genome by sequencing 2-4 individuals per population. Yeaman *et al.* (2016) performed GEA on each SNP using Spearman's ρ to aggregate data across sites. We downloaded the data for individual SNPs from the Dryad repository associated with Yeaman *et al.* (2016) (https://doi.org/10.5061/dryad.0t407). We converted Spearman's ρ *p*-values into empirical *p*-values and performed the WZA on the same genes analysed by Yeaman et al (2016). We also performed the top-candidate method, classifying SNPs with empirical *p-values* < 0.01 as outliers.

Data Availability

The simulation configuration files and code to perform the analysis of simulated data and generate the associated plots are available at github/TBooker/GEA/WZA. Tree-sequence files for the simulated populations are available at Dryad and all processed GEA files are available on (SomeCoolLocation).

Results

Simulations of local adaptation

We simulated meta-populations inhabiting and adapting to heterogeneous environments. We modelled the population structure in our simulations on an idealised conifer species. In conifers, strong isolation-by-distance has been reported and overall mean $F_{ST} < 0.10$ has been estimated in several species (Mimura and Aitken 2007; Mosca *et al.* 2014). In our simulations, population mean F_{ST} was 0.042 (Figure S1A). As expected under spatially restricted dispersal, our simulated stepping-stone populations exhibited strong patterns of isolation-by-distance with pairwise F_{ST} increasing with distance between demes (Figure S1A). Overall, population structure in our simulations was similar to what has been reported in natural conifer populations.

It has been reported that linkage disequilibrium (LD) decays rapidly in conifer species, with LD between pairs of SNPs decaying to background levels within 1,000bp or so (Pavy *et al.* 2012). We examined the decay of LD between pairs of SNPs located in the same gene in our simulations. LD decayed rapidly, with SNPs that were approximately 600bp apart having, on average, half the LD of immediately adjacent SNPs (Figure S1B). Linkage disequilibrium (LD) decayed to a background level of around 0.1 within demes in our simulations. Thus, patterns of LD decay in our simulations were similar to the patterns reported for conifers.

Simulated populations exhibited local adaptation to the maps of environmental heterogeneity that we implemented. The genome of our simulated species had 1,000 genes, of which 12 could potentially contribute to adaptation. When assuming directional selection, an average of 4.8, 5.15 and 4.10 genes contained genetic variants that established and contributed to local adaptation when assuming the *BC* map, the *Gradient* map and the *Truncated* map, respectively. In our simulations assuming stabilising selection, individuals' and population mean phenotypes closely matched the phenoypic optima of their local environment (Figure S2). The average numbers of genes contributing to local adaptation in individual replicates in these simulations were 7.15, 6.45 and 5.35 when assuming the *BC* map, the *Gradient* map and the *Truncated* map, respectively.

The distribution of WZA scores under neutrality

The WZA combines summary statistics from tests that are assumed to be statistically independent. Under the null hypothesis that all tests are non-significant, the distribution of the Z_W is expected to be the standard normal distribution (i.e. a mean of 0 and a standard deviation of 1). In this study, we propose using the weighted-Z test to aggregate results for GEA studies applied to
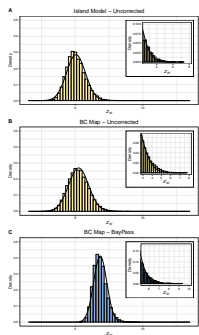
Figure 2 Density histograms of WZA scores for three cases. A) An finite island population, Z_W scores were calculated using uncorrected p-values obtained from Kendall's τ. B) The BC-map, Z_W scores were calculated using uncorrected p-values obtained from Kendall's τ. C). Inset in each panel is a histogram focussing on the upper tail of the Z_W distribution. The solid line in each panel is the normal distribution fit to the data.

genome-wide SNPs. Tightly linked SNPs obviously violate the assumption of statistical independence, so the distribution of Z_W scores from the WZA may not be normal.

To assess the statistical properties of the WZA, we first performed GEA analyses on populations structured according to an island model. While potentially unrealistic, the island model allowed us to determine the statistical properties of the WZA without spatial heterogeneity in the environment being confounded with population structure. We performed GEA using Kendall's τ on island model populations as if the data had been generated assuming the *BC* map. Applying the WZA to GEA results for populations modelled under the island model, we found that the distribution of Z_W scores slightly deviated from the null expectation; the mean Z_W was 0.00X and the standard deviation was 1.XXX. Figure 2A shows the distribution of Z_W scores was skewed with a fat right-hand tail.

We then applied the WZA to evolving neutrally stepping-stone populations and assumed the *BC* map. The deviation from the null expectation of Z_W scores was exaggerated when analysing data from stepping-stone simulations (Figure 2B-C). Figure 2B shows that the distribution of Z_W assuming Kendall's τ results has a thicker right-hand tail than expected under the null. Applying the WZA to population structure corrected GEA (BayPass) results, we also found a thicker right hand tail to the distribution of Z_W score than expected under the null.

If the distribution of Z_W scores were normal, scores for particular genomic regions could be converted into p-values. However, that the distribution was non-normal

even when assuming the island model, suggests that linkage effects cause the deviation. Thus, we did use Z_W scores to calculate parametric p-vaues, rather, we used the Z_W scores as an index for how much a particular genomic region deviates from the genome-wide pattern, similar to F_{ST} calculated in analysis windows.

Comparison of window-based and individual SNP-based GEA approaches

We performed GEA on our simulated data using either Kendall's τ, a rank correlation that does not model population structure, or *BayPass* an analysis that attempts to estimate correlations between environmental variables and allele frequencies incorporating a population covariance matrix. Our simulations incorporated three different maps of environmental heterogeneity (Figure 1) and modelled local adaptation under either directional or stabilising selection. When using either *BayPass* Bayes factors or p-values from Kendall's τ as input to the WZA, the distribution of Z_W scores for the genes that contributed to local adaptation separated from the distribution for neutrally evolving genes (Figure S4).

To assess the performance of the WZA at identifying locally adaptive genes, we examined the distribution of test statistics from the WZA, the top-candidate method proposed by (Yeaman *et al.* 2016)Z_W scores in individual simulation replicates.

Figure 3 shows how well the WZA, the top-candidate method and individual SNP approach perform when looking for the genes underlying local adaptation. There were notable differences in performance among the methods when simulations assumed directional or stabilising selection. Under directional selection, The biggest difference between *BayPass* and the uncorrected approach was seen when assuming the *Gradient* map. In that case, almost all of the true positives were detected when using the uncorrected p-value approach whereas around 50% were detected using *BayPass* (Figure 3).

From Figure 4 it is clear that none of the GEA methods that we implemented are

The performance of WZA when GEA results are statistically noisy

If either of you could come up with better sub-heading names, I'd be grateful

The WZA is a method to aggregate data across linked sites to identify loci that exhibit a GEA pattern that deviates from the genomic background. The WZA is analogous to combining estimates of F_{ST} for individual SNPs into analysis windows to identify signals in potentially noisy data (Hoban *et al.* 2016). For that reason, we postulated that the WZA would outperform the individual SNP approach when the underlying GEA
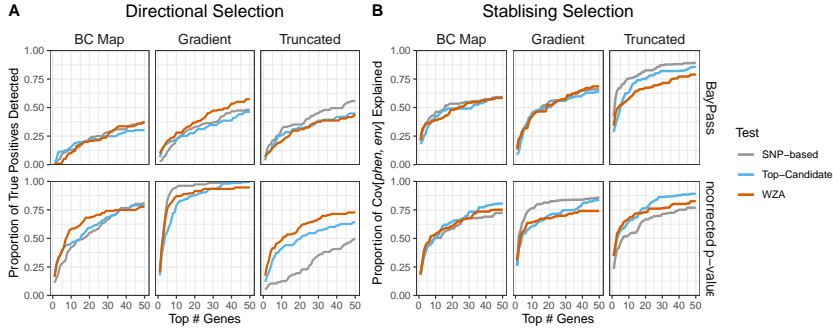
Figure 3 The proportion of true positives detected across a specified search effort when a model of local adaptation by directional selection was assumed. A) shows results obtained from simulations assuming directional selection. B) shows results obtained assuming stabilising selection. Lines represent the mean of 20 simulation replicates. Note the different quantities on the y-axes.
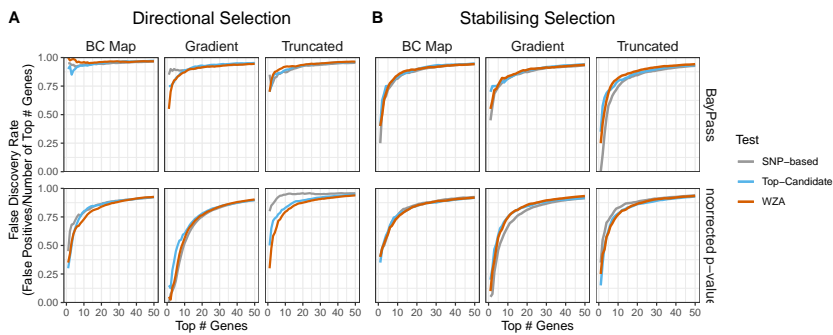

Figure 4 The proportion of true positives detected across a specified search effort when a model of local adaptation by stabilising selection was assumed.

was noisy. Here, we focus on three factors that may influence the "noisiness" of a GEA study: 1) When the environmental variable in question is only partially correlated with the true selection pressure, 2) when GEA is applied to a relatively small number of demes and 3) when allele frequency estimates are based on a small number of individuals.

