

1 **The accuracy of predicting maladaptation to new
2 environments with genomic data**

3 Brandon M. Lind*, Katie E. Lotterhos

4 Department of Marine and Environmental Sciences, Northeastern University
5 430 Nahant Road, Nahant, MA 01908, USA

6 24 July 2024

7 **Running Title:** Performance of genomic offsets

8 **Keywords:** genomic offset, environmental change, climate change, assisted gene flow,
9 genomic forecasting, random forests, redundancy analysis, risk of non-adaptedness

10 *Corresponding Author

11 Brandon M. Lind

12 Email: lind.brandon.m@gmail.com

13 Abstract

14 Rapid environmental change poses unprecedented challenges to species persistence. To
15 understand the extent that continued change could have, genomic offset methods have
16 been used to forecast maladaptation of natural populations to future environmental
17 change. However, while their use has become increasingly common, little is known
18 regarding their predictive performance across a wide array of realistic and challenging
19 scenarios. Here, we evaluate the performance of currently available offset methods
20 (gradientForest, the Risk-Of-Non-Adaptedness, redundancy analysis with and without
21 structure correction, and LFMM2) using an extensive set of simulated datasets that vary
22 demography, adaptive architecture, and the number and spatial patterns of adaptive
23 environments. For each dataset, we train models using either *all*, *adaptive*, or *neutral*
24 marker sets and evaluate performance using *in silico* common gardens by correlating
25 known fitness with projected offset. Using over 4,849,600 of such evaluations, we find that
26 1) method performance is largely due to the degree of local adaptation across the
27 metapopulation (*LA*), 2) *adaptive* marker sets provide minimal performance advantages,
28 3) performance within the species range is variable across gardens and declines when offset
29 models are trained using additional non-adaptive environments, and 4) despite (1),
30 performance declines more rapidly in globally novel climates (i.e., a climate without an
31 analog within the species range) for metapopulations with greater *LA* than lesser *LA*. We
32 discuss the implications of these results for management, assisted gene flow, and assisted
33 migration.

34 1 | Introduction

35 The impacts of climate change and habitat loss pose urgent challenges to the management
36 of species, communities, habitats, and ecosystem services (Bonan, 2008; Doney et al., 2012;
37 Hoegh-Guldberg & Bruno, 2010). Traditional methods used to infer such impacts, such as
38 reciprocal transplants and common gardens, require time and resources that may not be
39 available or feasible for many organisms of management concern, particularly for long-lived
40 organisms where reproductive stages occur after several decades of development. Ecological
41 forecasting models have therefore become increasingly germane to support environmental
42 decision making by managers across both terrestrial and marine systems.

43 In the context of population viability in the face of environmental change, many of these
44 models rely on theoretical expectations that the limits of species' distributions are primarily
45 determined by the distribution of environmental conditions (e.g., Good 1931), and that
46 occupancy of highly suitable habitat enables increased abundance through greater survival
47 and reproduction (i.e., fitness) of individuals (Brown, 1984). Such methods, termed species
48 distribution models or ecological niche models (see Elith & Leathwick, 2009 for a discussion
49 on terminology) are correlative approaches that are often used to predict (relative) habitat
50 suitability for a single species (Lee-Yaw et al., 2022). This information is used to understand
51 potential ecological impacts on the species from future climate change. However, these methods
52 often ignore aspects of the species' evolutionary history that could be important for predicting

53 long-term population persistence, such as the environmental drivers of local adaptation or
54 spatial patterns of adaptive genetic variation (Waldvogel et al., 2020).

55 Subsequent methods, termed genomic offsets (reviewed in Capblancq et al., 2020; Rellstab
56 et al., 2021), have attempted to address these shortcomings by modeling relationships between
57 environmental and genetic variation to predict maladaptation of natural populations to either
58 future climates *in situ*, or to predict the relative suitability of these populations for the specific
59 environment of a restoration site. Empirical attempts to confirm predictions from genomic
60 offset models are rare and, compared to evaluations *in silico* (Láruson et al., 2022), have found
61 relatively weaker relationships between predicted maladaptation to common garden climates
62 and the measurement of phenotypic proxies for fitness from individuals grown in these same
63 environments (e.g., Capblancq & Forester, 2021; Fitzpatrick et al., 2021; Lind et al., 2024).
64 This suggests that realized performance in natural systems may not meet expectations from
65 evaluations *in silico*. Even so, these empirical results have often shown the expected negative
66 relationship between predicted offset and common garden performance such as measures of
67 juvenile growth (e.g., Fitzpatrick et al. 2021; Lind et al. 2024) and even 52-year mortality at
68 multiple sites (Lind et al. 2024). Further, many of these studies found that genomic offsets
69 often perform better than climate or geographic distance alone (e.g., Capblancq & Forester,
70 2021; Fitzpatrick et al., 2021; Láruson et al., 2022; Lind et al., 2024).
71 Across empirical and *in silico* studies, little difference in performance was found between
72 models trained using only adaptive markers (i.e., known *in silico*, or candidates from empirical
73 genotype-environment [GEA] associations) and those chosen at random, suggesting that

74 genome-wide data may be sufficient to capture signals relevant to environmental adaptation
75 (Fitzpatrick et al., 2021; Lachmuth, Capblancq, Keller, et al., 2023; Láruson et al., 2022; Lind
76 et al., 2024). Together, these results suggest that genomic offset methods may provide valuable
77 insight for management without needing to identify adaptive loci, but the number of
78 evaluations has been relatively small with few comparisons among methods. Thus, little is
79 known about how robust these methods are across a wide array of realistic empirical scenarios
80 and the comparative performance among available methods. Other signals within marker data,
81 such as the degree to which allele frequencies are clinal across environmental gradients, also
82 require further exploration, particularly for methods that may algorithmically emphasize such
83 patterns over those more relevant to environmental selection, or for those methods that may
84 rely upon such clinal patterns to maintain accuracy in predictions. Indeed, concerns regarding
85 the accuracy of ecological forecasting models present a primary limitation towards
86 incorporating inferences from these models into management (Clark et al., 2001; Schmolke et
87 al., 2010) and genomic offset models are no exception. Major questions still remain about how
88 the performance of a method is affected by aspects of the evolutionary history of sampled
89 populations, whether the type of signals in putatively ideal datasets that may mislead offset
90 inference (e.g., clinal allele frequencies), how important it is to identify the environmental
91 drivers of local adaptation *a priori*, and how consistent predictive performance is across the
92 landscape. Finally, because novel climates with no recent analog are expected to increase in
93 the future (Lotterhos et al., 2021; Mahony et al., 2017) there is also uncertainty regarding the
94 performance of forecasting models when predictions are made to novel environments that

95 drastically differ from those used to train and build the models themselves (Fitzpatrick et al.,
96 2018; Lind et al., 2024).

97 While much uncertainty remains regarding the predictive performance of genomic offsets,
98 the domain of applicability (i.e., the circumstances under which a method is acceptably
99 accurate) for these methods can be more precisely defined using simulated data (Lotterhos et
100 al., 2022). Simulated data present ideal circumstances for understanding the opportunities and
101 limits of genomic offsets because there is no error in the measurement of allele frequencies,
102 environmental variables, individual fitness, or the drivers of local adaptation. To provide
103 relevant inference regarding the domain of applicability, simulations should capture the
104 complexities of empirical data with biological realism (e.g., clinal or patchy environments),
105 present contrasting cases of differing scenarios while controlling for important features of the
106 data (e.g., varying population connectivity but controlling for mean differentiation), and
107 challenge methods using adversarial scenarios that capture extreme characteristics of empirical
108 data (e.g., prediction to novel environments with no current analog available for model
109 training; Lotterhos et al. 2022).

110 Here, we used a wide array of previously published, biologically realistic, contrasting, and
111 adversarial simulations from Lotterhos (2023a) in an attempt to more precisely define the
112 limits of predictive performance of five implementations of four genomic offset methods (Table
113 1): gradientForest (GF_{offset} ; *sensu* Fitzpatrick & Keller, 2015), the Risk Of Non-Adaptedness
114 (RONA, Rellstab et al., 2016), Latent Factor Mixed Models ($LFMM2_{offset}$, *sensu* Gain &
115 François, 2021, and redundancy analysis with and without correction for population structure

116 (RDA_{offset}, *sensu* Capblancq & Forester, 2021). The main goal of this study was to understand
117 how the evolutionary and experimental parameters used in the training and evaluation of
118 offset methods affect the accuracy of the methods' projections of maladaptation under ideal
119 empirical scenarios (i.e., using data with no inherent error). Using these scenarios, we ask the
120 following six questions: 1) Which aspects of the past evolutionary history affect performance
121 of offset methods? 2) How is offset performance affected by the proportion of loci with clinal
122 alleles in the data? 3) Is method performance driven by causal loci or by genome-wide patterns
123 of isolation-by-environment? 4) What is the variation of model performance across the
124 landscape? 5) How does the addition of non-adaptive nuisance environments in training affect
125 performance? 6) How well do offset models extrapolate to novel environments outside the
126 range of environmental values used in training?

127 2 | Methods

128 Throughout this manuscript we cite analysis code used to carry out specific analyses in-line
129 with the text. Supplemental Text S1-S2 outlines and describes the sets of scripts or, most
130 often, jupyter notebooks, used to code analyses. Scripts and notebooks are both referenced as
131 Supplemental Code (SC) using a directory numbering system (e.g., SC 02.05). More
132 information regarding the numbering system, archiving, and software versions can be found
133 in the Data Availability section.

134 **2.1 / Explanation of Simulations and Training Data**

135 To train offset methods we used single nucleotide polymorphism (SNP) and environmental
136 data from a set of previously published simulations (225 levels with 10 replicates each) of a
137 Wright-Fisher metapopulation of 100 demes on a 10 x 10 grid evolving across a heterogeneous
138 landscape (Lotterhos, 2023a). Each dataset was simulated under a combination of the
139 following four evolutionary parameters: i) three landscapes (10 populations x 10 populations)
140 that varied in vicariance and environmental gradients (*Estuary - Clines; Stepping Stone -*
141 *Clines*; and *Stepping Stone - Mountain*), ii) five demographies that varied population size and
142 migration rates across the landscape, iii) three genic levels that varied in the effect size and
143 number of mutations underlying adaptation (mono-, oligo-, and polygenic), and iv) five
144 pleiotropy levels that varied the number of quantitative traits under locally stabilizing
145 selection ($n_{traits} \in \{1, 2\}$), presence of pleiotropy (when $n_{traits} = 2$), and variability of selection
146 strength across individual traits (see Fig. 1 in Lotterhos 2023a).

147 The adaptive trait(s) were under selection by a different environmental variable, where the
148 optimum trait value was given by the local environment on the landscape (Fig. S37). The
149 adaptive trait(s) undergoing selection responded to either a latitudinal temperature gradient
150 (*temp*; $n_{traits} = 1$), or to both *temp* and a longitudinal “*Env2*” gradient ($n_{traits} = 2$). *Env2*
151 represented distinct biological analogies depending on the context: in the *Stepping Stone -*
152 *Mountain* landscape *Env2* was analogous to elevation (e.g., as with tree species), whereas in
153 the *Estuary - Clines* landscape the *Env2* environment was analogous to gradients of salinity

154 within coastal inlets connected only by the outer marine (ocean) environment (e.g., as with
155 stickleback or oyster species).

156 Twenty independent linkage groups were simulated. Of these, mutations that had effects
157 on one or more phenotypes under selection (i.e., quantitative trait nucleotides, QTNs) were
158 allowed to evolve on only ten linkage groups, and neutral mutations were added to all 20
159 linkage groups with tree sequencing (for details see Lotterhos 2023a). Adaptive traits were
160 determined additively by effects of QTNs.

161 In all simulations, phenotypic clines evolved between each trait and the selective
162 environment (Lotterhos, 2023a), where populations became locally adapted to their
163 environment, measured at the metapopulation level as the mean difference of demes in
164 sympatry minus allopatry ($LA_{\Delta SA}$, Blanquart et al., 2013). $LA_{\Delta SA}$ equates to the average levels
165 of local adaptation at the deme level which can be calculated for each deme by both home-
166 away ($LA_{\Delta HA}$) and local-foreign ($LA_{\Delta LF}$) measures.

167 These simulations represent a wide array of realistic, contrasting, and adversarial scenarios
168 in which we could more precisely define the domain of applicability of offset methods. For
169 instance, in the *Stepping Stone - Mountain* landscape, geographic distance and environmental
170 distance were not strongly correlated, whereas in the *Stepping Stone - Clines* and *Estuary -*
171 *Clines* they were. Additionally, the proportion of mutations with monotonic frequency
172 gradients (i.e., allelic clines) underlying local adaptation varied across the simulated datasets
173 (Lotterhos, 2023a), which may also affect offset performance. These simulations also presented

174 demographic scenarios in which selection was confounded with genetic drift or population
175 genetic structure.

176 For each simulation, ten individuals were randomly chosen per population for a total of
177 1000 individuals. Individual genotypes were coded as counts of the derived allele. Alleles with
178 global minor allele frequency (MAF) < 0.01 were removed. Using all 100 populations,
179 population-level derived allele frequencies and current environmental values were used as input
180 to train offset methods.

181 In addition to the 2250 simulated Wright-Fisher datasets (225 levels * 10 replicates), we
182 also included a non-Wright-Fisher case with range expansion from three refugia and secondary
183 contact (Fig. 6 in Lotterhos 2023a). This simulation evolved variable degrees of admixture
184 across the landscape. Six moderately polygenic environmental traits ($n_{traits} = 6$) were under
185 selection from the environment. Environments were based on six weakly correlated
186 environmental variables taken from Bioclim environmental measures of western Canada. The
187 simulation evolved local adaptation at all six traits with unconstrained pleiotropy. For more
188 details on simulations, see (Lotterhos, 2023a).

189 ***2.2 / Evaluation of Offset Methods***

190 We investigated the performance of five implementations of four genomic offset methods
191 (Table 1): GF_{offset} , $RDA_{offset,}$, $LFMM2_{offset}$, and $RONA$. While GF_{offset} , RDA_{offset} , and
192 $LFMM2_{offset}$ can use multivariate environmental data to train models, $RONA$ is univariate
193 and can only account for a single environmental variable at one time (Table 1). Additionally,

194 while GF_{offset} and RONA do not apply correction for population genetic structure, LFMM2_{offset}
195 does by default, and structure correction with RDA_{offset} is optional. We thus evaluate RDA_{offset}
196 with (RDA-corrected) and without (RDA-uncorrected) population genetic structure correction
197 (Table 1). For additional specifics related to the implementation of each offset method, see
198 Supplemental Text S1.1-S1.4 and Fig. S1, Fig. S2, Fig. S3.

199 We varied construction of genomic offset training datasets for each replicate of the 1-, 2-,
200 and 6-trait simulations by varying the marker set used in model training (Fig. 1A, Table 2;
201 see *Q3* below). Each model was trained using genetic and environmental data from all 100
202 populations. The environmental variables used were only those imposing selection pressure.
203 We predicted offset from each model for each population to all 100 within-landscape common
204 gardens from a full factorial *in silico* reciprocal transplant design (Fig. 1B). For each common
205 garden, we quantified offset model performance as the rank correlation (Kendall's **T**) between
206 the population mean fitness (averaged over sampled individuals, Equation 3 in Lotterhos
207 2023a) and projected population offset (Fig. 1C). Strong negative relationships between fitness
208 and predicted offset indicate higher performance of the method (note y-axes of Kendall's **T**
209 are inverted within figures to show more intuitive performance relationships, Fig. 1C-11). We
210 refer to the preceding processing of data as the *Adaptive Environment* workflow (Fig. 1, Table
211 2). Across all workflows, all adaptive environments were included in training and were never
212 excluded.

213 To explore the impact of the choice of environmental variables used (see *Q5* below), we
214 used a workflow similar to the *Adaptive Environment* workflow, except instead of using only
215 adaptive environmental variables, we used additional non-adaptive (i.e., nuisance)
216 environmental variables in training and prediction (second row, Table 2). These nuisance
217 variables had relatively weak correlation structure with adaptive environments and each other
218 (Fig. S4). We refer to each of these nuisance levels by the number of traits under selection
219 and the number of nuisance environments used (e.g., *1-trait 3-nuisance*). We refer to this
220 workflow as the *Nuisance Environment* workflow .

221 Finally, to contrast with within- landscape evaluations, we explored predictive performance
222 of *Adaptive Environment* offset models in novel environments that are beyond the range of
223 values of those used in training (see *Q6* below). In these novelty cases, we use 11 common
224 gardens, each progressively more distant from the average environment used in training (i.e.,
225 climate center) and evaluate performance in each garden. We refer to this workflow as the
226 *Climate Novelty* workflow. See Supplemental Text S3 and Fig. S5 for details regarding the
227 choice of environmental values for novelty scenarios.

228 **2.3 / Study Questions**

229 *Q1 - Which aspects of the past evolutionary history affect within-landscape performance of
230 offset methods?*

231 For each offset method, we used a fixed-effects type II ANOVA model to test for significant
232 differences in the performance from 2-trait *Adaptive Environment* models trained using *all*
233 markers using the following factors: landscape (*Estuary - Clines*, *Stepping Stone - Clines*,
234 *Stepping Stone - Mountain*), demography (five levels describing population size and migration
235 patterns across the landscape), genic level of architecture (three levels from oligogenic to
236 polygenic), presence or absence of pleiotropy, proportion of loci with clinal allele frequencies
237 (as defined in Lotterhos, 2023a), degree of local adaptation (Δ SA), and common garden ID.
238 Specifically,

$$239 \quad Y_{ij} = L_i + D_i + GL_i + P_i + p_{cQTN,t,i} + p_{cNeut,t,i} + p_{cQTN,Env2,i} + p_{cNeut,Env2,i} + LA_{\Delta SA,i} + \\ 240 \quad G_j \\ 241 \quad \text{(Eq. 1)}$$

242 where Y_{ij} is the within-landscape performance (Kendall's τ) of a single method for garden j in
243 simulation i , with factors for landscape (L), demography (D), genic level (GL), presence of
244 pleiotropy (P), proportion of QTN or neutral alleles with *temp* clines (respectively $p_{cQTN,t,i}$
245 and $p_{cNeut,t,i}$), proportion of QTN or neutral alleles with *Env2* clines (respectively $p_{cQTN,Env2,i}$

246 and $p_{cNeut,Env2,i}$), degree of local adaptation ($LA_{\Delta SA}$), and garden ID (G). The first four factors
 247 are illustrated in Fig. 1 of Lotterhos (2023a).

248 Q2 - How is offset performance affected by the proportion of clinal alleles in the data?

249 Clinal alleles (i.e., alleles with monotonic gradients in frequency across space) that covary
250 with environmental clines could be weighted more heavily in offset models that emphasize loci
251 whose allele frequencies explain significant variation across local environmental values. Using
252 2-trait models trained using *all* markers from the *Adaptive Environment* workflow, we used
253 an ANOVA model (Eq. 2) to test the hypothesis that clinal alleles differentially impact model
254 performance, independent from the other factors from Eq. 1:

$$Y_{ij} = P_{cQTN,t,i} + P_{cNeut,t,i} + p_{cQTN,Env2,i} + P_{cNeut,Env2,i} \quad (\text{Eq. 2})$$

257 The factors representing clinal alleles in Eq. 2 are the same as those in Eq. 1.

258 Q3 - Is method performance driven by causal loci or by genome-wide patterns of Isolation By
259 Environment?

260 For each offset method and workflow, we varied the set of input markers for 1-, 2- and 6-
261 trait simulations that were used in training to determine if performance of a method was
262 driven by properties of the evolutionary forces shaping genotype-environment relationships: 1)
263 *adaptive* markers (i.e., QTNs with effects on at least one trait), 2) *neutral* markers (SNPs on
264 linkage groups without QTNs), and 3) *all* markers (union of *adaptive* and *neutral* markers, as

265 well as non-QTN markers on the same linkage groups as QTNs). Only loci that passed MAF
266 filtering were included in marker sets ($\overline{N_{adaptive}} = 188$, $\overline{N_{all}} = 33,169$, $\overline{N_{neutral}} = 16,520$). If
267 offset performance is determined solely by adaptive signals in genetic data, offsets trained
268 using *adaptive* markers should have better performance than *all* or *neutral* markers, and *all*
269 markers should have better performance than *neutral* markers.

270 If the marker set has little impact on offset performance, this could indicate that offset
271 methods are giving weight to genome-wide signals present in the data. Previously, some (e.g.,
272 Lachmuth, Capblancq, Keller, et al., 2023; Lind et al., 2024) have postulated that this signal
273 may be related to isolation by environment (IBE, i.e., when genetic and environmental
274 distances are positively correlated, independent of geographic distance; Wang & Bradburd,
275 2014).

276 If IBE is driving patterns of offset performance, we expect 1) performance to be similar
277 between offsets estimated using *adaptive* markers and those estimated using *neutral* markers;
278 2) a greater proportion of variation in performance to be explained by p_{cNeut} than p_{cQTN} (from
279 $Q2$); 3) a strong, positive relationship between performance and $LA_{\Delta SA}$; and 4) the difference
280 in IBE between two marker sets to be positively correlated with the difference in performance
281 of two models trained with those markers. We measured IBE as the rank correlation
282 (Spearman's ρ) between population pairwise F_{ST} (Weir & Cockerham, 1984) and Euclidean
283 climate distance of adaptive environmental variables.

284 *Q4 - What is the variation of model performance across the landscape?*

285 Within a landscape, offset methods may not have high predictive performance at every site
286 or every environment. Understanding variability in the predictive performance of offset models
287 across the landscape is particularly relevant when offsets are used for restoration or assisted
288 gene flow initiatives (i.e., ranking sources for a given site). If predictive performance is variable
289 across the landscape, this may limit the usefulness of genomic offsets for such purposes even
290 if model performance is validated in one common garden. Using the *Adaptive Environment*
291 workflow, we visualized variation of 1- and 2-trait within-landscape performance with boxplots
292 for each common garden for each method and landscape. To understand if variation in
293 predictive performance was a function of the model quality, we investigated the relationship
294 between a model's performance variability (i.e., standard deviation across 100 common
295 gardens) and the model's median performance.

296 *Q5 - How does the addition of non-adaptive nuisance environments in training affect*
297 *performance?*

298 In practice, the environments imposing selection are rarely known *a priori*. Additionally,
299 the inclusion of environmental measures that are not correlated with the main axes of selection
300 may reduce model performance compared to models trained using only causal environments.
301 To investigate the sensitivity of offset methods to environmental input we compared *Adaptive*
302 *Environment* workflow models from 1-, 2-, and 6-trait simulations – where only the adaptive
303 environment(s) are used in training (*0-nuisance*) – to models from the *Nuisance Environment*

304 workflow trained with the same data but with the addition of nuisance environments (N -
305 *nuisance*, where $N > 0$; Table 2).

306 We used nuisance environmental variables from Lotterhos (2023a) that were real BioClim
307 variables (*TSsd*, *PSsd*, and *ISO*) taken from British Columbia and Alberta, Canada, which
308 have minimal correlation with causal environments and each other (Fig. S4). These three
309 nuisance environments differ from previous implementations of such variables (Láurson et al.
310 2022) in that they are spatially autocorrelated whereas nuisance environments in Láurson et
311 al. (2022) were not. For 1-trait scenarios, *Env2* was also used as a nuisance environmental
312 variable. If offset methods are unaffected by the addition of nuisance environmental variables,
313 performance should not differ between *0-nuisance* and *N-nuisance* implementations.

314 Finally, in empirical settings the set of adaptive environments are not known *a priori*. We
315 also explored whether GF would rank adaptive environments higher than nuisance
316 environments using weighted importance output from GF.

317 *Q6 - How well do offset models extrapolate to novel environments outside the range of
318 environments used in training?*

319 Even if offset methods have high within-landscape performance, this does not directly
320 address situations where future environmental conditions are vastly different from the
321 environmental conditions used for training (i.e., novel environments). If performance decreases
322 with increasing environmental novelty relative to training data, this raises questions about the
323 utility of genomic offsets for predicting 1) relative *in situ* vulnerability of populations to future

324 climate change, and 2) the relative suitability of populations to restoration sites that differ
325 drastically than those used in training.

