

1 WZA: A window-based method for
2 characterizing genotype-environment
3 association

4 Tom R. Booker^{1,2,3,*}, Sam Yeaman¹, Michael C. Whitlock^{2,3}

5 1. Department of Biological Sciences, University of Calgary, Calgary, Canada

6 2. Department of Zoology, University of British Columbia, Vancouver, Canada

7 3. Biodiversity Research Centre, University of British Columbia, Vancouver, Canada

8 *Corresponding author: booker@zoology.ubc.ca

9

10

11 **Abstract**

12 Genotype environment association (GEA) studies have the potential to elucidate the
13 genetic basis of local adaptation in natural populations. Specifically, GEA approaches
14 look for a correlation between allele frequencies and putatively selective features of the
15 environment. Genetic markers with extreme evidence of correlation with the
16 environment are presumed to be tagging the location of alleles that contribute to local
17 adaptation. In this study, we propose a new method for GEA studies called the
18 weighted-Z analysis (WZA) that combines information from closely linked sites into
19 analysis windows in a way that was inspired by methods for calculating F_{ST} . We analyze
20 simulations modelling local adaptation to heterogeneous environments either using a
21 GEA method that controls for population structure or an uncorrected approach. In the
22 majority of cases we tested, the WZA either outperformed single-SNP based
23 approaches or performed similarly. The WZA outperformed individual SNP approaches
24 when the measured environment is not perfectly correlated with the true selection
25 pressure or when a small number of individuals or demes was sampled. We apply the
26 WZA to previously published data from lodgepole pine identified candidate loci that
27 were not found in the original study.

28

29 KEYWORDS: Local adaptation, population genetics, landscape genomics

30

31

32 Introduction

33 Studying local adaptation can provide a window into the process of evolution, yielding
34 insights about the nature of evolvability, constraints to diversification, and the how the
35 interplay between a species and its environment shapes its genome (e.g. Savolainen
36 2013). Understanding local adaptation can also benefit practical applications such as in
37 forestry where many species of economic interest exhibit pronounced trade-offs in
38 fitness across environments. Characterizing such trade-offs may help identify alleles
39 involved in local adaptation, revealing candidate genes important for breeding or
40 informing conservation management programs for buffering against the consequences
41 of anthropogenic climate change (Aitken and Whitlock 2013). Whatever the aim or
42 application, a first step in studying the basis of local adaptation is to identify the genes
43 that are driving it.

44 A potentially powerful method for identifying the genomic regions involved in local
45 adaptation is genotype-environment association (GEA) analysis, which has been widely
46 adopted in recent years. Alleles may vary in frequency across a species' range in
47 response to local environmental conditions that give rise to spatially varying selection
48 pressures (Haldane 1948). For that reason, genetic variants that exhibit strong
49 correlations with putatively selective features of the environment are often interpreted as
50 a signature of local adaptation (Coop et al. 2010). Genotype-environment association
51 (GEA) studies examine such correlations. Allele frequencies for many genetic markers,
52 typically single nucleotide polymorphisms (hereafter SNPs), are estimated in numerous
53 locations across a species' range. Correlations between allele frequency and
54 environmental variables are calculated then contrasted for sites across the genome. It is
55 assumed in GEA studies that current heterogeneity in the environment (whether biotic
56 or abiotic) reflects the history of selection.

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

66 The most straightforward way to perform a GEA analysis is to simply examine the
67 correlation between allele frequencies and environmental variables measured in
68 multiple populations, for example using rank correlations such as Spearman's ρ or

69 Kendall's τ . This simple approach may commonly lead to false positives, however, if
70 there is environmental variation across the focal species' range that is correlated with
71 patterns of gene flow or historical selection (Meirmans 2012; Novembre and Di Rienzo
72 2009). For example, consider a hypothetical species inhabiting a large latitudinal range.
73 If it had restricted migration and exhibited isolation-by-distance, neutral alleles may be
74 correlated with any environmental variable that happened to correlate with latitude, as
75 population structure would also correlate with latitude.

76 Several approaches have been proposed to identify genotype-environment correlations
77 above and beyond what is expected given an underlying pattern of population structure
78 and environmental variation. For example, the commonly used *BayPass* package
79 (Gautier 2015), an extension of *BayEnv* by Coop et al. (2010), estimates correlations
80 between alleles and environmental variables in a two-step process. First, a population
81 covariance matrix (Ω) is estimated from SNP data. Second, correlations between the

82 frequencies of individual SNPs and environmental variables are estimated treating Ω in
83 a manner similar to a random effect in a generalized mixed model. In a recent study,
84 Lotterhos (2019) compared several the most commonly used packages for performing
85 GEA on pooled-sequencing datasets; including *BayPass* (Gautier 2015), latent-factor
86 mixed models (LFMMs) as implemented in the LEA package (Frichot et al. 2013; Frichot
87 and François 2015), redundancy analysis (RDA; see Forester et al. 2016, 2018) and a
88 comparatively simple analysis calculating Spearman's ρ between allele frequency and
89 environment. Of the methods they tested, Lotterhos (2019) found that the GEA

90 approaches that did not correct for population structure (i.e. Spearman's ρ and RDA)
91 had higher power to detect local adaptation compared to *BayPass* or LFMMs. In their
92 standard application to genome-wide datasets, all of the GEA analysis methods provide
93 a summary statistic for each marker or SNP.

94 Individual SNPs may provide very noisy estimates of summary statistics, but closely
95 linked SNPs are not independently inherited and may have highly correlated
96 evolutionary histories. As a way to reduce noise, genome scan studies often aggregate
97 data across adjacent markers into analysis windows based on a fixed physical or
98 genetic distance or number of SNPs (Hoban et al. 2016). In the case of F_{ST} , the standard
99 measure of population differentiation, there are numerous methods for combining
100 estimates across sites (See Bhatia et al. (2013)). In Weir and Cockerham's (1984)
101 method, for example, estimates of F_{ST} for individual loci are combined into a single value
102 with each marker's contribution weighted by its expected heterozygosity.

103 In the context of GEA studies, each marker or SNP provides a test of whether a
104 particular genealogy is correlated with environmental variation. In the extreme case of a
105 non-recombining region, all SNPs present would share the same genealogy and thus
106 provide multiple tests of the same hypothesis. The SNPs that are the most informative

107 in this context are those with the highest heterozygosities as they contain the most
108 information about the shape of the underlying genealogy. For recombining portions of
109 the genome, however, linked sites will not have exactly the same genealogy, but

110 genealogies may be highly correlated. Similar to combining estimates of F_{ST} to decrease
111 statistical noise, combining GEA tests performed on individual markers may increase
112 the power of GEA studies to identify genomic regions that contribute to local adaptation.

113 In this study, we propose a general method for combining the results of single SNP
114 GEA scores into analysis windows that we call the weighted-Z analysis (WZA). We test
115 the efficacy of WZA using simulations. We generate datasets modelling a pooled-
116 sequencing experiment where estimates of allele frequency are obtained for numerous
117 populations across a species' range. Using our simulated data, we compare the

118 performance of WZA to Kendall's τ as well as *BayPass* (Gautier 2015), as it is a widely
119 used approach that corrects for population structure in GEA studies. Additionally, we
120 compare WZA to another window-based GEA approach that was proposed by Yeaman
121 et al. (2016). We found that the WZA is particularly useful when GEA analysis is
122 performed on small samples and when results for individual SNPs are statistically noisy.
123 We re-analyze previously published lodgepole pine data using the WZA and find several
124 candidate loci that were not identified using the methods of the original study.

125

126

127

The Weighted-Z Analysis

128 In this study, we propose the Weighted-Z Analysis (hereafter, the WZA) for combining
 129 information across linked sites in the context of GEA studies. The WZA uses the
 130 weighted-Z test from the meta-analysis literature that combines *p*-values from multiple
 131 independent hypothesis tests into a single score (Mosteller and Bush 1954; Liptak 1958;

132 Stouffer et al. 1949). In the weighted-Z test, each of the n independent tests is given a
 133 weight that is proportional to the inverse of its error variance (Whitlock 2005). In the

134 WZA, we use \bar{pq} , a marker's expected heterozygosity, to determine these weights. At a
 135 given polymorphic site, we denote the average frequency of the minor allele across

136 populations as \bar{p} (\bar{q} corresponds to the frequency of the major allele). Sites with lower

137 values of \bar{pq} will have a greater relative error in estimates of local allele frequency than

138 will sites with higher \bar{pq} , causing greater relative error and bias in estimates of the
 139 correlation between allele frequency and an environmental variable. In order to capture
 140 this effect approximately we use the same weights as used by Weir and Cockerham

141 (1984) (i.e., \bar{pq}) to combine estimates of F_{ST} across sites.

142 We combine information from biallelic markers (typically SNPs) present in a focal
 143 genomic region into a single weighted-Z score (Z_W). The genomic region in question

144 could be a gene or genomic analysis window. We calculate $Z_{W,k}$ for genomic region k ,
 145 which contains n SNPs as

$$146 Z_{W,k} = \frac{\sum_{i=1}^n \bar{p}_i \bar{q}_i z_i}{\sqrt{\sum_{i=1}^n (\bar{p}_i \bar{q}_i)^2}}, \quad (1)$$

147 where \bar{p}_i is the mean allele frequency across populations and z_i is the standard normal
 148 deviate calculated from the one-sided *p*-value for SNP i . A given *p*-value can be

149 converted into a z_i score using the *qnorm* function in the R programming language, for
 150 example.

151 Under the null hypothesis that there is no correlation between allele frequency and
 152 environment and no spatial population structure, the expected distribution of correlation

153 coefficients in a GEA would be normal about 0, with a uniform distribution of p -values.
154 However, as will often be the case in nature, there may be an underlying correlation
155 between population structure and environmental variation that will cause these genome-
156 wide distributions to deviate from this null expectation. The average effect of population
157 structure on individual SNP scores can be incorporated into an analysis by converting
158 an individual SNP's squared correlation coefficient or parametric p -value into empirical
159 p -values based on the genome-wide distribution (following the approach of Hancock et
160 al. [2011]). To calculate empirical p -values, we rank all values (from smallest to largest
161 in the case of p -values) and divide the ranks by the total number of tests performed (i.e.
162 the number of SNPs or markers analyzed). Note that in practice, we calculated
163 empirical p -values after removing SNPs with minor allele frequency less than 0.05 and
164 would recommend that others perform similar filtering. In empirical studies with varying
165 levels of missing data across the genome, it may be preferable to rank the parametric p -
166 values rather than the correlation coefficients themselves as there may be varying
167 power to calculate correlations across the genome. With the empirical p -value
168 procedure, aggregating the empirical p -values using the WZA will identify genomic
169 regions with a pattern of GEA statistics that deviate from the average genome-wide. A
170 feature of the WZA is that many tests can potentially be used as input as long as
171 individual p -values provide a measure for the strength of evidence against a null
172 hypothesis.

173 When we apply the WZA in this study, we used two different statistics as input:
174 empirical p -values calculated from the genome-wide distribution of parametric p -values
175 from Kendall's τ correlating the local environmental variable and local allele frequency
176 (referred to as $WZA\tau$), and empirical p -values calculated from the genome-wide
177 distribution of Bayes factors as obtained using the *BayPass* program (referred to as
178 WZA_{BP} ; see below). Note that Lotterhos (2019) identified Spearman's ρ as having
179 among the highest power of the GEA analyses that they had tested. We used Kendall's
180 τ as it calculates accurate p -values in the presence of tied datapoints.

181

183

Materials and Methods

184

Simulating local adaptation

185 We performed forward-in-time population genetic simulations of local adaptation to
 186 determine how well the WZA was able to identify the genetic basis of local adaptation.
 187 GEA studies are often performed on large spatially extended populations that may be
 188 comprised of hundreds of thousands of individuals. However, it is computationally
 189 infeasible to model selection and linkage in long chromosomal segments (>1Mbp) for
 190 such large populations. For that reason, we simulated relatively small populations
 191 containing 19,600 diploid individuals in total and scaled population genetic parameters
 192 so as to model a large population. We based our choice of population genetic
 193 parameters on estimates for conifer species. A representative set of parameters is given
 194 in Table S1 and in the Appendix we give a breakdown and justification of the
 195 parameters we simulated. All simulations were performed in *SLiM* v3.4 (Messer and
 196 Haller 2019).

197
 198 We simulated meta-populations inhabiting and adapting to heterogeneous environments
 199 and modelled the population structure on an idealized conifer species. In conifers,

200 strong isolation-by-distance has been reported and overall mean $F_{ST} < 0.10$ has been
 201 estimated in several species (Mimura and Aitken 2007; Mosca et al. 2014). We thus
 202 simulated individuals inhabiting a 2-dimensional stepping-stone population made up of

203 196 demes (i.e. a 14×14 grid). Each deme consisted of $N_d = 100$ diploid individuals.
 204 We assumed a Wright-Fisher model so demes did not fluctuate in size over time.
 205 Migration was limited to neighboring demes in the cardinal directions and the reciprocal

206 migration rate between demes (m) was set to 0.0375 in each possible direction to

207 achieve an overall F_{ST} for the metapopulation of around 0.04 (Figure S1). As expected
 208 under restricted migration, our simulations exhibited a strong pattern of isolation-by-
 209 distance (Figure S1). Additionally, we simulated metapopulations with no spatial
 210 structure (i.e., finite island models). In these simulations, we used the formula

$$m = \frac{1}{F_{ST}} - 1$$

$$m = \frac{1}{4N_d 196}$$

211
 212 (Charlesworth and Charlesworth 2010; pp319) to determine that a migration rate

213 between each pair of demes of $m = 4.12 \times 10^{-4}$ would give a target F_{ST} of 0.03.

214
 215 The simulated organism had a genome containing 1,000 genes uniformly distributed
 216 onto 5 chromosomes. We simulated a chromosome structure in *SLiM* by including

217 nucleotides that recombined at $r = 0.5$ at the hypothetical chromosome boundaries.
218 Each chromosome contained 200 segments of 10,000bp each. We refer to these
219 segments as genes for brevity, although we did not model an explicit exon/intron or
220 codon structure. It has been reported that linkage disequilibrium (LD) decays rapidly in
221 conifers, with LD between pairs of SNPs decaying to background levels within 1,000bp
222 or so in several species (Pavy et al. 2012). In our simulations, recombination within

223 genes was uniform and occurred at a rate of $r = 10^{-7}$ per base-pair, giving a

224 population-scaled recombination rate ($4N_{dr}$) of 0.0004. The recombination rate between
225 the genes was set to 0.005, effectively modelling a stretch of 50,000bp of intergenic
226 sequence. Given these recombination rates, LD decayed rapidly in our simulations with
227 SNPs that were approximately 600bp apart having, on average, half the LD of
228 immediately adjacent SNPs in neutral simulations (Figure S1). Thus, patterns of LD
229 decay in our simulations were broadly similar to the patterns reported for conifers.

230
231 We incorporated spatial variation in the environment into our simulations using a
232 discretized map of degree days below 0 (DD0) across British Columbia (BC). We
233 generated the discretized DD0 map by first downloading the map of DD0 for BC from
234 ClimateBC (<http://climatebc.ca/>; Wang et al. 2016; Figure 1A). Using Dog Mountain, BC
235 as the reference point in the South-West corner (Latitude = 48.37, Longitude = -122.97),
236 we extracted data in a rectangular grid with edges 3.6 degrees long in terms of both

237 latitude and longitude, an area of approximately $266 \times 400 \text{ km}^2$ (Figure 1A). We divided

238 this map into a 14×14 grid, calculated the mean DD0 scores in each grid cell,
239 converted them into standard normal deviates (i.e. Z-scores) and rounded up to the
240 nearest third. We used the number of thirds of a Z-score as phenotypic optima in our
241 simulations. We refer to this map of phenotypic optima as the BC map (Figure 1B).