Climatic or environmental variables may not be a perfect reflection of the selection pressures that species face in natural settings. It may be the case that a variable such as mean annual temperature is correlated with selection, but that that correlation is not perfect. Additionally, environmental variables likely include measurement error, contributing to an imperfect correlation with the true selection pressure. In the previous section, we conducted GEA assuming perfect knowledge of the phenotypic optima in each sampled deme. Here, we assess the performance of GEA methods when the measured environment (i.e. the 'E' in GEA) is not perfectly correlated with selection. We sampled data that were correlated with the true phenotypic optima to varying degrees and used them to perform GEA on the *BC* map simulations. We assumed the simulations modelling stabilising selection and, as above, compared the performance of the WZA, top-candidate and the SNP-based GEA methods using *BayPass* or Kendall's τ.

YOYOYOYOYO We found that the window-based GEA methods outperformed single SNP approaches when the measured environment is not perfectly correlated with the true selection pressure. As might be expected, when the correlation between the measured environment and selection was very weak (i.e. a correlation of 0.1), few true positives were present in the top 50 genes with or without population structure correlation and those present explained only a small proportion of Cov(Phen. Env.)(Figure 5). With a correlation of 0.3 between the measured environment and true selection, the WZA substantially outperformed both the top-candidate method and single-SNP based approaches using BayPass or Kendall's τ (Figure 5). With a correlation of 0.5 or 0.8 between the measured environment and true selection, there were only small differences in performance between window-based and the single-SNP approaches when using BayPass (Figure 5). When using the uncorrected approach, however, the WZA and the top-candidate method outperformed the single-SNP approach (Figure 5).

In a hypothetical world where one had perfect knowledge of allele frequency variation across a species' range, a SNP-based approach would be the best way to perform a GEA analysis, as one would be able to determine the true correlation between genetic and environmental variation. Indeed, when one has perfect knowledge of allele frequencies, a SNP-based GEA outperforms the

WZA and top-candidate methods for each of the maps we tested (Figure S7). However, such a situation is obviously unrealistic and empirical GEA studies will likely always be limited to samples from population of interest.

So far we have only modelled a situation where 40 demes were sampled across the simulated metapopulations, with 50 individuals sampled in each location. We re-sampled the simulated populations, and extracted 50 individuals from 20 or 10 demes, the configuration of samples is shown in Figure S8. We compared the performance of the WZA to the top-candidate and SNP-based methods in these smaller datasets using the Kendall's τ p-values. As expected, reducing the number of sampled demes from 40 to 20 or 10, reduced the proportion of true positives identified (Figure S7). However, we found that the WZA outperformed the top-candidate and SNP-based methods in datasets with 10 or 20 demes for each of the three maps we simulated (Figure S7).

In such cases, smaller numbers of demes and/or fewer individuals may be sampled. In either case, reducing sample sizes will reduce power of GEA, and potentially introduce statistical noise. Indeed, reducing the number of sampled demes from 40 to 20 or 10 reduced the proportion of true positives detected across all analysis methods and maps of environmental heterogeneity (Figure REFERENCE). Notably, though, the WZA outperformed the top-candidate and SNP-based method when the

We compared the performance of the WZA to the top-candidate and SNP-based methods when a fewer than 40 demes were sampled. Reducing the number of sampled demes reducedFigure XXX shows

Application of the WZA to data from lodgepole pine

We re-analysed a previously published (Yeaman *et al.* 2016) lodgepole pine (*Pinus contorta*) dataset comprising 256 populations sampled across North Western North America.Yeaman *et al.* (2016) performed GEA on their data using numerous environmental variables, but here we only analyse DD0. Yeaman *et al.* (2016) phenotyped the lodgepole pine populations they studied for response to injury from freezing temperatures. They found that the cold-injury response phenotype for a population was strongly negatively correlated with the estimated DD0, suggesting that DD0 is correlated with a true local adaptation pressure.

We downloaded the data associated with the Yeaman *et al.* (2016) study, which included tables containing allele frequency and GEA results (Spearman's ρ and the associated p-value) for genome-wide SNPs. Yeaman *et al.* (2016) applied the top-candidate method to these data. Here, we replicated that analysis but also applied the WZA, using genes as the unit of analysis. Unlike our analysis of simulation data, where we divided the genome into analysis windows of a fixed physical size, Yeaman *et al.* (2016) analysed gene sequences.
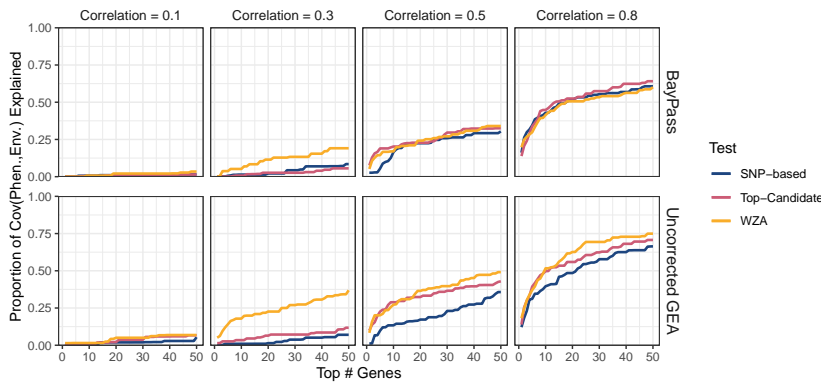
Figure 5 The proportion of true positives recovered when the measured environment is correlated with selection pressure to varying degrees. The correlation between environment and selection pressure is shown above the plots. Results are shown assuming the BC Map and the model of stabilising selection.
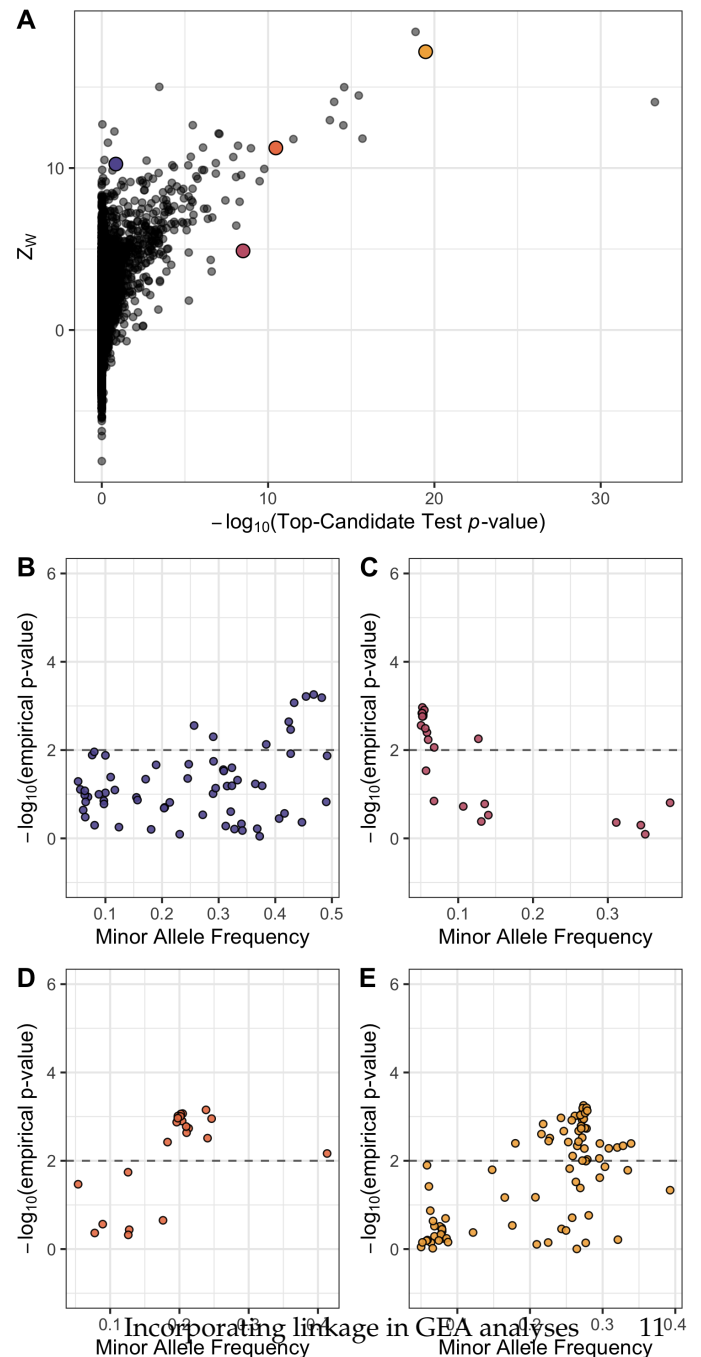

Across the lodgepole pine genome, Z_W scores had a mean of 0.013 and $\sigma = 1.67$. Additionally, the distribution of Z_W had a fat right-hand tail (Figure S10B). Figure 6A shows the relationship between WZA scores and the $-log_{10}(p-values)$ from the top-candidate method. The scores from the two tests were positively correlated (Kendall's $\tau = 0.245$, p-value $< 10^{-16}$). Figures 6C-D show the relationship between allele frequency and the empirical p-value for SNPs present in two genes that had extreme scores from both the top-candidate method and the WZA. There were several genes that had Z_W scores greater than 10 (approximately 6σ), but very modest top-candidate scores (Figure 6A). Figure 6B shows that for one such region, there were several SNPs with high mean allele frequency that have small p-values. This particular region had a high p-value from the top-candidate method. Conversely, Figure 6C shows a region that only had a $Z_W \approx 5$, but an extremely small p-value from the top-candidate method. In this case, there were numerous SNPs that passed the top-candidate significance threshold, but they were mostly at low allele frequency.