326 To understand if offset performance degrades with environmental novelty relative to
327 training data, we predicted offset to 10 novel environmental scenarios for the 1-, 2-, and 6-
328 trait simulations using the *Climate Novelty* workflow (Table 2). The novel environmental
329 scenarios were a set of common garden environments, z_E , extending outward from the training
330 populations and exceeding values observed on the landscape for all adaptive environmental
331 variables (Supplemental Text S3). We represent these scenarios as standard deviations from
332 the center of environmental values used in training: $z_E \in \{1.72, 2.35, 2.74, 3.13, 3.53, 3.92,$
333 $4.31, 4.70, 5.09, 5.48, 5.88\}$. Fitness in novel environments was estimated assuming that the
334 phenotypic optimum continues to have a linear relationship with the environmental variable
335 (Equation 3 in Lotterhos 2023a).

336 3 | Results

337 *Q1 - Which aspects of the past evolutionary history affect within-landscape performance of*
338 *offset methods?*

339 The ANOVA model (Eq. 1) indicated that the degree of local adaptation of the
340 metapopulation ($LA_{\Delta SA}$) was the primary factor influencing offset performance, followed by
341 common garden location, demography, and landscape (Table S1; Fig. S6). Within the
342 simulations, $LA_{\Delta SA}$ was impacted by pleiotropy, the relative strength of selection, and

343 landscape, (Fig. S7; see also Figs. S2A, S2B in Lotterhos, 2023a), so there may be some
344 confounding among these factors.

345 In line with the ANOVA model, the performance of specific offset methods generally
346 increased with increasing $LA_{\Delta SA}$ (Fig. 2), but there were some interesting differences among
347 methods. For instance, GF_{offset}, LFFM2_{offset}, RDA-uncorrected, and RONA_{temp} all improved
348 as $LA_{\Delta SA}$ increased, while RDA-corrected and RONA_{Env2} showed relatively weaker
349 relationships.

350 Across landscapes, offset methods generally had higher performance in *Stepping Stone -*
351 *Clines* landscapes than *Stepping Stone - Mountain* landscapes (Fig. 2B) despite similar levels
352 of $LA_{\Delta SA}$ (Fig. 2A). Offset methods also generally performed better in the two *Stepping Stone*
353 landscapes than the *Estuary - Clines* landscape (Fig. 2B). However, there were some
354 interactions between method and landscape (Fig. 2C). For instance, RDA-corrected performed
355 better in the *Estuary - Clines* compared to the two *Stepping Stones* landscapes, while the
356 RDA-uncorrected showed the opposite pattern: performance was higher in the two *Stepping*
357 *Stones* landscapes compared to *Estuary - Clines*.

358 The performance of methods was similar across genic levels but increased slightly as the
359 number of QTNs underlying adaptation became more polygenic (Fig. S8). Additionally, while
360 demography primarily influenced population differentiation across the landscape with little
361 impact on $LA_{\Delta SA}$ within simulations (Table S2 in Lotterhos 2023a), migration breaks between
362 populations and latitudinal clines in population size generally decreased offset performance for
363 LFMM2_{offset}, GF_{offset}, and RDA- uncorrected (Fig. S9).

364 *Q2 - How is offset performance affected by the proportion of clinal alleles in the data?*

365 The sum of squares from Eq. 1 indicated that the proportion of clinal alleles did not account
366 for meaningful variation in offset performance ([Table S1](#)). Even so, results from an ANOVA
367 model with just the proportion of clinal loci as explanatory variables (Eq. 2) indicated that
368 p_{cNeut} accounted for 4.14–9.65 times the variation than did p_{cQTN} for GF_{offset}, LFMM2_{offset}, and
369 RDA-corrected. For GF_{offset} and RDA-uncorrected, $p_{cNeut,Env2}$ accounted for >16% of the sum
370 of squares (Table S2, Fig. S10).

371 Overall, relationships between performance and p_{cNeut} (second column, Fig. S11) were
372 stronger than between performance and p_{cQTN} (first column, Fig. S11). However, sometimes
373 performance increased with p_{cNeut} and sometimes it decreased, depending on the method (Fig.
374 S11), indicating that each method is differentially sensitive to clinal alleles in the data.
375 Ultimately, strong population genetic structure along environmental clines in 2-trait
376 simulations (Fig. S12) drove relationships with p_{cNeut} (Fig. S13) which in turn drove
377 relationships with performance (Fig. S14, Fig. S11).

378 *Q3 - Is method performance driven by causal loci or by genome-wide patterns of Isolation-By-
379 Environment?*

380 Overall, 1- and 2-trait *Adaptive Environment* models had relatively similar performance
381 among marker sets. For instance, models trained using *all* or *neutral* markers had similar
382 performance while models trained using *adaptive* markers performed slightly higher than the
383 other sets. The median increase in performance from *adaptive* compared to *all* or *neutral*

models was less than 3%. In total, using *adaptive* markers outperformed 68% of models using *neutral* markers and 67% of models using *all* markers, while 74% of models using *all* markers outperformed *neutral* models (Fig. 3A-C). For RDA-corrected the *neutral* markers performed slightly better than either *adaptive* or *all* markers in 2-trait evaluations (Fig. 3E). *Adaptive* markers from 6-trait evaluations provided varied performance advantages across methods (Fig. 4; Fig. S37).

The *adaptive* marker sets had relatively elevated levels of *IBE* compared to sets of *neutral* or *all* markers in 1- and 2-trait simulations, but levels of *IBE* were nonetheless quite similar between marker sets (Fig. S15). Consequently, performance of models trained with *adaptive* markers generally had stronger relationships with *IBE* than $LA_{\Delta SA}$ but this was not the case for models trained with either *all* or *neutral* markers (Fig. S16).

Intriguingly, levels of *IBE* found within a landscape (Fig. S17A) did not correspond to the degree of $LA_{\Delta SA}$ that developed (Fig. S17B). Even so, while *IBE* was generally unrelated to $LA_{\Delta SA}$ across all simulations, there were generally positive relationships between *IBE* and $LA_{\Delta SA}$ when controlling for the number of traits and differences in strengths of selection (Fig. S18). As such, *IBE* from *all* markers explained very little variation in performance when added as a factor to the ANOVA model from Eq. 1 (SC 02.02.01), but accounted for some variation in ANOVA models with only $LA_{\Delta SA}$ and *IBE* as explanatory variables (0-34% for *IBE* vs 0-74% for $LA_{\Delta SA}$; Table S3). Except for RONA, the differences in performance between two models trained with different marker sets was generally unrelated to the differences in *IBE* between the two marker sets used to train the models (Fig. S19).

405 *Q4 - What is the variation of model performance across the landscape?*

406 All 1- and 2-trait models exhibited variation in the predictive performance across gardens
407 within a landscape, from essentially no predictive performance to very high predictive
408 performance (Fig. S20, Fig. S21, Fig. S22, Fig. S23). Variation in performance was also
409 observed for 6-trait models (Fig. 4).

410 While there was variability in predictive performance of 1- and 2-trait models within each
411 landscape, in many cases the best performing models had the lowest levels of performance
412 variation (Figs. S24, S25, S26). Ultimately, we found no strong indicator for predicting when
413 a model will be highly variable. Indeed, while performance generally increased with $LA_{\Delta SA}$
414 (Fig. 2), variability in performance was not strongly related to the variability in deme-level
415 LA on the landscape (Figs. S27, S28, S29). Despite $LA_{\Delta SA}$ driving performance more generally
416 (from Q1), this indicates that variation in model performance across the landscape was not
417 strongly driven by metapopulation levels of, nor deme-level variation in, LA .

418 *Q5 - How does the addition of non-adaptive nuisance environments in training affect
419 performance?*

420 Training offset models with the addition of non-adaptive nuisance environmental variables
421 generally reduced offset method performance (Fig. 5). This decline was most dramatic for
422 offset trained on 1-trait simulations (Fig. 5A) compared to the decline observed for 2-trait
423 (Fig. 5B) and 6-trait (Fig. 5C) simulations. The only instances for which median performance

424 did not decrease monotonically with nuisance level were for 2-trait simulations evaluated with
425 GF_{offset} (Fig. S30).

426 Overall, landscape had the most influence over performance differences due to non-adaptive
427 nuisance environments (Fig. S30), whereas there was little difference across other simulation
428 parameters (not shown except in SC 02.02.06). Even so, *adaptive* markers seemed to provide
429 some advantages in the presence of nuisance environments, particularly for 1-trait datasets
430 where the advantages were more substantial compared to 2-trait datasets (Fig. S31, Fig. S32).

431 In some cases, the rankings of weighted environmental importance output from GF ranked
432 nuisance variables higher than at least one adaptive environment (Table S4). Across 1- and
433 2-trait *N-nuisance* models trained with *all* markers, GF incorrectly ranked environmental
434 drivers in 26.9% (133/495) of the cases. Rankings improved somewhat for models trained with
435 *adaptive* markers, incorrectly ranking environmental variables in 20.6% (102/495) of the cases
436 (Table S4).

437 *Q6 - How well do offset models extrapolate to novel environments outside the range of
438 environments used in training?*

439 The datasets that had the greatest within-landscape performance (i.e., those with higher
440 levels of $LA_{\Delta SA}$) were also those that experienced the steepest decline in performance with
441 increasing climate novelty (red shade, Fig. 6). Importantly, declines in performance for
442 datasets with greater $LA_{\Delta SA}$ were not due to instances where all populations had zero fitness
443 (and thus performance was undefined and manually set to 0; Supplemental Text S4, Fig. S33).

444 Despite little change in the median performance for datasets with low levels of LA, most
445 performance scores from these datasets did not exceed Kendall's $\tau=-0.5$, and therefore had
446 little predictive value in novelty scenarios. Performance of individual methods followed the
447 overall trend presented in Fig. 6 (Fig. S35).

448 Advantages of *adaptive* marker sets were much less prevalent across methods for *Climate*
449 *Novelty* scenario performance than either *Adaptive Environment* or *Nuisance Environment*
450 scenarios (Fig. S34).

451 4 | Discussion

452 In the last decade, genomic offset methods have been identified as a complement to other
453 ecological forecasting models because they incorporate intraspecific variation (Keller &
454 Fitzpatrick, 2015; Capblancq et al., 2020; Rellstab et al., 2021). Our evaluations show that
455 offset methods may be impacted by both the evolutionary history of sampled populations as
456 well as the decisions made during model training. Our analyses emphasize that performance
457 of offset methods is highest when applied to locally adapted populations with known drivers
458 of environmental selection, and when restricting offset projections to climates similar to those
459 used in training. These results suggest that there may be a number of scenarios for which
460 genomic offset methods may provide valuable insight for management, but also highlight that
461 they will not yield accurate results in every situation. Below, we discuss the implications of
462 these findings towards restoration, conservation, and the management of biodiversity.

463 **4.1 / The importance of local adaptation**

464 A basic assumption of genomic offset methods is that the sampled populations are adapted
465 to their local environment (Rellstab et al., 2016, 2021), but this assumption has not been
466 formally tested. Our analyses show that indeed the degree of local adaptation ($LA_{\Delta SA}$) is one
467 of the primary factors that determine model performance for most methods. A value of $LA_{\Delta SA}$
468 ~ 0.5 indicates that fitness in demes is on average 50% higher in sympatry than allopatry.
469 Values of $LA_{\Delta SA}$ represent the average deme-level magnitudes of $LA_{\Delta HA}$ and $LA_{\Delta LF}$ across the
470 metapopulation (Blanquart et al., 2013). Previous metaanalyses of studies measuring local
471 adaptation of natural populations have used different measures of LA from the ones we
472 calculate here, but do show that some species evolve large fitness differences among
473 populations (Hereford, 2009; Leimu & Fischer, 2008). Given the prevalence of LA found
474 previously (Hereford, 2009; Leimu & Fisher, 2010), we may therefore expect some genomic
475 offset methods to do reasonably well when predictions are made for environmental values most
476 similar to those used in training, and when local adaptation in the metapopulation is high
477 (e.g., when $LA_{\Delta SA} > 0.5$, the best performing methods, RDA-uncorrected and GF_{offset} , had an
478 average correlation between offset and fitness of ~ -0.6). However, even moderately high levels
479 of local adaptation had substantial decline in performance (e.g., when $LA_{\Delta SA} \sim 0.3$, the best
480 performing methods, RDA-uncorrected and GF_{offset} , had an average correlation between GO
481 and common garden fitness of ~ -0.2). Future studies should seek to demonstrate evidence for
482 strong local adaptation in the system prior to applying genomic offset methods, but ultimately

483 the level of predictable performance that is acceptable will depend on the stakeholders and
484 management goals in the system. In the absence of evidence for local adaptation for the target
485 system, additional data that can be used to verify offset predictions (such as common garden[s],
486 demographic data, or remote sensing data) should be used to verify offset predictions (Lind et
487 al., 2024).

488 **4.2 / The importance of the signals within genomic marker sets**

489 Initial implementations of genomic offset models focused on putatively adaptive markers
490 where the signal of local adaptation may be strongest (Fitzpatrick & Keller, 2015; Rellstab et
491 al., 2016). More recently, investigators have varied the set of markers used to train models
492 but have found little influence on performance (Fitzpatrick et al., 2021; Lachmuth, Capblancq,
493 Keller, et al., 2023; Láruson et al., 2022; Lind et al., 2024) and our results are consistent with
494 these studies. We found that the *adaptive* marker sets provided a slight advantage over *all* or
495 *neutral* marker sets, but not universally or by great margins.

496 One hypothesis put forth as to why adaptive marker sets perform similarly to all markers
497 is that genome-wide data captures sufficient signatures of IBE (Lachmuth, Capblancq, Keller,
498 et al., 2023; Lind et al., 2024). Our analysis found weak positive relationships between
499 performance and levels of *IBE* within marker sets. Even so, and except for RONA, there were
500 no universal relationships within methods between the difference in *IBE* of marker sets and
501 the difference in performance of the models trained with these markers. Together these results
502 indicate that while higher degrees of local adaptation may lead to increased levels of *IBE* in

503 the genome, the signal of *IBE* of input markers generally has minimal and varied impact on
504 performance differences for the scenarios evaluated here. Alternatively, the levels of *IBE*
505 present in the simulated genomes may exceed a minimum threshold of *IBE*, beyond which
506 differences in performance between marker sets are minimized.

507 While we found little impact of levels of *IBE* on overall performance, the way in which we
508 measured IBE may have masked causative relationships. For instance, we measured IBE as
509 the correlation between pairwise environmental distance and pairwise F_{ST} , and did not
510 incorporate geographic distance. In doing so, our measure of IBE distilled genetic distance
511 down to a single value from a large number of loci and gave less weight to loci with rare alleles.
512 In future studies, creating a fourth marker set based on loci with the highest IBE signals to
513 compare with other marker sets offers another opportunity to understand the impact of IBE
514 on performance. Empirical datasets will also be able to specifically address geographical
515 distances while quantifying IBE (e.g., Bradburd et al., 2013).

516 While measures of IBE are one signal remaining to be explored in future analyses, the
517 proportion of clinal neutral loci within marker sets was shown to have varied impacts on
518 performance, sometimes being positively related to performance and sometimes negatively
519 depending on the context. In addition to the effect of clinal allele patterns, our results also
520 suggest that nonmonotonic relationships between environmental gradients and allele
521 frequencies will be an important element to consider during model building, as the methods
522 that strictly modelled linear relationships (RONA, LFMM2) had on average poorer
523 performance than other methods we evaluated (GF_{offset} , RDA_{offset}). In contrast, we found very

524 similar performance across *adaptive*, *neutral*, and *all* loci. Though, while the similarity in
525 performance among marker sets evaluated here may indicate that it is often unnecessary to
526 subset loci for model training to those that are putatively adaptive, testing multiple marker
527 sets during model exploration will be important to understand model sensitivity (Lind et al.,
528 2024). These and other signals within data that could improve or mislead offset models,
529 including the choice of how and when to correct for population structure, also warrant further
530 investigation.

531 ***4.3 / The importance of adaptive environmental variables***

532 In empirical settings, the environmental drivers of local adaptation are rarely known *a*
533 *priori*. Even so, our results emphasize the importance of identifying these variables before
534 training offset models, as there were often declines in performance between models trained
535 using only adaptive environmental variables (*O-nuisance*) and those trained using additional
536 non-adaptive nuisance environmental variables (*N-nuisance*).

537 The importance of identifying these selective environments may be particularly germane
538 to two general empirical scenarios. In the first empirical scenario, sparsely sampling an
539 environmentally heterogeneous range may enrich genetic signals (e.g., coincident population
540 structure) most correlated to environmental variables that maintain a gradient across this
541 extent, and miss signals relevant to more local scales. In the second empirical scenario,
542 identifying the environmental variables underlying selection is particularly important when a
543 specific genomic offset method is ill-suited to differentiate importance among input variables.

544 For instance, RDA (and therefore $\text{RDA}_{\text{offset}}$) assumes that the environmental variables used
545 to build models are not collinear; (as implemented here; Capblancq & Forester, 2021; Legendre
546 & Legendre, 2012). Because of this, empirical datasets must be limited to a subset of available
547 environmental measures. The process of excluding environmental variables in this way may
548 omit signals of adaptive drivers (particularly when true drivers are not well measured), or
549 perhaps incorporate environmental variables that do not coincide with drivers of selection. In
550 these cases, performance is likely to decline. As such, this may indicate that methods such as
551 $\text{RDA}_{\text{offset}}$ are likely to perform worse in, or less uniformly across, realistic empirical settings
552 than what our current findings suggest.

553 On the other hand, users of GF may be tempted to include a large number of environmental
554 variables in training, hoping that GF can accurately attribute the correct environmental
555 variation to adaptive genetic structure. Our results show that it is not necessarily the case
556 that GF will give the highest importance values to the true adaptive environmental variables.
557 Indeed, weighted feature importance scores from GF models still incorrectly ranked the
558 adaptive environments below neutral environments in 20%-27% of the datasets, depending on
559 which marker set was used. These importance values ultimately affect the model predictions.
560 Including all available environmental variables may therefore negatively impact $\text{GF}_{\text{offset}}$
561 performance and could have weakened overall performance in previous empirical evaluations
562 that used a large number of environmental measures in training (e.g., Lind et al., 2024).

563 There are some differences between the nuisance environmental variables implemented here
564 and those that have been implemented previously. For instance, Láruson et al. (2022) created

565 nuisance variables by randomly sampling a multivariate normal distribution. In contrast to
566 findings here, Láruson et al. (2022) found that model performance was relatively unaffected
567 with the addition of nuisance variables. The minimal influence of nuisance variables on
568 performance found by Láruson et al. (2022) may differ from the performance declines reported
569 here because the nuisance variables we used were spatially autocorrelated, while those from
570 Láruson et al. (2022) were not. Inclusion of nuisance variables that are spatially autocorrelated
571 may mislead offset models more generally than variables with little spatial autocorrelation
572 because of the spurious relationship between environmental structure and genetic structure.

573 ***4.4 / The effect of environmental novelty***

574 While within-landscape performance generally increased with $LA_{\Delta SA}$, the datasets with the
575 greatest levels of $LA_{\Delta SA}$ were also the datasets where performance declined most readily with
576 climate novelty. This occurred because locally adapted metapopulations were under strong
577 selection to be fine-tuned to their environment, and as a result most individuals suffered severe
578 fitness declines with environmental change. In contrast, less locally adapted metapopulations
579 were under weaker selection, and suffered less steep fitness declines with environmental change.
580 This result highlights an interesting paradox: offset methods that have the highest performance
581 in common garden transplants under current climates (because of strong local adaptation)
582 may have the lowest performance in predicting “genomic vulnerability” as the range of climate
583 variables become more novel compared to the ranges used in training the model.

584 Thus, it will be important for future studies to clearly show what offset predictions have
585 been extrapolated to environments that have no analog in the data used to train the model.
586 This is particularly relevant for applications of offset methods that attempt to estimate the *in*
587 *situ* risk of climate change where the environment is expected to be increasingly novel. Global
588 climate novelty is often measured with respect to historical variability across a species range
589 (e.g., Lotterhos et al., 2021; Mahony et al., 2017; Williams et al., 2007). Terrestrial systems
590 could experience change in excess of three standard deviations relative to historic values
591 (Williams et al., 2007), with similar indices in marine systems indicating potential for even
592 greater novelty (Lotterhos et al., 2021). We observed performance declines below the analogous
593 $z_E=3.13$ standard deviations *Climate Novelty* scenario, indicating offset predictions could be
594 inaccurate in many real-world climate change predictions. These issues are also germane to
595 measures derived from offset values (Gougherty et al., 2021; Lachmuth, Capblancq, Keller, et
596 al., 2023; Lachmuth, Capblancq, Prakash, et al., 2023), which currently do not consider the
597 degree of climate novelty in the prediction (but see DeSaix et al., 2022 for an empirical example
598 that highlights model extrapolations).

599 Our results present a best-case scenario for predicting performance in novel environments,
600 as in many cases there will be biological reasons as to why climate-fitness relationships will
601 differ in future environments from relationships measured within the contemporary climate
602 space (see Fig. 5 in Capblancq et al., 2020). The simulations used here assumed a linear
603 relationship between environment (whether current or novel) and optimal trait value, without
604 a cost for the production of extreme traits. Thus, by assuming linearity in the simulations, we

605 presented methods with a straightforward relationship to extrapolate to novel climates. In
606 empirical settings, the relationship between the selective environment and optimal trait value
607 could be nonlinear. Developing a clearer understanding of the relationship between the
608 environment and optimal trait value, as well as various non-linear methods for extrapolation
609 (e.g., see Supplemental Text S1.1), are important areas of future work.

610 **4.5 / Genomic offsets in practice**

611 Our evaluations show that genomic offset methods hold promise for predicting
612 maladaptation to environmental change for metapopulations that evolve strong local
613 adaptation, and within environments similar to those in the data used to train the model.
614 However, our analyses also emphasize the limits of these methods in some scenarios and the
615 variability of performance across methods. Although GF_{offset} and RDA-uncorrected generally
616 had greater performance than RDA-corrected, LFMM2, or RONA, there was no single method
617 that outperformed the others across all situations. For instance, unlike other methods that
618 performed best in *Stepping Stones – Clines* landscapes, RDA-corrected had highest
619 performance in *Estuary – Clines* landscapes. RDA-corrected also had greater performance than
620 other methods in many of the datasets that developed low levels of $LA_{\Delta SA}$, but this was not
621 the case when local adaptation was high. In practice, species that are locally adapted to
622 measurable environmental variables will be best suited for offset methods when predicting the
623 relative performance of populations in a contemporary common garden, but paradoxically

624 these species may be least suited to using these methods to predict their vulnerability to novel
625 climates.

626 Together, these results indicate that some genomic offset methods may be suited to guide
627 initiatives such as near-term assisted gene flow, where targeted restoration sites within a
628 species range have climates that are similar to those used to train offset models. Even so, our
629 results also show that the performance of these methods are often variable across a landscape,
630 indicating that high performance at one site does not mean the offset model will perform well
631 at another. While genomic offset methods may be suitable for assisted gene flow initiatives,
632 they may be less suited for assisted migration programs where populations are moved outside
633 of their native range and environments differ from training data.

634 Before genomic offsets can be incorporated into management plans, considerable thought
635 must be put into the sensitivity of model outcomes from input data (e.g., from marker sets
636 and the populations used; Lind et al., 2024), the uncertainty inherent in environmental or
637 climate forecasts (Lachmuth, Capblancq, Keller, et al., 2023), as well as the degree of novelty
638 of future climates (DeSaix et al., 2022, this study). While accurate predictions are limited for
639 novel climates of the future, these offset methods could still be used to guide management in
640 the intervening time in a stepwise manner where experiments can be used to validate model
641 performance in practice. Using simulations tailored to the life history of target species also
642 presents a promising avenue to understand limitations of these methods for specific
643 management cases.

644

645 **Acknowledgements**

646 This research was funded by NSF-2043905 (KEL) and Northeastern University. The funding
647 bodies did not have any role in the design of the study, analysis, interpretation of results, or
648 in writing of the manuscript. We thank four anonymous reviewers for helpful comments that
649 improved the clarity and presentation of our results.