242
243 We used data from the BC map to generate two additional maps of environmental

244 variation. First, we ordered the data from the BC map along one axis of the 14×14 grid
245 and randomised optima along the non-ordered axis. We refer to this re-ordered map as
246 the *Gradient* map (Figure 1C). Second, we generated a map where selection differed
247 over only a small portion of the environmental range. For some species, fitness optima
248 may differ only beyond certain environmental thresholds, leading to a non-normal
249 distribution of phenotypic optima. To model such a situation, we set the phenotypic
250 optimum of 20 demes in the top-right corner of the meta-population to +3 and set the
251 optimum for all other populations to -1. We chose 20 demes as it represented
252 approximately 10% of the total population. We refer to this map as the *Truncated* map
253 (Figure 1D).

254

255 We simulated local adaptation using models of either directional or stabilizing selection.
256 In both cases, there were 12 causal genes distributed evenly across four simulated
257 chromosomes that potentially contributed to local adaptation. With directional selection,
258 mutations affecting fitness could only occur at a single nucleotide position in the center

259 of the 12 potentially selected genes. Directionally selected mutations had a spatially
260 antagonistic effect on fitness. In deme d with phenotypic optimum θ_d , the fitness of a
261 selected allele was calculated as $1 + s_a \theta_d$ for an individual homozygous for a locally
262 beneficial allele (selected alleles were semi-dominant). The fitness affecting alleles had
263 a mutation rate of 3×10^{-7} in simulations modelling directional selection and a fixed s_a
264 = 0.003 (see *Appendix*).

265
266 Under stabilizing selection, the mutations that occurred in the 12 genes had a normal
267 distribution of phenotypic effects, with variance $\sigma_a^2 = 0.5$. Phenotype-affecting mutations
268 occurred at a rate of 10^{-10} per base-pair in the 12 genes, and could occur at any of the
269 10,000 sites within a given gene. An individual's phenotype was calculated as the sum
270 of the effects of all phenotype-affecting mutations. We calculated an individual's fitness
271 using the standard expression for Gaussian stabilizing selection,

272

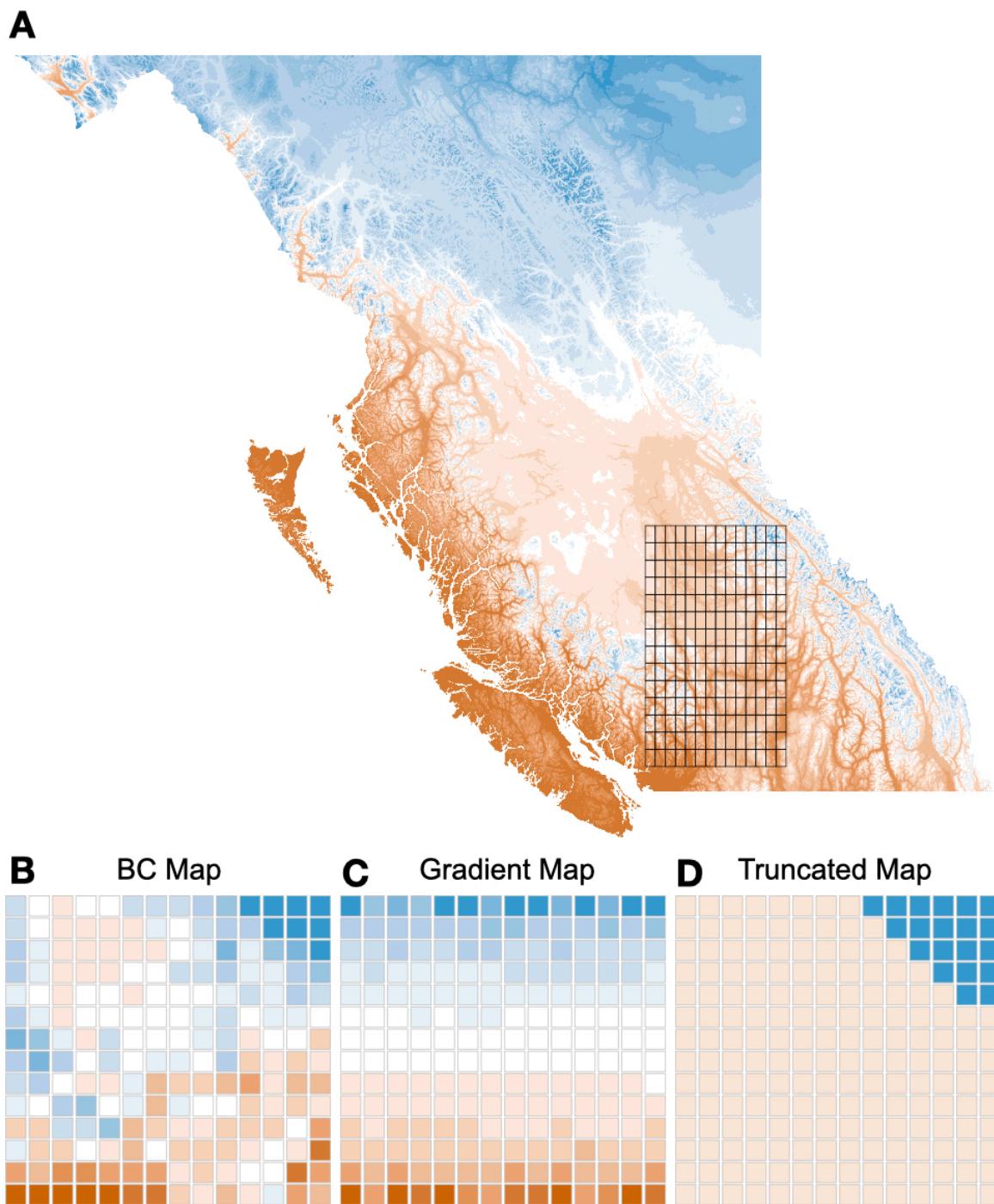
$$W_{z_{i,j}} = \exp \left[\frac{-(f_{i,j} - \theta_d)^2}{2V_s} \right],$$

273 where f_i is the phenotype of the i^{th} individual in environment j and V_s is the variance of
274 the Gaussian fitness function (Walsh and Lynch 2018). We set $V_s = 196$ so that there
275 was a 40% fitness difference between individuals perfectly adapted to the two extremes
276 of the distribution of phenotypic optima. This was motivated by empirical studies of local
277 adaptation that have demonstrated such fitness differences in numerous species
278 (Hereford 2009; Bontrager et al. 2020); see *Appendix*.

279
280 We ran simulations for a total of 200,102 generations. The 19,600 individuals initially
281 inhabited a panmictic population that evolved neutrally. After 100 generations, the
282 panmictic population divided into a 14×14 stepping-stone population and evolved
283 strictly neutrally (when modelling directional selection), or with a phenotypic optimum of
284 0 for all demes (when modelling stabilizing selection). After 180,000 generations, we
285 imposed the various maps of phenotypic optima and simulated for a further 20,000
286 generations. For selected mutations, we used the "*f*" option for *SLiM*'s mutation stack
287 policy, so only the first mutational change was retained. Using the tree-sequence option
288 in *SLiM* (Haller et al. 2019) we tracked the coalescent history of each individual in the
289 population. At the end of each simulations, neutral mutations were added at a rate of
290 10^{-8} using *PySLiM* (<https://pyslim.readthedocs.io/en/latest/>). For each combination of
291 map and mode of selection, we performed 20 replicate simulations.

292

293



296 **Figure 1** A) Degree days below zero across British Columbia, the overlain grid in A
 297 shows the locations we used to construct phenotypes for our simulated populations. B)
 298 A discretized map of DD0 in Southern British Columbia, we refer to the map in B as the
 299 *BC map*. C) A 1-dimensional gradient of phenotypic optima, we refer to this as the

300 Gradient map. D) A model of selection acting on a small proportion of the population,
301 we refer to this map as the Truncated map.

302

303 *Classifying simulated genes as locally adapted*

304 To evaluate the performance of different GEA methods, we needed to identify which
305 genes contribute to local adaptation and which do not in our simulated data. While there
306 were only 12 genes that were allowed to affect fitness in our simulations, not all of those
307 need be used by the evolving population to result in local adaptation. As described
308 above, our simulations incorporated a stochastic mutation model so from replicate to
309 replicate the genes that contributed to local adaptation varied and, in the case of
310 stabilizing selection, so did the effect size of the alleles in those genes.

311 For simulations modelling directional selection, we identified locally adapted genes
312 based on the mean fitness of their alleles. For a given gene containing directionally
313 selected alleles, our measure of local adaptation was the covariance between the mean
314 fitness contributed by the selected allele in each population and the environment.

315 For simulations modelling stabilizing selection, we identified locally adapted genes
316 based on the covariance of the environment and the phenotypic effects of their alleles,
317 summed across all variant sites within each gene. For a given gene, we summed the
318 additive phenotypic effects of all non-neutral variants and took the average for each
319 population. Our measure of local adaptation for each gene was the covariance between
320 that average additive phenotypic effect and environmental variation (we refer to this as
321 $\text{Cov}(\text{Phen}, \text{Env})$).

322 For both selection regimes, we defined locally adapted genes as those with a
323 covariance between environment and allelic effect (in fitness or phenotypic terms)
324 greater than 0.005. When assuming directional selection, an average of 6.35, 6.50 and
325 5.80 genes (out of 12) contained genetic variants that established and contributed to
326 local adaptation for the *BC map*, the Gradient map and the Truncated map,
327 respectively. In our simulations assuming stabilizing selection, individuals' and
328 population mean phenotypes closely matched the phenotypic optima of their local
329 environment (Figure S2). The average numbers of genes contributing to local
330 adaptation in individual replicates in these simulations were 7.15, 6.45 and 5.35 for the
331 *BC map*, the Gradient map and the Truncated map, respectively. However, when
332 analyzing stabilizing selection simulations, we calculated the proportion of the total
333 $\text{Cov}(\text{Phen}, \text{env})$ explained by a particular set of genes rather the number of true
334 positives.

335 *Analysis of simulation data*

336 We performed GEA on our simulated data using either Kendall's τ -b (hereafter
337 Kendall's τ), a rank correlation that does not model population structure, or *BayPass*,
338 which corrects for a population covariance matrix (Gautier 2015). For all analyses,
339 except where specified, we analyzed data for a set of 40 randomly selected demes and
340 sampled 50 individuals from each to estimate allele frequencies. We sampled

341 individuals from the same set of demes for all analyses, shown in Figure S3. Each
342 simulation replicate included 1,000 genes, and after excluding alleles with a minor allele
343 frequency less than 0.05 there was an average of 23.3 SNPs per gene.

344

345

346 We ran *BayPass* following the "worked example" in section 5.1.2 of the manual
347 provided with the software.

348 We used three different methods to summarize the GEA results for each gene in a
349 given simulation replicate: a single SNP-based approach, the WZA and the top-
350 candidate method developed by Yeaman et al. (2016). For all three tests, we used

351 either the *p*-values from Kendall's τ or *BayPass*.

352 • For the implementation of the single SNP-based approach, the SNPs with the
353 most extreme test statistic (i.e. smallest *p*-value or largest Bayes factor) for each
354 gene were recorded and other SNPs in the gene were subsequently ignored.
355 This was done to prevent multiple outliers that are closely linked from being
356 counted as separate hits. The single-SNP based method is perhaps most similar
357 to how GEA analyses are typically interpreted, as it relies upon the evidence from
358 the most strongly associated SNP to assess significance for a closely linked
359 gene.

360 • We implemented a simplified version of the top-candidate method proposed by
361 Yeaman et al (2016). The top-candidate method attempts to identify regions of
362 the genome involved in local adaptation under the assumption that such regions
363 may contain multiple sites that exhibit strong correlation with environmental
364 variables. The top-candidate method asks whether there is a significant excess
365 of "outlier" SNPs in a region compared to what one would expect given the
366 genome wide distribution. The number of outliers in a given genomic region is
367 compared to the expected number of outliers based on the genome-wide
368 proportion of SNPs that are outliers, using a binomial test. The *p*-value from the
369 binomial test is used as a continuous index.

370 • For the implementation of the WZA, we converted the *p*-values (from Kendall's τ)
371 or Bayes factors (from *BayPass*), into empirical *p*-values. For each of the n SNPs
372 present in a gene, empirical *p*-values were converted into z scores and used to
373 calculate WZA scores using Equation 1.

374 We examined effect of variation in recombination on the properties of the WZA by
375 manipulating the tree-sequences that we recorded in *SLiM*. In our simulations, genes
376 were 10,000 bp long, so to model genomic regions of low recombination rate, we
377 extracted the coalescent trees that corresponded to the central 1,000bp or 100bp of

378 each gene. For the 1,000bp and 100bp intervals, we added mutations at 10 \times and 100 \times
379 the standard mutation rate, respectively.

380 All SNPs present in each 10,000bp gene in our simulations were analyzed together.
381 However, to explore the effect of window size on the performance of the WZA, we
382 calculated WZA scores for variable numbers of SNPs. In these cases, we calculated
383 WZA scores for all adjacent sets of g SNPs and retained the maximum WZA score for
384 all sets of SNPs in the gene.

385 Tree sequences were manipulated using the *tskit* package. Mutations were added to
386 trees using the *msprime* (REF), *tskit* and *PySLiM* workflow (version). F_{ST} and r^2 (an
387 estimator of linkage disequilibrium) were calculated using custom Python scripts that
388 invoked the *scikitallel* package (REF).

389

390 **Analysis of data from lodgepole**
391 **pine**

392

393

394 We re-analyzed a previously published population genomic dataset for lodgepole pine,
395 *Pinus contorta*, a conifer that is widely distributed across the Northwest of North
396 America. Briefly, Yeaman et al. (2016) collected samples from 254 populations across
397 British Columbia and Alberta, Canada and Northern Washington, USA. The lodgepole
398 pine genome is very large (20Gbp), so Yeaman et al. (2016) used a sequence capture
399 technique based on the *P. contorta* transcriptome. Allele frequencies were estimated for
400 many markers across the captured portion of the genome by sequencing 1-4 individuals

401 per population. Yeaman et al. (2016) performed GEA on each SNP using Spearman's ρ
402 and used their top-candidate method (see above) to aggregate data across sites within
403 genes. We downloaded the data for individual SNPs from the Dryad repository
404 associated with Yeaman et al. (2016) (<https://doi.org/10.5061/dryad.0t407>). We

405 converted Spearman's ρ p -values into empirical p -values and performed WZA on the
406 same genes analyzed by Yeaman et al (2016). We also repeated the top-candidate
407 method, classifying SNPs with empirical p -values < 0.01 as outliers. However, as
408 above, we use the p -value from the top-candidate method as a continuous index.

409 **Data and Code Availability**

410 The simulation configuration files and code to perform the analysis of simulated data
411 and generate the associated plots are available at github/TBooker/GEA/WZA. Analyses
412 were performed using a combination of R and Python. All plots were made using

413 *ggplot2* (REF). Tree-sequence files for the simulated populations are available at Dryad
414 and all processed GEA files are available on (SomeCoolLocation).