Incorporating linkage in GEA analyses 11

Figure 6 The WZA applied to GEA results Lodgepole

Discussion

In this study, we have shown that combining information across linked sites in GEA analyses is a potentially powerful way to identify genomic loci involved in local adaptation. The method we propose, the WZA, was often more powerful than looking at individual sites in isolation, particularly when working with small samples or when the environmental variation being analysed is only weakly correlated with selection. An advantage of the WZA is that it provides a continuous metric that naturally lends itself to comparison among species. For example, one could compare Z_W scores for a particular genes that were orthologous among distantly related species as a way to identify convergent evolution.

In many of the cases we examined, the performance of the WZA and the top-candidate method of Yeaman *et al.* (2016) were fairly similar (Figure 3). However, there are philosophical reasons as to why the WZA should be preferred over the top-candidate method. First, the top-candidate method assumes that there is a fraction of genetic markers analysed that are tagging causal variants (i.e. that there are true positives in the dataset). This is undesirable, however, because genuine genotype-environment correlations may be very weak and GEA may simply be an underpowered approach to identify alleles that contribute to local adaptation. If there were no detectable signal of local adaptation, ascribing significance to a fraction of the genome may lead to false positives. Secondly, the top-candidate method gives equal weight to all SNPs that have exceeded a fixed significance threshold. For example, with a significance threshold of $\alpha = 0.01$, genomic regions with only a single outlier are treated in the same way whether that outlier has a p-value of 0.009 or 10^{-10}. It seems desirable to retain information about particularly strong outliers.

It should be kept in mind that the WZA (and the top-candidate method for that matter) does not explicitly test for local adaptation and only provides an indication of whether a particular genomic region has a pattern that deviates from the genome-wide average. Indeed, numerous processes other than local adaptation may cause excessive correlation between environmental variables and allele frequencies in particular genomic regions. For example, population expansions can cause allelic surfing, where regions of the genome "surf" to high frequency at leading edge of an expanding population. Allelic surfing can leave a heterogeneous patterns of variation across a species range leaving signals that may resemble local adaptation (Novembre and Di Rienzo 2009; Klopfstein *et al.* 2006).

As stated above, a feature of the WZA is that a wide range of summary statistics could be used as input. We converted p-values from Kendall's τ correlation or Bayes factors from *BayPass* into empirical p-values and used these as input to the WZA. The use of empirical p-values to account for genome-wide false positives is not unique to this study, we got the idea from Hancock *et al.* (2011) who used empirical p-values from GEA results to identify loci involved in climate adaptation in *Arabidopsis thaliana*. While the empirical p-value approach corrects for false positives due to population structure genome-wide, it throws away the null model for our summary statistic of interest. A GEA approach that produced parametric p-values that was adequately controlled for population structure may provide a more powerful input statistic to the WZA. Alternatively, with the advent of methods for reconstructing ancestral recombination graphs from population genomic data (Hejase *et al.* 2020), perhaps a GEA method could be developed that explicitly analyses inferred genealogies rather than individual markers in a manner similar to regression of phenotypes on genealogies proposed by Ralph *et al.* (2020).

The width of analysis windows

When performing a genome scan using a windowed approach a question that inevitably arises is, how to to decide on the width of analysis windows? In our simulation study, we used analysis windows that were 10,000bp long and within that distance LD had fully decayed to background levels (Figure S1). LD can be considered as the correlation in coalescence times between pairs of sites (McVean 2002; Wakeley 2007), so high LD is suggestive of highly correlated evolutionary histories. If the width of an analysis window were greater than the average distance over which LD decayed the window would presumably encompass a variety of evolutionary histories, some more correlated than others. Applying the WZA to GEA results in 10,000bp windows in our analysis of simulated data led to a distribution of Z_W scores that was close to the null expectation of the standard normal distribution, but with a thicker right hand tail (Figure 2). The deviation from normality presumably arose because some SNPs in the analysis windows had highly correlated histories, violating the independence assumption of the weighted-Z test.

On the other hand, if analysis windows were narrower than the average distance over which LD decayed most SNPs would presumably have highly correlated evolutionary histories. Random drift may cause genealogies in some regions of the genome to correlate with environmental variables more than others. Many of the SNPs present in an analysis window that consisted of genealogies that were highly correlated with the environment may be highly significant in a GEA analyses, leading to a large Z_W value. This effect would lead to a larger variance in Z_W for analysis windows that were narrower than the average LD decay. We down sampled the tree-sequences we recorded for our simulated populations to model analysis windows present in low

recombination regions and performed the WZA on the resulting data. As expected, we found that the variance of the distribution of Z_W scores was greater when there was a lower recombination rate (Figure S11).

Theoretical studies of local adaptation suggest that we should expect regions of the genome subject to spatially varying selection pressures to exhibit elevated linkage disequilibrium (LD) relative to the genomic background for a number of reasons. Under local adaptation, alleles are subject to positive selection in some parts of a species's range, but not in others. As a locally adaptive allele spreads in the locations where it is beneficial, it may cause some linked neutral variants to hitchhike along with it (Sakamoto and Innan 2019). Non-beneficial genetic variants introduced to local populations via gene flow may be removed, with a result being a build up of LD between selected alleles and linked neutral sites. This process can be thought of as a local barrier to gene flow (Barton and Bengtsson 1986). Beyond this hitchhiking signature, there is a selective advantage for alleles that are involved in local adaptation to cluster together, particularly in regions of low recombination (Rieseberg 2001; Noor *et al.* 2001; Kirkpatrick and Barton 2006; Yeaman 2013). For example, in sunflowers and *Littorina* marine snails, there is evidence that regions of suppressed recombination cause alleles involved in local adaptation to be inherited together (Morales *et al.* 2019; Todesco *et al.* 2020). The processes we have outlined are not mutually exclusive, but overall, genomic regions containing strongly selected alleles that contribute to local adaptation may have elevated LD and potentially exhibit GEA signals at multiple linked sites.

Window size would ideally be informed by the extent by which LD decays at neutral sites and how much LD is induced by selection. Determining that is likely very difficult however as understanding the effects of migration-selection balance on linkage disequilbrium requires estiamtes of selection and dispersal which are difficult to determine. Additionally, recombination rates vary widely among taxa but also within the genome (Stapley *et al.* 2017). Such variation causes LD to decay over greater or shorter distances in different locations across the genome so window size may need to be dynamic with respect to recombination rate variation. Random drift may cause some regions of the genome to correlate with environmental variables more than others. If such drift happens in regions of low recombination, multiple linked SNPs may exhibit strong GEA signals. This would lead to a distribution of Z_W scores that had a higher variance in regions of low recombination and generate statistical artefacts as we outlined in our previous study (Booker *et al.* 2020).

Literature Cited

Aitken, S. N. and M. C. Whitlock, 2013 Assisted Gene Flow to Facilitate Local Adaptation to Climate Change. Annu. Rev. Ecol. Evol. Syst. **44**: 367–388.

Barton, N. and B. O. Bengtsson, 1986 The barrier to genetic exchange between hybridising populations. Heredity **57**: 357–376.

Bhatia, G., N. Patterson, S. Sankararaman, and A. L. Price, 2013 Estimating and interpreting FST: The impact of rare variants. Genome Research **23**: 1514–1521.

Bontrager, M., C. D. Muir, C. Mahony, D. E. Gamble, R. M. Germain, *et al.*, 2020 Climate warming weakens local adaptation.

Booker, T. R., S. Yeaman, and M. C. Whitlock, 2020 Variation in recombination rate affects detection of outliers in genome scans under neutrality. Molecular Ecology **29**: 4274–4279.

Charlesworth, B. and D. Charlesworth, 2010 *Elements of Evolutionary Genetics*. Roberts & Company, Greenwood Village, Colorado.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard, 2010 Using environmental correlations to identify loci underlying local adaptation. Genetics **185**: 1411–1423.

Forester, B. R., M. R. Jones, S. Joost, E. L. Landguth, and J. R. Lasky, 2016 Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular Ecology **25**: 104–120.

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban, 2018 Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. Molecular Ecology **27**: 2215–2233.

Frichot, E. and O. François, 2015 LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution **6**: 925–929.

Frichot, E., S. D. Schoville, G. Bouchard, and O. François, 2013 Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Molecular Biology and Evolution **30**: 1687–1699.

Gautier, M., 2015 Genome-wide scan for adaptive divergence and association with population-specific covariates. Genetics **201**: 1555–1579.

Haldane, J. B., 1948 The theory of a cline. Journal of Genetics **48**: 277–284.

Haller, B. C., J. Galloway, J. Kelleher, P. W. Messer, and P. L. Ralph, 2019 Tree-sequence recording in SLiM opens new horizons for forward-time simulation of whole genomes. Mol. Ecol. Resour. **19**: 552–566.

Hancock, A. M., B. Brachi, N. Faure, M. W. Horton, L. B. Jarymowycz, *et al.*, 2011 Adaptation to climate across the Arabidopsis thaliana genome. Science **334**: 83–86.

Hejase, H. A., N. Dukler, and A. Siepel, 2020 From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection.

Hereford, J., 2009 A quantitative survey of local adaptation and fitness trade-offs.

Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd, *et al.*, 2016 Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. American Naturalist **188**: 379–397.

Kirkpatrick, M. and N. Barton, 2006 Chromosome inversions, local adaptation and speciation. Genetics **173**: 419–434.

Klopfstein, S., M. Currat, and L. Excoffier, 2006 The fate of mutations surfing on the wave of a range expansion. Molecular Biology and Evolution .

Legendre, P. and L. Legendre, 2012 *Numerical Ecology, Volume 24*. Elsevier, third edition.