650 **Author Contributions**

651 KEL received funding. KEL and BML conceptualized the project and methodology. With
652 input, editing, and feedback from KEL, BML wrote code to train and evaluate offset models,
653 created figures, curated coding and records for archiving, and wrote the manuscript.

654 **Conflict of Interests**

655 The authors declare no conflicts of interest.

656 **Data Availability**

657 We reference the analysis code in the text of our documents by designating Supplemental
658 Code (SC) using a directory numbering system from our servers (as opposed to the order
659 listed in the manuscript). Supplemental Code includes both executable scripts (*.R, *.py) as
660 well as jupyter notebooks (*.ipynb). For example, for Script 3 in Directory 1, we refer to SC
661 01.03; for Notebook 5 in Subfolder 3 of Directory 2, we will refer to SC 02.03.05. Each
662 Directory will be archived on Zenodo.org and include a citation below, which will also link
663 to the GitHub repository. Notebooks are best viewed within a local jupyter or jupyter lab
664 session (to enable cell output scrolling / collapsing), but can also be viewed at
665 nbviewer.jupyter.org using the web link in the archive's README on GitHub. Analyses
666 were carried out primarily using python v3.8.5 and R v3.5.1 and v4.0.3. Exact package and
667 code versions are available at the top of each notebook. More information on coding
668 workflows and coding environments can be found in Supplemental Text S1-S2. Data used
669 for analysis has been archived previously (Lotterhos 2023b).

670

671 Code has been archived on Zenodo (Lind, 2024), which mirrors the GitHub repository at
672 <https://github.com/ModelValidationProgram/MVP-offsets>

673 **References**

- 674 Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to
675 measuring local adaptation. *Ecology Letters*, 16(9), 1195–1205.
676 <https://doi.org/10.1111/ele.12150>
- 677 Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the
678 Climate Benefits of Forests. *Science*, 320(5882), 1444–1449.
679 <https://doi.org/10.1126/science.1155121>
- 680 Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the Effects of
681 Geographic and Ecological Isolation on Genetic Differentiation. *Evolution*, 67(11),
682 3258–3273. <https://doi.org/10.1111/evo.12193>
- 683 Brown, J. H. (1984). On the Relationship between Abundance and Distribution of
684 Species. *The American Naturalist*, 124(2), 255–279.
685 <https://doi.org/10.1086/284267>
- 686 Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R.
687 (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future
688 Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*,
689 51(1), 245–269. <https://doi.org/10.1146/annurev-ecolsys-020720-042553>
- 690 Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for
691 landscape genomics. *Methods in Ecology and Evolution*.
692 <https://doi.org/10.1111/2041-210x.13722>
- 693 Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge,
694 D. M., Pascual, M., Jr., R. P., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A.,
695 Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological
696 Forecasts: An Emerging Imperative. *Science*, 293(5530), 657–660.
697 <https://doi.org/10.1126/science.293.5530.657>
- 698 DeSaix, M. G., George, T. L., Seglund, A. E., Spellman, G. M., Zavaleta, E. S., &
699 Ruegg, K. C. (2022). Forecasting climate change response in an alpine specialist
700 songbird reveals the importance of considering novel climate. *Diversity and
701 Distributions*, 28(10), 2239–2254. <https://doi.org/10.1111/ddi.13628>

- 702 Doney, S. C., Ruckelshaus, M., Duffy, J. E., Barry, J. P., Chan, F., English, C. A.,
703 Galindo, H. M., Grebmeier, J. M., Hollowed, A. B., Knowlton, N., Polovina, J.,
704 Rabalais, N. N., Sydeman, W. J., & Talley, L. D. (2012). Climate Change
705 Impacts on Marine Ecosystems. *Annual Review of Marine Science*, 4(1), 11–37.
706 <https://doi.org/10.1146/annurev-marine-041911-111611>
- 707 Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological
708 Explanation and Prediction Across Space and Time. *Annual Review of Ecology,*
709 *Evolution, and Systematics*, 40(1), 677–697.
710 <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- 711 Fitzpatrick, M. C., Blois, J. L., Williams, J. W., Nieto-Lugilde, D., Maguire, K. C., &
712 Lorenz, D. J. (2018). How will climate novelty influence ecological forecasts?
713 Using the Quaternary to assess future reliability. *Global Change Biology*, 24(8),
714 3575–3586. <https://doi.org/10.1111/gcb.14138>
- 715 Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021).
716 Experimental support for genomic prediction of climate maladaptation using the
717 machine learning approach Gradient Forests. *Molecular Ecology Resources*.
718 <https://doi.org/10.1111/1755-0998.13374>
- 719 Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level
720 modelling of biodiversity: mapping the genomic landscape of current and future
721 environmental adaptation. *Ecology Letters*, 18(1), 1–16.
722 <https://doi.org/10.1111/ele.12376>
- 723 Gain, C., & François, O. (2021). LEA 3: Factor models in population genetics and
724 ecological genomics with R. *Molecular Ecology Resources*, 21(8), 2738–2748.
725 <https://doi.org/10.1111/1755-0998.13366>
- 726 Good, R. D. (1931). A THEORY OF PLANT GEOGRAPHY. *New Phytologist*, 30(3),
727 149–149. <https://doi.org/10.1111/j.1469-8137.1931.tb07414.x>
- 728 Gougherty, A. V., Keller, S. R., & Fitzpatrick, M. C. (2021). Maladaptation, migration
729 and extirpation fuel climate change risk in a forest tree species. *Nature Climate
730 Change*, 1–15. <https://doi.org/10.1038/s41558-020-00968-6>
- 731 Hereford, J. (2009). A Quantitative Survey of Local Adaptation and Fitness Trade-Offs.
732 *The American Naturalist*, 173(5), 579–588. <https://doi.org/10.1086/597611>

- 733 Hoegh-Guldberg, O., & Bruno, J. F. (2010). The Impact of Climate Change on the
734 World's Marine Ecosystems. *Science*, 328(5985), 1523–1528.
735 <https://doi.org/10.1126/science.1189930>
- 736 Lachmuth, S., Capblancq, T., Keller, S. R., & Fitzpatrick, M. C. (2023). Assessing
737 uncertainty in genomic offset forecasts from landscape genomic models (and
738 implications for restoration and assisted migration). *Frontiers in Ecology and
739 Evolution*, 11, 1155783. <https://doi.org/10.3389/fevo.2023.1155783>
- 740 Lachmuth, S., Capblancq, T., Prakash, A., Keller, S. R., & Fitzpatrick, M. C. (2023).
741 Novel genomic offset metrics account for local adaptation in climate suitability
742 forecasts and inform assisted migration. *BioRxiv*, 2023.06.05.541958.
743 <https://doi.org/10.1101/2023.06.05.541958>
- 744 Láruson, Á. J., Fitzpatrick, M. C., Keller, S. R., Haller, B. C., & Lotterhos, K. E.
745 (2022). Seeing the forest for the trees: Assessing genetic offset predictions from
746 gradient forest. *Evolutionary Applications*, 15(3), 403–416.
747 <https://doi.org/10.1111/eva.13354>
- 748 Lee-Yaw, J. A., McCune, J. L., Pironon, S., & Sheth, S. N. (2022). Species distribution
749 models rarely predict the biology of real populations. *Ecography*, 2022(6).
750 <https://doi.org/10.1111/ecog.05877>
- 751 Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (Vol. 24). Elsevier.
752 Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants.
753 *PLoS ONE*, 3(12), e4010. <https://doi.org/10.1371/journal.pone.0004010.s001>
- 754 Lind, B. M. (2024) GitHub.com/ModelValidationProgram/MVP-offsets. Revision release
755 (v1.0.1). Zenodo. DOI: <https://doi.org/10.5281/zenodo.11209812>
- 756 Lind, B. M., Candido-Ribeiro, R., Singh, P., Lu, M., Vidakovic, D. O., Booker, T. R.,
757 Whitlock, M., Isabel, N., Yeaman, S., & Aitken, S. N. (2024). How useful is
758 genomic data for predicting maladaptation to future climate? *Global Change
759 Biology*. <https://doi.org/10.1111/gcb.17227>
- 760 Lotterhos, K. E. (2023a). The paradox of adaptive trait clines with nonclinal patterns in
761 the underlying genes. *Proceedings of the National Academy of Sciences*, 120(12).
762 <https://doi.org/10.1073/pnas.2220313120>
- 763 Lotterhos, K. (2023b) Output model data from paradox of adaptive trait clines with

- 764 non-clinal patterns in the underlying genes (Model Validation Program project).
765 Biological and Chemical Oceanography Data Management Office (BCO-DMO).
766 (Version 1) Version Date 2023-02-13 [if applicable, indicate subset used].
767 doi:10.26008/1912/bco-dmo.889769.1 [access date]
- 768 Lotterhos, K. E., Fitzpatrick, M. C., & Blackmon, H. (2022). Simulation Tests of
769 Methods in Evolution, Ecology, and Systematics: Pitfalls, Progress, and
770 Principles. *Annual Review of Ecology, Evolution, and Systematics*, 53(1), 113–
771 136. <https://doi.org/10.1146/annurev-ecolsys-102320-093722>
- 772 Lotterhos, K. E., Láruson, Á. J., & Jiang, L.-Q. (2021). Novel and disappearing climates
773 in the global surface ocean from 1800 to 2100. *Scientific Reports*, 11(1), 15535.
774 <https://doi.org/10.1038/s41598-021-94872-4>
- 775 Mahony, C. R., Cannon, A. J., Wang, T., & Aitken, S. N. (2017). A closer look at novel
776 climates: new methods and insights at continental to landscape scales. *Global
777 Change Biology*, 23(9), 3934–3955. <https://doi.org/10.1111/gcb.13645>
- 778 Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of
779 genomic offset in conservation management. *Evolutionary Applications*, 14(5),
780 1202–1212. <https://doi.org/10.1111/eva.13205>
- 781 Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., Bodénès, C.,
782 Sperisen, C., Kremer, A., & Gugerli, F. (2016). Signatures of local adaptation in
783 candidate genes of oaks (*Quercus* spp.) with respect to present and future
784 climatic conditions. *Molecular Ecology*, 25(23), 5907–5924.
785 <https://doi.org/10.1111/mec.13889>
- 786 Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models
787 supporting environmental decision making: a strategy for the future. *Trends in
788 Ecology & Evolution*, 25(8), 479–486. <https://doi.org/10.1016/j.tree.2010.05.001>
- 789 Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C.,
790 Kofler, R., Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O.,
791 Bergland, A., Flatt, T., Guillaume, F., & Pfenninger, M. (2020). Evolutionary
792 genomics can improve prediction of species' responses to climate change.
793 *Evolution Letters*, 4(1). <https://doi.org/10.1002/evl3.154>
- 794 Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular
795 Ecology*, 23(23), 5649–5662. <https://doi.org/10.1111/mec.12938>

- 796 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of
797 population structure. *Evolution*, 38(6), 1358–1370.
- 798 Williams, J. W., Jackson, S. T., & Kutzbach, J. E. (2007). Projected distributions of
799 novel and disappearing climates by 2100 AD. *Proceedings of the National
800 Academy of Sciences*, 104(14), 5738–5742.
801 <https://doi.org/10.1073/pnas.0606292104>
- 802

Figures for:

The accuracy of predicting maladaptation to new environments with genomic data

Brandon M. Lind^{*}, Katie E. Lotterhos

Department of Marine and Environmental Sciences, Northeastern University
430 Nahant Road, Nahant, MA 01908, USA

24 July 2024

Running Title: Limits of genomic offsets

Keywords: genomic offset, environmental change, climate change, assisted gene flow, genomic forecasting, restoration

***Corresponding Author**

Brandon M. Lind

Email: lind.brandon.m@gmail.com

Method	abbr.	Multivariate?	Structure correction?
gradientForest ¹	GF _{offset}	Yes	No
Redundancy Analysis ² with population structure correction	RDA-corrected	Yes	Yes, with axes loadings from PCA*
Redundancy Analysis ² without population structure correction	RDA-uncorrected	Yes	No
Latent factor mixed model from Landscape and Ecological Association Studies R package ³	LFMM2 _{offset}	Yes	Yes, with latent factors
Risk Of Non-Adaptedness ⁴	RONA	No	No

* principal component analysis

Table 1 Genomic offset methods used for evaluation. Genomic offset methods differ in their capability to use multivariate environmental data in training as well as whether a correction for population genetic structure is applied. Superscripts apply to the following reference citations: 1 - Fitzpatrick & Keller, 2015; 2 - Capblancq & Forester, 2021; 3 - Gain & François, 2021; 4 - Rellstab et al., 2016.

Workflow	n_{traits}	(1) Simulations Levels (replicates per level)	(3, 7) Environmental Data	Training and Prediction?	(6) Within- landscape Evaluation?	(9) Total Performance Evaluations
Adaptive Environment (AE)	1-trait	45 (10)	temp	Yes	Yes	675,000
	2-trait	180 (10)	temp + Env2			3,240,000
	6-trait	1 (1)	MAT + MTwetQ + MTDQ + PDM + PwarmQ + PWM			3,000
Nuisance Environment (NE)**	1-trait	45 (1)	[AE _{1-trait} environments + Env2] +/- [ISO + PSsd] +/- [TSsd]	Yes	Yes	175,500
	2-trait	180 (1)	[AE _{2-trait} environments + ISO + PSsd] +/- [TSsd]			432,000*
	6-trait	1 (1)	[AE _{6-trait} environments + ISO + PSsd + TSsd]			1,200*
Climate Novelty (CN)*	1-trait	45 (10)	AE _{1-trait} environments	Prediction only (using AE training models)	No	64,800*
	2-trait	180 (10)	AE _{2-trait} environments			259,200*
	6-trait	1 (1)	AE _{6-trait} environments			144*

* excludes RONA

** The set of population values for each unique nuisance environment were the same across traits and landscapes

† includes evaluation of climate center and eleven Climate Novelty scenarios

Table 2 Workflows used to process simulation data for the evaluation of genomic offset methods. Numbers given in column names refer to locations in schematic of Fig. 1. The *Adaptive Environment* workflow processes all population data from 1- and 2-trait (example shown in Fig. 1) as well as 6-trait simulations using only adaptive environmental variables in training, and evaluates performance in each garden on the metapopulation landscape. The *Nuisance Environment* workflow processes 1-, 2-, and 6-trait simulations similarly to the *Adaptive Environment* workflow, except in addition to adaptive environmental variables used in training, non-adaptive (i.e., nuisance) environmental variables are also used - each bracketed set of environmental variables indicate a distinct nuisance level (e.g., “1-trait 1-nuisance” = [AE_{1-trait} environments + Env2] and “1-trait 4-nuisance” = [AE_{1-trait} environments + Env2 + ISO + PSsd + TSsd]). The *Climate Novelty* workflow uses trained models from the *Adaptive Environment* workflow (Fig. 1A-5) and evaluates offset in 11 novel environments relative to the range of environments used in training. See Supplemental Note S3 for details regarding the choice of *Climate Novelty* environmental values and visualizations of climate data in principal component space. See Supplemental Notes S1-S2 for descriptions of coding workflows. Total evaluations = 4,849,644. Counts of evaluations were tabulated in SC 02.10.01.

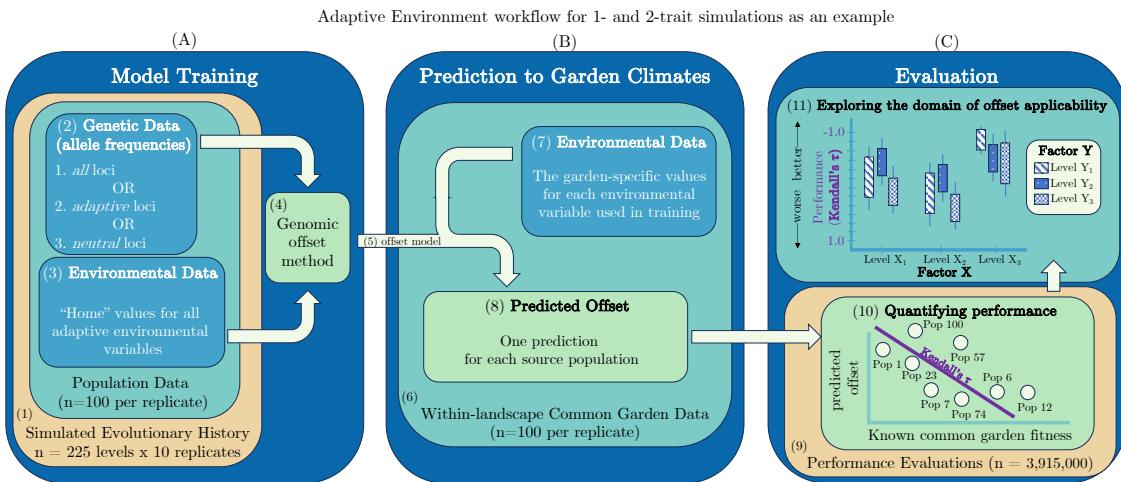
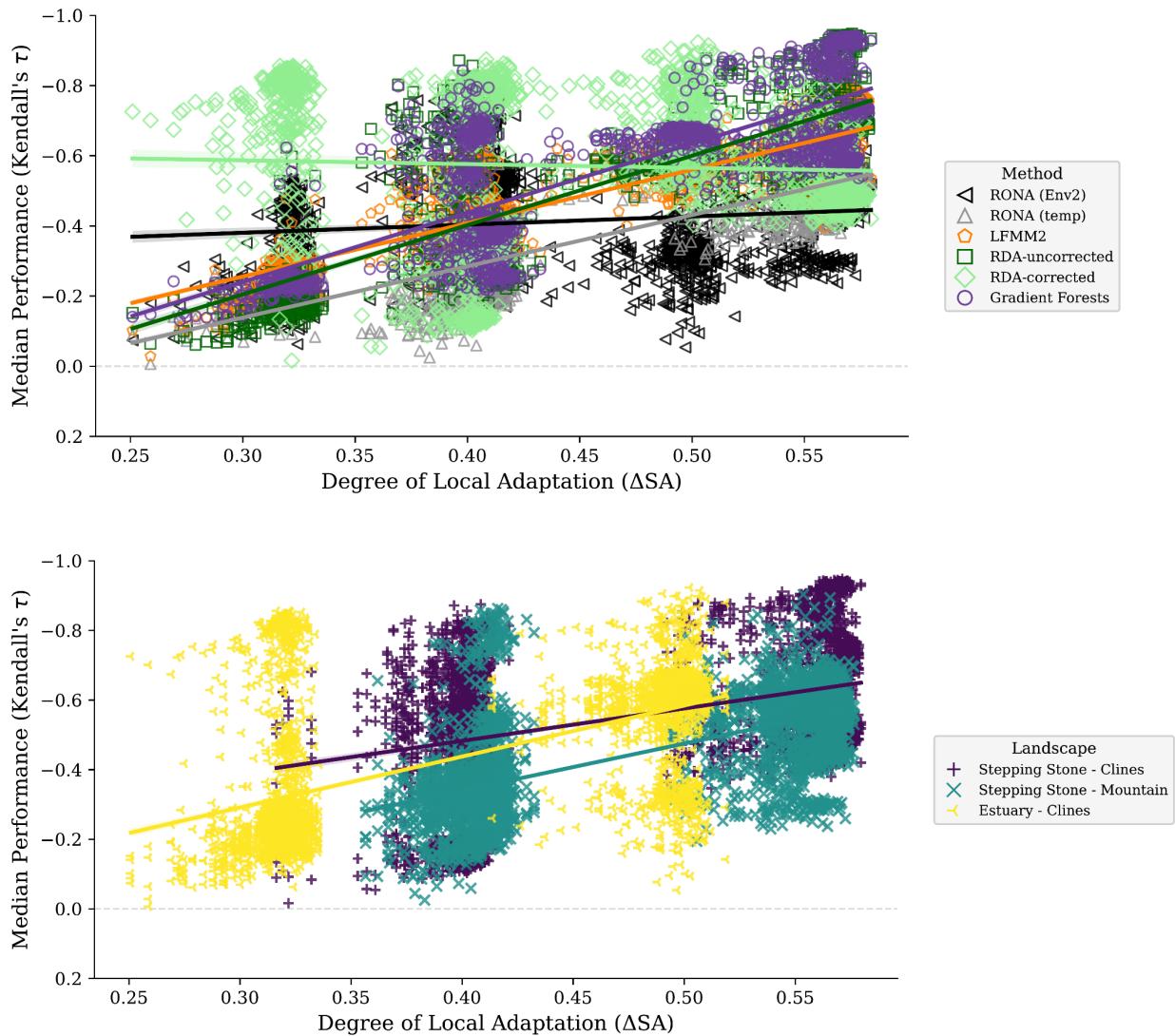


Figure 1 Analysis of 1-, 2-, and 6-trait simulations included three main phases: A) model training, B) model prediction, and C) evaluation of models. The *Adaptive Environment* workflow is shown as an example of the processing of 1- and 2-trait simulation data for genomic offset evaluation. In total, three general workflows are used to evaluate genomic offset methods (Table 2). Subpanels of this schematic are numbered for referencing in Table 2 and the main text.

(Fig. 2)



(Fig. 2 continued)

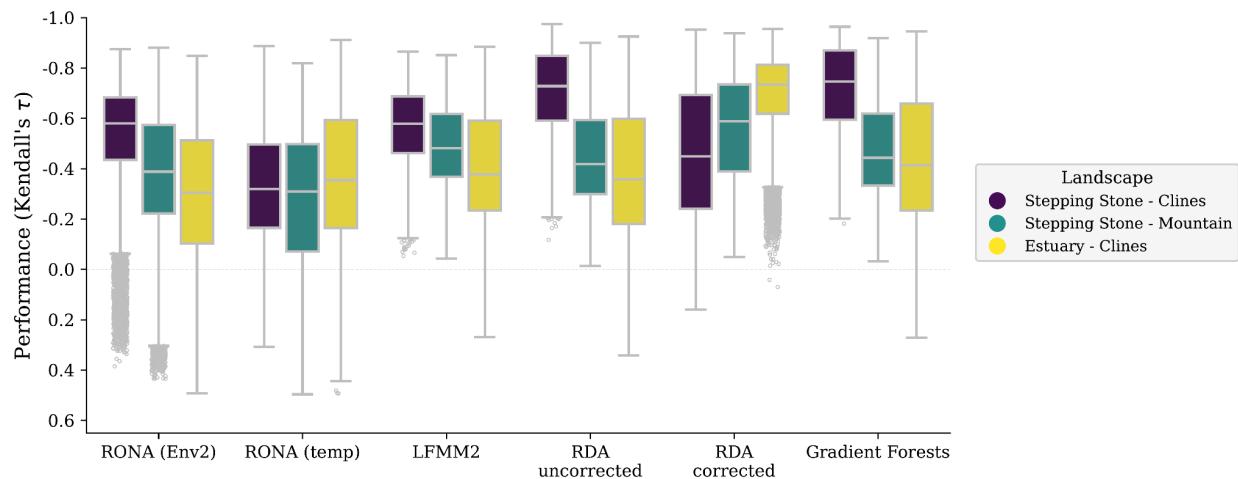


Figure 2 Predictive performance of genomic offset models (y-axes) is driven by the degree of local adaptation (A) and the spatial patterns of adaptive environments across the landscape (B, C). For each model, a median value from performance scores from 100 common gardens is shown for A and B; C shows scores across all common gardens for each model (note that y-axes are inverted, as more negative values have higher performance). Data included in these figures was processed through the *Adaptive Environment* workflow but only includes models trained using 2-trait simulations and *all* loci. Code to create (A) and (B) can be found in SC 02.02.02; code to create (C) can be found in SC 02.02.01.