415

416

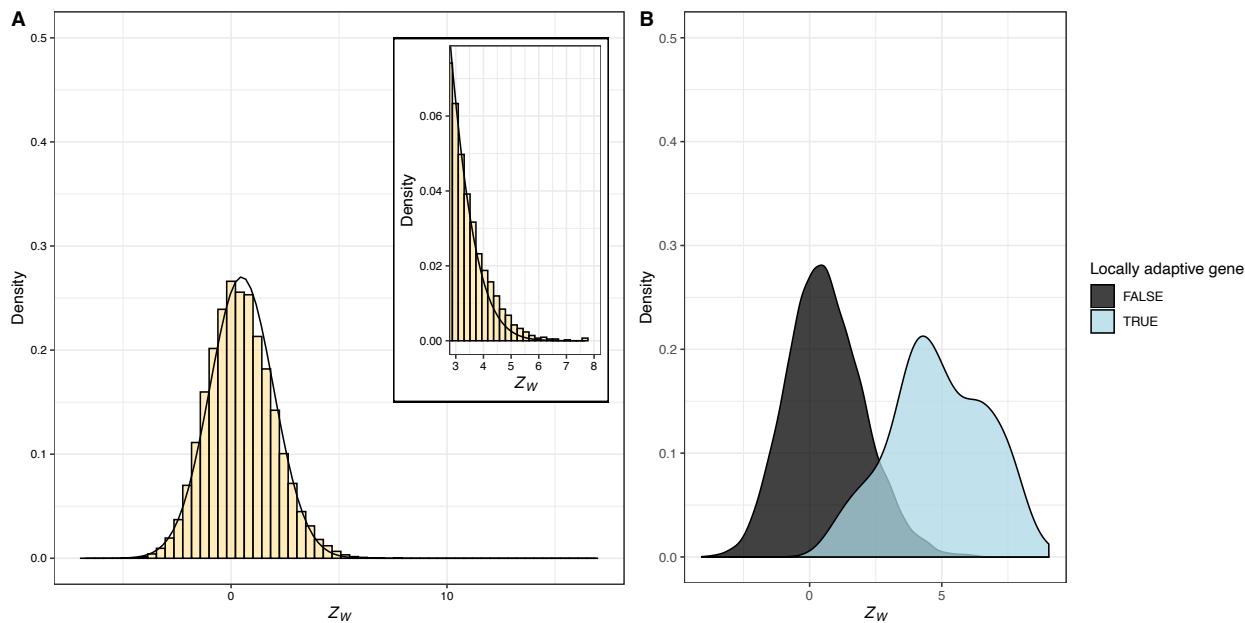
417

Results

418

The statistical properties of the WZA

419 To assess the statistical properties of the WZA, we first performed GEA analyses on
 420 populations that were evolving neutrally. Figure 2A shows the distribution of $WZA\tau$
 421 scores for stepping-stone populations simulated under the *BC Map*. The null
 422 expectation for WZA scores is the standard normal distribution (mean of 0 and standard
 423 deviation of 1), but we found that the distribution of $WZA\tau$ scores deviated slightly from
 424 this even under neutrality, where the mean and standard deviation of $WZA\tau$ scores from
 425 individual simulation replicates were approximately 0.089 and 1.38, respectively.
 426 Additionally, the inset histogram in Figure 2A shows that distribution of $WZA\tau$ scores
 427 had a thicker right-hand tail than expected under the normal distribution.



428

429 **Figure 2.** The distribution of WZA scores under neutrality and a model of local
 430 adaptation. A) A histogram of $WZA\tau$ scores under strict neutrality across a set of 20
 431 replicate simulations, inset is a close-up view of the upper tail of the distribution of Z_W
 432 scores. B) A density plot showing the separation of $WZA\tau$ scores for genes that are
 433 locally adaptive versus evolving neutrally across the genome of 20 simulation replicates.
 434 GEA was performed on 40 demes sampled from the *BC Map*.

435

436 The deviation from the standard normal distribution is driven by non-independence of
 437 SNPs within the analysis windows we used to calculate WZA scores. To demonstrate
 438 this, we re-calculated WZA scores, but permuted the locations of SNPs across the

439 genome, effectively erasing the signal of linkage within genes. The distribution of WZA τ
440 scores in this permuted dataset closely matched the null expectation and did not have a
441 thick right-hand tail (Figure S4; shuffled); each of 20 simulation replicates had mean a
442 WZA τ indistinguishable from 0 with a standard deviation very close to 1. It is worth
443 noting that we modelled populations that did not change in size over time. Non-
444 equilibrium population dynamics such as population expansion may influence the
445 distribution of WZA scores.

446 When evolution includes selection, WZA can often clearly distinguish regions of the
447 genome containing loci that contribute to local adaptation from those that do not. Figure
448 2B shows clear separation of WZA τ scores for genes that contribute to local adaptation
449 from those that are evolving neutrally (similar results were found for both the *Gradient*
450 and *Truncated* maps; Figure S5). The distributions of WZA τ scores for locally adapted
451 genes when modelling stabilizing selection was broader than when modelling directional
452 selection (Figure S5), consistent with differences in the distributions of effect size for the
453 genes involved in local adaptation under the two selection models (Figure S6). The
454 separation of the distributions of WZA τ scores for locally adaptive genes versus
455 neutrally evolving genes indicates that it may be a powerful method for identifying the
456 genetic basis of local adaptation.

457

458 Comparison of the WZA with other GEA approaches

459 We compared WZA to two other methods for identifying genomic regions that contribute
460 to local adaptation from GEA data (Figure 3). To assess the performance of the different

461 methods, we examined the top 1, 2, 3,... 50 genes in terms of WZA τ scores, $-\log_{10}(p$ -
462 values) from the top-candidate method, or the single SNP Kendall's τ approach and
463 calculated the proportion of all true positives that were identified in each case. In our
464 simulations, there were 1,000 genes in total with around 6 locally adapted genes in
465 each replicate (see Methods). For visualization purposes, we include Figure S7, which

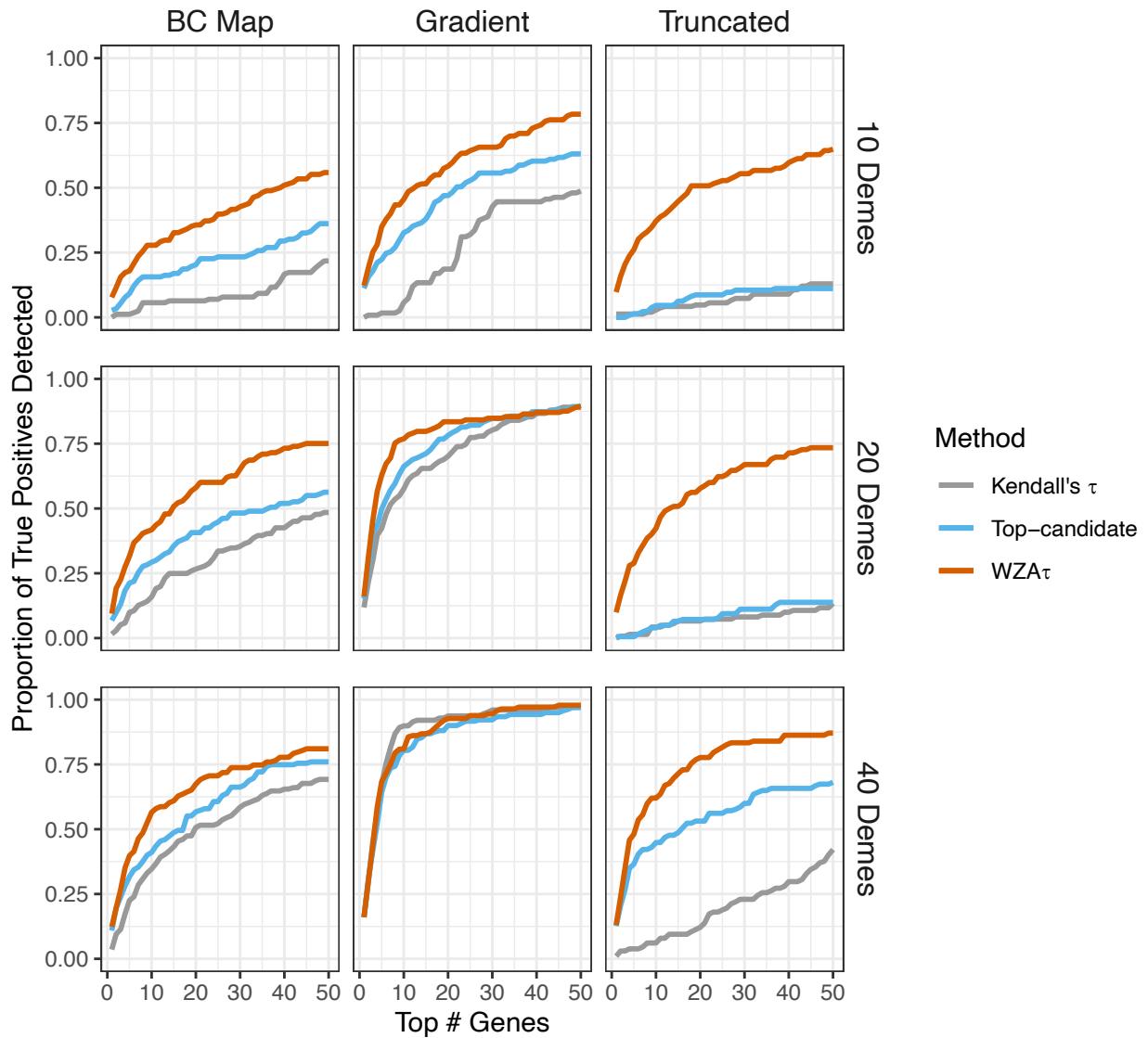
466 shows the $-\log_{10}(p$ -values) from Kendall's τ represented as a Manhattan plot for
467 individual simulation replicates, WZA τ and top-candidate scores calculated from those
468 data and the proportion of true positives detected using the three different analysis
469 methods.

470

471

472 Figure 3 compares the performance of the GEA methods across the three different
473 maps of environmental variation that we simulated. Empirical GEA studies may vary
474 substantially in terms of how many populations or demes are sampled. For each of the
475 three maps we simulated, we analyzed samples of 10, 20 or 40 demes where allele
476 frequencies were estimated from 50 individuals sampled in each location; Figure SF
477 shows the specific demes we sampled in each case. Figure 3 shows that WZA τ
478 substantially outperformed both the top-candidate and single SNP-based Kendall's τ
479 analyses in many cases. When analyzing simulations that used the *BC* map or the
480 *Truncated* map, WZA τ always outperformed the top-candidate and SNP-based
481 methods, but particularly so when fewer demes were sampled (Figure 3). When
482 simulations assumed the *Gradient* map, WZA τ outperformed the other GEA methods
483 when the sample was restricted to 10 demes, but with larger samples, the tests were
484 more similar (Figure 3). This suggests that WZA τ is a powerful method for identifying
485 regions of the genome that contribute to local adaptation in empirical analyses, but
486 particularly so when they are performed on small samples.

487 An additional source of variation in GEA studies comes from the number of individuals
488 sampled in each location. We also examined the effect that reduced sampling of
489 individuals within each deme had on the performance of the methods. Figure S8 shows
490 that the WZA outperforms the top-candidate and SNP-based methods when a small
491 number of individuals is used to estimate allele frequencies. Note that this is not strictly
492 a test of how well pooled-seq will perform with small sample sizes, however. With small
493 numbers of individuals in sequencing pools, differential amounts of DNA from each
494 individual may add to error in allele frequency estimation (Schlötterer et al. 2014).



495

496 **Figure 3** A comparison of three methods to identify outliers in GEA analysis conducted
 497 on simulations modelling local adaptation via directional selection. Plots show the
 498 proportion of true positives identified by examining the genes with the top ranked scores
 499 across the genome for the three GEA methods. The rows of the plot show results
 500 obtained from samples of 10, 20 or 40 demes as indicated by the labels on the right-
 501 hand side. Lines represent the means of 20 simulation replicates.
 502

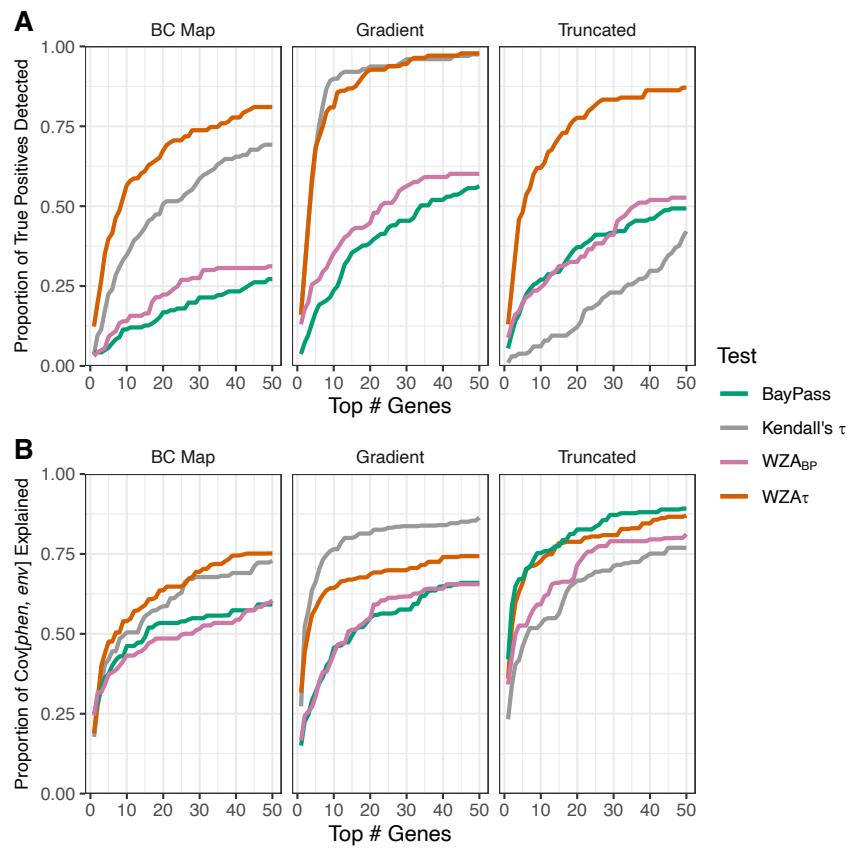
Effects of population structure correction

504 In each of the maps of environmental variation that we simulated, there was a strong
 505 correlation between environmental variables and gene flow. Environmental variation in
 506 each map was autocorrelated along a major axis: the diagonal axis from the bottom-left
 507 corner to the top right-corner in the case of the BC map (Figure 1B), the vertical axis in
 508 the case of the Gradient map (Figure 1C), and the top-right corner versus the rest of the

509 landscape in the case of the *Truncated* map (Figure 1D). There was also a strong
 510 pattern of isolation-by-distance in our simulated populations (Figure S1). These two
 511 factors may make it difficult to identify genes involved in local adaptation in GEA studies
 512 (Meirmans 2012).

513 We compared the performance of the WZA to a widely adopted method for performing
 514 GEA that corrects for the confounding effects of population structure, *BayPass* (Gautier
 515 2015). In all cases, WZA performed as well, or better than, *BayPass* (Figure 4). WZA
 516 performed much better than *BayPass* when selection was directional, but WZA was also
 517 significantly more likely to identify the genes underlying local adaptation with stabilizing
 518 selection.

519 Notably, even the single SNP analyses based on Kendall's τ in most cases
 520 outperformed *BayPass*, even though the Kendall's τ analysis did not adjust for spatial
 521 population structure. (The exception was the case with stabilizing selection on the
 522 *Truncated* map.) The discriminatory power of GEAs does not seem to be improved
 523 consistently by careful accounting of the underlying pattern of genetic structure.