Lotterhos, K. E., 2019 The Effect of Neutral Recombination Variation on Genome Scans for Selection. G3 **9**: 1851–1867.

Meirmans, P. G., 2012 The trouble with isolation by distance. Molecular Ecology **21**: 2839–2846.

Mimura, M. and S. N. Aitken, 2007 Adaptive gradients and isolation-by-distance with postglacial migration in Picea sitchensis. Heredity **99**: 224–232.

Morales, H. E., R. Faria, K. Johannesson, T. Larsson, M. Panova, *et al.*, 2019 Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. Science Advances **5**: eaav9963.

Mosca, E., S. C. González-Martínez, and D. B. Neale, 2014 Environmental versus geographical determinants of genetic structure in two subalpine conifers. New Phytologist **201**: 180–192.

Noor, M. A., K. L. Gratos, L. A. Bertucci, and J. Reiland, 2001 Chromosomal inversions and the reproductive isolation of species. Proceedings of the National Academy of Sciences of the United States of America **98**: 12084–12088.

Novembre, J. and A. Di Rienzo, 2009 Spatial patterns of variation due to natural selection in humans.

Pavy, N., M. C. Namroud, F. Gagnon, N. Isabel, and J. Bousquet, 2012 The heterogeneous levels of linkage disequilibrium in white spruce genes and comparative analysis with other conifers. Heredity **108**: 273–284.

Ralph, P., K. Thornton, and J. Kelleher, 2020 Efficiently summarizing relationships in large samples: A general duality between statistics of genealogies and genomes. Genetics **215**: 779–797.

Rieseberg, L. H., 2001 Chromosomal rearrangements and speciation.

Sakamoto, T. and H. Innan, 2019 The evolutionary dynamics of a genetic barrier to gene flow: From the establishment to the emergence of a peak of divergence. Genetics **212**: 1383–1398.

Schlötterer, C., R. Tobler, R. Kofler, and V. Nolte, 2014 Sequencing pools of individuals-mining genome-wide polymorphism data without big funding.

Stapley, J., P. G. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja, 2017 Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences **372**.

Todesco, M., G. L. Owens, N. Bercovich, J. S. Légaré, S. Soudi, *et al.*, 2020 Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature **584**: 602–607.

Walsh, B. and M. Lynch, 2018 *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Wang, T., A. Hamann, D. Spittlehouse, and C. Carroll, 2016 Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America. PLOS ONE **11**: e0156720.

Weir, B. S. and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution **38**: 1358–1370.

Whitlock, M. C., 2005 Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. Journal of Evolutionary Biology **18**: 1368–1373.

Yeaman, S., 2013 Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proceedings of the National Academy of Sciences of the United States of America **110**: E1743–E1751.

Yeaman, S., A. C. Gerstein, K. A. Hodgins, and M. C. Whitlock, 2018 Quantifying how constraints limit the diversity of viable routes to adaptation. PLoS Genetics **14**: e1007717.

Yeaman, S., K. A. Hodgins, K. E. Lotterhos, H. Suren, S. Nadeau, *et al.*, 2016 Convergent local adaptation to climate in distantly related conifers. Science **353**: 1431–1433.

Zhou, X., P. Carbonetto, and M. Stephens, 2013 Polygenic Modeling with Bayesian Sparse Linear Mixed Models. PLoS Genetics **9**: 1003264.

Appendix

Parametrising simulations of local adaptation

Consider a hypothetical species of conifer inhabiting British Columbia. There may be many hundreds of millions of individuals in this hypothetical species distributed across the landscape. It would be computationally intractable to simulate all individuals forward-in-time, incorporating adaptation to environmental heterogeneity across the landscape and recombining chromosomes. In our simulations we scaled several population genetic parameters to model a large population when simulating a much smaller one. In the following sections, we outline and justify the approach we used to scale pertinent population genetic parameters.

Mutation rate

The neutral mutation rate was set such that there would be an average of around 30 SNPs that had a minor allele frequency greater than 0.05. We aimed at around 20-30 SNPs per gene as that is the number identified for lodgepole pine by Yeaman *et al.* (2016).

Recombination rates

In conifers, recombination rates (r) have been estimated to be on the order of 0.05 cM/Mbp; more than $10\times$ lower than the average for humans (Stapley *et al.* 2017). The pattern of LD decay in a panmictic population can be predicted by the population scaled recombination parameter ($\rho = 4N_e r$; Charlesworth and Charlesworth 2010), but the pattern of LD decay in structured is less well described. However, LD decays very rapidly in conifers (Pavy *et al.* 2012). In conifers, $\rho = 0.005$ has been estimated (Pavy *et al.* 2012). This implies a very large effective population size of roughly $\frac{0.005}{4 \times 0.5 \times 10^{-8}} = 2.5 \times 10^6$, much larger than is feasible to simulate. We simulated meta-populations of 19,600 individuals

We thus chose a recombination rate that gave us a pattern of LD decay that was similar to what has been observed in conifers. We set the per base pair recombination $r = 1 \times 10^{-7}$.

to simulate populations with a recombination rate parameter of $4N_d r = 0.00004$ as we found this gave a pattern of LD-decay in our stepping-stone populations (Figure S1). Note that the that was very similar to what has been observed in conifers. To achieve levels of LD-decay in our simulations that are similar to those expected in our hypothetical organism, we set $4N_d r = 0.00004$, but with only 100 individuals per deme that gave a per base pair recombination of 1×10^{-7}.

Selection coefficients

It is difficult to choose a realistic set of selection parameters for modelling local adaptation because there are, at present, no estimates of the distribution of fitness effects for mutations that have spatially divergent effects. However, common garden studies of a variety of taxa have estimated fitness differences of up between 35-45% between populations grown in home-like conditions versus away-like conditions (Hereford 2009; Bontrager *et al.* 2020). Motivated by such studies, we chose to parametrise selection using the maximum possible fitness difference between home versus away environments.

When modelling directional selection, our simulations contained 12 loci that could mutate to generate a locally beneficial allele. The phenotypic optima that we simulated ranged from -7 to 7 and we modelled selection on a locus as $1 + s_a\theta$ for a homozygote and $1 + hs_a\theta$ for a heterozygote, where s_a is the selection coefficient, θ is the phenotypic optimum and h is the dominance coefficient. With a selection coefficient of $s_a = 0.003$, the maximum relative fitness was $(1 + 7 \times s_a)^{12} = 1.28$ for an individual homozygous for all locally beneficial alleles. An individual homozygous for those alleles, but in the oppositely selected environment (i.e. present in the wrong deme) had a fitness of $(1 - 7 \times 0.003)^{12} = 0.775$. However, in preliminary simulations, we found that directional selection on locally beneficial alleles with $s_a = 0.003$ resulted in such strong patterns of local adaptation that all GEA methods identified all adapted loci with ease. We thus chose to reduce the strength of selection in the directional selection simulations to $s_a = 0.003$, this resulted in a maximum of 35% difference in fitness between individuals grown in home-like conditions versus away-like conditions.

As stated the main text, for stabilising selection simulations we chose $V_s = 192$ as this gave a maximum of 50% difference in fitness between individuals grown in home-like conditions versus away-like conditions.

Migration rate

For the migration rate, we worked backwards. We set out to achieve F_{ST} across the metapopulation of approximately 0.05, as has been reported for widely distributed conifer species such as lodgepole pine and interior spruce (Yeaman *et al.* 2016). For the stepping-stone simulations, we chose a migration rate of $\frac{7.5}{2N_d}$ as we found that this gave a mean F_{ST} of 0.04. For an island model, we used the analytical formulae given in the main text to set m to achieve a mean F_{ST} of 0.03.

Table S1 Population genetic parameters of a hypothetical organism, and how they are scaled in the simulations. The meta-population inhabits a 14×14 2-dimensional stepping stone. Parameters are shown for a population with 12 loci subject to directional selection.