(Fig. 3)

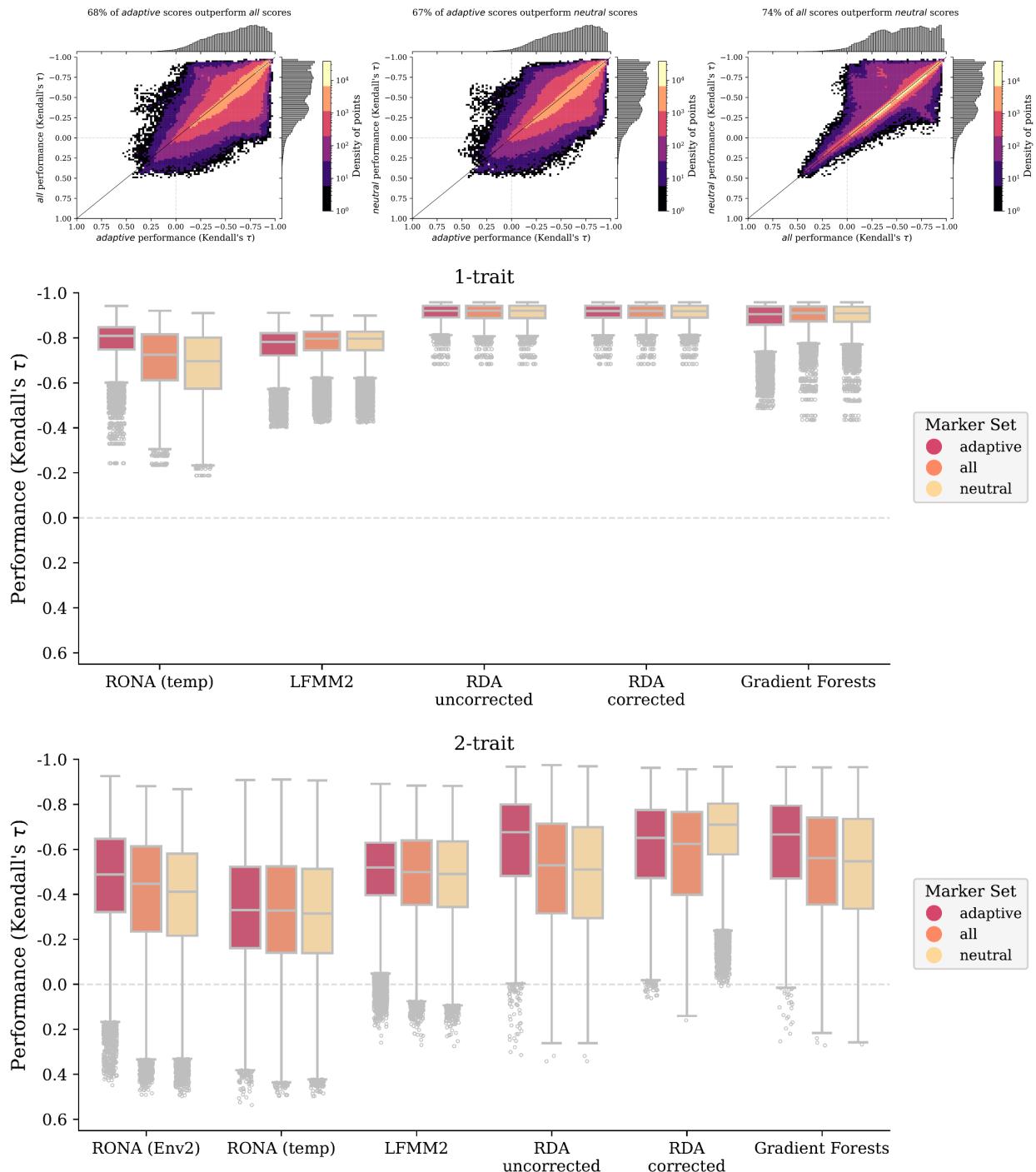


Figure 3 Comparison of marker choice across genomic offset methods for 1- and 2-trait simulations. A-C are scatterplots of pairwise comparisons of performance between marker sets (histograms in each margin) from both 1- and 2-trait models where density of points is indicated by color in legend (note color scale is different for each figure to accentuate patterns in data). D-E are boxplots from the same data in A-C separated by individual traits. Data included in these figures is from all 1- and 2-trait models from the *Adaptive Environment* workflow. Code to create these figures can be found in SC 02.02.03.

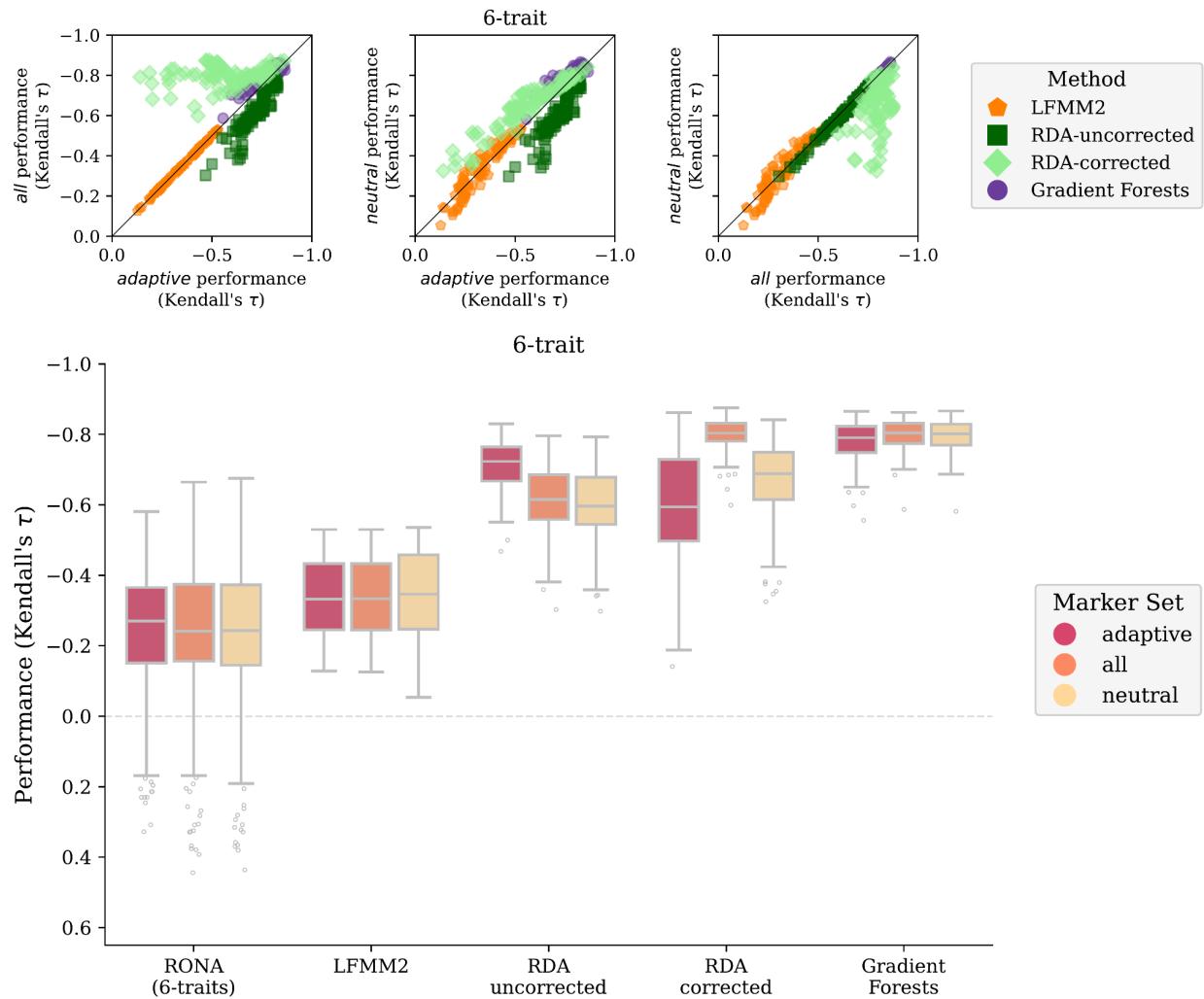


Figure 4 Comparison of marker choice across genomic offset methods for the 6-trait simulation. A-C are scatterplots of pairwise comparisons of performance between marker sets (RONA is not shown, except in SN 02.05.10). D are boxplots from the same data in A-C (RONA_{6-trait} is the combined performance across all six environmental models). Data included in this figure is from the 6-trait models processed through the *Adaptive Environment* workflow. Note there is only one 6-trait replicate, and variation within figures represents the performance across 100 common gardens for each method. Code to create these figures can be found in SN 02.05.10.

(Fig. 5)

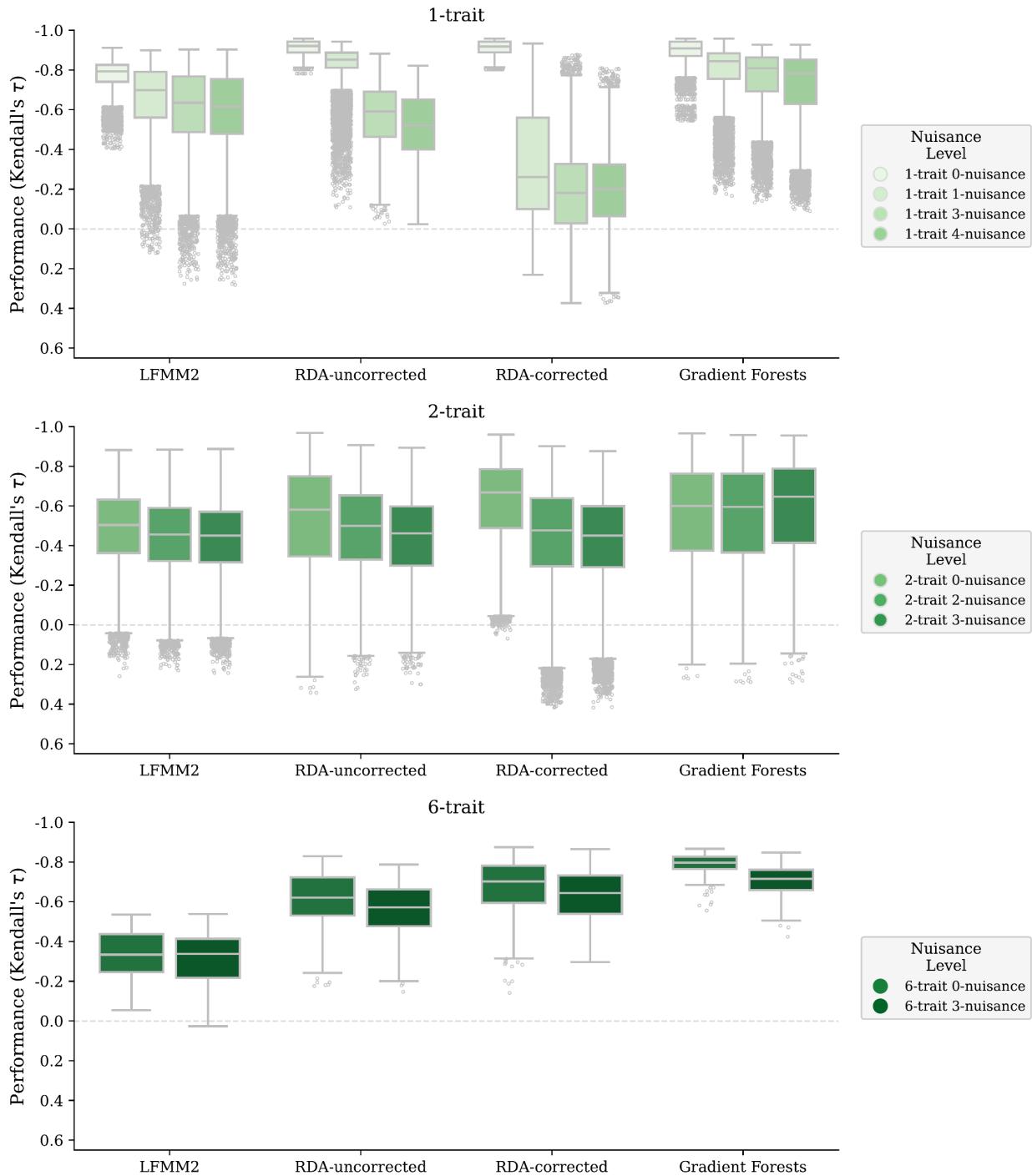


Figure 5 Effect of non-adaptive nuisance environmental variables on offset performance. Shown are evaluations of offsets from 1- and 2-trait models trained using only adaptive environments (*0-nuisance*) or with adaptive environments and the addition of $N>0$ non-adaptive environmental variables (*N-nuisance*). RONA is not shown because it is univariate with respect to environmental variables. Nuisance variables are listed in Table 2. Code to create figures can be found in SC 02.02.06 and SC 02.02.08.

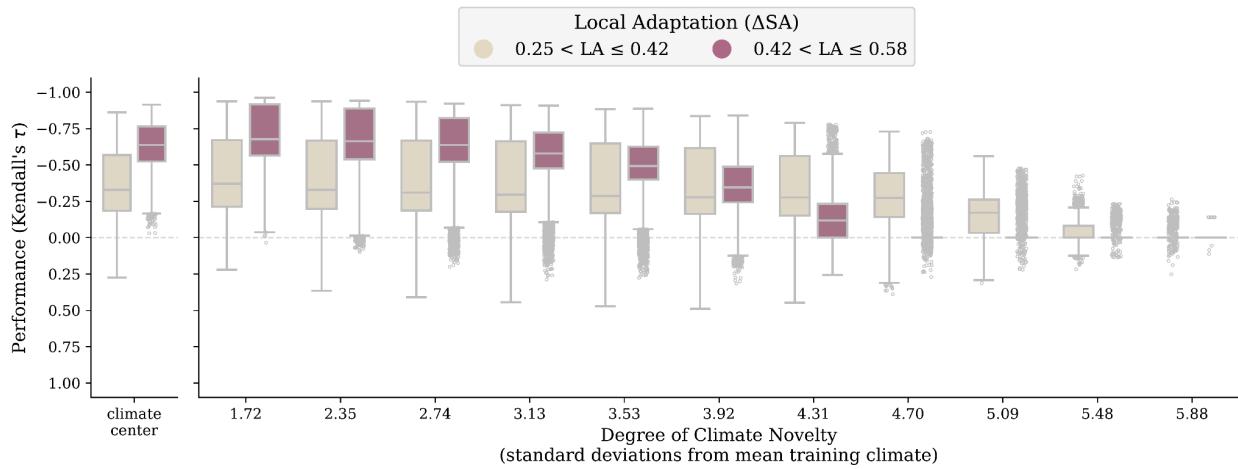


Figure 6 Performance decays with climate novelty relative to training data. Shown is model performance (y-axes) across methods at climate center (A) and across common gardens each representing increasing degrees of climate novelty relative to training data (x-axis of B) where all 100 populations have been transplanted. The standard deviation values (x-axis, B) are applicable to all environments for all landscapes except for *Env2* in the *Stepping Stone - Mountain* landscape; the corresponding standard deviation values are 1.55, 2.12, 2.47, 2.82, 3.18, 3.53, 3.88, 4.24, 4.60, 4.95, 5.3. When fitness for all transplanted individuals was zero, a model's performance was undefined and manually set to 0; no method predicted a single offset value for all populations in these situations. Setting undefined performance to 0 did not substantially impact patterns between performance and climate novelty, and is explored in Supplemental Text S3. Data included in this figure are from models trained using 1- and 2-trait simulations from the *Climate Novelty* workflow, which excludes both RONA_{temp} and RONA_{Env2}. Code used to create this figure can be found in SC 02.04.05.

1 Supplemental Information

The accuracy of predicting maladaptation to new environments with genomic data

Brandon M. Lind^{*}, Katie E. Lotterhos

Department of Marine and Environmental Sciences, Northeastern University
430 Nahant Road, Nahant, MA 01908, USA

24 July 2024

8 Running Title: *Performance of genomic offset methods*

Keywords: genomic offset, environmental change, climate change, assisted gene flow, genomic forecasting, restoration

11 *Corresponding Author

12 Brandon M. Lind

13

14 Email: lind.brandon.m@gmail.com

15 Table of Contents

16	Supplemental Notes	6
17	S1 - Implementation of Offset Methods.....	6
18	1.1 gradientForest.....	6
19	1.2 The Risk Of Non-Adaptedness	7
20	1.3 Landscape and Ecological Association (LEA) Studies R package	8
21	Fig S1 Distribution of K used for the lfmm2 genetic.offset function for 1- and 2-trait simulations.....	9
22	Fig S2 Percent variance explained from principal component (PC) axes from principal component analysis of SNP data from the 6-trait simulation.....	10
23	1.4 Redundancy Analysis	11
24	Fig S3 Performance of RDA-outlier markers are on par with other marker sets for (A) 1-trait, (B) 2-trait, and (C) 6-trait evaluations of offset estimated with (RDA-corrected) or without (RDA-uncorrected) population structure correction..	
25	13
30	S2 - Coding workflows	14
31	1.1 The Adaptive Environment coding workflow.....	14
32	1.1.1 1- and 2-trait simulations.....	15
33	1.1.2 6-trait simulation	16
34	1.2 The Climate Novelty coding workflow	16
35	1.3 The Nuisance Environment coding workflow	17
36	1.4 Misc.....	17
37	S3 - Defining Climate Novelty scenarios	18
38	Fig S5 Differentiation of <i>Climate Novelty</i> environments (blue stars, including climate center) from within-landscape environments (black circles) using Principal Component Analysis (PCA) of environmental data.....	19
41	S4 - Missing data in Climate Novelty evaluations.....	20
42	Fig S33 The effect of simulation parameters on missing data for <i>Climate Novelty</i> scenarios.....	25
44	Supplemental Tables.....	26
45	Table S1 Results from Type II ANOVAs from regressing simulation factors on offset performance (see Equation 1 of the main text)..	26
47	Table S2 Results from Type II ANOVAs from regressing the proportion of clinal QTNs (cor_TPR_tmp and cor_TPR_sal) and clinal neutral alleles (cor_FPR_temp_neutSNPs, cor_FPR_sal_neutSNPs) on offset performance (see Equation 2 of the main text).....	27
51	Table S3 Results from Type II ANOVAs regressing two factors - degree of local adaptation (final_LA) and levels of isolation-by-environment in all marker sets)	

53	on offset performance.....	28
54	Table S4 Gradient Forests (GF) sometimes incorrectly identifies the	
55	environments driving adaptation.....	29
56	Supplemental Figures	30
57	Fig S4 Correlation (Spearman's rho) among environmental variables faceted by	
58	landscape.....	31
59	Fig S6 Percent sum of squares of the various factors from the ANOVA model in	
60	Table S1.....	34
61	Fig S7 Effect of the degree of local adaptation (x-axes) on method performance	
62	(y-axes) colored by the relative strength of selection on the two traits..	35
63	Fig S8 Effect of polygenicity on performance of offset methods trained using all	
64	markers on simulations with two adaptive traits.....	36
65	Fig S9 Effect of demography on performance of offset methods trained using all	
66	markers on simulations with two adaptive traits.....	37
67	Fig S10 Stacked bar plot of the percent sum of squares from Type II ANOVAs	
68	from regressing the proportion of clinal QTNs and clinal neutral alleles on offset	
69	performance (see Equation 2 of the main text)..	38
70	Fig S11 Impact on method performance (y-axes) from the proportion of QTNs	
71	with clinal relationships with temp (first column) or Env2 (second column).....	43
72	Fig S12 Stacked bar plot showing correlation between environmental variables	
73	and axes of population genetic structure (Principal Component Analysis axes [PC	
74	axes])..	44
75	Fig S13 Relationship between the proportion of clinal neutral loci for <i>temp</i> (y-	
76	axes, first row) or <i>Env2</i> (y-axes, second row) with the strength of the relationship	
77	between environmental variables and axes of population genetic structure.	
78	Purple = <i>Stepping Stone - Clines</i> ; teal = <i>Stepping Stone - Clines</i> ; yellow = <i>Estuary -</i>	
79	<i>Clines</i> . Data included in this figure is from all 2-trait simulations..	45
80	Fig S14 Relationship between median performance and absolute correlation	
81	(Pearson's <i>r</i>) between environmental variables and axes of population genetic	
82	structure (principal component analysis axes).....	51
83	Fig S15 Adaptive markers contain greater levels of isolation-by-environment	
84	(<i>IBE</i>) than other marker sets.....	52
85	Fig S16 The relationship between the degree of local adaptation (LA_{ASA}), levels	
86	of <i>IBE</i> within marker sets, and median performance of models trained with one	
87	of the three marker sets: (A) <i>adaptive</i> , (B) <i>all</i> , and (c) <i>neutral</i> marker sets.....	54
88	Fig S17 Levels of isolation-by-environment in marker sets vary across	
89	landscapes (A) and the degree of local adaptation reached by metapopulations	
90	on these landscapes (B).....	55
91	Fig S18 The levels of isolation-by-distance in marker sets (panels) are weakly	
92	correlated with the degree of local adaptation (LA_{ASA}) within simulation levels..	56

93	Fig S19 Differences in levels of <i>IBE</i> between marker sets used to train models is generally unrelated to differences in model performances.....	60
94		
95	Fig S20 A map of Garden ID (unbolded entries) across each landscape for 1-, 2- and 6-trait simulations (latitudinal and longitudinal grids are bolded).....	61
96		
97	Fig S21 Genomic offset methods have variable performance across the <i>Stepping-Stone - Clines</i> landscape.....	64
98		
99	Fig S22 Genomic offset methods have variable performance across the <i>Stepping-Stone - Mountain</i> landscape.....	67
100		
101	Fig S23 Genomic offset methods have variable performance across the <i>Estuary - Clines</i> landscape.....	70
102		
103	Fig S24 Variability of genomic offset performance (y-axes) for a given model (+) often decreases with increasing median performance (x-axes).....	72
104		
105	Fig S25 Variability across evaluations of genomic offsets often decreases with increasing average performance across marker sets.....	77
106		
107	Fig S26 Variability across evaluations of genomic offsets often decreases with increasing average performance across marker set.....	74
108		
109	Fig S27 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations.....	75
110		
111	Fig S28 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations.....	76
112		
113	Fig S29 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations.....	77
114		
115	Fig S30 Effect of non-adaptive nuisance environmental variables on offset performance faceted by landscape.....	78
116		
117	Fig S31 Effect of non-adaptive nuisance environmental variables on offset performance faceted by marker set.....	79
118		
119	Fig S32 Pairwise comparison of performance differences between marker sets for <i>Nuisance Environment</i> scenarios.....	84
120		
121	Fig S34 Pairwise comparison of performance differences between marker sets for <i>Climate Novelty</i> scenarios.....	86
122		
123	Supplemental References	87
124		

125 **Supplemental Notes**

126 **S1 - Implementation of Offset Methods**

127

128 See Supplemental Note S2 for specific citations of code.

129

130 **1.1 | gradientForest**

131 For a given set of input loci (*all*, *adaptive*, or *neutral*; see Q3 in Methods), and for all

132 workflows, `gradientForest` (`GFOffset`) is trained using `ntree=500`, `corr.threshold=0.5`,

133 and $\text{maxLevel}=(0.368 * \frac{N}{2})$, where N is the number of populations. Using default linear

134 extrapolation, the trained model is projected onto the landscape using the `'predict'` function

135 and the same environmental values used in training. This creates the “current” projection used

136 to calculate offset below. We considered a linear extrapolation to be appropriate because of the

137 linear relationships between fitness and environmental optima within simulations for both

138 current and future environments (discussed further in Section 4.4 of the Discussion).

139 The trained models are then fit to the climate of each of 100 common gardens on the

140 landscape for the *Adaptive Environment* and *Nuisance Environment* scenarios, or to each of the

141 11 *Climate Novelty* scenarios. Specifically, for each garden, the `'predict'` function is used to

142 take the trained model and the garden’s climate to create a projection similar to that using

143 current climate data (previous paragraph). Then the Euclidean distance is taken between the

144 current and future projections to calculate offset.