524

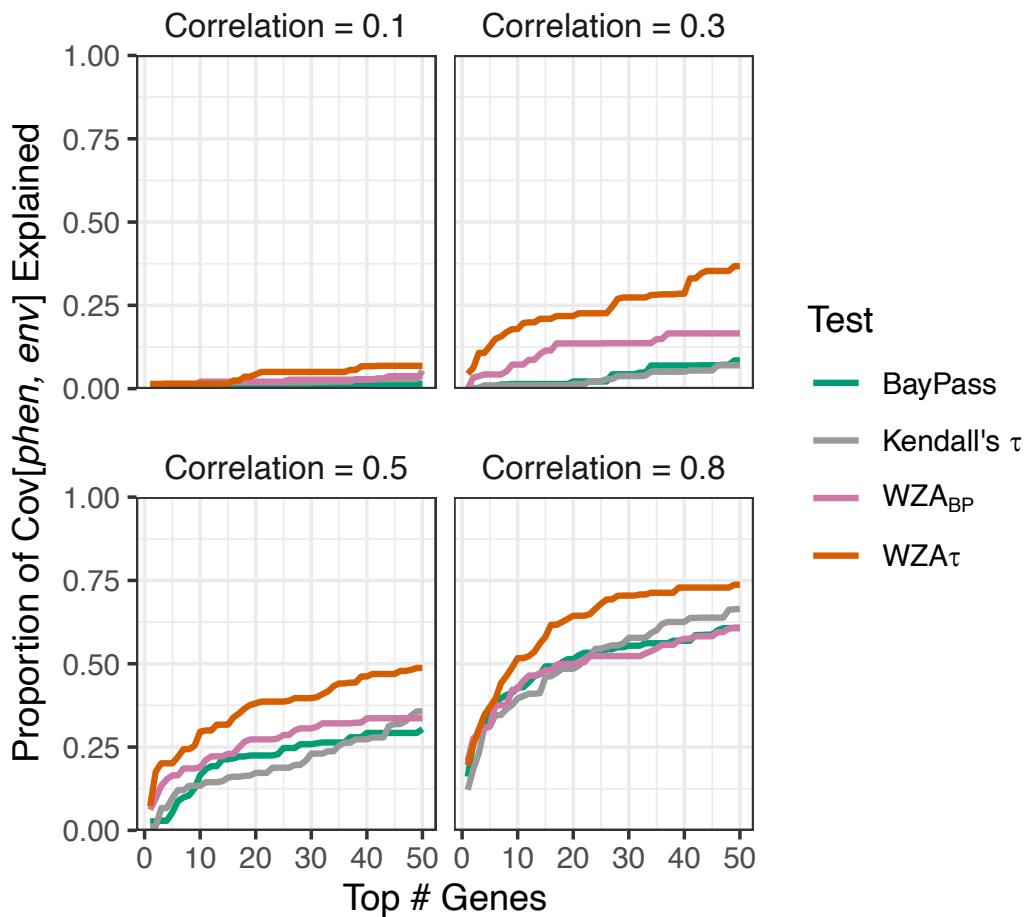
525 **Figure 4** The performance of population structure correction. A) Results for simulations
 526 modelling directional selection and b) results for simulations modelling stabilizing
 527 selection. Lines represent the mean of 20 simulation replicates.

528
529

530 *The performance of WZA when environmental variables are*
531 *weakly correlated with selection pressure*

532
533 In the previous section, we conducted GEA assuming perfect knowledge of the
534 phenotypic optima in each sampled deme. However, in empirical GEA analyses
535 researchers will be probably always be limited to studying environmental variables that
536 are imperfect proxies for historical selection patterns. Additionally, environmental
537 variables are often obtained via interpolation and/or may be measured with error. These
538 factors mean that the “E” in GEA will probably always be imperfectly correlated with
539 historical selection. The strength of that correlation will, of course, determine power in
540 GEA studies. Using the simulations modelling local adaptation on the *BC* map via
541 stabilizing selection, we compared the performance of WZA against the single-SNP
542 GEA methods when the measured environment is imperfectly correlated with the
543 phenotypic optima.

544



545

546 **Figure 5** The proportion of true positives recovered when the measured environment is
 547 imperfectly correlated with phenotypic optima. The correlation between environment
 548 and selection pressure is shown above each panel. Results are shown assuming the
 549 *BC Map* and the model of stabilising selection. Line indicate the means from 20
 550 simulation replicates, and each is based on samples of 50 individuals from each of 40
 551 demes.

552 We found that the WZA outperformed single SNP approaches (Kendall's τ or BayPass)
 553 when the measured environment was not perfectly correlated with phenotypic optima.
 554 We analyzed a sample of 40 demes from the population with 50 individuals taken in
 555 each location (Figure S2) but added random noise to the phenotypic optima from these
 556 locations to simulate environmental variables that were variably correlated with
 557 selection pressures and with which to conduct GEA. As might be expected, when the
 558 correlation between the measured environment and phenotypic optima was very weak
 559 (i.e., a correlation of 0.1), very few true positives were present in the top 50 genes under
 560 any of the methods we used, and those genes present only accounted for a small
 561 proportion of the covariance between phenotype and environment (Figure 5). With a
 562 correlation of 0.3 between the measured environment and true selection, WZA τ
 563 outperformed WZA_{BP} and the single-SNP approaches (Figure 5). With a correlation of
 564 0.5 or 0.8 between the measured environment and phenotypic optima WZA τ

565 outperformed all other methods, with only relatively small differences in performance
566 between WZA_{BP} and the single-SNP approaches. Overall, this result suggests that
567 WZA_T outperforms the single-SNP approaches when the measurement of the
568 environment is a poor proxy for historical selection.

569

570 The width of analysis windows and recombination rate variation

571 Random drift may cause genealogies in some regions of the genome to correlate with
572 environmental variables more than others. Many of the SNPs present in an analysis
573 window that consisted of genealogies that were highly correlated with the environment
574 may be highly significant in a GEA analysis, leading to a large WZA score. This effect
575 would lead to a larger variance in WZA scores for analysis windows that were present in
576 regions of low recombination. To demonstrate this, we down-sampled the tree-
577 sequences we recorded for our simulated populations to model analysis windows
578 present in low recombination regions and performed the WZA on the resulting data. As
579 expected, we found that the variance of the distribution
580 o

581 [REDACTED]
582 f WZA scores was greater when there was a lower recombination rate (Figure S9). This
583 is the same effect we described in a previous paper focusing on F_{ST} (Booker et al.
584 2020).

585

586 [REDACTED]

587 [REDACTED]

588 Application of the WZA to data from lodgepole pine

589 We re-analyzed a previously published (Yeaman et al. 2016) lodgepole pine (*Pinus*
590 *contorta*) dataset comparing the WZA to the top-candidate method, which had been
591 developed for the original study.

592 [REDACTED]

593 Overall, the WZA and top candidate statistic were broadly correlated and identified
594 many of the same genes as the most strongly associated loci, but also differed in
595 important ways. Across the lodgepole pine genome, there was a mean
596 WZA

597 [REDACTED]

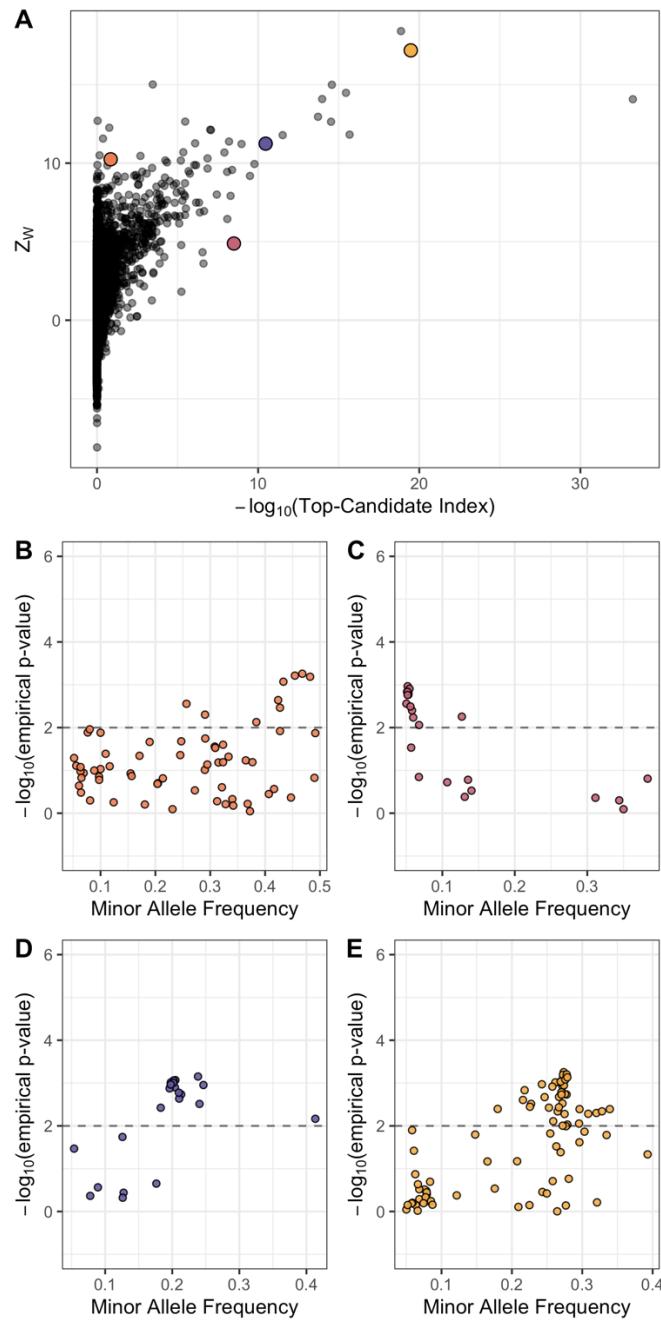
598 score of 0.013 with a standard deviation $\sigma = 1.67$, and a fat right-hand tail (Figure S11).

599 Figure 6A shows the relationship between WZA scores and the $-\log_{10}(p\text{-value})$ from
600 the top-candidate method, which were positively correlated (Kendall's $\tau = 0.245$, $p\text{-value}$
601 $< 10^{-16}$). When many of the SNPs in a gene had strongly associated statistics, both
602 methods would tend to yield high scores (Figure 6D-E). When there were many SNPs
603 with marginally significant empirical p -values (*i.e.* $0.05 < p < 0.10$) at relatively high
604 frequencies, the WZA method would tend to yield a high score but the top candidate
605 method would not (Figure 6B). By contrast, if the most strongly associated SNPs tended
606 to have low allele frequencies, the top candidate method would tend to yield a high
607 score but the WZA would not (Figure 6C). There were several genes that had
608 WZA

609

610 scores greater than 10 (approximately 6σ), but very modest top-candidate scores
611 (Figure 6A). Figure 6B shows that for one such region, there were several SNPs with
612 high mean allele frequency that have small p -values. This particular region had a high
613 score from the top-candidate method. Conversely, Figure 6C shows a region that only
614 had a $Z_W \approx 5$, but an extreme score from the top-candidate method. In this case, there
615 were numerous SNPs that passed the top-candidate outlier threshold, but they were
616 mostly at low allele frequency. Figures 6C-D show the relationship between allele
617 frequency and the empirical p -value for SNPs present in two genes that had extreme
618 scores from both the top-candidate method and the WZA.

619



622 **Figure 6** The WZA applied to GEA results on Lodgepole Pine for degree days below 0
 623 (DD0). A) Z_w scores compared to scores from the top-candidate method for each of the
 624 genes analyzed by Yeaman et al. (2016). Panels B-E show the results for $-\log_{10}(p\text{-}$
 625 values) for Spearman's ρ applied to individual SNPs against minor allele frequency
 626 (MAF) for the colored points in A. The dashed horizontal line in B-D indicates the

627 significance threshold used for the top-candidate method (i.e. $\boxed{}$ percentiles of GEA
628 $-\log_{10}$ (*p*-values) genome-wide).

629

630 **Discussion**

631 In this study, we have shown that combining information across linked sites in GEA
632 analyses is a potentially powerful way to identify genomic loci involved in local
633 adaptation. The method we propose, the WZA, was usually more powerful than looking
634 at individual sites in isolation, particularly when working with small samples or when the
635 environmental variation being analyzed is only weakly correlated with selection (Figures
636 3 and 5). The WZA outperformed the other window-based method we examined, the
637 top-candidate method (Figure 3). In a hypothetical world where one had perfect
638 knowledge of allele frequency variation across a species' range for all sites across the
639 genome, a single marker approach would likely be the best way to perform a GEA
640 analysis, as one would be able to determine the true correlation between genetic and
641 environmental variation for each site in the genome. Indeed, we found that when we
642 had perfect knowledge of allele frequencies in all locations, the SNP-based GEA always
643 outperformed or matched the WZA and top-candidate methods (Figure S13). However,
644 such a situation is unrealistic, and empirical GEA studies will likely always be limited to
645 finite samples from populations of interest. Thus, leveraging the correlated information
646 present among closely linked sites in GEA studies may provide a powerful method for
647 identifying the genetic basis of local adaptation.

648 Theoretical studies of local adaptation suggest that we should expect regions of the
649 genome subject to spatially varying selection pressures to exhibit elevated linkage
650 disequilibrium (LD) relative to the genomic background for a number of reasons. Under
651 local adaptation, alleles are subject to spatial fluctuation in the direction of selection. As
652 a locally adaptive allele spreads in the locations where it is beneficial, it may cause
653 some linked neutral variants to hitchhike along with it (Sakamoto and Innan 2019). LD
654 can be increased further as non-beneficial genetic variants introduced to local
655 populations via gene flow are removed by selection. This process can be thought of as
656 a local barrier to gene flow acting in proportion to the linkage with a selected site
657 (Barton and Bengtsson 1986). Beyond this hitchhiking signature, there is a selective
658 advantage for alleles that are involved in local adaptation to cluster together, particularly
659 in regions of low recombination (Rieseberg 2001; Noor et al. 2001; Kirkpatrick and
660 Barton 2006; Yeaman 2013). For example, in sunflowers and *Littorina* marine snails,
661 there is evidence that regions of suppressed recombination cause alleles involved in
662 local adaptation to be inherited together (Morales et al. 2019; Todesco et al. 2020). The
663 processes we have outlined are not mutually exclusive, but overall, genomic regions
664 containing strongly selected alleles that contribute to local adaptation may have
665 elevated LD and potentially exhibit GEA signals at multiple linked sites. Window-based
666 GEA scans can potentially take advantage of the LD that is induced by local adaptation,
667 aiding in the discovery of locally adaptive genetic variation.

668 The two window-based GEA methods we compared in this study, the WZA and the top-
669 candidate method of Yeaman et al. (2016), were fairly similar in power in some cases,
670 but WZA was often better (Figure 3). Moreover, there are philosophical reasons as to
671 why WZA should be preferred over the top-candidate method. Firstly, the top-candidate

672 method requires the use of an arbitrary significance threshold. This is undesirable,
673 however, because genuine genotype-environment correlations may be very weak and
674 GEA may simply be an underpowered approach to identify alleles that contribute to
675 local adaptation. If there were no detectable signal of local adaptation, ascribing
676 significance to a fraction of the genome may lead to false positives. Secondly, the top-
677 candidate method gives equal weight to all SNPs that have exceeded the significance
678 threshold. For example, with a threshold of $\alpha = 0.01$, genomic regions with only a single
679 outlier are treated in the same way whether that outlier has a *p*-value of 0.009 or 10^{-10} .
680 It is desirable to retain information about particularly strong outliers. It should be kept in
681 mind, however, that the WZA (and the top-candidate method for that matter) does not
682 explicitly test for local adaptation and only provides an indication of whether a particular
683 genomic region has a pattern that deviates from the genome-wide average. Indeed,
684 numerous processes other than local adaptation may cause excessive correlation
685 between environmental variables and allele frequencies in particular genomic regions.
686 For example, population expansions can cause allelic surfing, where regions of the
687 genome "surf" to high frequency at leading edge of an expanding population. Allelic
688 surfing can leave heterogeneous patterns of variation across a species range leaving
689 signals across the genome that may resemble local adaptation (Novembre and Di
690 Rienzo 2009; Klopstein, Currat, and Excoffier 2006).

691 When performing a genome-scan using a windowed approach a question that inevitably
692 arises is, how to choose the width of analysis windows? If analysis windows were too
693 narrow, there may be little benefit in using a windowed approach over a single-SNP
694 approach. In all the results presented above, 10,000bp analysis windows were used for
695 the WZA. We found that the performance of the WZA when analysis windows that were
696 narrower than 10,000bp was intermediate between the 10,000bp case and the single-
697 SNP approach (Figure S12). Of course, if analysis windows were too wide, the signal
698 of local adaptation may be diluted and the WZA would have little power. It seems like the
699 ideal width for analysis windows would be informed by the the pattern of recombination
700 rate variation, LD decay and SNP density across a species genome. In practice, it may
701 be useful to perform the WZA on groups of SNPs, such as genes as in the Yeaman et al
702 (2016) study. Future study is required to determine the optimal size for analysis
703 windows.