Parameter	Hypothetical Biological Value	Scaled Parameter	Unscaled (simulation)
Global population size (N_e)	10^6	-	19,600
Number of demes (d)	196	-	196
Local population size (N_d)	5,100	-	100
Recombination rate (r)	10^{-8}	$4N_d r = 2.04 \times 10^{-4}$	5.10×10^{-7}
Selection coefficient (s_{Max})	0.0001	$2N_d s_{Max} = 0.6$	0.003
Migration rate (m)	9.80×10^{-4}	$2N_d m = 10$	0.05
Functional mutation rate (μ_α)	2×10^{-10}	$4N_e \mu_\alpha = 0.0008$	10^{-8}
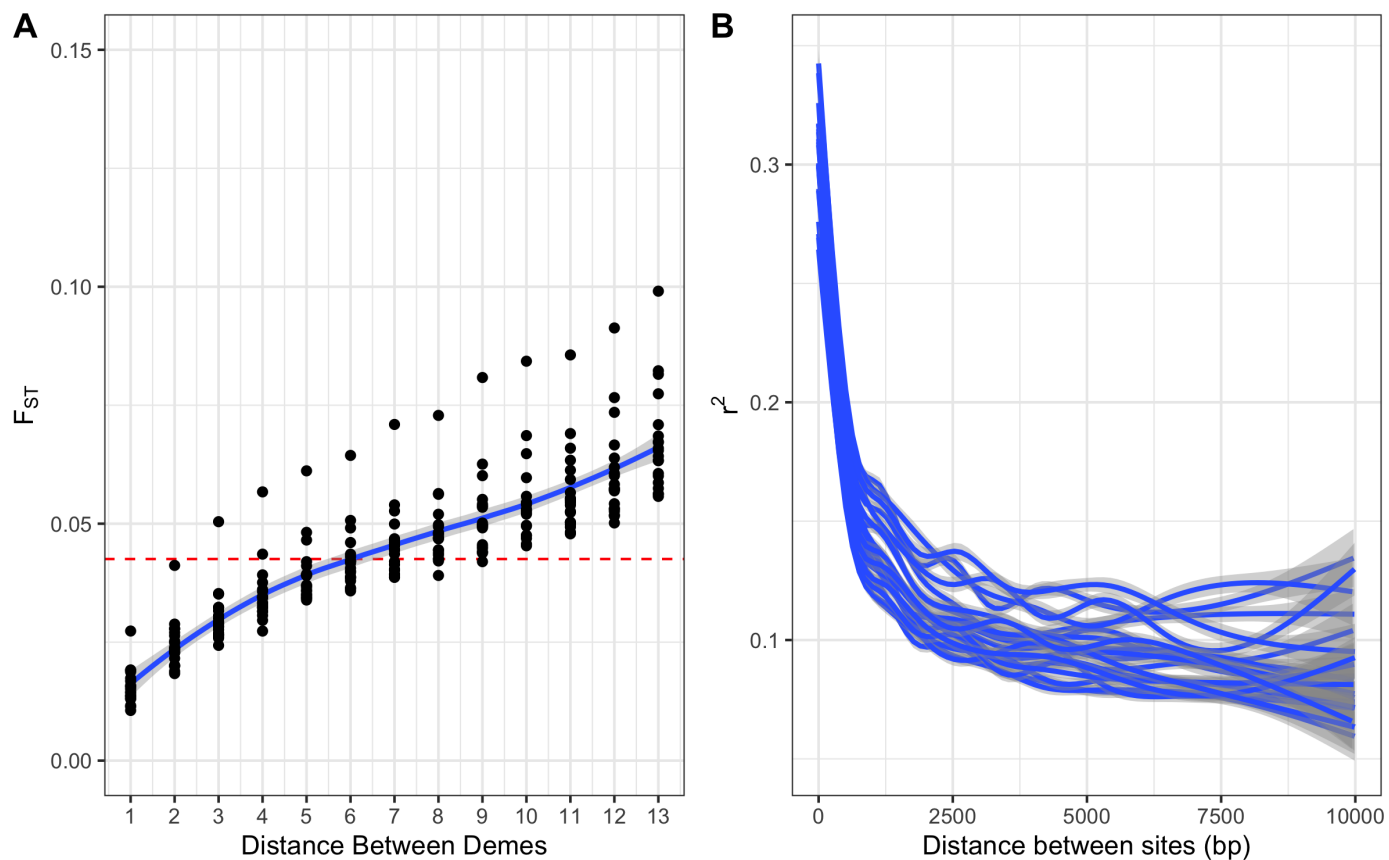
Figure S1 Summary statistics from simulations. A) shows the F_{ST} between pairs of demes in stepping-stone populations, the average across replicates is . B) shows LOESS smoothed LD, as measured by r^2, between pairs of SNPs, each line corresponds to a single simulation replicate.
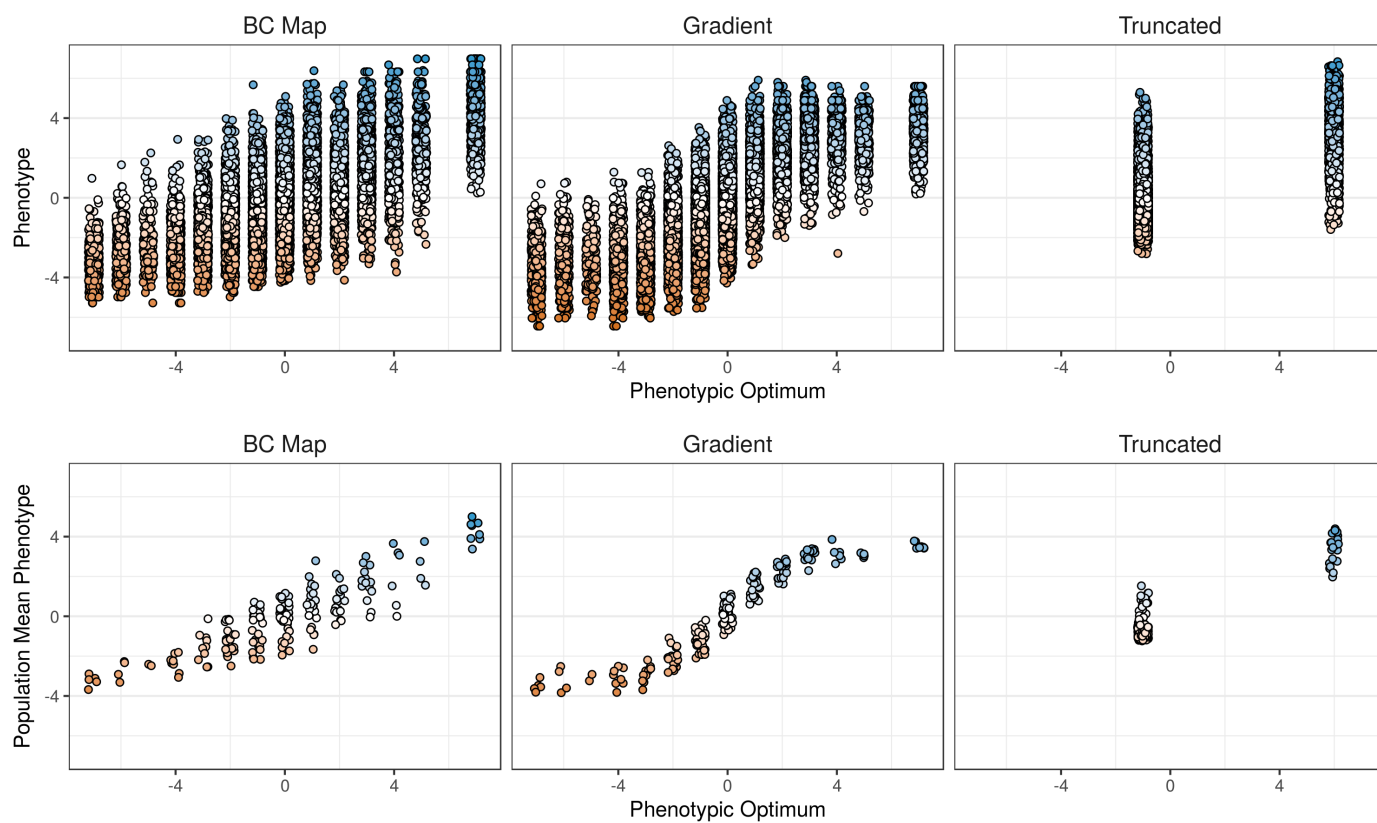
Figure S2 Individual and population mean phenotypes observed in representative simulations for each of the environment maps simulated. A small amount of horizontal jitter was added to points for visualisation purposes