145 **1.2 | The Risk Of Non-Adaptedness**

146 For a given set of input loci (*all*, *adaptive*, or *neutral*; see Q3 in Methods), we first
 147 discarded any locus that did not have significant ($p \leq 0.05$) linear models relating population-level
 148 allele frequencies with environmental variables. p-values were not corrected for multiple testing.
 149 For each common garden, and once for each environmental variable, RONA offset for each
 150 population was calculated by averaging the absolute allele frequency difference between the
 151 population's current frequency and that predicted by using each locus' linear model fit using
 152 climate of the common garden,

$$\text{RONA} = \frac{1}{n} \sum_{i=1}^n |(S_{\text{present}_i} * \text{EF}_{\text{future}} + I_{\text{present}_i}) - \text{AAF}_{\text{present}_i}|$$

153
 154 where n is the total number of loci with significant linear models; S_{present} and I_{present} are
 155 respectively the slope and intercept from the linear model for locus_i relating current climate and
 156 allele frequencies from all populations; $\text{AAF}_{\text{present}}$ is the current allele frequency for the
 157 population under consideration; and $\text{EF}_{\text{future}}$ is the environmental value for the common garden.
 158 RONA can only be calculated for a single population and environmental variable at a time.

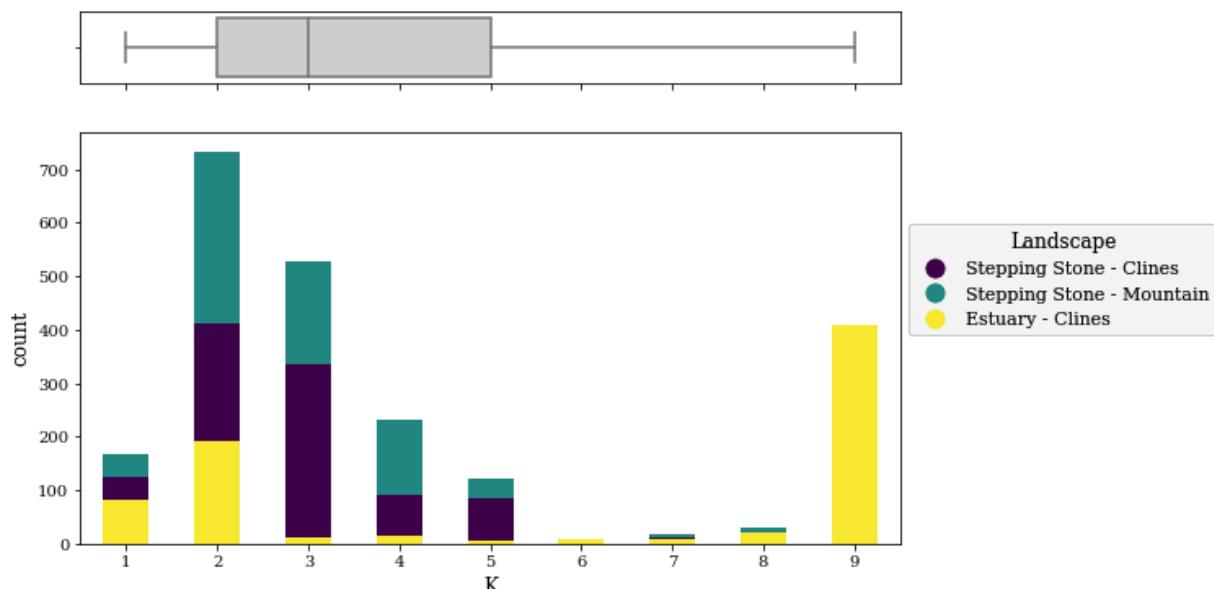
159 RONA was excluded from *Nuisance Environment* and *Climate Outlier* workflows because
 160 of its poor (Fig. 2A) and variable (Fig. 2C, Fig. 4) performance from evaluations from the *Adaptive*
 161 *Environment* workflow.

162 Of note, in some instances, particularly *Adaptive Environment* datasets simulated with
 163 oligogenic architectures, there were no loci with significant linear relationships with

164 environmental variables and these instances were given NA performance values (i.e., excluded
 165 from analyses).

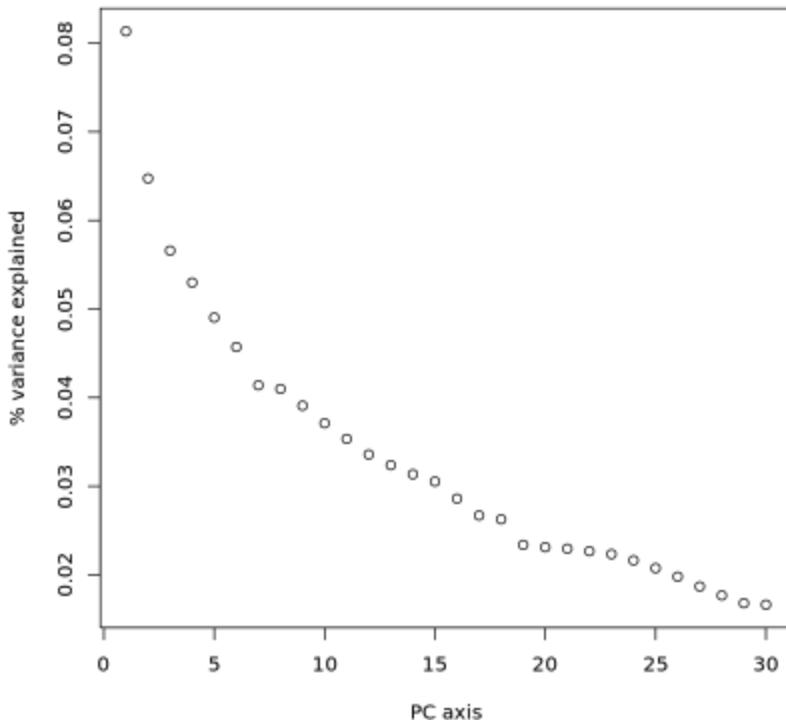
166 **1.3 | Landscape and Ecological Association (LEA) Studies R package**

167 We used the `genetic.offset` function in the LEA package to estimate $\text{LFMM2}_{\text{offset}}$ for
 168 each workflow (Fig. 1). The `genetic.offset` function was used with default settings, except
 169 for K , the number of subdivisions within the data. To determine K needed for the
 170 `genetic.offset` function for 1- and 2-trait simulations, we first used filtered SNP data (see
 171 Section 2.1) to estimate 21 principal components (PCs) using principal component analysis (PCA).
 172 Then we equated K to the number of PC axes that explain greater than 1.3x the variation of the
 173 next subsequent axis (see line 677-697 of [c-AnalyzeSimOutput.R](#) from Lotterhos, 2023). This
 174 resulted in varied K across simulation levels and replicates (Fig S1). For the 6-trait simulation, it
 175 was never the case that a PC axis explained >1.3x the variation explained by the previous axis, so
 176 we used the elbow rule to estimate $K=7$ (Fig S2).



177

178 **Fig S1** Distribution of K used for the Ifmm2 genetic.offset function for 1- and 2-trait
179 simulations. K was estimated by determining the number of principal component axes that
180 explain at least 1.3x times the amount of variation of the subsequent axis. Code used to create
181 this figure can be found in SC 02.09.01.

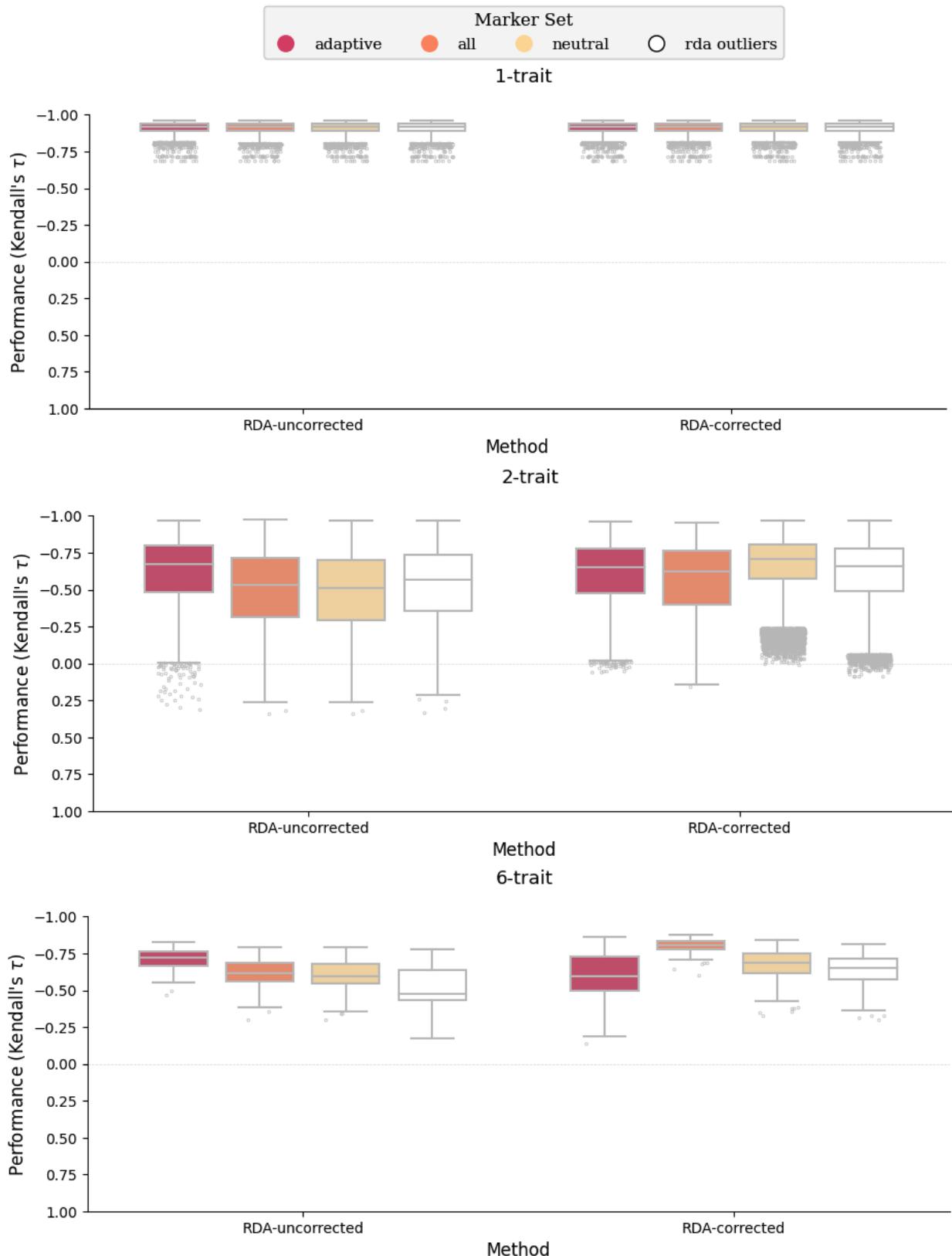


182

183 **Fig S2** Percent variance explained from principal component (PC) axes from principal
184 component analysis of SNP data from the 6-trait simulation. The “elbow rule” was used to
185 estimate $K=7$ for this simulation. Code used to create this figure can be found in SC 02.05.11.

186 1.4 | Redundancy Analysis

187 RDA_{offset} was implemented as in Capblancq & Forester (2021). Note that the
188 environmental variables used here across workflows had minimal correlation, as required by RDA
189 (Fig. S4). In addition to the three marker sets used as input (*all*, *adaptive*, or *neutral*; see Q3 in
190 Methods), we also used *RDA-outliers* as input to RDA offset estimation. *RDA-outlier* loci were
191 those from separate RDA models trained using *all* loci and adaptive environments, and were
192 included in this study because of their use in the original implementation of RDA_{offset} by
193 Capblancq & Forester (2021). *RDA-outliers* were identified as in Capblancq et al. (2018) for loci
194 with q-values < 0.05. For each 1-, 2-, and 6-trait simulation replicate, RDA_{offset} was estimated with
195 (RDA-corrected) and without (RDA-uncorrected) correction for population genetic structure.
196 When correcting for structure, the loadings for the first two PCs from PCA estimated with *all* loci
197 were used. We chose to use *all* markers because in many empirical settings researchers will not
198 be able to precisely know which markers are truly neutral, but are likely to choose a set of
199 markers that captures genome-wide structure. Because *RDA-outliers* performed on par with or
200 worse than other marker sets in 1-trait (Fig S2A), 2-trait (Fig S3B), and 6-trait (Fig S3C) evaluations
201 from the *Adaptive Environment* workflow (Fig. 1) we focus on *all*, *neutral*, and *adaptive* marker
202 sets for the main text.



203

204

205

206 **Fig S3** Performance of RDA-outlier markers are on par with other marker sets for (A) 1-trait, (B)
207 2-trait, and (C) 6-trait evaluations of offset estimated with (RDA-corrected) or without (RDA-
208 uncorrected) population structure correction. Data in this figure is from the *Adaptive*
209 *Environment* workflow. Code to create this figure can be found in SC 02.06.02.

210 **S2 - Coding workflows**

211 Below we reference the scripts (*.R, *.py) and notebooks (*.ipynb) used to analyze
 212 data in this manuscript using the naming convention described in the Data Availability section
 213 (e.g., SC 05.02). Scripts are often written using only functions, instead of a linear development of
 214 code. This allows the functions to be imported/sourced in other scripts or notebooks to avoid
 215 code redundancy. At the top of all script files are detailed instructions for use. The “main”
 216 function in many script files gives a general outline for the code and calls all other functions.

217 All python scripts and notebooks are run in the “mvp_env” (python v3.8) Anaconda
 218 environment. All GF scripts are run in R within the “r35” (R v3.5) Anaconda environment . All
 219 other R code is run within the “MVP_env_R4.0.3” (R v4.0.3) Anaconda environment. All
 220 Anaconda environments can be recreated using their .yml files found in the code archive. These
 221 files contain all package and library versions at the time of saving. Package and library versions
 222 that were used are found at the top of each notebook - look for “Click to view session information”
 223 (python notebooks) or printouts from `sessionInfo ()` (R notebooks).

224 1- and 2-trait simulations are often processed separately from the 6-trait simulation.
 225 Descriptions of coding workflows reflect this.

226 All scripts referenced by name are in the SC 01 directory.

227 Notebooks used to create figures and tables are not described here (but see coding
 228 archive README). Instead, these notebooks are referenced within the caption of all figures and
 229 tables, or in the main text when appropriate. These notebooks (mainly within SC 02.02 directory)
 230 rely on data processed through the coding workflows described below. Similarly, code previously
 231 described in Supplemental Note S1 is not redescribed here.

232 Simulation data used below within scripts and notebooks has been processed from SLiM
 233 output separately by Lotterhos (2023) into more user-friendly forms - see here for more
 234 information: [https://github.com/ModelValidationProgram/MVP-
 235 NonClinalAF/tree/main/sim_output_20220428_metadata](https://github.com/ModelValidationProgram/MVP-NonClinalAF/tree/main/sim_output_20220428_metadata)

236 **1.1 | The Adaptive Environment coding workflow**

237 The *Adaptive Environment* workflow represents the general pipeline for processing
 238 simulations and running genomic offset methods, most other processing code is built on top of
 239 this main pipeline (i.e., scripts and notebooks source/import functions from these scripts to avoid
 240 code redundancy).

241 **1.1.1 / 1- and 2-trait simulations**

242 The *Adaptive Environment* pipeline is kicked off using SC 01.00, which allows the user to
 243 decide which method to run. All analyses were generally run in batches of 225 simulation levels
 244 (one replicate per level). SC 01.00 can call SC 01.01 (for GF), SC 01.05 (for RONA), SC 01.10 (for
 245 LFMM), SC 01.07 (for pairwise F_{ST}), or scripts related to RDA (more details below).

246 $\text{GF}_{\text{offset}}$: SC 01.01 processes the data into formats suitable for GF input. This includes
 247 converting genotype data into derived allele frequencies, asserting MAF cutoffs, and
 248 reformatting environmental data. This script creates .sh files for the slurm HPC and trains GF
 249 models using `MVP_gf_training_script.R`. The slurm .sh files call SC 01.02, which takes
 250 the trained GF model and predicts offset to each of the 100 environments (population sources)
 251 on the landscape using `MVP_gf_fitting_script.R`. Performance of GF offset predictions
 252 are then validated using SC 01.03. Performance results are saved in a nested dictionary.
 253 Environmental importance is extracted from each GF model using SC 01.04 within SC 02.10.02.

254 RONA : Using files created from SC 01.01, SC 01.05 creates files suitable for RONA analyses
 255 and calculates RONA itself. Performance of RONA is validated with SC 01.06. As with GF,
 256 performance results are saved in a nested dictionary.

257 LFMM_{offset} : SC 01.10 creates files suitable for LFMM in R and submits jobs to the slurm
 258 HPC to train LFMM with `MVP_process_lfmm.R`. SC 01.10 also submits SC 01.11 to validate
 259 LFMM offsets. `MVP_watch_for_failure_of_train_lfmm2_offset.py` watches
 260 for failed jobs and reruns them. Performance of LFMM is validated in SC 01.11. As with GF and
 261 RONA, performance results are saved in a nested dictionary.

262 RDA_{offset} : `MVP_pooled_pca_and_rda.R` creates principal component analysis data
 263 and RDA objects using allele frequencies of *all* loci; it also creates additional files needed
 264 downstream. Next, SC 01.12 is run to estimate RDA offset. Performance of RDA_{offset} is validated
 265 with SC 01.13. As with GF, RONA, and LFMM, performance results are saved in a nested
 266 dictionary.

267 Nested dictionaries containing validation results from each method are reformatted and
 268 combined into a single object in notebooks within the SC 02.01.00 directory. These combined
 269 objects are used throughout remaining analyses in jupyter notebooks found in subdirectories of
 270 SC 02.

271 1.1.2 / 6-trait simulation

272 The 6-trait simulation was processed through the *Adaptive Environment* workflow using
 273 code found in the SC 02.05 directory. 6-trait simulations needed extra formatting in order to be
 274 comparable to the 1- and 2-trait evaluations. First, SC 02.05.00 assigns individuals to populations
 275 using a gridded system. Population-level environmental values are the average climate from
 276 assigned individuals on the landscape (each environmental variable is averaged independently).
 277 Genetic and environmental data was formatted as with 1- and 2-trait simulations. Fitness for
 278 each population in each environment was calculated using
 279 `MVP_climate_outlier_fitness_calculator.R`. The script
 280 `MVP_climate_outlier_fitness_calculator.R` was validated against previous
 281 fitness estimates from 1- and 2-trait simulations in SC 02.05.01.

282 GF was trained using the same script as 1- and 2-trait simulations
 283 (`MVP_gf_training_script.R`). GF offset was predicted manually in SC 02.05.02, and

284 validated manually in SC 02.05.03. In SC 02.05.04 - 02.05.05 LFMM was trained and validated
 285 manually. Similarly, RDA was trained and validated in SC 02.05.06 - SC 02.05.07, and RONA
 286 trained and validated in SC 02.05.08 - SC 02.05.09.

287 **1.2 | The Climate Novelty coding workflow**

288 Fitness was calculated for 1- and 2-trait populations within the *Climate Novelty* scenarios
 289 (x-axis, Fig. 6, Supplemental Note S3) using
 290 `MVP_climate_outlier_fitness_calculator.R` in 02.04.01.

291 Using 1- and 2-trait offset models output from the *Adaptive Environment* workflow, the
 292 following code predicted offset to *Climate Novelty* scenarios (GF: SC 01.14; LFMM: SC 01.16; RDA:
 293 SC 01.18; and RONA: SC 01.20) which was subsequently validated against known fitness (GF: SC
 294 01.15; LFMM: SC 01.17; RDA: SC 01.19; and RONA: SC 01.21). A few examples of code executions
 295 are shown in SC 02.04.03.

296 Fitness of 6-trait populations for *Climate Novelty* scenarios was calculated in SC 02.04.06.
 297 6-trait GF models used the same scripts as 1- and 2-trait runs (SC 01.14 - SC 01.15); executed from
 298 SC 02.04.07. Commands to train LFMM were created in SC 02.04.07, which called on
 299 `MVP_complex_sims_process_1fmm.R`. RDA was trained manually in SC 02.04.07. Offset
 300 from both LFMM and RDA were validated manually in SC 02.04.08.

301 **1.3 | The Nuisance Environment coding workflow**

302 Environmental files for *Nuisance Environment* scenarios were created in SC 02.07.02.02.

303 Files for 1- and 2-trait simulations were created in SC 02.07.02.01 to train GF using
 304 `MVP_gf_training_script.R`. SC 01.02 and SC 01.03 are used for predicting and
 305 validating GF offset, respectively, executed in SC 02.07.02.02. Code for LFMM was executed in SC
 306 02.07.02.07 and used `MVP_process_1fmm.R` for training and SC 01.11 for validation.
 307 Commands for RDA were created in SC 02.07.02.06 similarly to *Adaptive Environment* workflow
 308 (calling `MVP_pooled_pca_and_rda.R`) and used `MVP_nuisance_RDA_offset.R`
 309 for training and `MVP_nuisance_rda_validation.py` for validation.

310 6-trait sims were processed for GF exactly as they were for 6-trait data in the *Adaptive*
 311 *Environment* workflow (with updated environmental files) and executed in SC 02.07.02.03, SC
 312 02.07.02.04, and validated manually in SC 02.07.02.10. Code to train both LFMM and RDA was
 313 executed in SC 02.07.02.12, which called on `MVP_complex_sims_process_1fmm.R` for
 314 LFMM. LFMM was validated in SC 02.07.02.13; RDA was validated in SC 02.07.02.14.

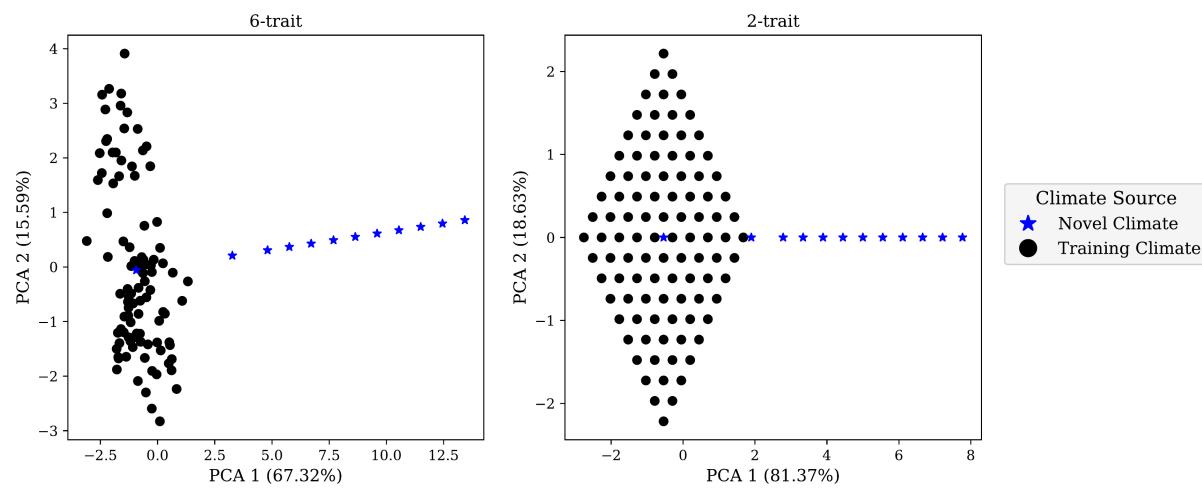
315 **1.4 | Misc**

316 `MVP_summary_functions.py` contains much of the API used within notebooks for loading
 317 and filtering data as well as creating figures. It is often imported using the alias `mvp` within
 318 python scripts and notebooks.

319 **S3 - Defining Climate Novelty scenarios**

320 To understand if genomic offset models maintained predictive performance in
 321 environments differentiated from training environments, we created 11 climates, each
 322 progressively more distant from the mean training environment. Specifically, for each
 323 environmental variable, we used a standardized set of z-scores ($z_E \in \{1.72, 2.35, 2.74, 3.13, 3.53,$
 324 $3.92, 4.31, 4.70, 5.09, 5.48, 5.88\}$) to calculate corresponding environmental values. In other
 325 words, we used the distribution of the within-landscape values from which to identify the
 326 appropriate value for a given z-score for each environmental variable independently. The *temp*
 327 environment and all six of the 6-trait environments were given positive values for *Climate Novelty*
 328 scenarios, and *Env2* was given negative values.