704 A striking result from our comparison of the various GEA methods we tested in this
705 study was the low power of *BayPass* compared to Kendall's τ (Figure 4). As mentioned
706 in the Introduction, Lotterhos (2019) obtained a similar result in a previous study, though
707 they had used Spearman's ρ rather than Kendall's τ . This presumably occurs because
708 genome-wide population genetic structure is oriented along a similar spatial axis as
709 adaptation, and the correction in *BayPass* therefore causes a reduction in the signal of
710 association at genes involved in adaptation. In such cases, the use of simple rank
711 correlations such as Spearman's ρ or Kendall's τ , which assume that all demes are
712 independent, may often yield a skewed distribution of *p*-values. Such a distribution
713 would lead to a large number of false positives if a standard significance threshold is
714 used (Meirmans 2012). Here, we avoid standard significance testing, and instead make

715 use of an attractive quality of the distribution of *p*-values: SNPs in regions of the
716 genome that contribute to adaptation tend to have extreme *p*-values, relative to the
717 genome-wide distribution. By converting them to empirical *p*-values, we retain the
718 information contained in the rank-order of *p*-values, but reduce the inflation of their
719 magnitude, which increases the power of the test (Figure S12). While the empirical *p*-
720 value approach may partially and indirectly correct for false positives due to population
721 structure genome-wide, it loses information contained in the raw *p*-value that represents
722 the deviation of the data from the null model for our summary statistic of interest. A GEA
723 approach that produced parametric *p*-values that was adequately controlled for population
724 structure may provide a more powerful input statistic to the WZA.

725 Perhaps more striking was how underpowered these GEA methods were at identifying
726 the genes involved in local adaptation. In our simulations, around 6 locally adaptive
727 genes established in each replicate in each of the cases we tested. When analyzing our
728 simulations, we examined the true positives present in the top 1, 2, 3, ..., 50 genes, but
729 in most cases, the proportion of all true positives identified did not reach 1.0 (Figures 3-
730 5), indicating high false discovery rates. Each simulation replicate included 1,000 genes,
731 so the top 50 represents the 95th percentile of the genome-wide distribution. Examining
732 the upper percentiles of the empirical distribution of GEA scores is an approach taken in
733 empirical analyses (e.g. Shi et al. 2021; Leigh et al. 2021), though it would perhaps be
734 preferable to have a threshold that was applied to an appropriate null distribution. One
735 drawback of the WZA is that since the distribution of WZA scores was non-normal even
736 under neutrality (Figure 2), we cannot compute a parametric *p*-value for each analysis
737 window tested. *BayPass*, on the other hand, returns a Bayes factor for each analyzed
738 SNP, the log-transformed ratio of the likelihoods under the alternate and null
739 hypotheses. A general rule of thumb for the interpretation of Bayes factors is that BFs >
740 20 are considered strong evidence against the null hypothesis (i.e., Jeffrey's rule).
741 Overall, applying a stringent Bayes factor threshold to *BayPass* may result in a test with
742 low false positive rates, the WZA may provide a more sensitive test at the cost of
743 specificity.

744 Ultimately, performing GEA analyses using analysis windows is an attempt to leverage
745 information from closely linked sites. With the advent of methods for reconstructing
746 ancestral recombination graphs from population genomic data (Hejase et al. 2020),
747 perhaps a GEA method could be developed that explicitly analyzes inferred genealogies
748 rather than individual markers in a manner similar to regression of phenotypes on
749 genealogies proposed by Ralph et al. (2020). Such a method would require large
750 numbers of individuals with phased genome sequences, which may now be feasible
751 given recent technological advances (Meier et al. 2021).

752 In conclusion, theoretical models of local adaptation suggest that we should expect
753 elevated LD in genomic regions subject to spatially varying selection pressures. For that
754 reasons, GEA analyses may gain power by making use of information encoded in
755 patterns of tightly linked genetic variation. The method we proposed in this study, the
756 WZA, outperforms single-SNP approaches in a range of settings so provides
757 researchers with a powerful tool to characterize the genetic basis of local adaptation in
758 population and landscape genomic studies.

759 **Acknowledgements**

760 Thanks to Pooja Singh for many helpful discussions, to Tongli Wang for help with BC
761 climate data and to Simon Kapitza for help with wrangling raster files. Thanks to Finlay
762 Booker for moral support throughout the course of this project. TRB is supported by
763 funding from Genome Canada, Genome Alberta and NSERC Discovery Grants
764 awarded to MCW and SY. SY is supported by an AIHS research chair and NSERC
765 Discovery Grant. MCW is supported by an NSERC Discovery Grant. Computational
766 Support was provided by Compute Canada. This study is part of the CoAdapTree
767 project which is funded by Genome Canada (241REF), Genome BC and 16 other
768 sponsors (<http://coadaptree.forestry.ubc.ca/sponsors/>).
769

770

771 **Bibliography**

- 772 Aitken, Sally N, and Michael C Whitlock. 2013. "Assisted Gene Flow to Facilitate Local
773 Adaptation to Climate Change." *Annu. Rev. Ecol. Evol. Syst.* 44 (1): 367–88.
- 774 Barton, Nick, and Bengt Olle Bengtsson. 1986. "The barrier to genetic exchange
775 between hybridising populations." *Heredity* 57 (3): 357–76.
776 <https://doi.org/10.1038/hdy.1986.135>.
- 777 Bhatia, Gaurav, Nick Patterson, Sriram Sankararaman, and Alkes L. Price. 2013.
778 "Estimating and interpreting FST: The impact of rare variants." *Genome Research* 23
779 (9): 1514–21. <https://doi.org/10.1101/gr.154831.113>.
- 780 Bontrager, M., C. D. Muir, C. Mahony, D. E. Gamble, R. M. Germain, A. L. Hargreaves,
781 E. J. Kleynhans, K. A. Thompson, and A. L. Angert. 2020. "Climate warming weakens
782 local adaptation." bioRxiv. <https://doi.org/10.1101/2020.11.01.364349>.
- 783 Booker, Tom R., Sam Yeaman, and Michael C. Whitlock. 2020. "Variation in
784 recombination rate affects detection of outliers in genome scans under neutrality."
785 *Molecular Ecology* 29 (22): 4274–9. <https://doi.org/10.1111/mec.15501>.
- 786 Charlesworth, B, and D Charlesworth. 2010. *Elements of Evolutionary Genetics*. Book.
787 Greenwood Village, Colorado: Roberts & Company.
- 788 Coop, Graham, David Witonsky, Anna Di Rienzo, and Jonathan K. Pritchard. 2010.
789 "Using environmental correlations to identify loci underlying local adaptation." *Genetics*
790 185 (4): 1411–23. <https://doi.org/10.1534/genetics.110.114819>.
- 791 Forester, Brenna R., Matthew R. Jones, Stéphane Joost, Erin L. Landguth, and Jesse
792 R. Lasky. 2016. "Detecting spatial genetic signatures of local adaptation in
793 heterogeneous landscapes." *Molecular Ecology* 25 (1): 104–20.
794 <https://doi.org/10.1111/mec.13476>.
- 795 Forester, Brenna R., Jesse R. Lasky, Helene H. Wagner, and Dean L. Urban. 2018.
796 "Comparing methods for detecting multilocus adaptation with multivariate genotype-
797 environment associations." *Molecular Ecology* 27 (9): 2215–33.
798 <https://doi.org/10.1111/mec.14584>.
- 799 Fritchot, Eric, and Olivier François. 2015. "LEA: An R package for landscape and
800 ecological association studies." Edited by Brian O'Meara. *Methods in Ecology and*
801 *Evolution* 6 (8): 925–29. <https://doi.org/10.1111/2041-210X.12382>.
- 802 Fritchot, Eric, Sean D. Schoville, Guillaume Bouchard, and Olivier François. 2013.
803 "Testing for Associations between Loci and Environmental Gradients Using Latent

- 804 Factor Mixed Models." *Molecular Biology and Evolution* 30 (7): 1687–99.
805 <https://doi.org/10.1093/molbev/mst063>.
- 806 Gautier, Mathieu. 2015. "Genome-wide scan for adaptive divergence and association
807 with population-specific covariates." *Genetics* 201 (4): 1555–79.
808 <https://doi.org/10.1534/genetics.115.181453>.
- 809 Haldane, J. B. S. 1948. "The theory of a cline." *Journal of Genetics* 48 (3): 277–84.
810 <https://doi.org/10.1007/BF02986626>.
- 811 Haller, Benjamin C, Jared Galloway, Jerome Kelleher, Philipp W Messer, and Peter L
812 Ralph. 2019. "Tree-sequence recording in SLiM opens new horizons for forward-time
813 simulation of whole genomes." *Mol. Ecol. Resour.* 19 (2): 552–66.
- 814 Hancock, Angela M., Benjamin Brachi, Nathalie Faure, Matthew W. Horton, Lucien B.
815 Jarymowycz, F. Gianluca Sperone, Chris Toomajian, Fabrice Roux, and Joy Bergelson.
816 2011. "Adaptation to climate across the *Arabidopsis thaliana* genome." *Science* 334
817 (6052): 83–86. <https://doi.org/10.1126/science.1209244>.
- 818 Hejase, Hussein A., Noah Dukler, and Adam Siepel. 2020. "From Summary Statistics to
819 Gene Trees: Methods for Inferring Positive Selection." Elsevier Ltd.
820 <https://doi.org/10.1016/j.tig.2019.12.008>.
- 821 Hereford, Joe. 2009. "A quantitative survey of local adaptation and fitness trade-offs."
822 The University of Chicago Press. <https://doi.org/10.1086/597611>.
- 823 Hoban, Sean, Joanna L. Kelley, Katie E. Lotterhos, Michael F. Antolin, Gideon
824 Bradburd, David B. Lowry, Mary L. Poss, Laura K. Reed, Andrew Storfer, and Michael
825 C. Whitlock. 2016. "Finding the genomic basis of local adaptation: Pitfalls, practical
826 solutions, and future directions." *American Naturalist* 188 (4): 379–97.
827 <https://doi.org/10.1086/688018>.
- 828 Kirkpatrick, Mark, and Nick Barton. 2006. "Chromosome inversions, local adaptation
829 and speciation." *Genetics* 173 (1): 419–34. <https://doi.org/10.1534/genetics.105.047985>.
- 830 Klopfstein, Seraina, Mathias Currat, and Laurent Excoffier. 2006. "The fate of mutations
831 surfing on the wave of a range expansion." *Molecular Biology and Evolution*.
832 <https://doi.org/10.1093/molbev/msj057>.
- 833 Legendre, P., and L. Legendre. 2012. *Numerical Ecology*, Volume 24. 3rd Englisch.
834 Elsevier. <https://www.elsevier.com/books/numerical-ecology/legendre/978-0-444-53868-0>.
- 836 Lotterhos, Katie E. 2019. "The Effect of Neutral Recombination Variation on Genome
837 Scans for Selection." *G3* 9 (6): 1851–67.
- 838 Meirmans, Patrick G. 2012. "The trouble with isolation by distance." *Molecular Ecology*
839 21 (12): 2839–46. <https://doi.org/10.1111/j.1365-294X.2012.05578.x>.

- 840 Mimura, M., and S. N. Aitken. 2007. "Adaptive gradients and isolation-by-distance with
841 postglacial migration in *Picea sitchensis*." *Heredity* 99 (2): 224–32.
842 <https://doi.org/10.1038/sj.hdy.6800987>.
- 843 Morales, Hernán E., Rui Faria, Kerstin Johannesson, Tomas Larsson, Marina Panova,
844 Anja M. Westram, and Roger K. Butlin. 2019. "Genomic architecture of parallel
845 ecological divergence: Beyond a single environmental contrast." *Science Advances* 5
846 (12): eaav9963. <https://doi.org/10.1126/sciadv.aav9963>.
- 847 Mosca, Elena, Santiago C. González-Martínez, and David B. Neale. 2014.
848 "Environmental versus geographical determinants of genetic structure in two subalpine
849 conifers." *New Phytologist* 201 (1): 180–92. <https://doi.org/10.1111/nph.12476>.
- 850 Noor, M. A. F., K. L. Gratos, L. A. Bertucci, and J. Reiland. 2001. "Chromosomal
851 inversions and the reproductive isolation of species." *Proceedings of the National
852 Academy of Sciences of the United States of America* 98 (21): 12084–8.
853 <https://doi.org/10.1073/pnas.221274498>.
- 854 Novembre, John, and Anna Di Rienzo. 2009. "Spatial patterns of variation due to natural
855 selection in humans." *Nat Rev Genet.* <https://doi.org/10.1038/nrg2632>.
- 856 Pavly, N., M. C. Namroud, F. Gagnon, N. Isabel, and J. Bousquet. 2012. "The
857 heterogeneous levels of linkage disequilibrium in white spruce genes and comparative
858 analysis with other conifers." *Heredity* 108 (3): 273–84.
859 <https://doi.org/10.1038/hdy.2011.72>.
- 860 Ralph, Peter, Kevin Thornton, and Jerome Kelleher. 2020. "Efficiently summarizing
861 relationships in large samples: A general duality between statistics of genealogies and
862 genomes." *Genetics* 215 (3): 779–97. <https://doi.org/10.1534/genetics.120.303253>.
- 863 Rieseberg, Loren H. 2001. "Chromosomal rearrangements and speciation." Elsevier.
864 [https://doi.org/10.1016/S0169-5347\(01\)02187-5](https://doi.org/10.1016/S0169-5347(01)02187-5).
- 865 Sakamoto, Takahiro, and Hideki Innan. 2019. "The evolutionary dynamics of a genetic
866 barrier to gene flow: From the establishment to the emergence of a peak of divergence."
867 *Genetics* 212 (4): 1383–98. <https://doi.org/10.1534/genetics.119.302311>.
- 868 Schlötterer, Christian, Raymond Tobler, Robert Kofler, and Viola Nolte. 2014.
869 "Sequencing pools of individuals-mining genome-wide polymorphism data without big
870 funding." Nature Publishing Group. <https://doi.org/10.1038/nrg3803>.
- 871 Stapley, Jessica, Philine G. D. Feulner, Susan E. Johnston, Anna W. Santure, and
872 Carole M. Smadja. 2017. "Variation in recombination frequency and distribution across
873 eukaryotes: Patterns and processes." *Philosophical Transactions of the Royal Society
874 B: Biological Sciences* 372 (1736). <https://doi.org/10.1098/rstb.2016.0455>.
- 875 Todesco, Marco, Gregory L. Owens, Natalia Bercovich, Jean Sébastien Légaré,
876 Shaghayegh Soudi, Dylan O. Burge, Kaichi Huang, et al. 2020. "Massive haplotypes