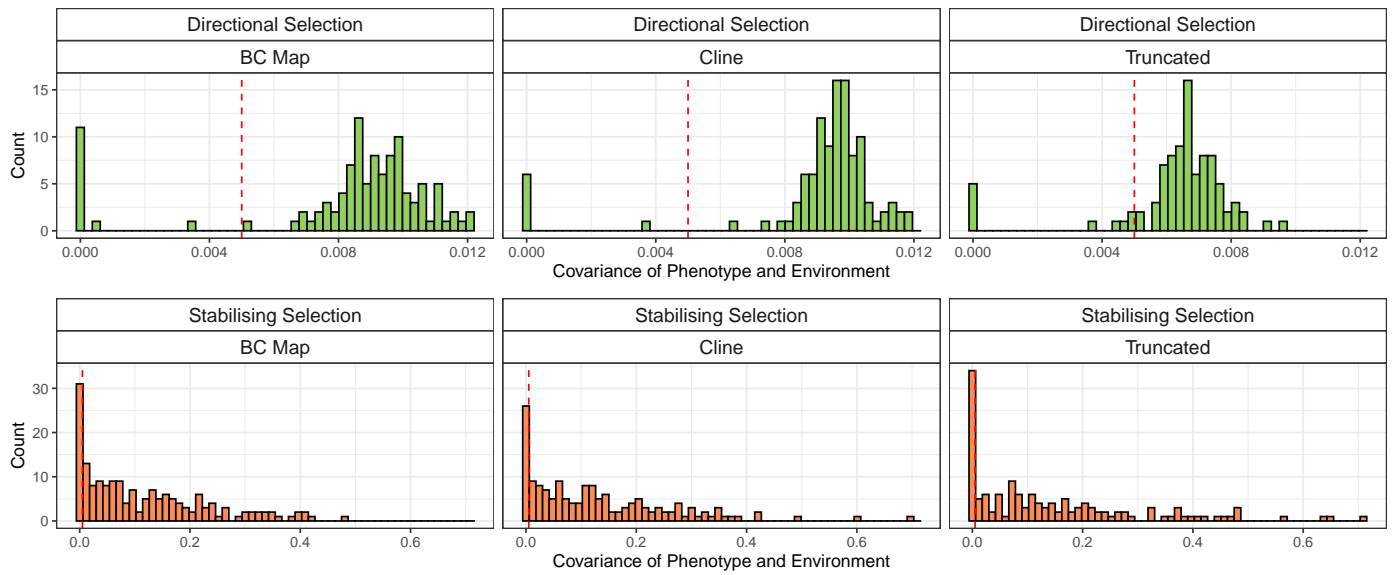
Figure S3 The effect size distribution from simulations of local adaptation. The vertical line indicates Cov = 0.005, the threshold we used to determine whether a gene was considered as important for local adaptation.
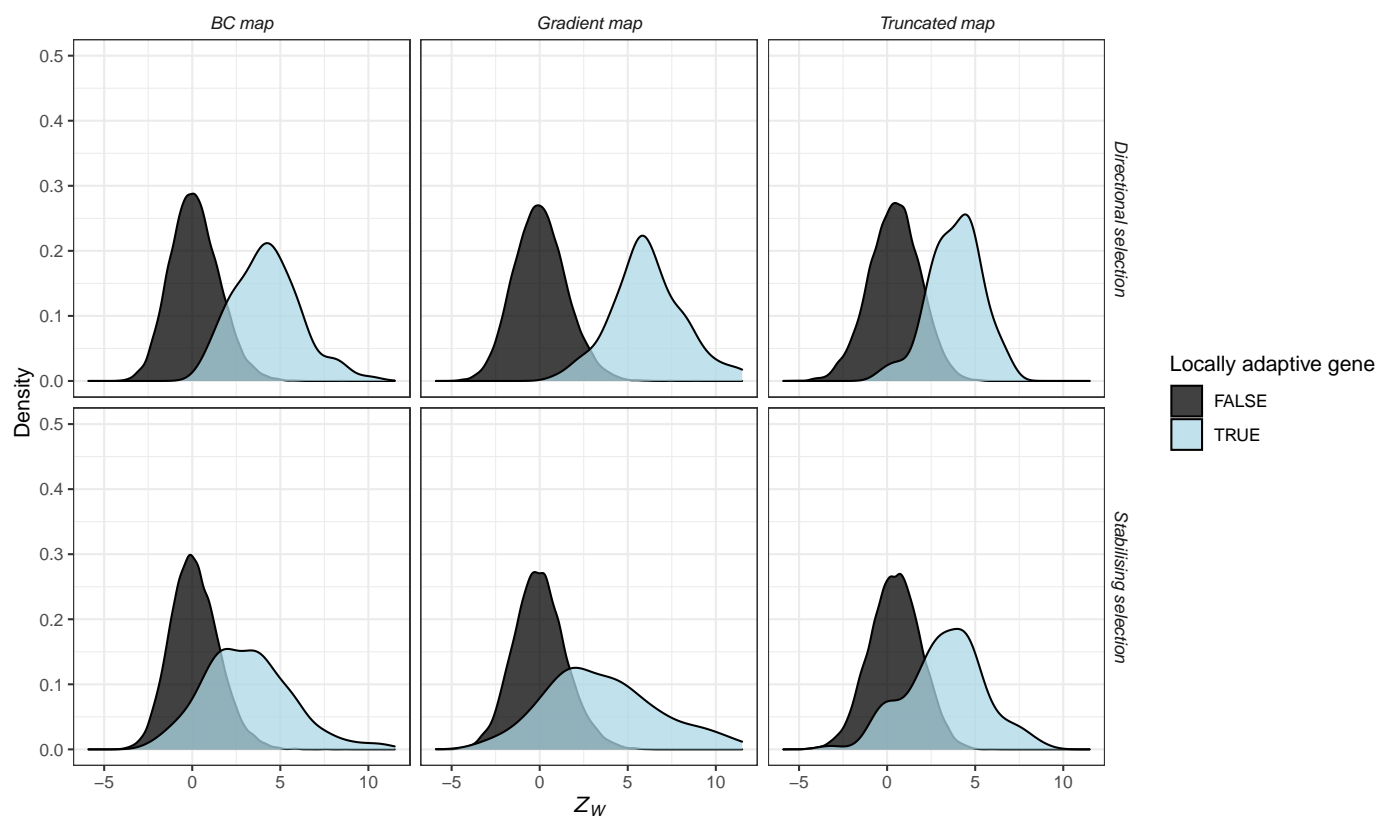
Figure S4 The distribution of WZA scores from simulations of local adaptation. Note, the plot does not indicate the relative frequency of genes that are or are not locally adaptive.
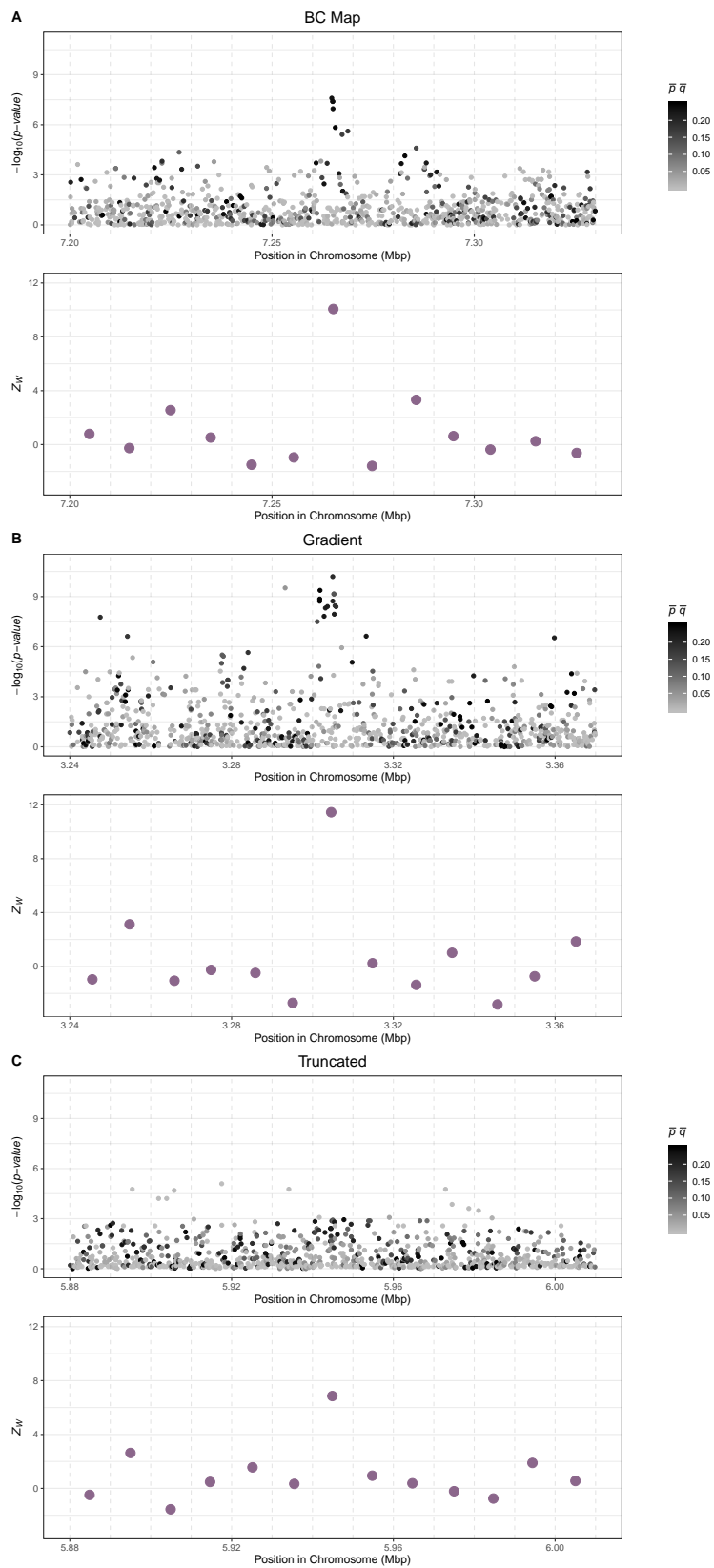
Figure S5 Plots demonstrating the genomic landscape of genotype-environment correlations around a true positive (i.e. a gene that contributes to local adaptation) for the three maps of environmental heterogeneity we simulated. In each panel, the upper plot shows the p-value of Kendall's τ correlation between population allele frequencies and the environment, the lower plot shows the Z_W score for each gene. The boundaries of genes are indicated by the dashed grey vertical lines. Note that the boundaries of genes recombined at a rate that was $50\times$ higher than the background rate.