329 Novelty climates for 6-trait and 2-trait evaluations are shown Fig S5A and B, respectively.
 330 In this and other figures related to performance in *Climate Novelty* scenarios, we also include
 331 $z_E=0.00$ for comparison of novelty climates to the mean training climate (i.e., climate center). We
 332 chose z-scores over Mahalanobis distances because of 1) the reduced correlation structure
 333 among environmental variables (where z-scores and Mahalanobis distances should be roughly
 334 equivalent; Fig. S4), and 2) the large number of combinations of values from environmental
 335 variables that could be used for a given Mahalanobis distance. The standard deviation values that
 336 we used are applicable to all environments and for all landscapes except for *Env2* in the *Stepping*
 337 *Stone - Mountain* landscape; the corresponding standard deviations for this case are $z_E \in \{1.55,$
 338 $2.12, 2.47, 2.82, 3.18, 3.53, 3.88, 4.24, 4.60, 4.95, 5.3\}$.



339

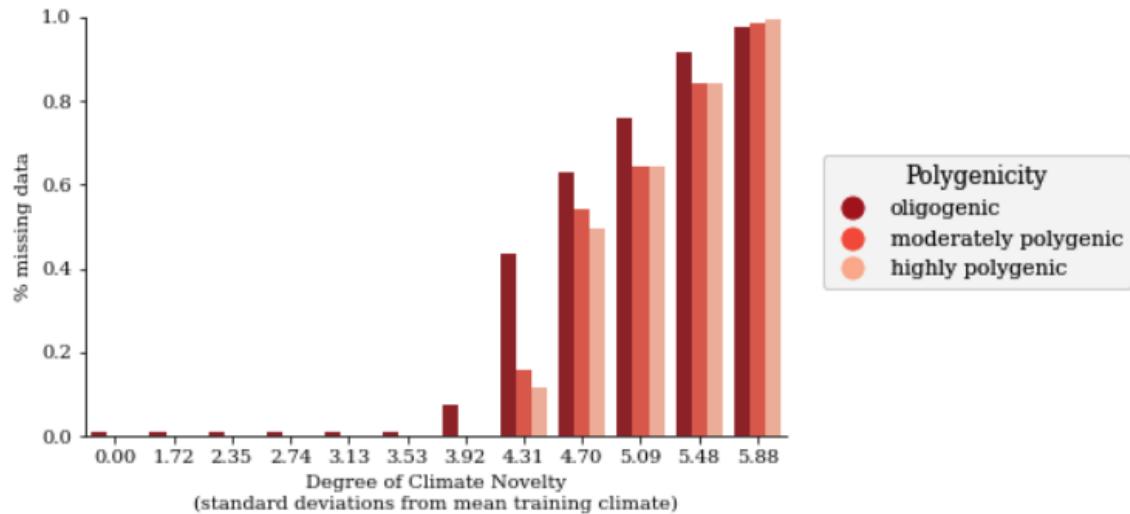
340 **Fig S5** Differentiation of *Climate Novelty* environments (blue stars, including climate center)
 341 from within-landscape environments (black circles) using Principal Component Analysis (PCA) of
 342 environmental data. Environmental data is centered and standardized relative to the within-
 343 landscape environmental values. Scatter plots show the first two principal components (PCs) of
 344 environmental data used to evaluate 6-trait (A) and 2-trait (B) *Climate Novelty* scenarios. There
 345 is no figure for 1-trait evaluations because there would only be one PC axis. Code to create these
 346 figures can be found in SC 02.04.10.

347 S4 - Missing data in *Climate Novelty* evaluations

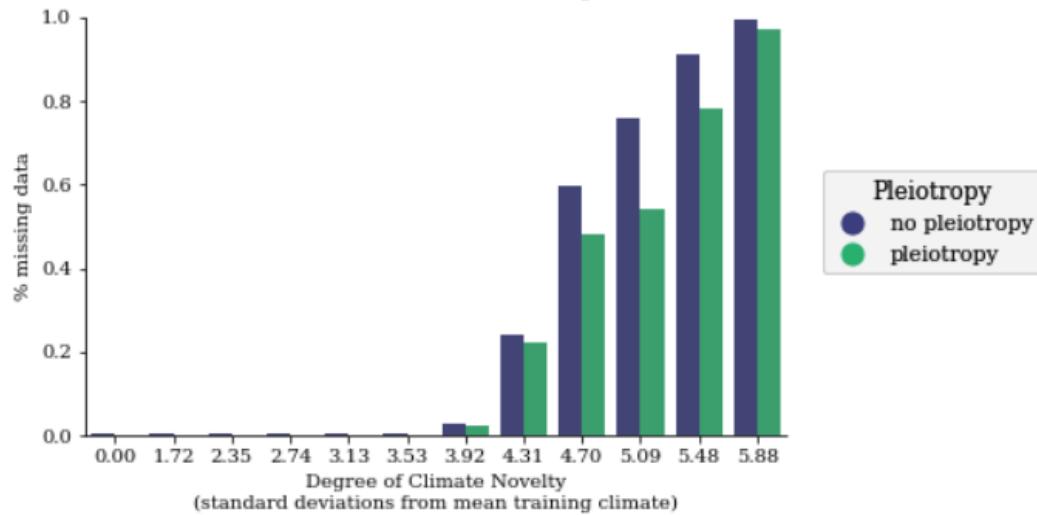
348 When calculating fitness of populations in *Climate Novelty* scenarios, it could be the case
349 that all populations have zero fitness because of the extremity of the novel climate. In these cases
350 the calculation of performance is technically undefined due to the lack of variability in one of the
351 vectors (i.e., the code returns “NAN”), but for Figure 7 we replaced these undefined values with
352 0 (because there was no predictive performance of the offset method). We refer to these cases
353 as missing data below. It is therefore important to explore the effect of these missing data points
354 on patterns observed between performance and climate novelty (i.e., in the context of Fig. 7 of
355 the main text) to ensure patterns before and after setting missing data to 0 do not affect
356 inferences.

357 To understand impacts of missing data, we created figures that grouped simulation and
358 experimental levels across novelty scenarios (Fig. S33). We also printed out specific scenarios in
359 the code (SC 02.04.05). Importantly, missing data is not substantial until *Climate Novelty (CN)*
360 *Scenario 4.31*, which is preceded by the drop in performance from datasets with elevated LA_{ASA} .
361 After *CN Scenario 4.31* missing data begins to increase because of climate novelty, first with
362 datasets where high levels of LA_{ASA} take place through oligogenic architectures, then missing data
363 is more uniform across simulation and experimental parameters for the remaining *CN Scenarios*
364 (Fig. S33). (Before *CN Scenario 4.31*, missing data is not due to all populations having zero fitness
365 - instead missing data is primarily due to 1-trait oligogenic scenarios evaluated by RONA where
366 there are no *adaptive* alleles with significant clines with *temp* in the *Estuary - Clines* landscape
367 (Fig. S33; SC 02.04.05).) Finally, we also explored patterns presented in Fig. 6 before setting
368 undefined performance scores to zero and found nearly identical trends (not shown).

369 (Fig S33)

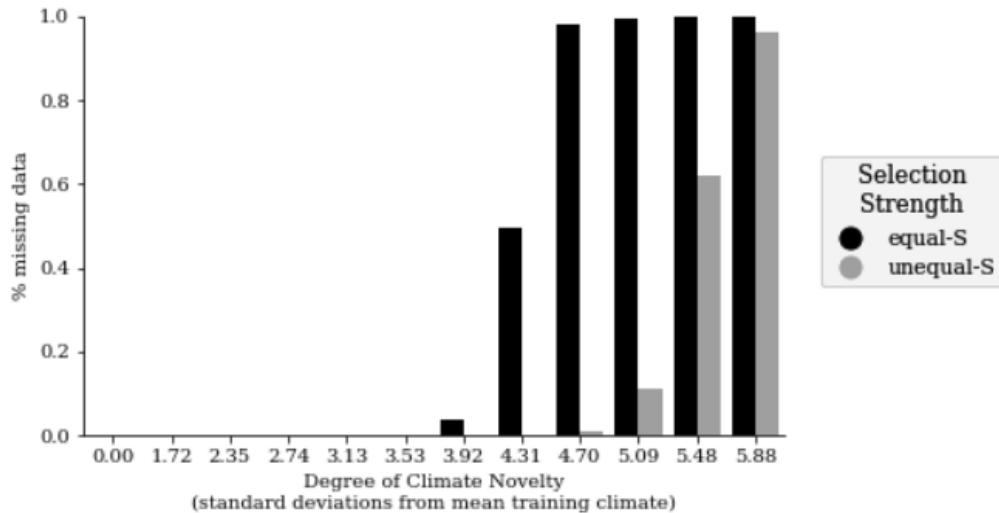


370

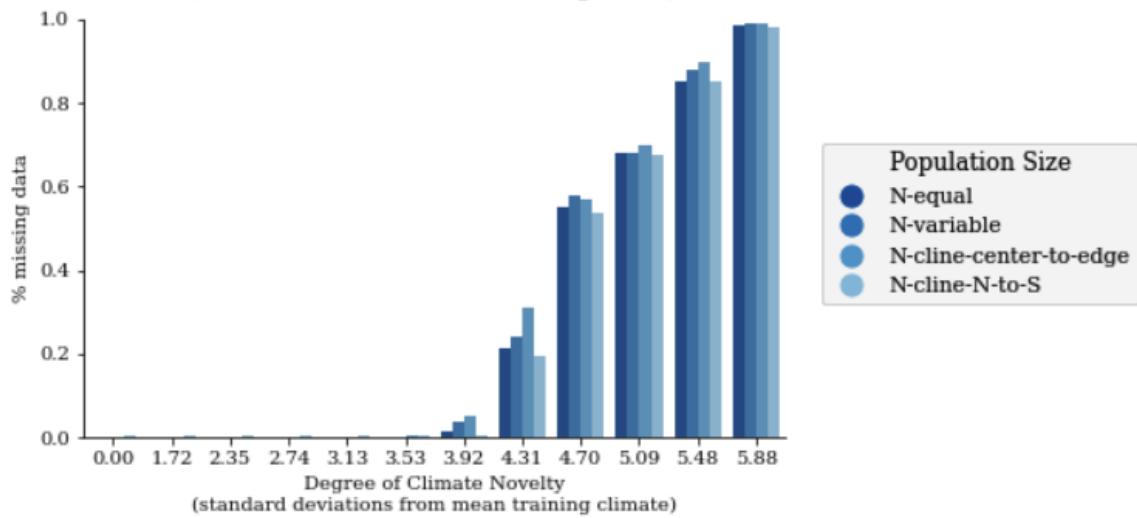


371

372 (Fig S33 continued)

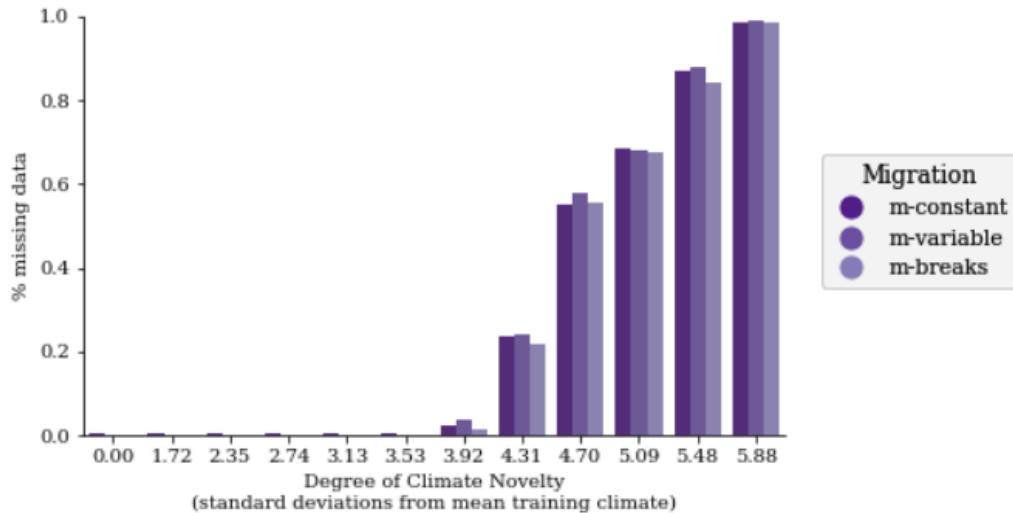


373

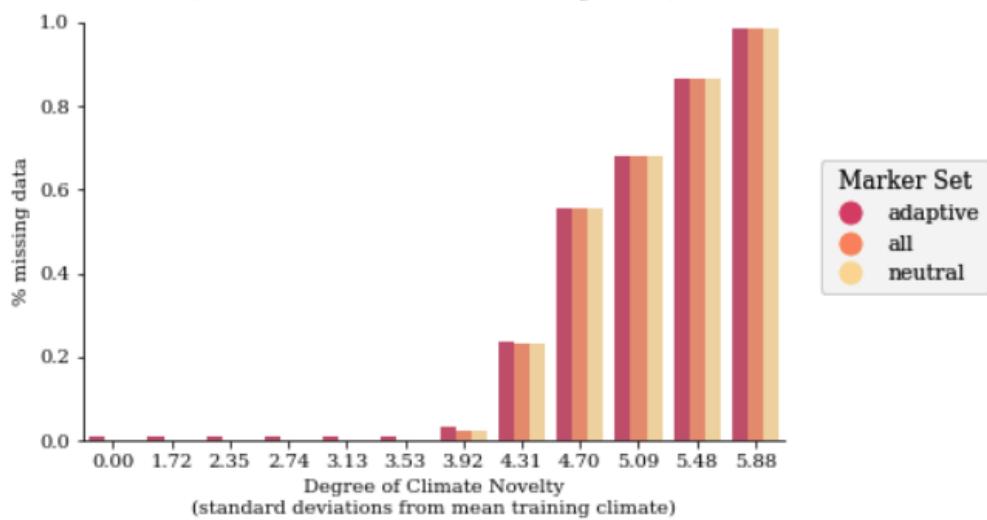


374

375 (Fig S33 continued)

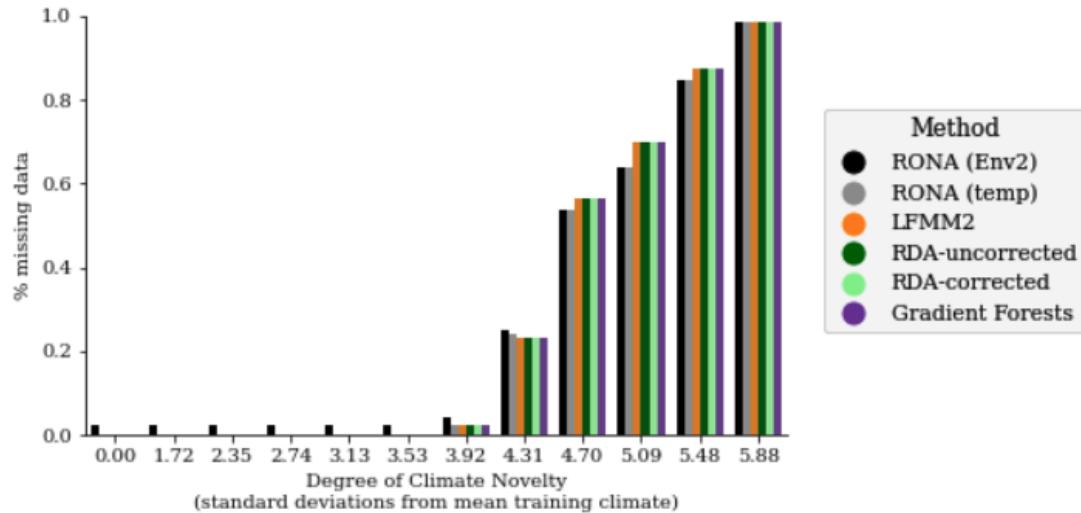


376

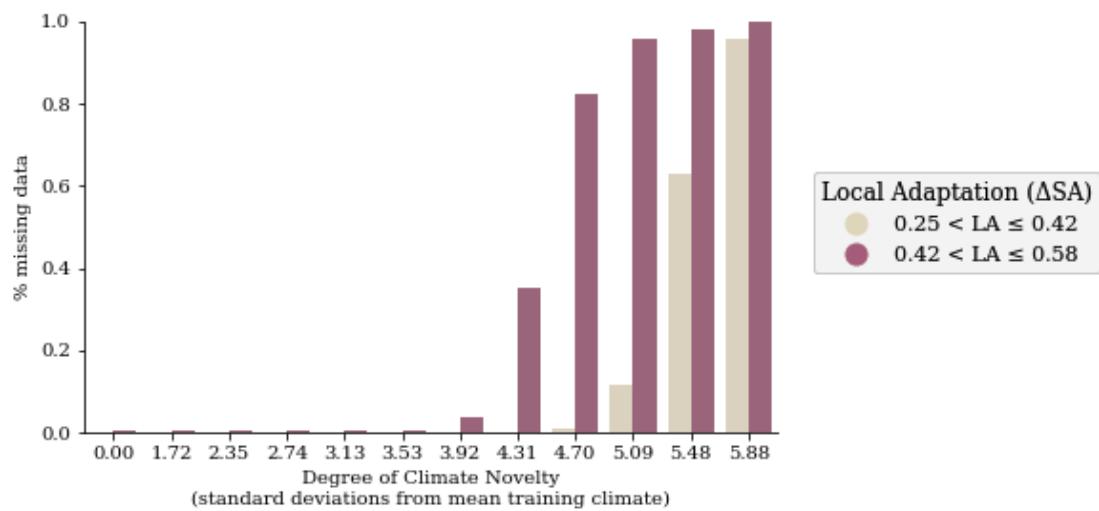


377

378 (Fig S33 continued)

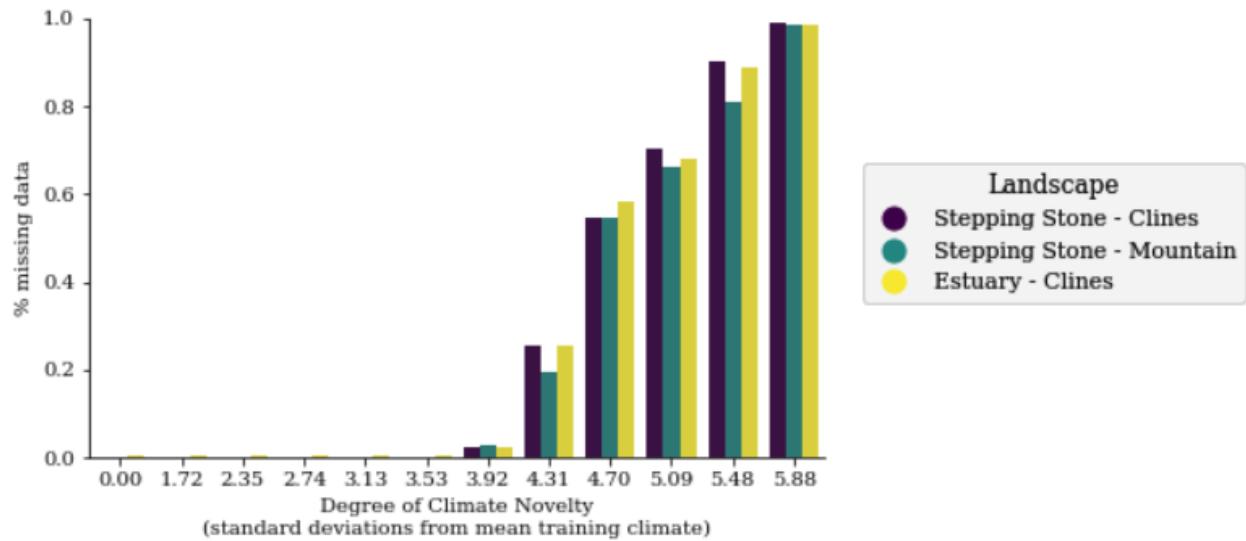


379



380

381 (Fig S33 continued)



382

383 **Fig S33** The effect of simulation parameters on missing data for *Climate Novelty* scenarios.
 384 Shown are the percent missing data (y-axes) due to experimental and simulation parameters
 385 (legends). Missing data is when all populations have zero fitness in a given novelty scenario, and
 386 thus performance cannot be defined (though we manually set it to zero for other figures). Data
 387 included in these figures are from 1- and 2-trait evaluations of *Climate Novelty* scenarios. Code
 388 to create these figures can be found in SC 02.04.05.

389 Supplemental Tables

RONA-sal_opt

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	15.864775	2.0	210.709735	3.953193e-92	0.17
landscape	149.664742	2.0	1987.788561	0.000000e+00	1.59
demography	433.301763	4.0	2877.472267	0.000000e+00	4.60
plevel_pleio	0.033619	1.0	0.893039	3.446564e-01	0.00
C(garden)	1941.690095	99.0	520.985219	0.000000e+00	20.61
cor_TPR_temp	0.724989	1.0	19.258049	1.142528e-05	0.01
cor_TPR_sal	3.874653	1.0	102.923263	3.536502e-24	0.04
cor_FPR_temp_neutSNPs	0.897974	1.0	23.853067	1.040651e-06	0.01
cor_FPR_sal_neutSNPs	14.829886	1.0	393.929492	1.433616e-87	0.16
final_LA	87.744615	1.0	2330.779320	0.000000e+00	0.93
Residual	6771.995848	179886.0	NaN	NaN	71.88

RONA-temp_opt

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	3.264399	2.0	99.857527	4.534018e-44	0.03
landscape	74.659979	2.0	2283.838967	0.000000e+00	0.64
demography	26.679201	4.0	408.056627	0.000000e+00	0.23
plevel_pleio	1.505757	1.0	92.121798	8.249029e-22	0.01
C(garden)	4962.472456	99.0	3066.694571	0.000000e+00	42.52
cor_TPR_temp	1.035753	1.0	63.367084	1.725510e-15	0.01
cor_TPR_sal	0.021308	1.0	1.303608	2.535568e-01	0.00
cor_FPR_temp_neutSNPs	2.155381	1.0	131.865626	1.640714e-30	0.02
cor_FPR_sal_neutSNPs	0.039487	1.0	2.415797	1.201186e-01	0.00
final_LA	3659.697867	1.0	223899.355101	0.000000e+00	31.35
Residual	2940.287212	179886.0	NaN	NaN	25.19

390

GF

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	3.109424	2.0	158.728388	1.336264e-69	0.05
landscape	344.465503	2.0	17584.110829	0.000000e+00	5.24
demography	104.816048	4.0	2675.299841	0.000000e+00	1.59
plevel_pleio	0.373620	1.0	38.144788	6.582481e-10	0.01
C(garden)	392.397617	99.0	404.665183	0.000000e+00	5.97
cor_TPR_temp	3.305288	1.0	337.453511	2.682308e-75	0.05
cor_TPR_sal	0.498155	1.0	50.859112	9.961078e-13	0.01
cor_FPR_temp_neutSNPs	46.646231	1.0	4762.349106	0.000000e+00	0.71
cor_FPR_sal_neutSNPs	37.556091	1.0	3834.290889	0.000000e+00	0.57
final_LA	3881.402690	1.0	396271.989655	0.000000e+00	59.02
Residual	1761.946397	179886.0	NaN	NaN	26.79

lfmm2

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	0.648503	2.0	26.071108	4.776417e-12	0.01
landscape	69.824926	2.0	2807.098902	0.000000e+00	1.34
demography	75.595816	4.0	1519.549983	0.000000e+00	1.45
plevel_pleio	0.088264	1.0	7.096738	7.723125e-03	0.00
C(garden)	391.270017	99.0	317.774174	0.000000e+00	7.50
cor_TPR_temp	3.333636	1.0	268.037430	3.359626e-60	0.06
cor_TPR_sal	1.434571	1.0	115.345148	6.738112e-27	0.03
cor_FPR_temp_neutSNPs	18.408493	1.0	1480.114981	1.679823e-322	0.35
cor_FPR_sal_neutSNPs	0.000221	1.0	0.017772	8.939475e-01	0.00
final_LA	2419.263042	1.0	194518.232194	0.000000e+00	46.37
Residual	2237.278977	179886.0	NaN	NaN	42.88

391

rda-nocorr

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	2.722620	2.0	114.252516	2.583753e-50	0.04
landscape	351.685552	2.0	14758.193717	0.000000e+00	5.17
demography	86.983434	4.0	1825.093984	0.000000e+00	1.28
plevel_pleio	0.043391	1.0	3.641764	5.634878e-02	0.00
C(garden)	384.324179	99.0	325.815086	0.000000e+00	5.65
cor_TPR_temp	4.653284	1.0	390.542466	7.801493e-87	0.07
cor_TPR_sal	1.660051	1.0	139.325332	3.842628e-32	0.02
cor_FPR_temp_neutSNPs	57.083425	1.0	4790.917538	0.000000e+00	0.84
cor_FPR_sal_neutSNPs	39.340402	1.0	3301.774980	0.000000e+00	0.58
final_LA	3728.556367	1.0	312931.576881	0.000000e+00	54.83
Residual	2143.328255	179886.0	NaN	NaN	31.52

rda-structcorr

	sum_sq	df	F	PR(>F)	perc_sum_sq
glevel	19.270968	2.0	303.886012	1.763727e-132	0.29
landscape	54.926356	2.0	866.139726	0.000000e+00	0.81
demography	644.354351	4.0	5080.447174	0.000000e+00	9.54
plevel_pleio	2.784965	1.0	87.832851	7.201110e-21	0.04
C(garden)	184.307173	99.0	58.714343	0.000000e+00	2.73
cor_TPR_temp	13.822608	1.0	435.940440	1.078989e-96	0.20
cor_TPR_sal	0.004197	1.0	0.132380	7.159771e-01	0.00
cor_FPR_temp_neutSNPs	1.803296	1.0	56.872745	4.671127e-14	0.03
cor_FPR_sal_neutSNPs	45.534733	1.0	1436.084373	5.262511e-313	0.67
final_LA	85.096985	1.0	2683.807346	0.000000e+00	1.26
Residual	5703.746286	179886.0	NaN	NaN	84.43

392

393 **Table S1** Results from Type II ANOVAs from regressing simulation factors on offset performance
 394 (see Equation 1 of the main text). In this table, the common garden ID was included as a
 395 categorical factor (n=100 per simulation). Code to create these tables can be found in SC
 396 02.02.01.