- 877 underlie ecotypic differentiation in sunflowers." *Nature* 584 (7822): 602–7.
878 <https://doi.org/10.1038/s41586-020-2467-6>.
- 879 Walsh, B., and M. Lynch. 2018. *Evolution and Selection of Quantitative Traits*. Oxford
880 University Press. <https://global.oup.com/academic/product/evolution-and-selection-of->
881 [quantitative-trait-9780198830870?cc=ca&lang=en&](#).
- 882 Wang, Tongli, Andreas Hamann, Dave Spittlehouse, and Carlos Carroll. 2016. "Locally
883 Downscaled and Spatially Customizable Climate Data for Historical and Future Periods
884 for North America." Edited by Inés Álvarez. *PLOS ONE* 11 (6): e0156720.
885 <https://doi.org/10.1371/journal.pone.0156720>.
- 886 Weir, Bruce S, and C Clark Cockerham. 1984. "Estimating F-statistics for the analysis of
887 population structure." *Evolution* 38 (6): 1358–70.
- 888 Whitlock, M. C. 2005. "Combining probability from independent tests: the weighted Z-
889 method is superior to Fisher's approach." *Journal of Evolutionary Biology* 18 (5): 1368–
890 73. <https://doi.org/10.1111/j.1420-9101.2005.00917.x>.
- 891 Yeaman, Sam. 2013. "Genomic rearrangements and the evolution of clusters of locally
892 adaptive loci." *Proceedings of the National Academy of Sciences of the United States of
893 America* 110 (19): E1743–E1751. <https://doi.org/10.1073/pnas.1219381110>.
- 894 Yeaman, Sam, Aleeza C. Gerstein, Kathryn A. Hodgins, and Michael C. Whitlock. 2018.
895 "Quantifying how constraints limit the diversity of viable routes to adaptation." *PLoS
896 Genetics* 14 (10): e1007717. <https://doi.org/10.1371/journal.pgen.1007717>.
- 897 Yeaman, Sam, Kathryn A. Hodgins, Katie E. Lotterhos, Haktan Suren, Simon Nadeau,
898 Jon C. Degner, Kristin A. Nurkowski, et al. 2016. "Convergent local adaptation to
899 climate in distantly related conifers." *Science* 353 (6306): 1431–3.
900 <https://doi.org/10.1126/science.aaf7812>.
- 901 Zhou, Xiang, Peter Carbonetto, and Matthew Stephens. 2013. "Polygenic Modeling with
902 Bayesian Sparse Linear Mixed Models." *PLoS Genetics* 9 (2): 1003264.
903 <https://doi.org/10.1371/journal.pgen.1003264>.
- 904
- 905
- 906

907

908 **Appendix**

909 **Parametrizing simulations of local adaptation**

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

919 **Mutation rate**

920 We set the neutral mutation rate such that there would be an average of around 20
921 SNPs in each gene with a minor allele frequency threshold greater than 0.05. This
922 number was motivated by the average number of SNPs per gene in the lodgepole pine
923 dataset described by (Yeaman et al. 2016). We found that a neutral mutation rate

924 (μ_{neu}) of 10^{-8} in our simulations achieved an average of 23.3. Note that this μ_{neu} gave
925 a very low population-mutation rate within demes, $4N_d\mu_{neu} = 4.0 \times 10^{-8}$

926 There are no estimates available of the mutation rate for locally adaptive alleles. As
927 such, we had no empirical estimates to base our simulations on. Instead, we opted to
928 use mutation rates that resulted in multiple locally beneficial alleles establishing in our

929 simulations. For directional selection, we found that a mutation rate of $\mu_{alpha} = 3 \times 10^{-7}$
930 resulted in an average of 6.XX locally adaptive genes establishing. For stabilizing

931 selection, a mutation rate of $\mu_{alpha} = 1 \times 10^{-8}$ resulted in similar numbers of genes
932 establishing. Note that in our model of directional selection, only a single nucleotide in
933 each of 12 genes could mutate to a locally beneficial allele. In the case of stabilizing
934 selection, all 1,000bp in the simulated gene could give rise to mutations that affected
935 phenotype.

936 **Recombination rates**

937 We based our choice of recombination rate on patterns of LD decay reported for
938 conifers. The pattern of LD decay in a panmictic population can be predicted by the

939 population-scaled recombination parameter ($\rho = 4N_e r$) Charlesworth and Charlesworth
940 2010), but the pattern of LD decay in structured populations is less well described. In
941 conifers, LD decays very rapidly in conifers and $\rho \approx 0.005$ has been estimated (Pavy et
942 al. 2012). However, per basepair recombination rates (r) in conifers are extremely low,
943 estimated to be on the order of 0.05 cM/Mbp - more than 10^x lower than the average for
944 humans (Stapley et al. 2017). This implies a very large effective population size of
945 roughly $\frac{0.005}{4 \times 0.05 \times 10^{-8}} = 2.5 \times 10^6$, much larger than is feasible to simulate. To achieve a
946 similar number of recombination events through time in our simulated populations, we
947 needed to increase r above what has been empirically estimated. We chose a
948 recombination rate that gave us a pattern of LD decay that was similar to what has been
949 observed in conifers. We found that a per base pair recombination $r = 1 \times 10^{-7}$ (i.e.
950 roughly $200 \times$ greater than in natural populations) gave a pattern of LD in our simulated
951 populations that was similar to what has been reported for conifers.

952 Selection coefficients

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

960 When modelling directional selection, our simulations contained 12 loci that could
961 mutate to generate a locally beneficial allele. The phenotypic optima that we simulated
962 ranged from -7 to 7 and we modelled selection on a locus as $1 + s_a \theta$ for a homozygote
963 and $1 + h s_a \theta$ for a heterozygote, where s_a is the selection coefficient, θ is the
964 phenotypic optimum and h is the dominance coefficient. With a selection coefficient of
965 $s_a = 0.003$, the maximum relative fitness was $(1 + 7 \times s_a)^{12} = 1.28$ for an individual
966 homozygous for all locally beneficial alleles. An individual homozygous for those alleles,
967 but in the oppositely selected environment (i.e. present in the wrong deme) had a
968 fitness of $(1 - 7 \times s_a)^{12} = 0.775$. Thus, there would be approximately 40% difference in

970 fitness between well locally adapted individuals at home versus away in the most
971 extreme case.

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

975 **Migration rate**

976 We wanted to model populations with F_{ST} across the metapopulation of approximately
977 0.05, as has been reported for widely distributed conifer species such as lodgepole pine
978 and interior spruce (Yeaman et al. 2016). For the stepping-stone simulations, we chose

979 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

980 model, we used the analytical formulae given in the main text to set m to achieve a
981 mean F_{ST} of 0.03.

982

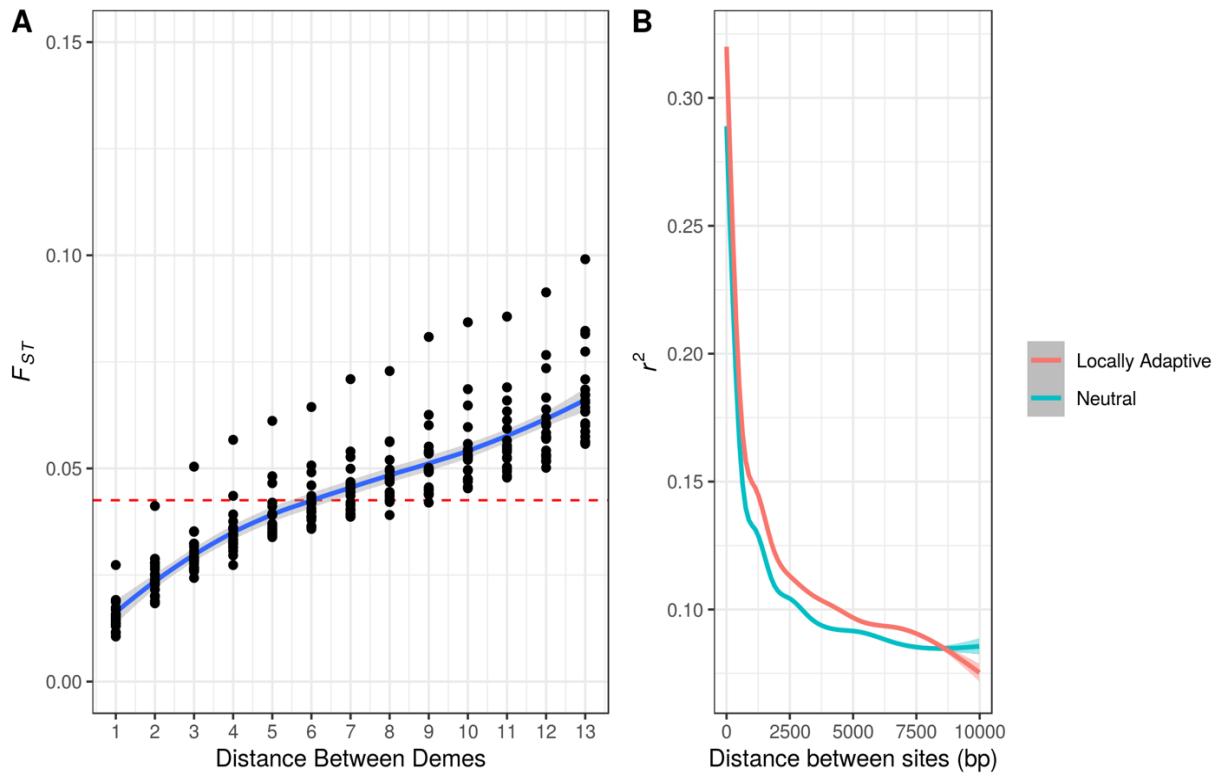
983 **Table S1** Population genetic parameters of a hypothetical organism, and how they are
 984 scaled in the simulations. The meta-population inhabits a 14×14 2-dimensional
 985 stepping stone. Parameters are shown for a population with 12 loci subject to directional
 986 selection.

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)	2.00×10^{-9}	$4N_d r = 0.00004$	1×10^{-7}
Selection coefficient (s_a)	0.0001	$2N_d s_a = 0.6$	0.003
Migration rate (m)	7.35×10^{-4}	$2N_d m = 7.5$	0.0375
Neutral mutation rate (μ_{neu})	2×10^{-10}	$4N_e \mu_{neu} = 0.000004$	10^{-8}
Functional mutation rate (μ_α)	2×10^{-9}	$4N_e \mu_\alpha = 0.00004$	3×10^{-7}

987

988

989

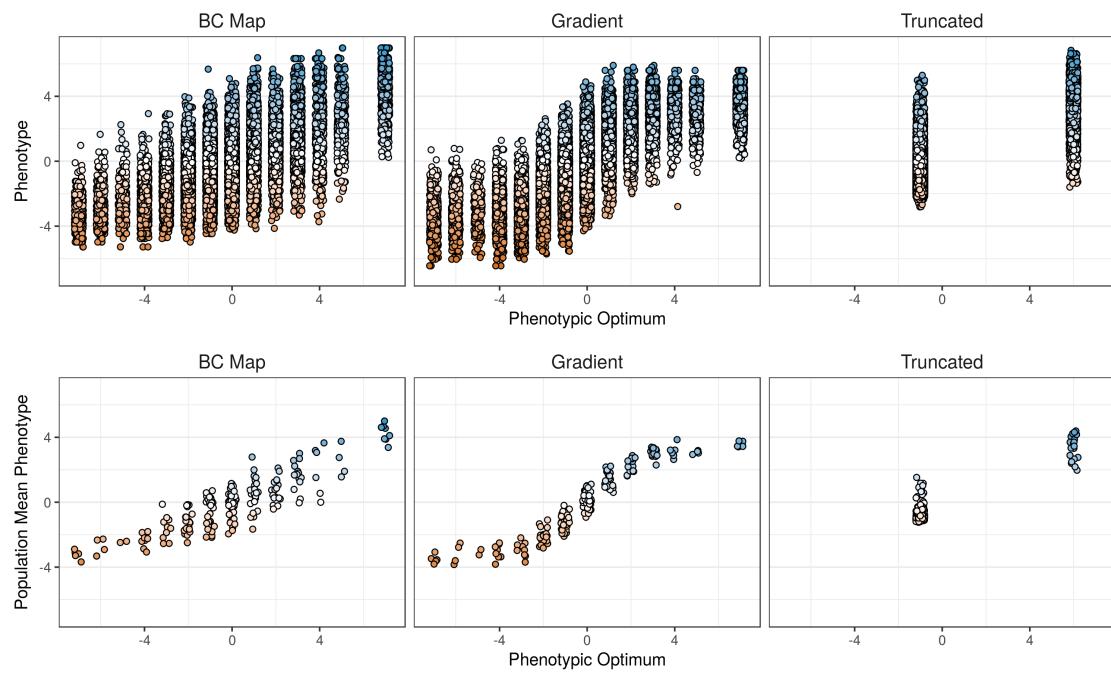


990

991 **Figure S1** Summary statistics from neutral simulations. A) F_{ST} between pairs of demes in
 992 stepping-stone populations. The average across replicates is 0.042. B) LOESS
 993 smoothed LD, as measured by r^2 , between pairs of SNPs in genes that are either
 994 evolving neutrally or locally adapting as indicated by the color. Smoothing was
 995 performed using the ggplot2 package in R.

996

997

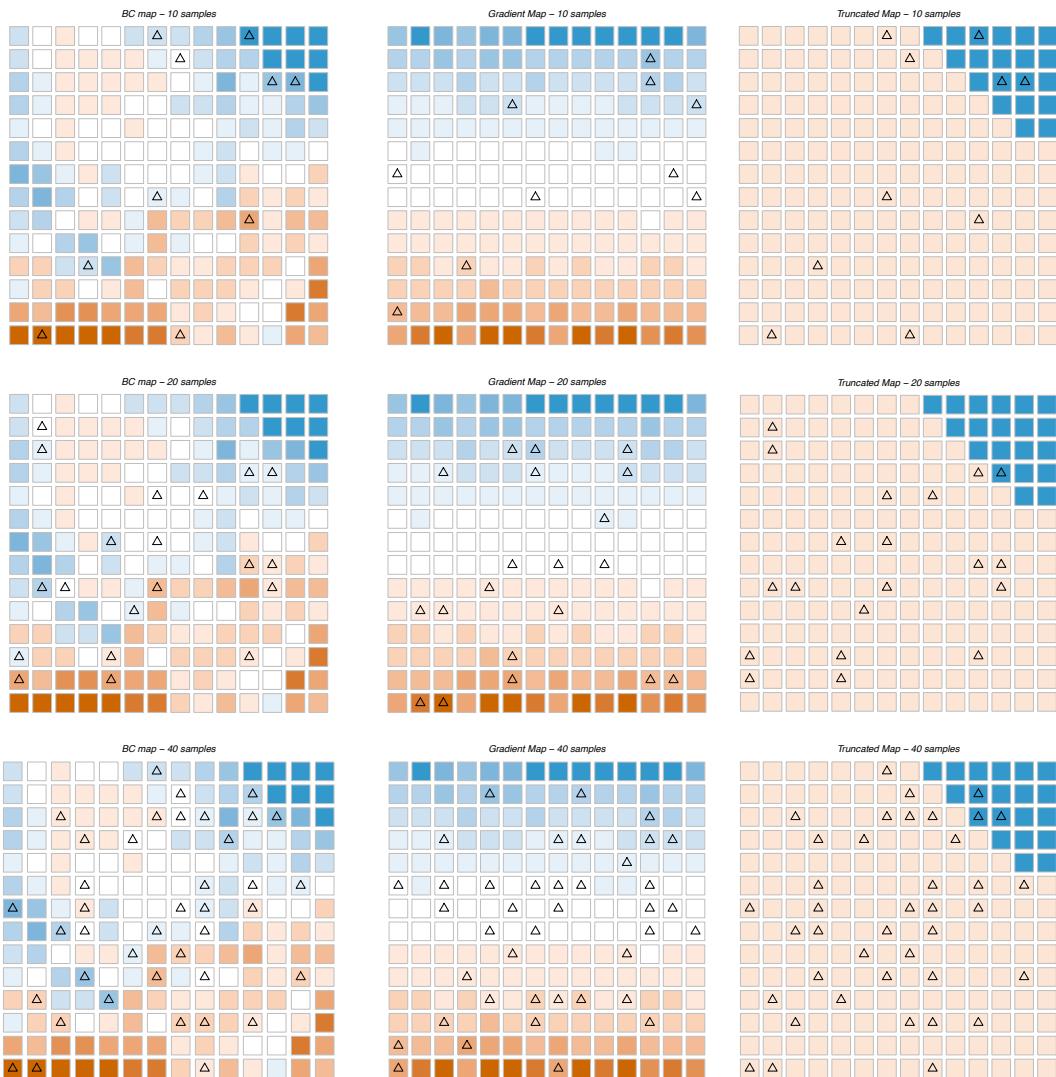


998

999 **Figure S2** Individual and population mean phenotypes observed in representative
1000 simulations for each of the environment maps simulated. A small amount of horizontal
1001 jitter was added to points for visualization purposes. Colors are for visualization
1002 purposes only.