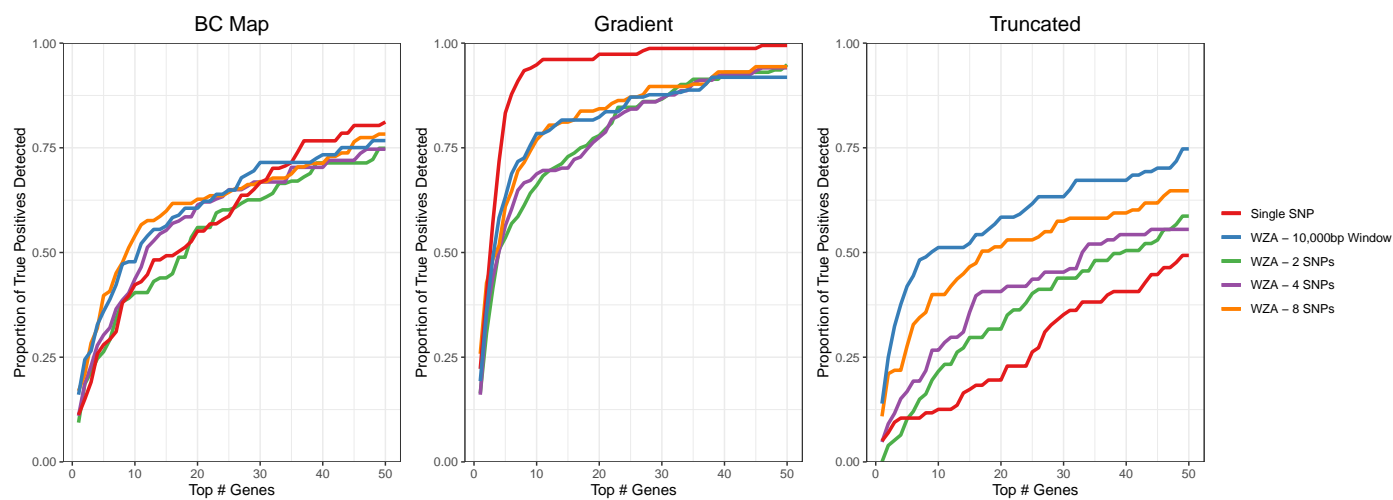
Figure S6 Comparing the performance of the WZA genes identified using the WZA, using analysis windows analysing a fixed number of SNPs.
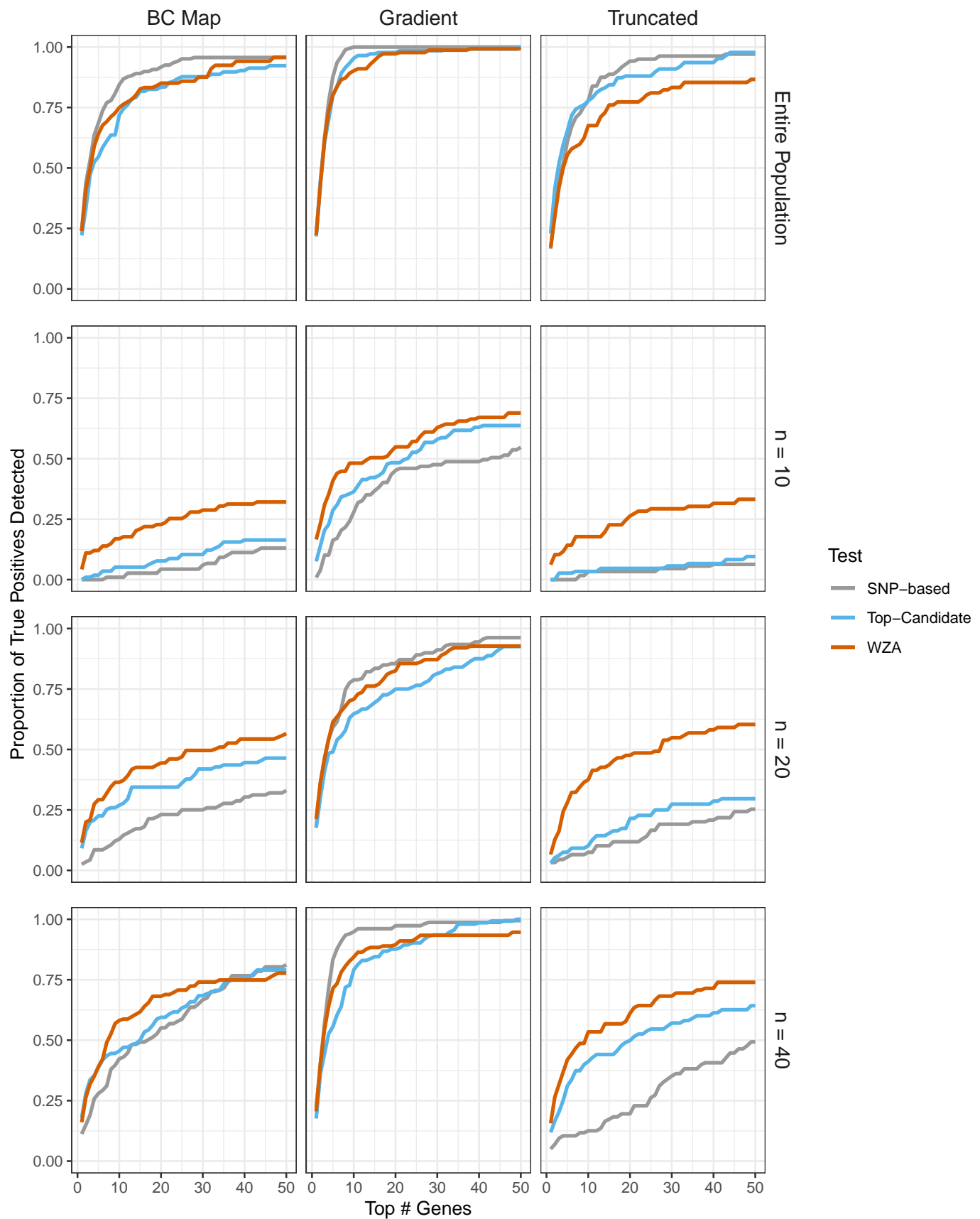
Figure S7 Comparison of the WZA, the top-candidate and the single-SNP approaches with varying numbers of sampled demes. Simulations shown assumed the *BC* map and directional selection. Note that the *n = 40* panel is shown as part of Figure 3 in the main text. Lines represent the mean of 20 simulation replicates.
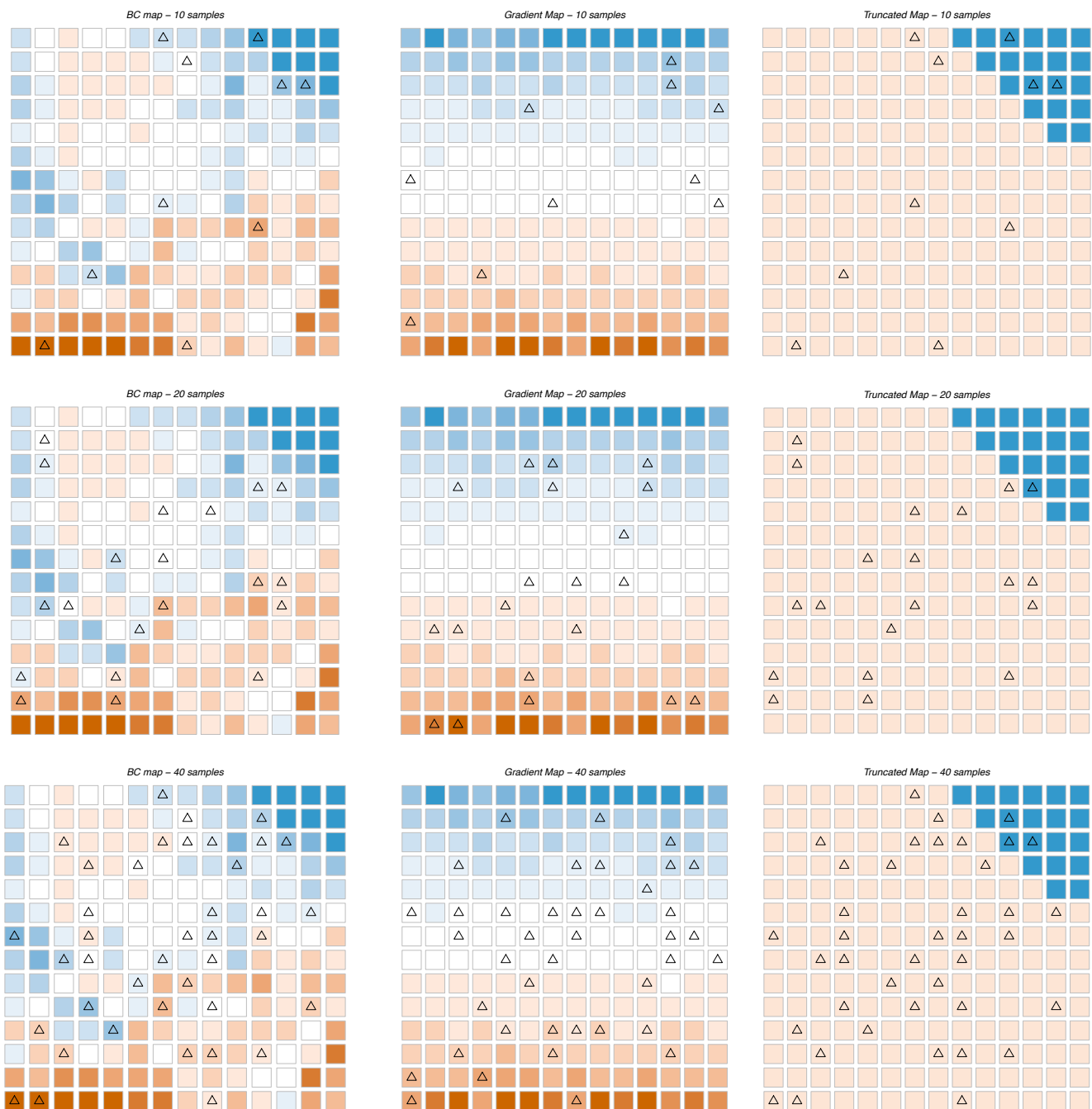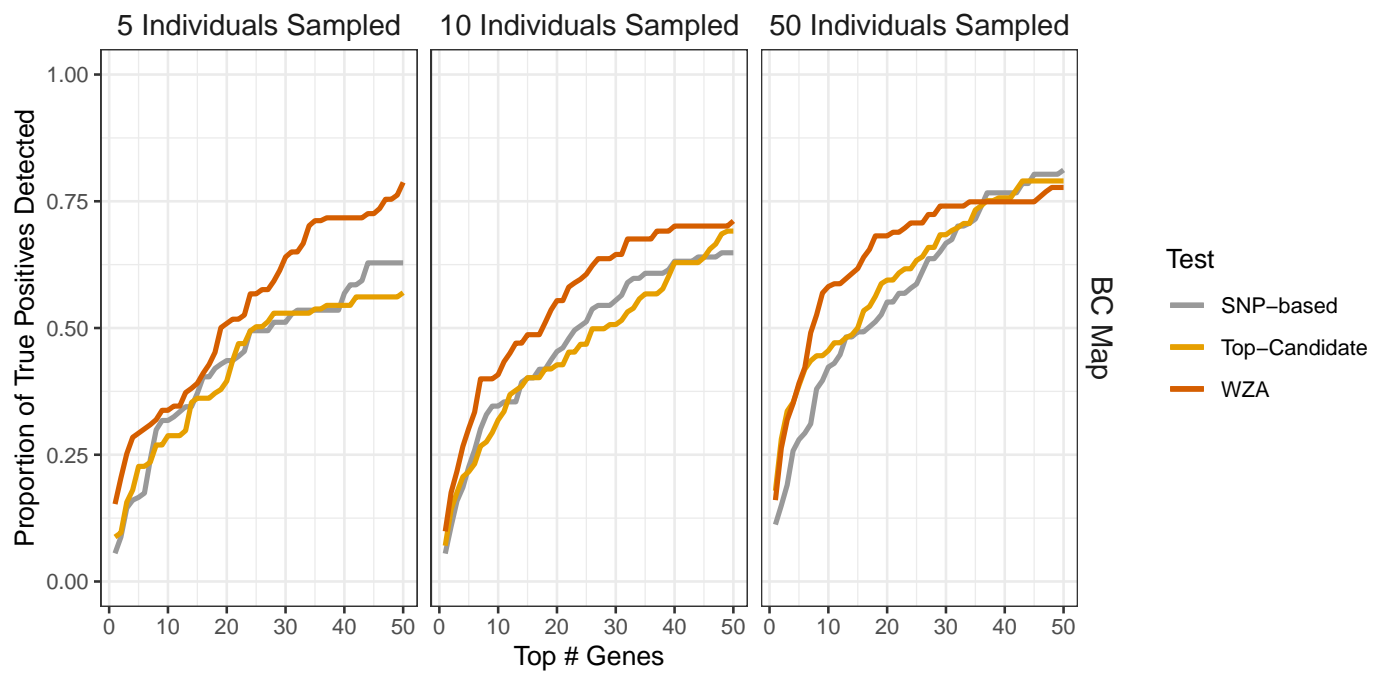
Figure S8 Comparison of the WZA, the top-candidate and the single-SNP approaches with varying numbers of sampled demes. Simulations shown assumed the *BC* map and directional selection. Note that the *n = 40* panel is shown as part of Figure 3 in the main text. Lines represent the mean of 20 simulation replicates.