RONA (Env2)						RONA (temp)					
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq
<i>PcQTN, temp</i>	11.076279	1.0	203.959434	3.027995e-46	0.11	<i>PcQTN, temp</i>	461.435097	1.0	6854.545092	0.000000e+00	3.54
<i>PcQTN, Env2</i>	7.130868	1.0	131.308340	2.171984e-30	0.07	<i>PcQTN, Env2</i>	321.669143	1.0	4778.344045	0.000000e+00	2.47
<i>PcNeut, temp</i>	350.186388	1.0	6448.358576	0.000000e+00	3.40	<i>PcNeut, temp</i>	21.729532	1.0	322.788745	4.136405e-72	0.17
<i>PcNeut, Env2</i>	165.769820	1.0	3052.497975	0.000000e+00	1.61	<i>PcNeut, Env2</i>	119.950781	1.0	1781.849799	0.000000e+00	0.92
Residual	9774.859470	179995.0	NaN	NaN	94.82	Residual	12116.925221	179995.0	NaN	NaN	92.91

397

RDA-uncorrected

RDA-uncorrected						RDA-corrected					
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq
<i>PcQTN, temp</i>	326.478563	1.0	8168.479995	0.0	3.29	<i>PcQTN, temp</i>	36.345449	1.0	970.987128	1.344472e-212	0.48
<i>PcQTN, Env2</i>	171.117641	1.0	4281.356235	0.0	1.72	<i>PcQTN, Env2</i>	26.680765	1.0	712.790185	1.002191e-156	0.35
<i>PcNeut, temp</i>	589.314088	1.0	14744.613849	0.0	5.93	<i>PcNeut, temp</i>	281.267846	1.0	7514.213388	0.000000e+00	3.70
<i>PcNeut, Env2</i>	1649.284886	1.0	41265.038898	0.0	16.61	<i>PcNeut, Env2</i>	512.076667	1.0	13680.388309	0.000000e+00	6.74
Residual	7194.056785	179995.0	NaN	NaN	72.45	Residual	6737.472479	179995.0	NaN	NaN	88.72

398

Gradient Forests

Gradient Forests						LFMM2					
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq
<i>PcQTN, temp</i>	350.833368	1.0	9184.348121	0.0	3.69	<i>PcQTN, temp</i>	256.979852	1.0	8357.260709	0.0	3.72
<i>PcQTN, Env2</i>	210.540691	1.0	5511.673549	0.0	2.21	<i>PcQTN, Env2</i>	173.645491	1.0	5647.137802	0.0	2.51
<i>PcNeut, temp</i>	375.213154	1.0	9822.578289	0.0	3.94	<i>PcNeut, temp</i>	718.499204	1.0	23366.365591	0.0	10.40
<i>PcNeut, Env2</i>	1702.524610	1.0	44569.816059	0.0	17.89	<i>PcNeut, Env2</i>	227.994551	1.0	7414.627593	0.0	3.30
Residual	6875.637914	179995.0	NaN	NaN	72.26	Residual	5534.718859	179995.0	NaN	NaN	80.08

399

400 **Table S2** Results from Type II ANOVAs from regressing the proportion of clinal QTNS
 401 (cor_TPR_tmp and cor_TPR_sal) and clinal neutral alleles (cor_FPR_temp_neutSNPs,
 402 cor_FPR_sal_neutSNPs) on offset performance (see Equation 2 of the main text). Code to create
 403 these tables can be found in SC 02.02.05.

404

RONA-sal_opt						RONA-temp_opt							
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq		
	all	13.066056	1.0	939.445081	2.543708e-166	34.04		all	0.086501	1.0	8.324025	0.003959	0.18
	final_LA	0.321152	1.0	23.090730	1.673759e-06	0.84		final_LA	30.251825	1.0	2911.137332	0.000000	61.72
	Residual	24.993162	1797.0	Nan	Nan	65.12		Residual	18.673983	1797.0	Nan	Nan	38.10

1fmm2

GF

405

1fmm2						GF							
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq		
	all	3.673454	1.0	873.091702	9.874623e-157	8.64		all	8.505224	1.0	749.172746	3.638160e-138	10.54
	final_LA	31.290136	1.0	7436.913517	0.000000e+00	73.58		final_LA	51.787680	1.0	4561.657568	0.000000e+00	64.18
	Residual	7.560714	1797.0	Nan	Nan	17.78		Residual	20.401019	1797.0	Nan	Nan	25.28

rda-nocorr

rda-structcorr

406

rda-nocorr						rda-structcorr							
	sum_sq	df	F	PR(>F)	perc_sum_sq		sum_sq	df	F	PR(>F)	perc_sum_sq		
	all	10.763364	1.0	827.118396	6.020460e-150	12.54		all	3.192007	1.0	78.914320	1.537690e-18	4.20
	final_LA	51.668736	1.0	3970.521018	0.000000e+00	60.21		final_LA	0.062105	1.0	1.535395	2.154664e-01	0.08
	Residual	23.384518	1797.0	Nan	Nan	27.25		Residual	72.686890	1797.0	Nan	Nan	95.71

407

Table S3 Results from Type II ANOVAs regressing two factors - degree of local adaptation (final_LA) and levels of isolation-by-environment in *all* marker sets) on offset performance. Code to create these tables can be found in SC 02.02.11.

410

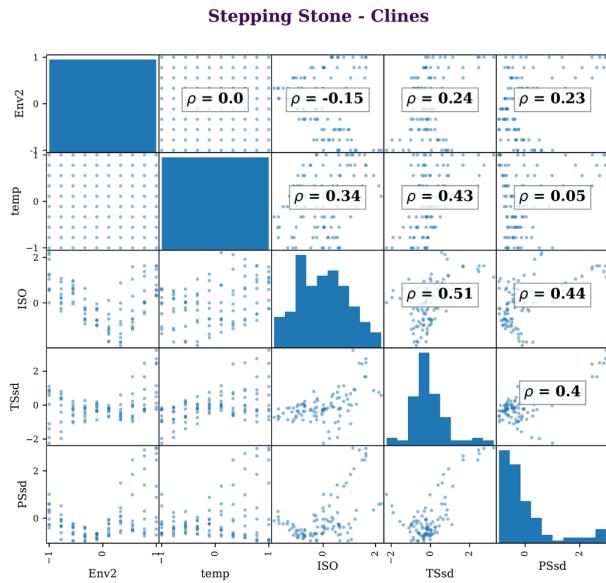
Nuisance Level	<i>Adaptive</i> models	<i>All</i> models	<i>Neutral</i> models
1-trait 1-nuisance	45/45	45/45	45/45
1-trait 3-nuisance	45/45	43/45	38/45
1-trait 4-nuisance	43/45	36/45	35/45
2-trait 2-nuisance	120/180	119/180	119/180
2-trait 3-nuisance	140/180	119/180	119/180

411 **Table S4** gradientForest (GF) sometimes incorrectly identifies the environments driving
 412 adaptation. Shown are the proportions of simulation levels (N 1-trait = 45 levels; N 2-trait = 180
 413 levels; one replicate each) where weighted feature importance output from GF correctly
 414 identified the adaptive environments in the top-most ranks. The Nuiance Level refers to the
 415 number of non-adaptive environmental variables included in training; column names refer to the
 416 marker set used in training. If at least one nuisance environment was ranked above an adaptive
 417 environment this was counted as incorrect. Data used to create this table is from the GF models
 418 output from the *Nuisance Environment* workflow. Code used to create this table can be found in
 419 SC 02.10.02.

420 **Supplemental Figures**

421 Figs. S1-S3 are in Supplemental Note S1.

422



423

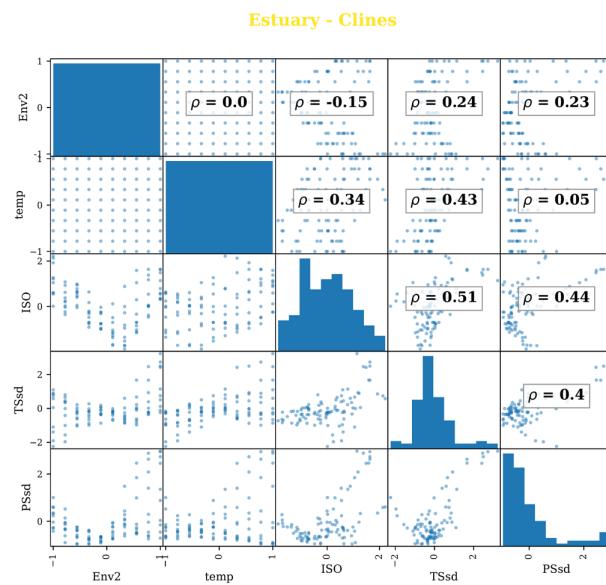
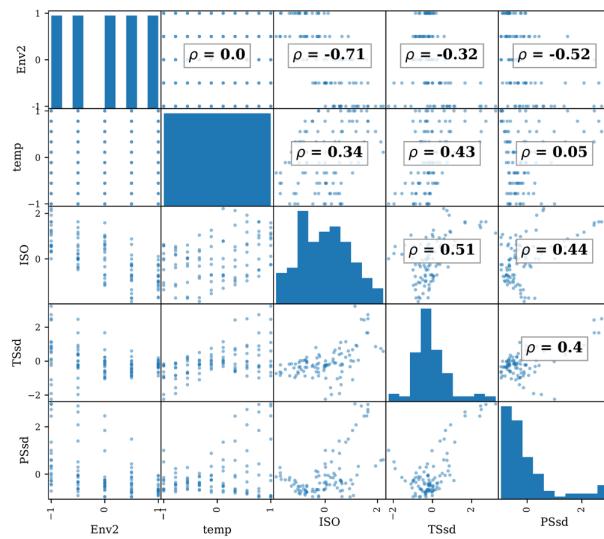
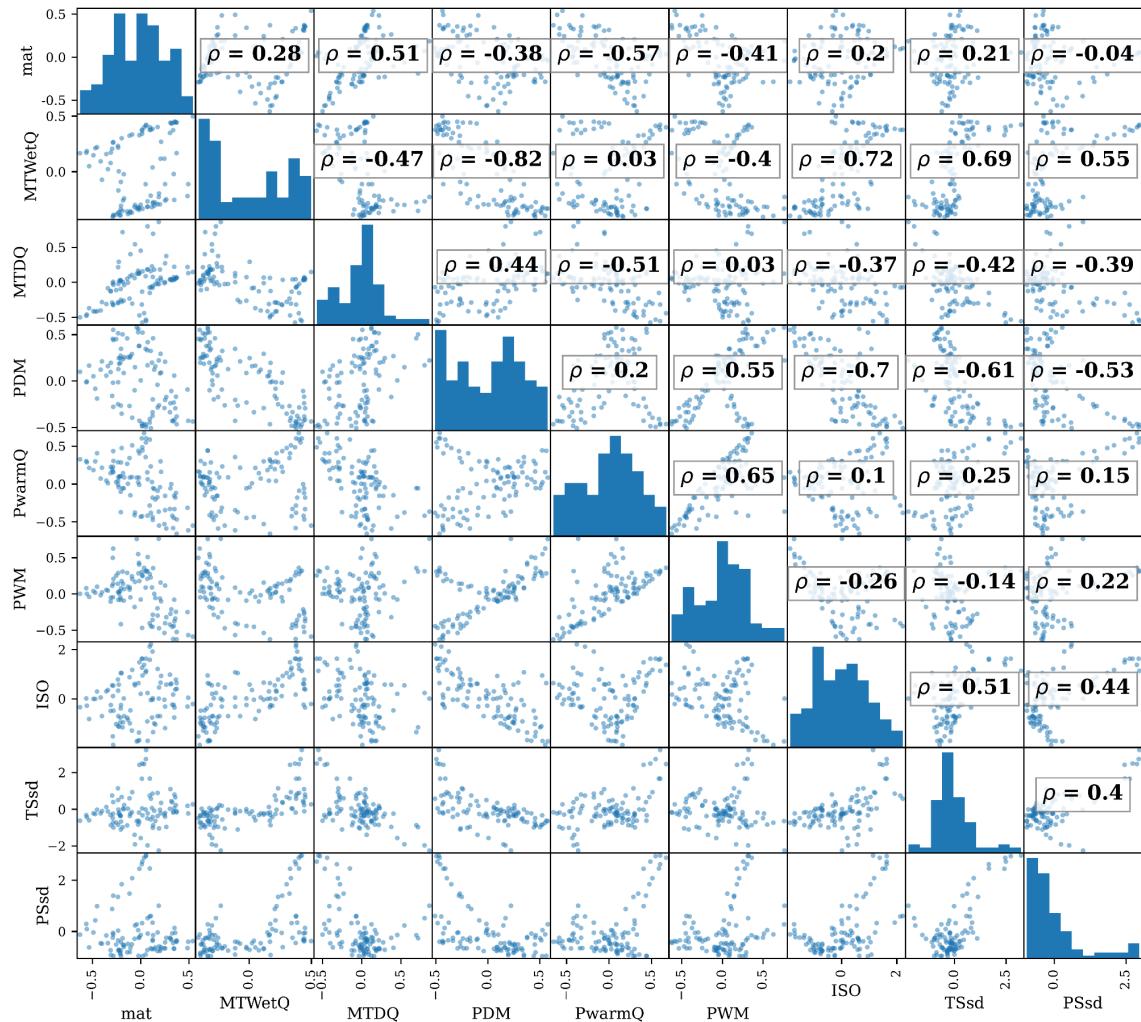
**Stepping Stone - Mountain**

Fig S4 Correlation (Spearman's rho) among environmental variables faceted by landscape. On-diagonal entries are histograms of environmental values, off-diagonal entries are scatter plots between pairwise variables. Included are environmental variables from 1-trait (*temp*), 2-trait (*temp*, *Env2*), and 6-trait simulations (MAT, MTWetQ, MTDQ, PDM, PwarmQ, PWM), as well as nuisance environmental variables (*ISO*, *PSsd*, *TSsd*). Note *Estuary - Clines* and *Stepping Stone - Clines* have the same correlation structure; *Stepping Stones - Mountain* only differs from these two landscapes with *Env2*. Figure continues on the next page. Code to create these figures can be found in SC 02.07.02.11.

439 (Fig S4 continued)

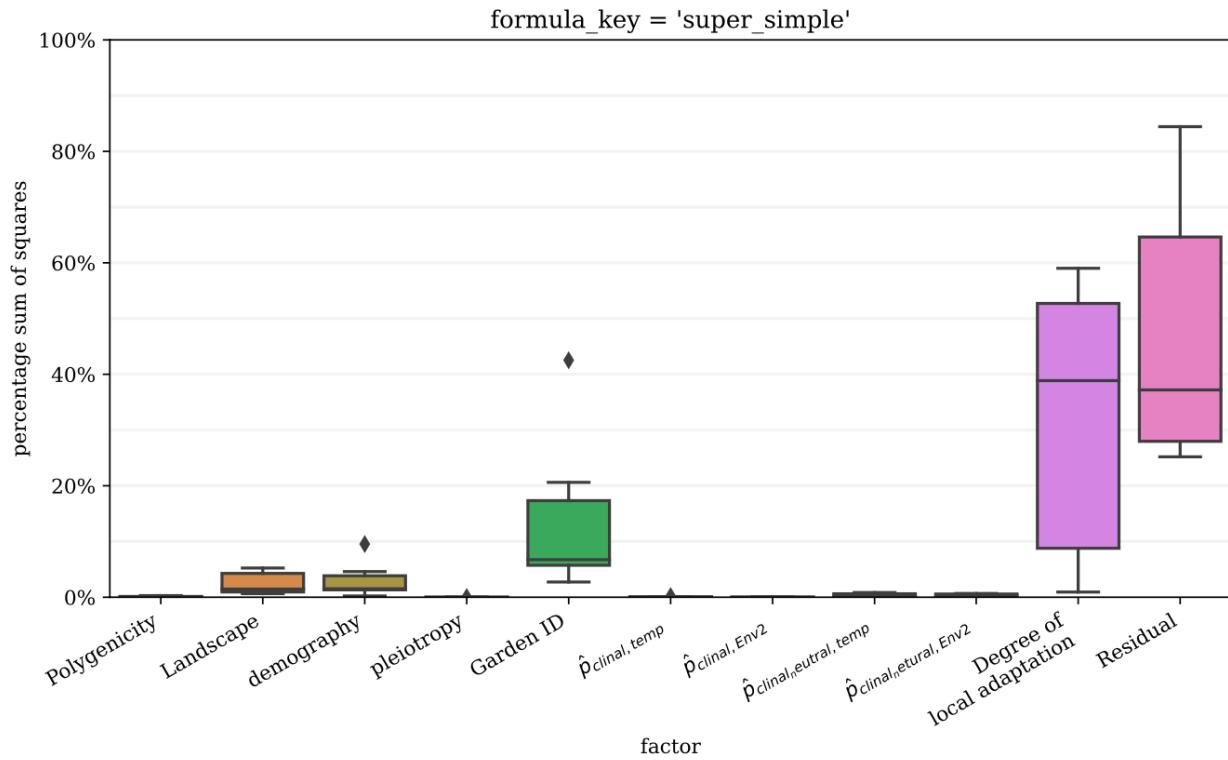
440

6-trait



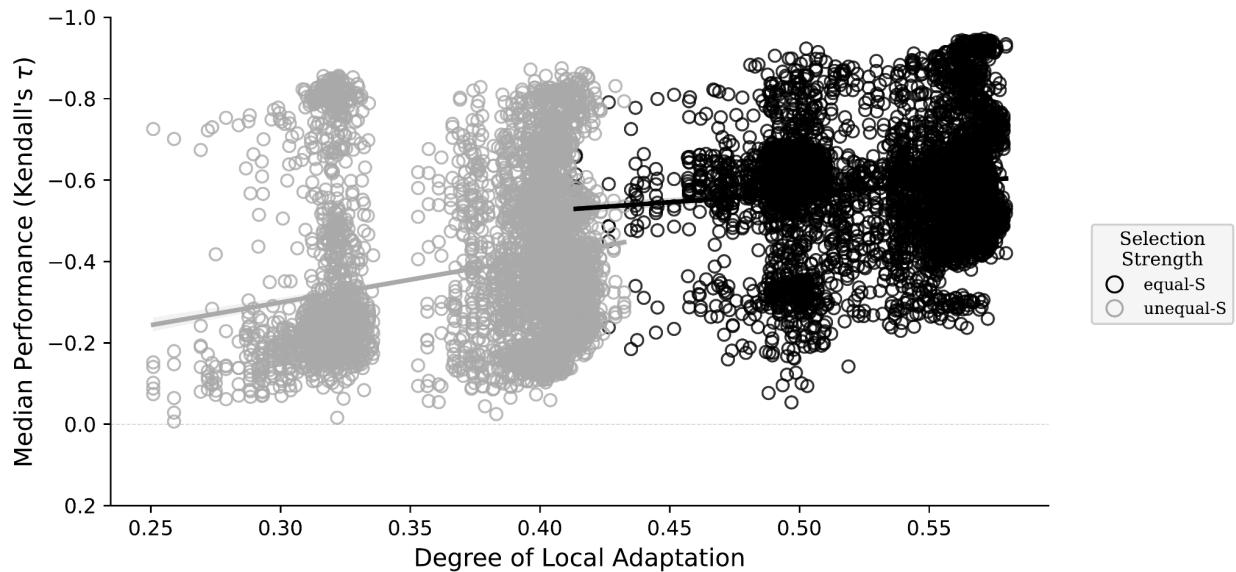
441

442 Fig S5 is in Supplemental Note S3



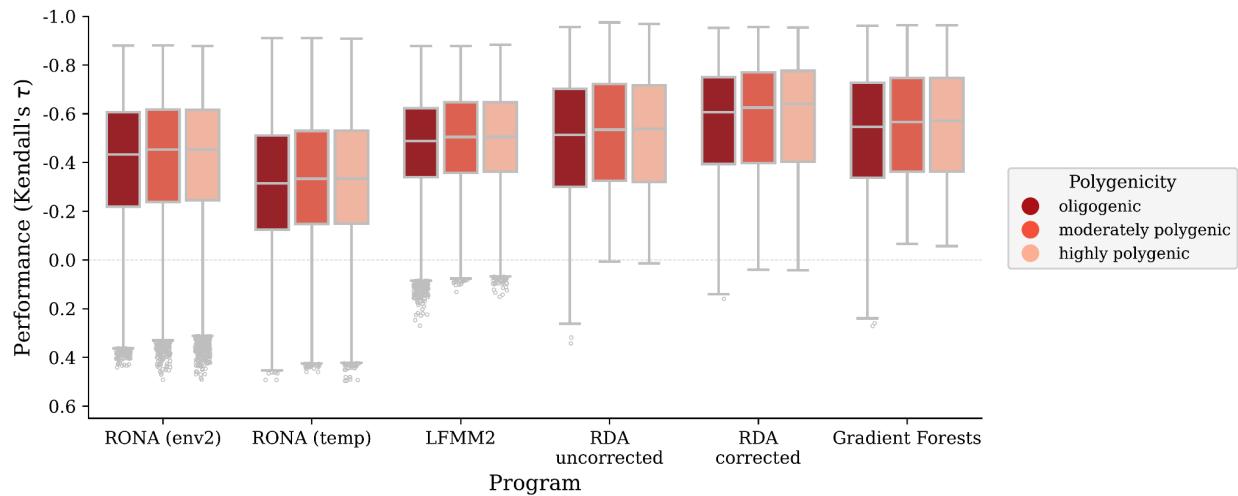
443

444 **Fig S6** Percent sum of squares of the various factors from the ANOVA model in Table S1.
 445 Boxplots are created from the percent sum of squares from each method's individual ANOVA
 446 model. Data included in this figure are from models trained using all markers and simulations
 447 with two selective environments with performance evaluated in all 100 common gardens. Code
 448 to create this table is in 02.02.01.