1003

1004

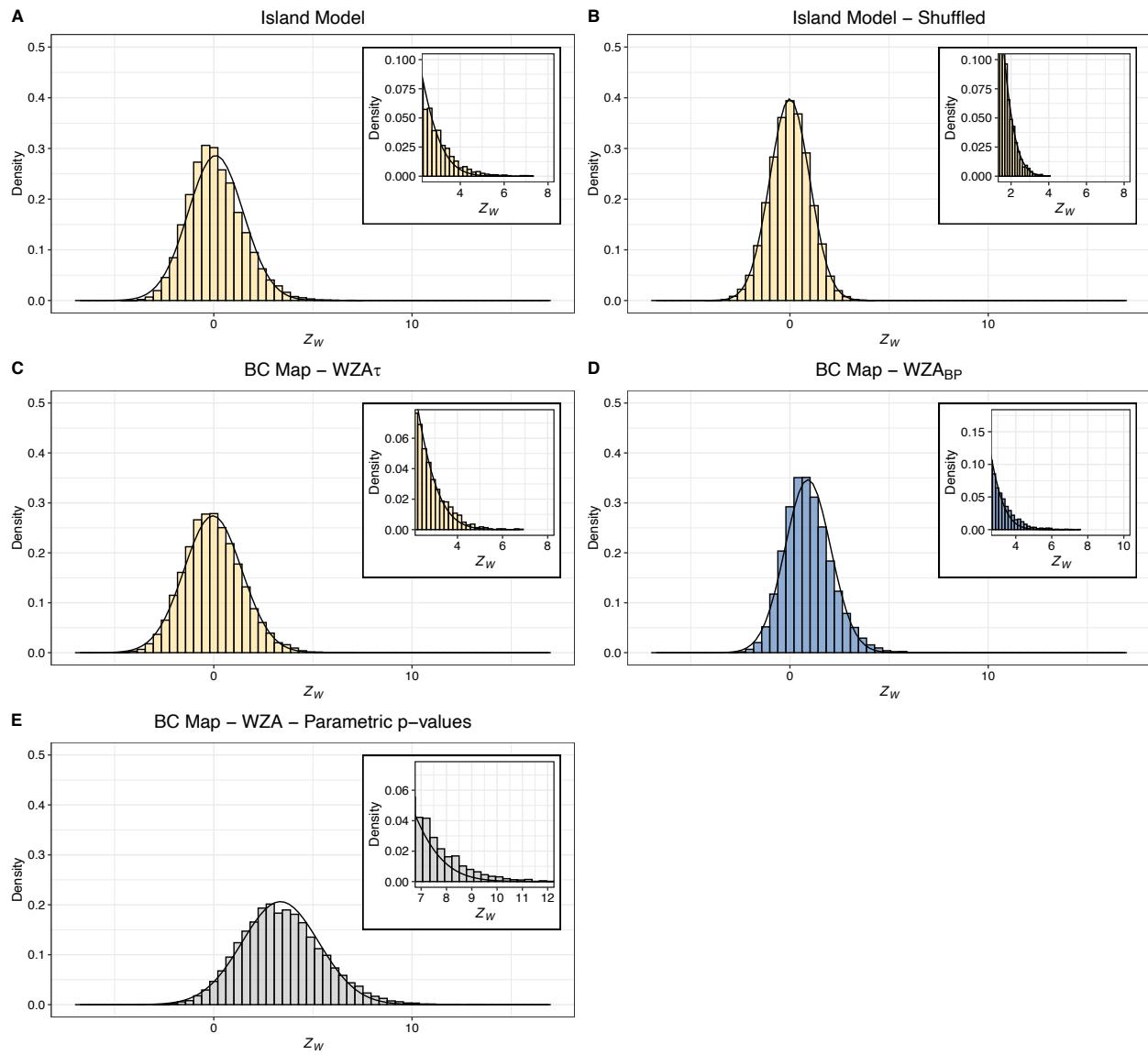


1005

1006 **Figure S3** Locations of sampled demes on the maps of environmental variation we
1007 assumed in the simulations. Triangles indicate the locations where individuals were
1008 sampled in each case. Colors represent the optimal phenotype in each population the
1009 same as Figure 1 in the main text.

1010

1011

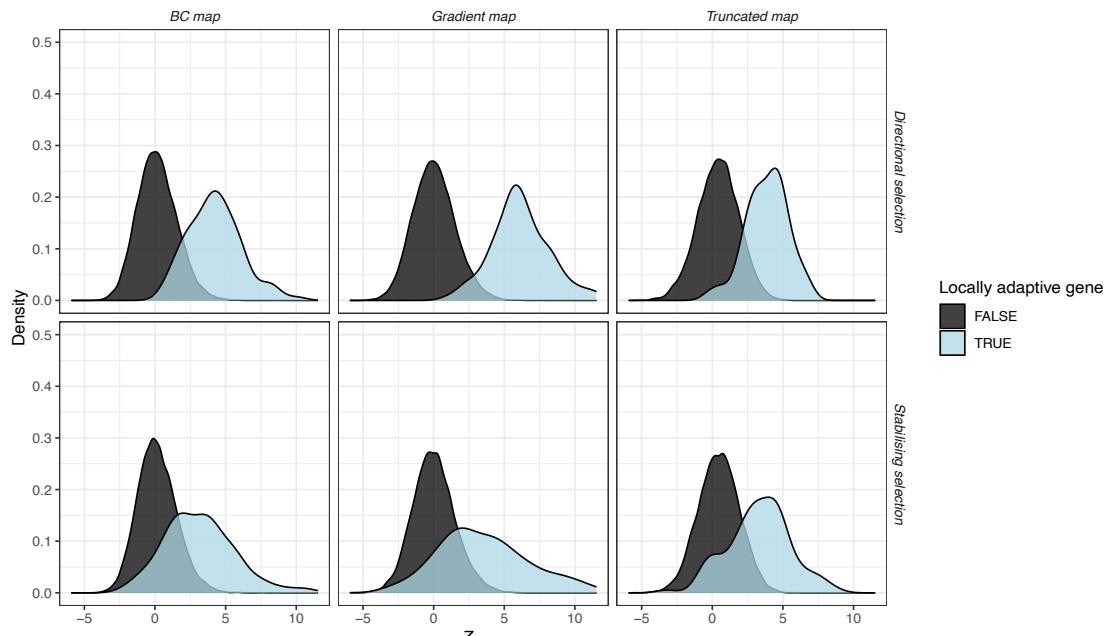


1012

1013 **Figure S4** The distribution of WZA scores from neutral simulations with details of the
 1014 right tail in the insets. Overlaid on each panel is the normal distribution fitted to each
 1015 dataset. In all cases, results from 20 simulation replicates are plotted together.

1016

1017

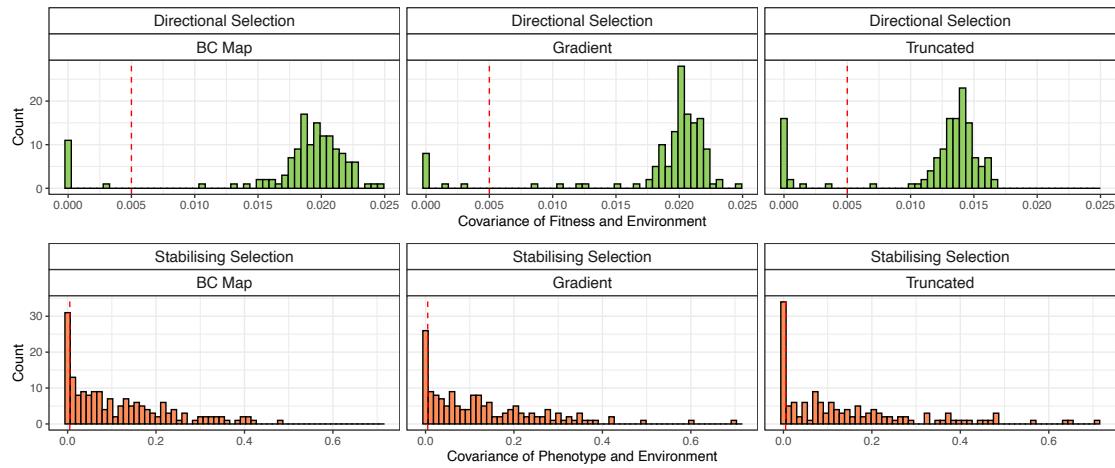


1018

1019 **Figure S5** The distribution of WZA scores from simulations of local adaptation. Note,
1020 the plot does not indicate the relative frequency of genes that are or are not locally
1021 adaptive.

1022

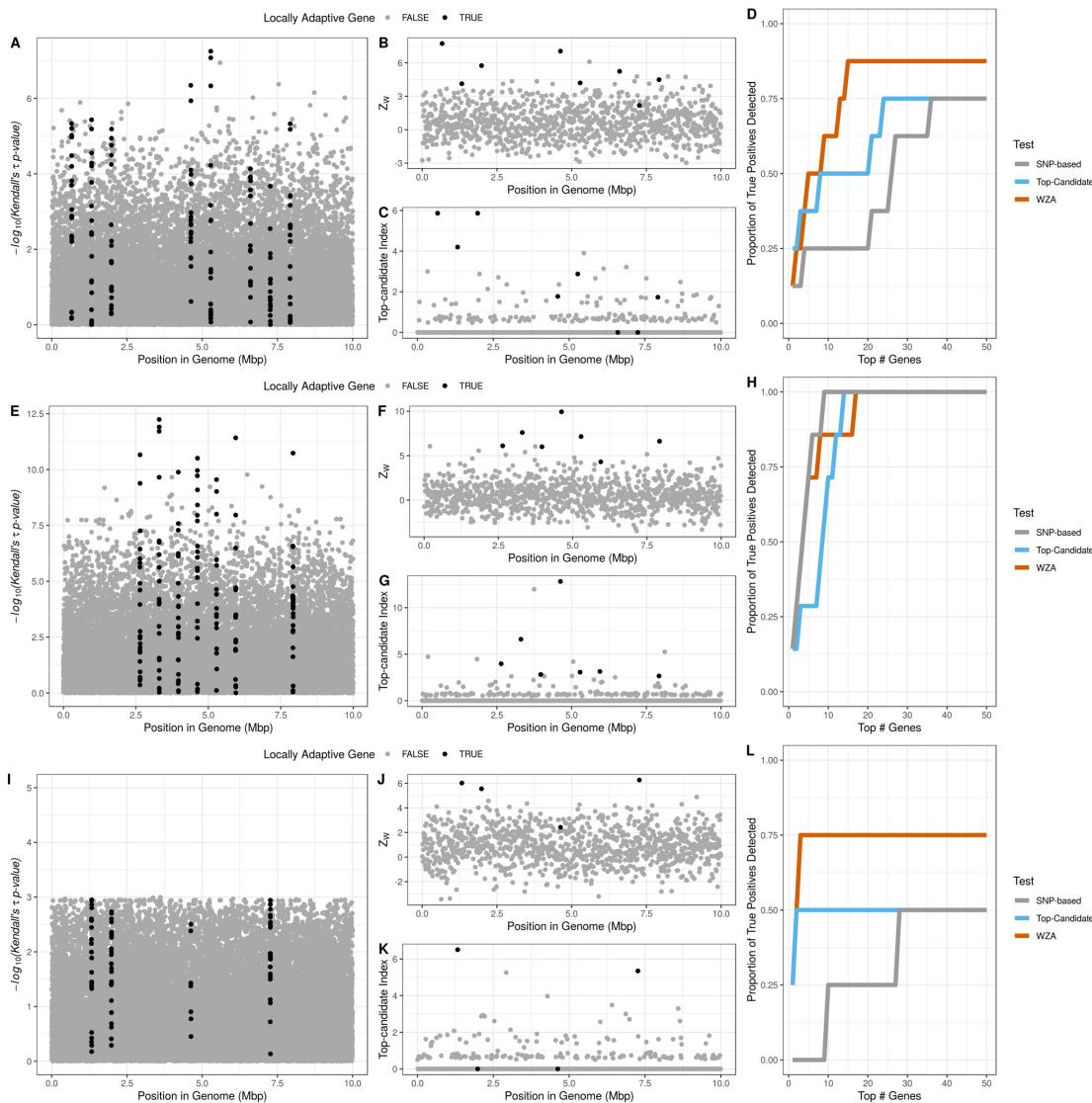
1023



1024

1025 **Figure S6** The effect size distribution from simulations of local adaptation. The vertical
1026 line indicates the threshold we applied to the simulated data to classify genes as locally
1027 adaptive or not.

1028

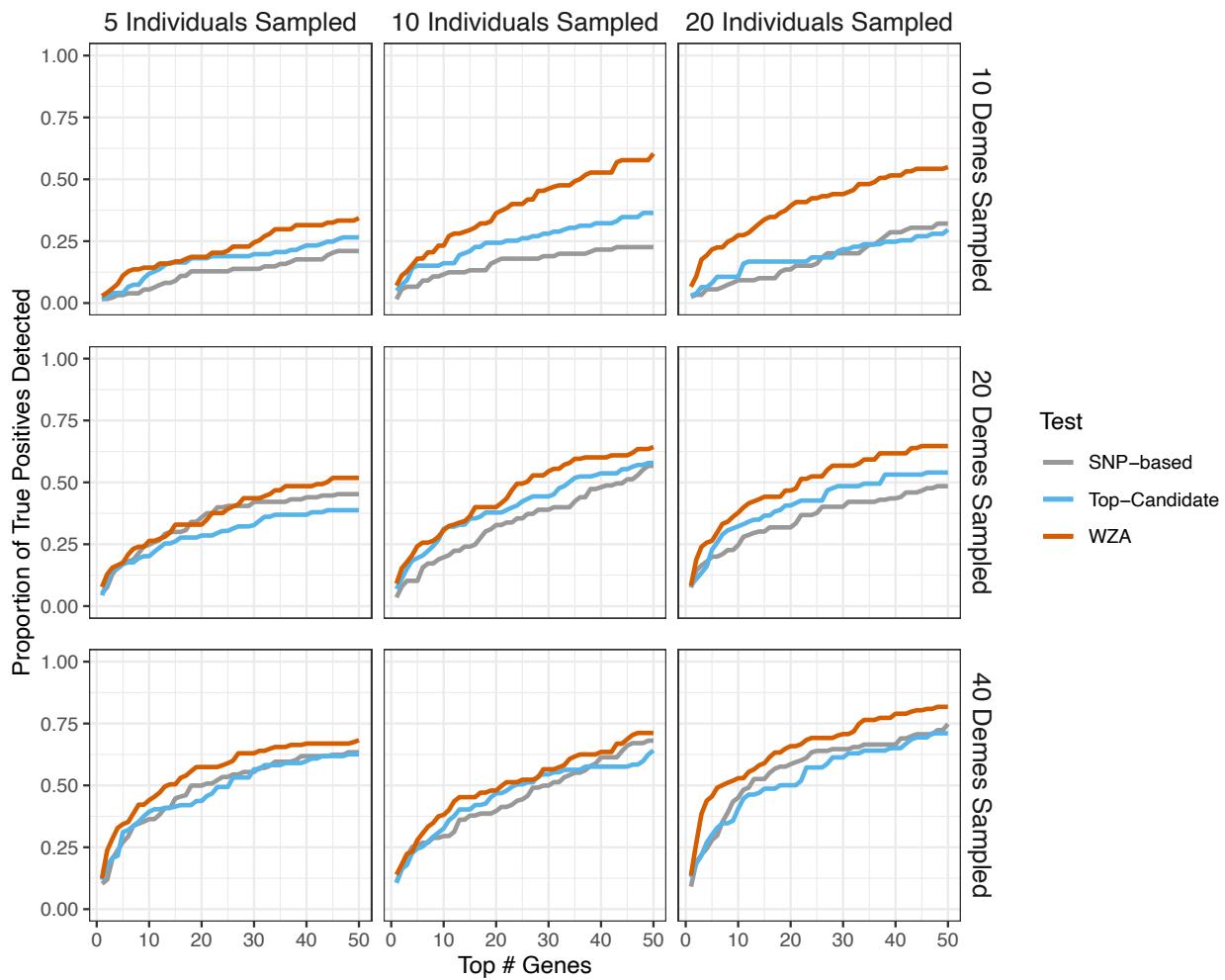


1031 **Figure S7** Plots demonstrating the genomic landscape of genotype-environment
 1032 correlations for a single replicate for each of the three maps of environmental variation
 1033 we simulated. From top to bottom, the three rows correspond to the *BC Map* (panels A-
 1034 D), the gradient map (panels E-H) and the truncated map (panels I-L), respectively. The
 1035 leftmost panel in each row shows the Manhattan plot of $-\log_{10}$ (p-values) from
 1036 Kendall's τ (panels A, E and I). The central panels in each row show the distribution of
 1037 Z_W scores from the WZA across the genome (B, F and J) and the distribution of results
 1038 from the top-candidate method (C, G and K). The rightmost panels show the proportion
 1039 of locally adapted genes identified using the three different tests for an increasing
 1040 number of genes in the search effort. Results are shown for directional selection

1041 simulations. Note that only SNPs with a minor allele frequency > 0.05 are shown in
1042 panels (A, E and I).