Figure S9 Comparison of the WZA, the top-candidate and the single-SNP approaches with varying numbers of sampled demes. Simulations shown assumed the *BC* map and directional selection. Note that the *n = 40* panel is shown as part of Figure 3 in the main text. Lines represent the mean of 20 simulation replicates.
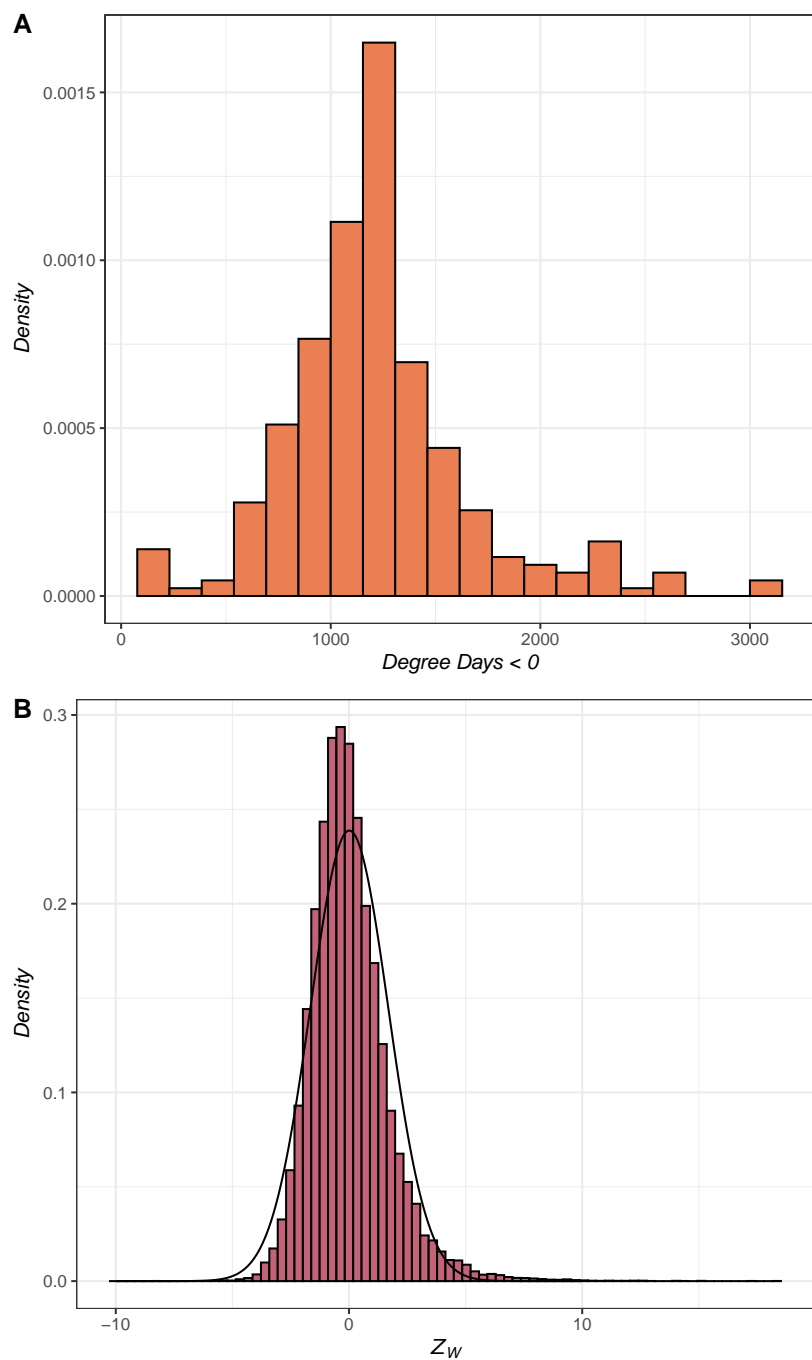
Figure S10 The distribution of A) degree days < 0 (DD0) across the populations of *P. contorta* sampled by Yeaman et al (2016) and B) Z_W scores for the GEA on DD0. Note that the DD0 values in A) are unscaled. In B) the curve shows a normal distribution fitted to the data.
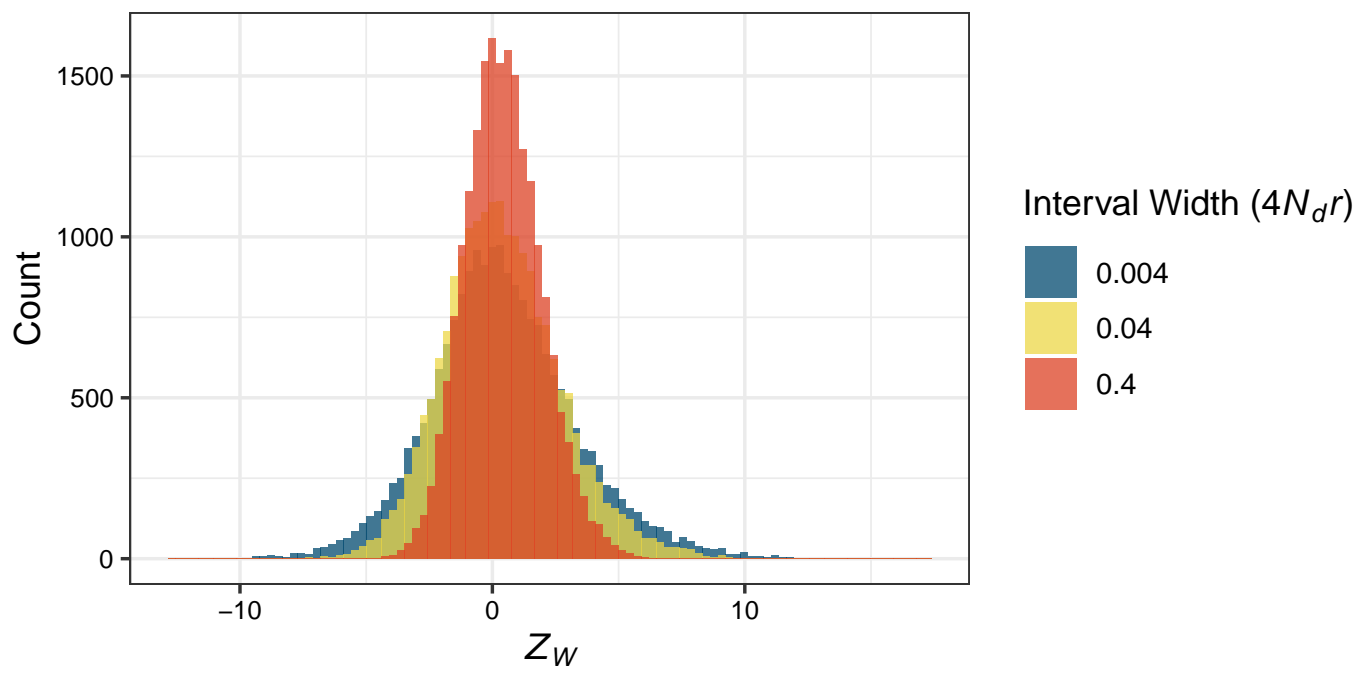
Figure S11 The distribution of Z_W scores under different recombination rates.