1043

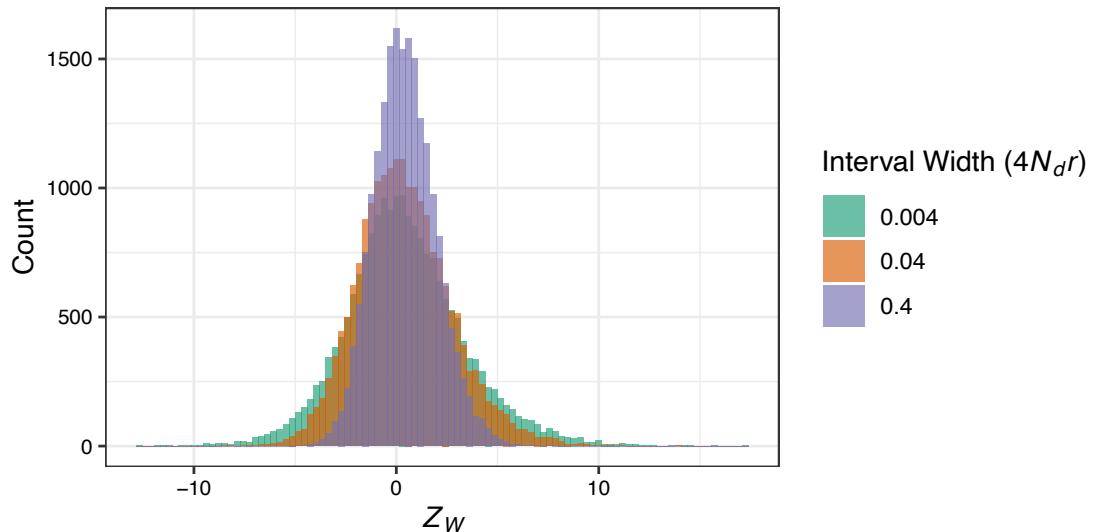
1044



1045

1046 **Figure S8** Comparison of the WZA, the top-candidate and the single-SNP approaches
1047 with varying numbers of individuals sampled per deme. Simulations shown used the BC
1048 map and directional selection. Lines represent the mean of 20 simulation replicates.

1049

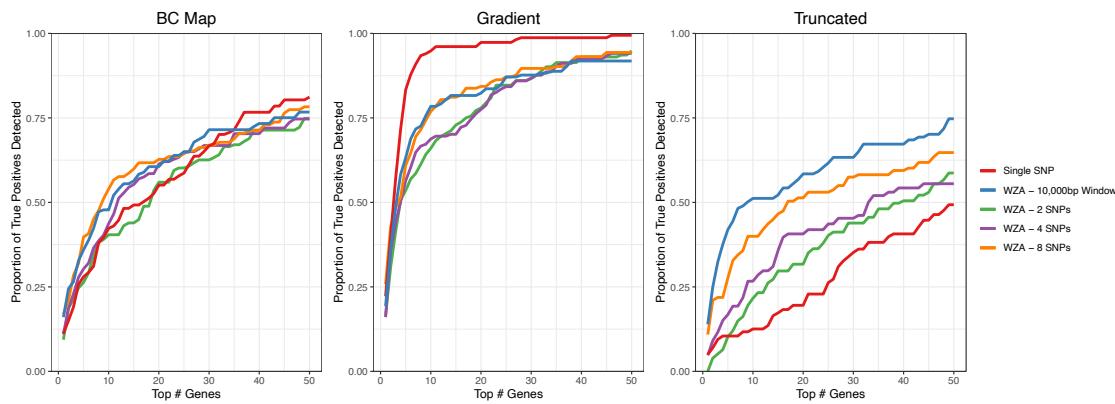


1050

1051 **Figure S9** The distribution of Z_W scores under different recombination rates.

1052

1053

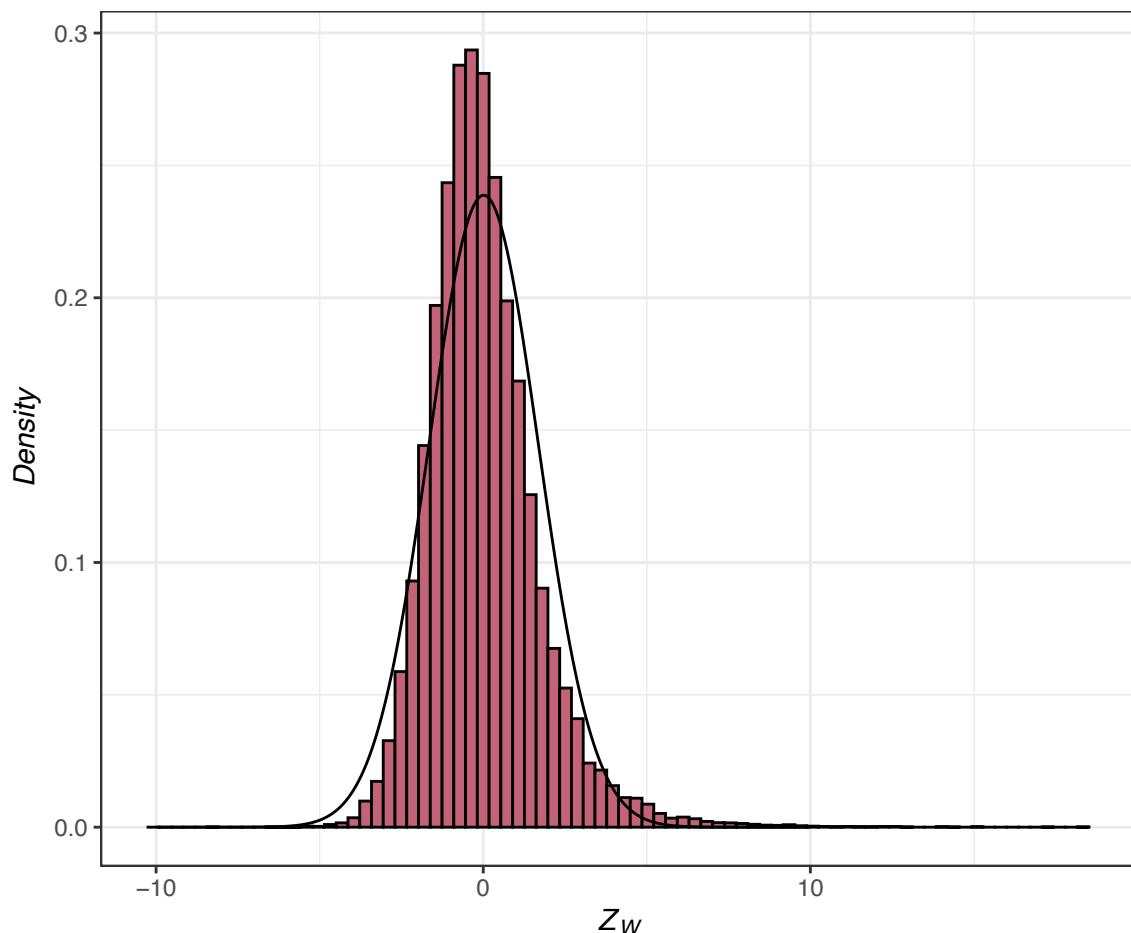


1054

1055 **Figure S10** Comparing the performance of the WZA genes identified using the WZA,
1056 using analysis windows analyzing a fixed number of SNPs. Lines represent the means
1057 of 20 replicates.

1058

1059

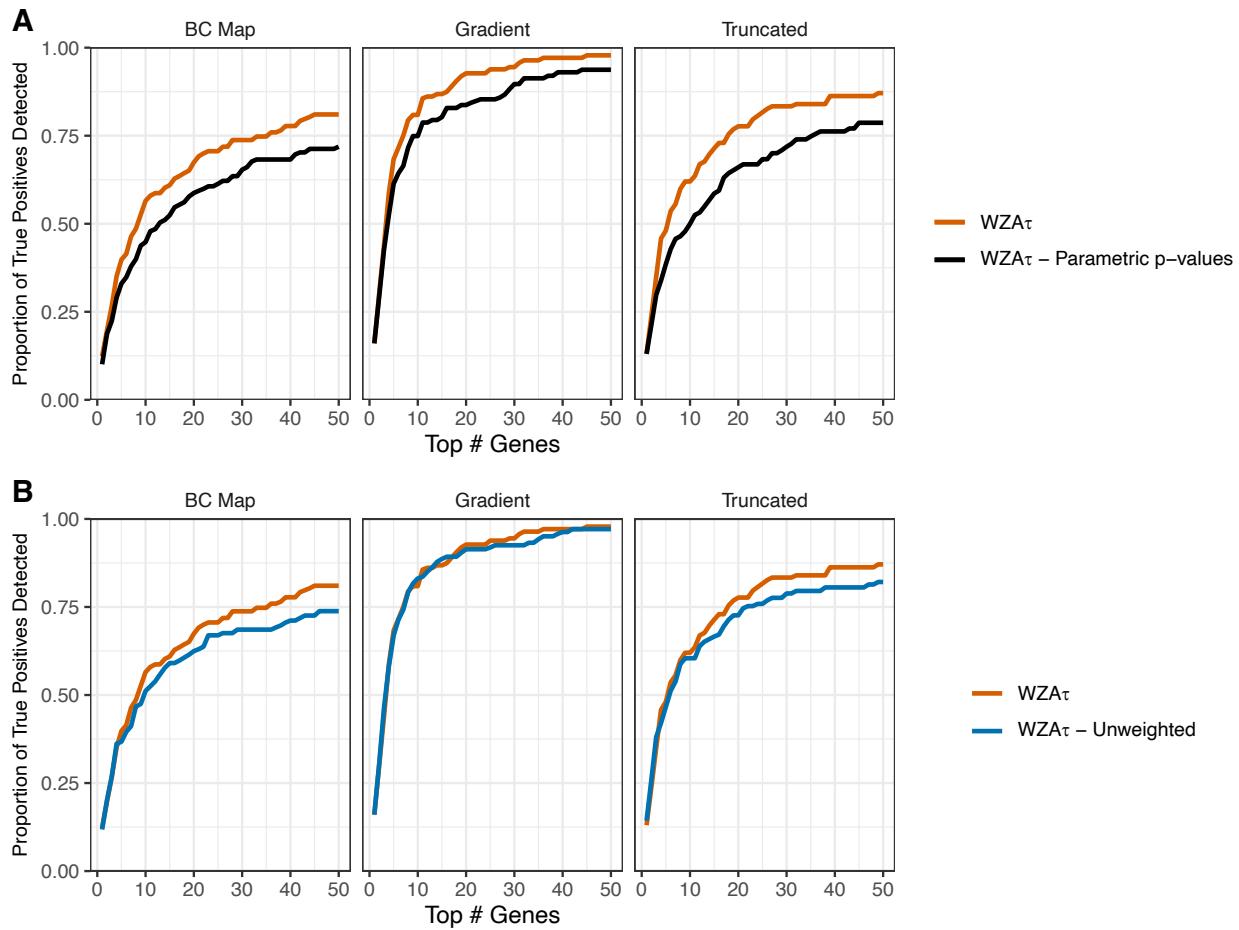


1060

1061 **Figure S11** The distribution of A) degree days < 0 (DD0) across the populations of P.
1062 contorta sampled by Yeaman et al (2016) and B) Z_w scores for the GEA on DD0. Note
1063 that the DD0 values in A) are unscaled. In B) the curve shows a normal distribution
1064 fitted to the data.

1065

1066



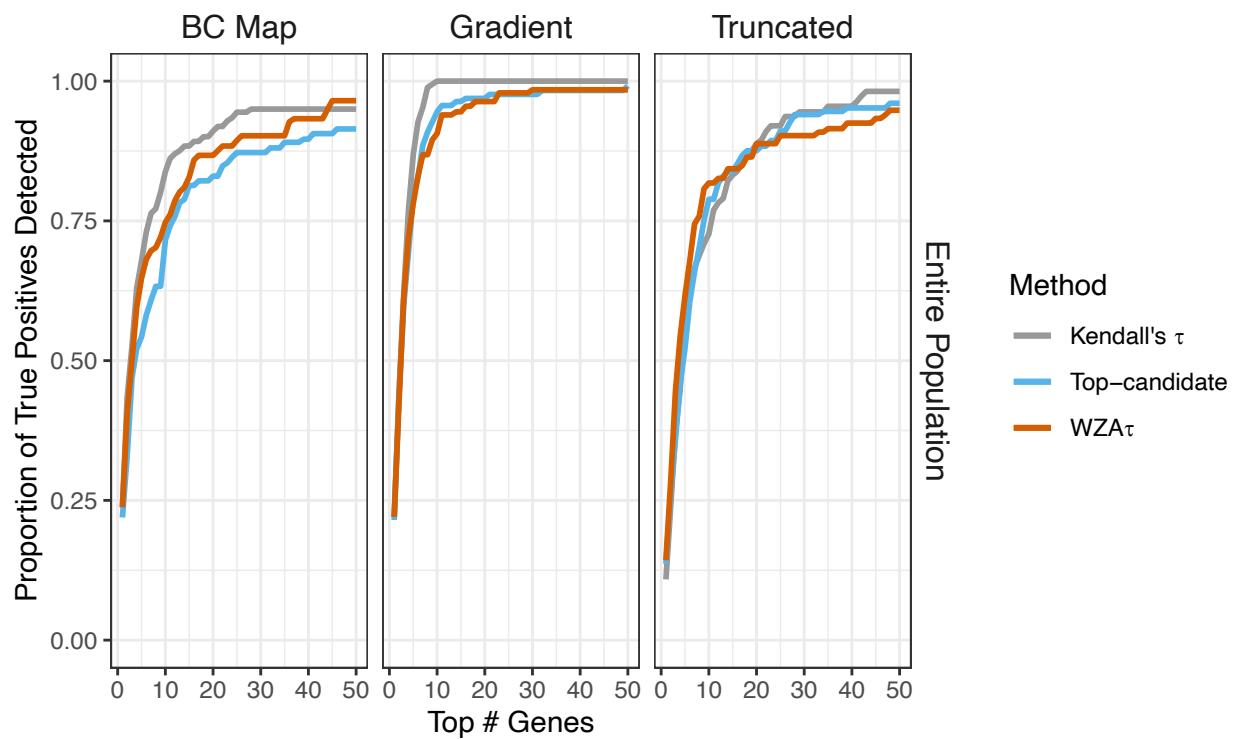
1067

1068

Figure S12

1069

1070



1071

1072 **Figure S13** A comparison of three methods to identify the genetic basis of local
1073 adaptation when one has complete information on all aspects of the metapopulation.