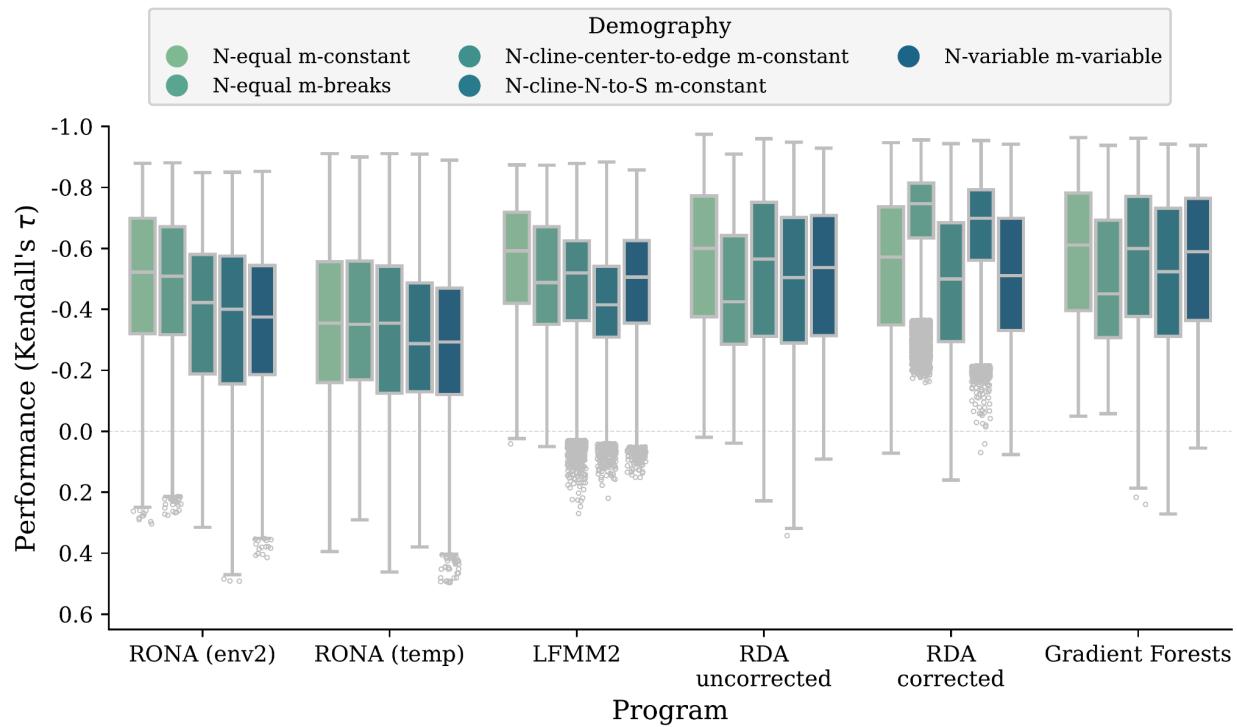
449

450 **Fig S7** Effect of the degree of local adaptation (x-axes) on method performance (y-axes) colored
 451 by the relative strength of selection on the two traits. Shown are the linear relationships between
 452 the median validation scores (circles, taken from validation scores across all 100 common
 453 gardens on the landscape) and the simulation's mean level of local adaptation (taken across all
 454 100 populations). Data included in this figure are from models trained using all markers and
 455 simulations with two selective environments. Code to create this figure can be found in SC
 456 02.02.02.



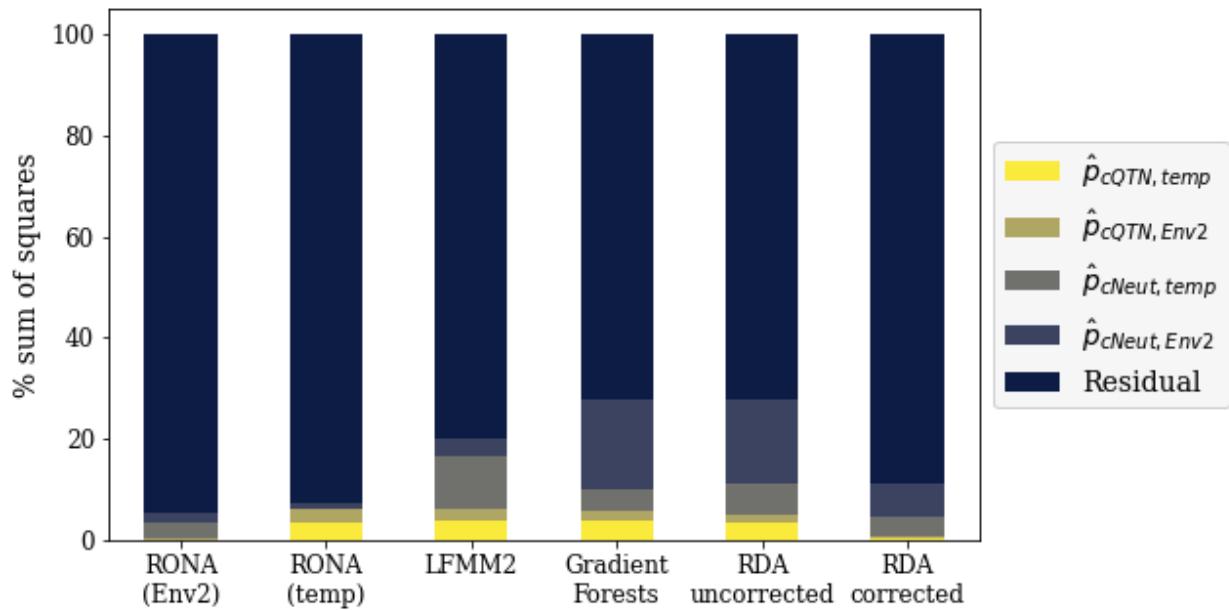
457

458 **Fig S8** Effect of polygenicity on performance of offset methods trained using all markers on
459 simulations with two adaptive traits. Code to create this figure can be found in SC 02.02.01.



460

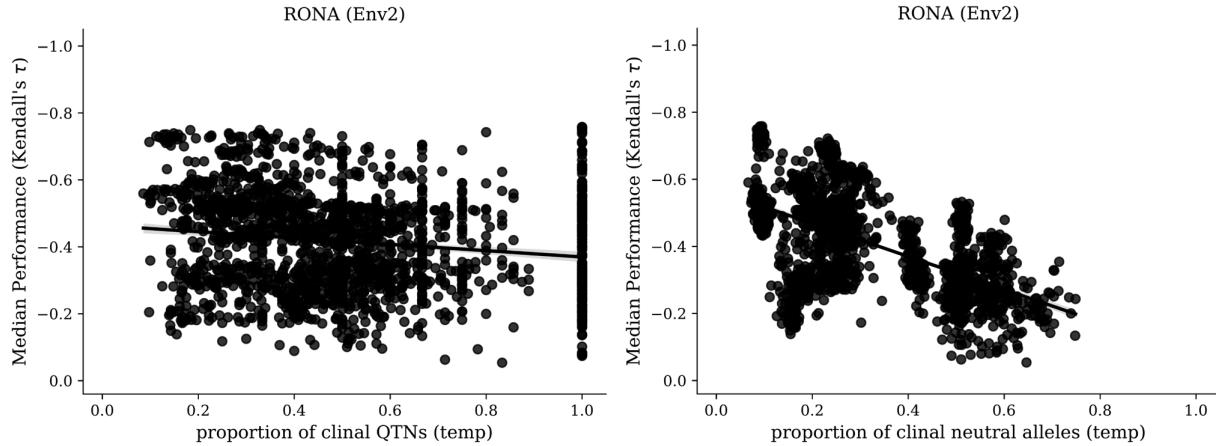
461 **Fig S9** Effect of demography on performance of offset methods trained using all markers on
 462 simulations with two adaptive traits. Code to create this figure can be found in SC 02.02.01.



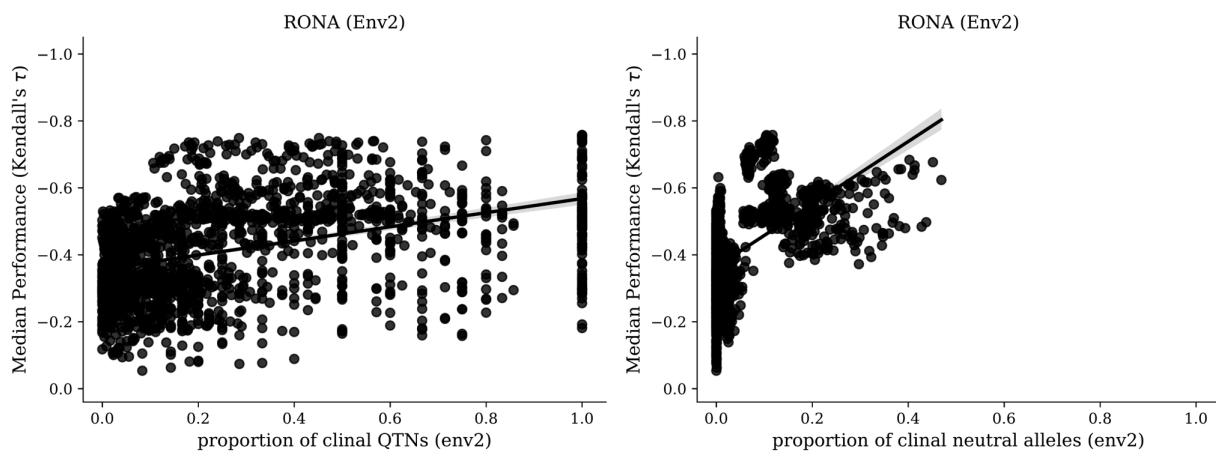
463

464 **Fig S10** Stacked bar plot of the percent sum of squares from Type II ANOVAs from regressing
 465 the proportion of clinal QTNs and clinal neutral alleles on offset performance (see Equation 2 of
 466 the main text). Code to create this table is in 02.02.05.

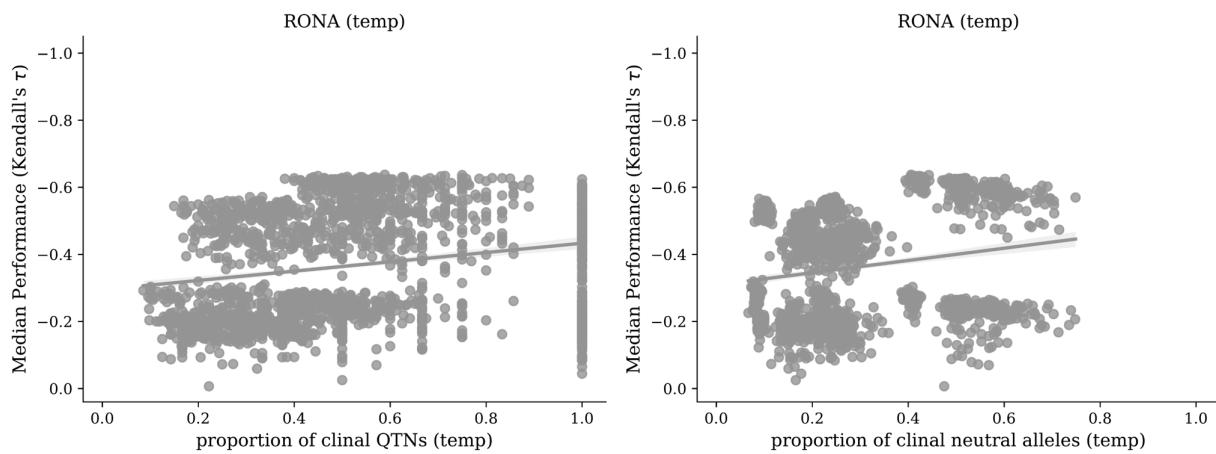
467



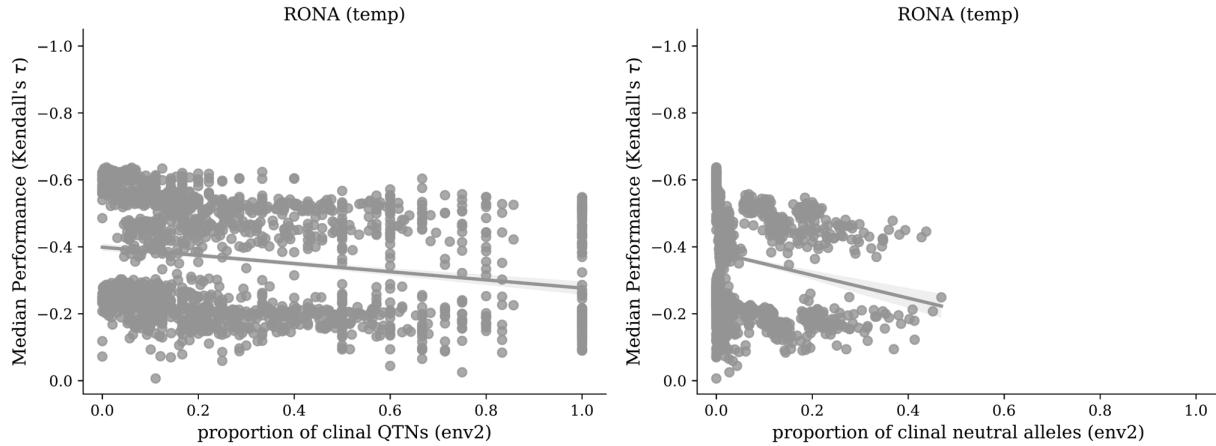
468



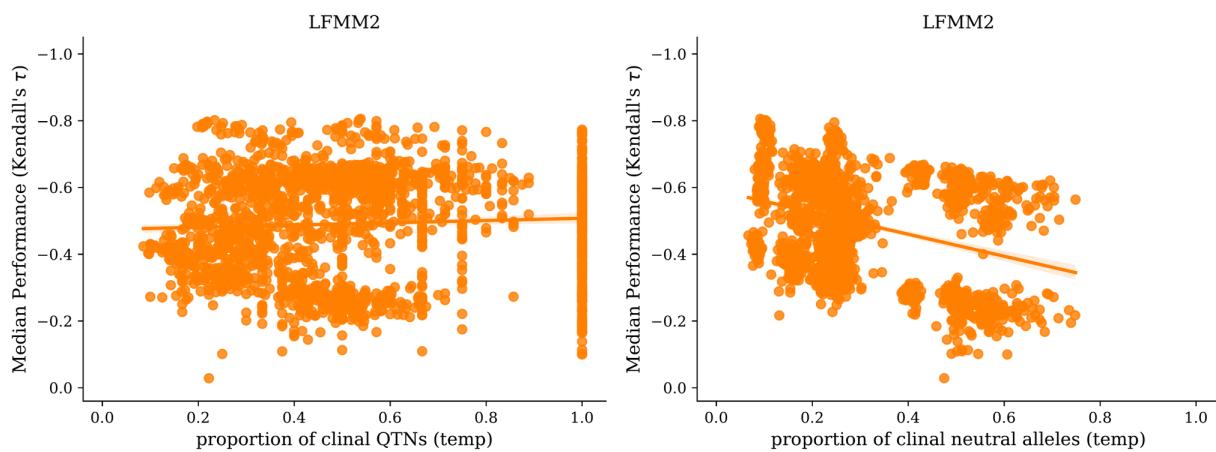
469



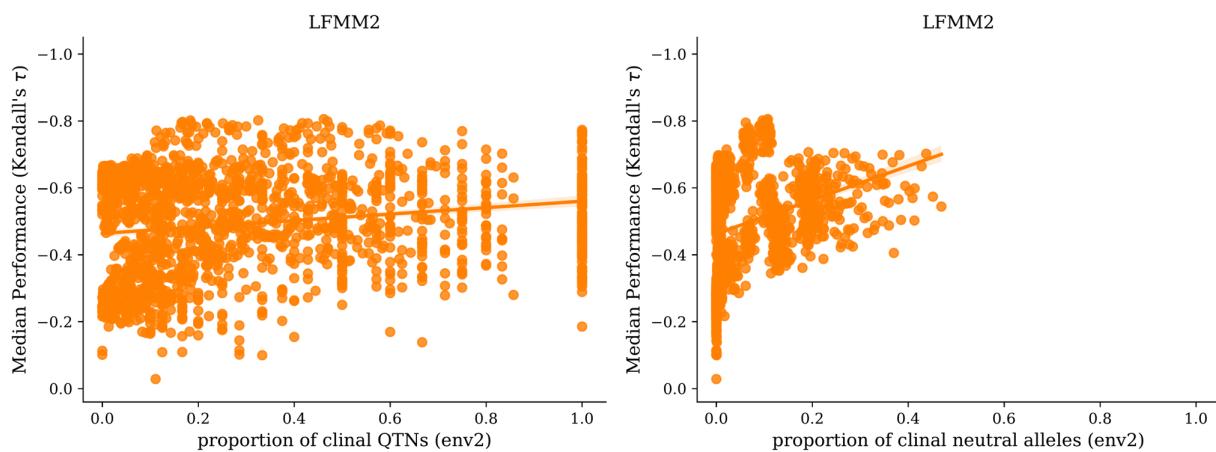
470



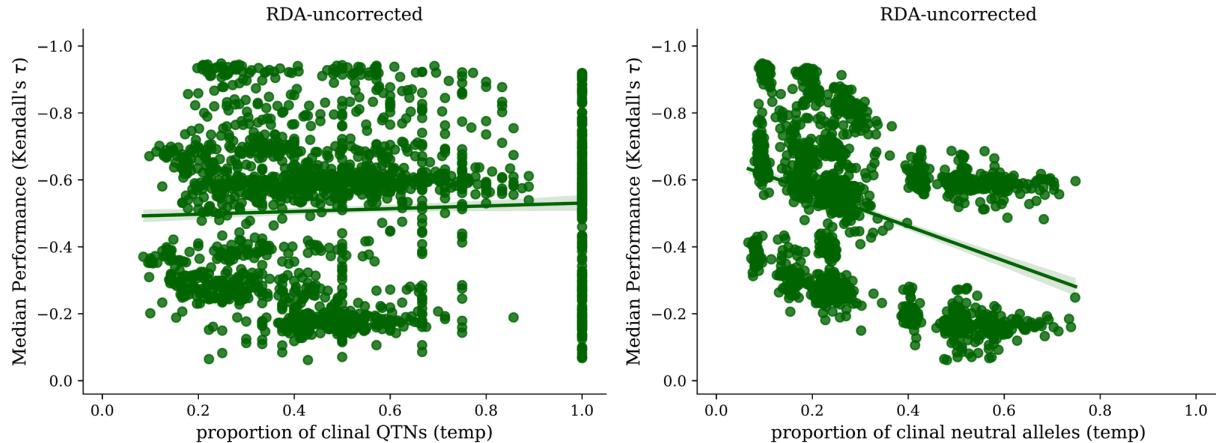
471



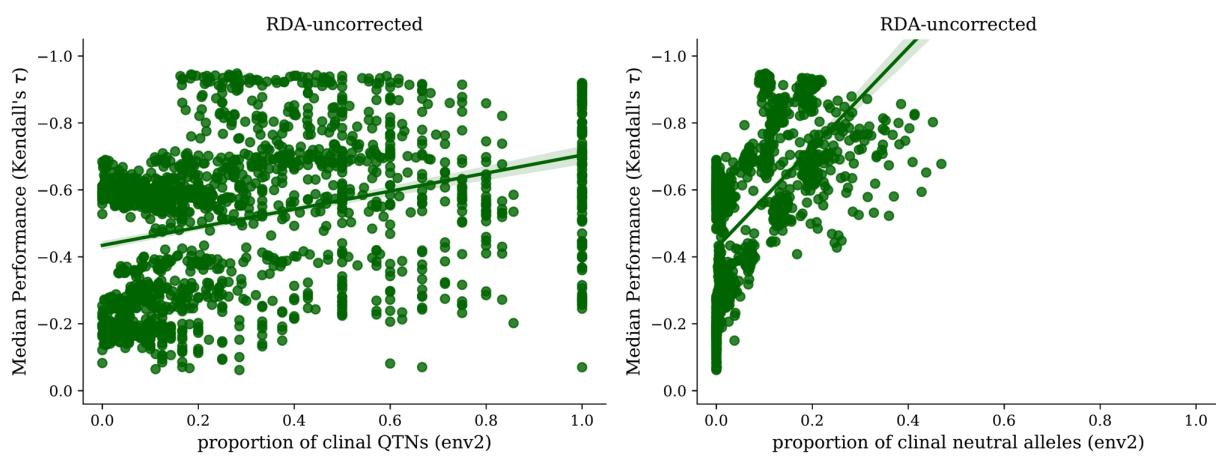
472



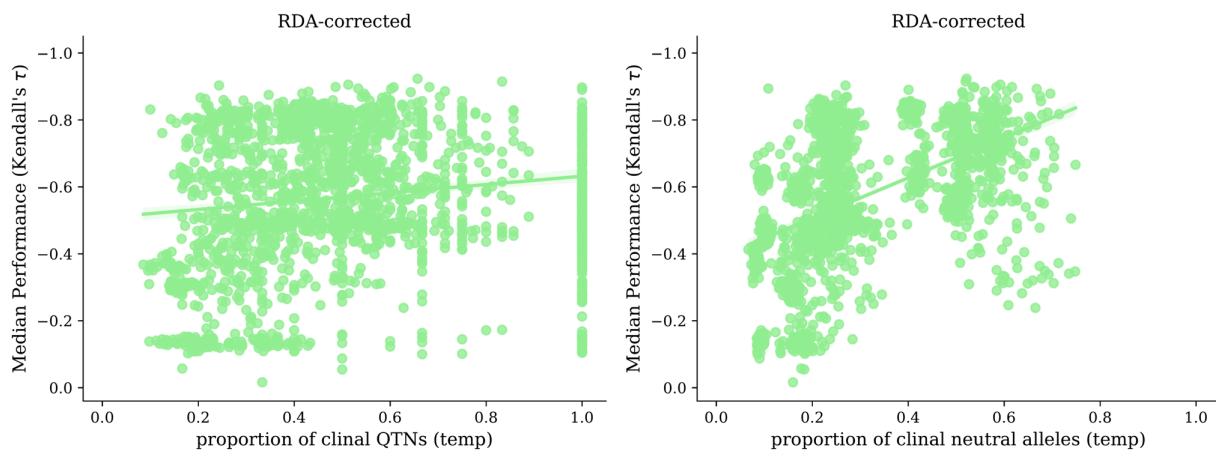
473



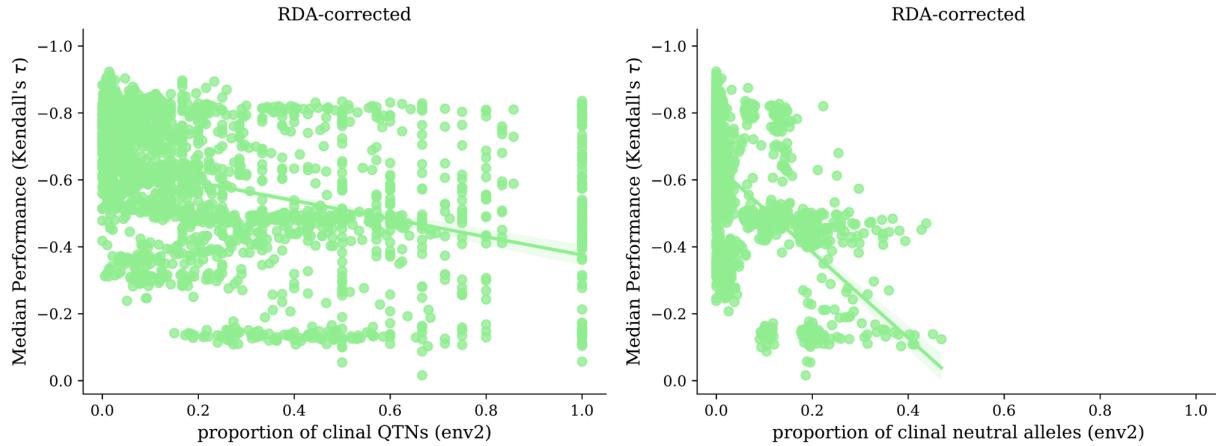
474



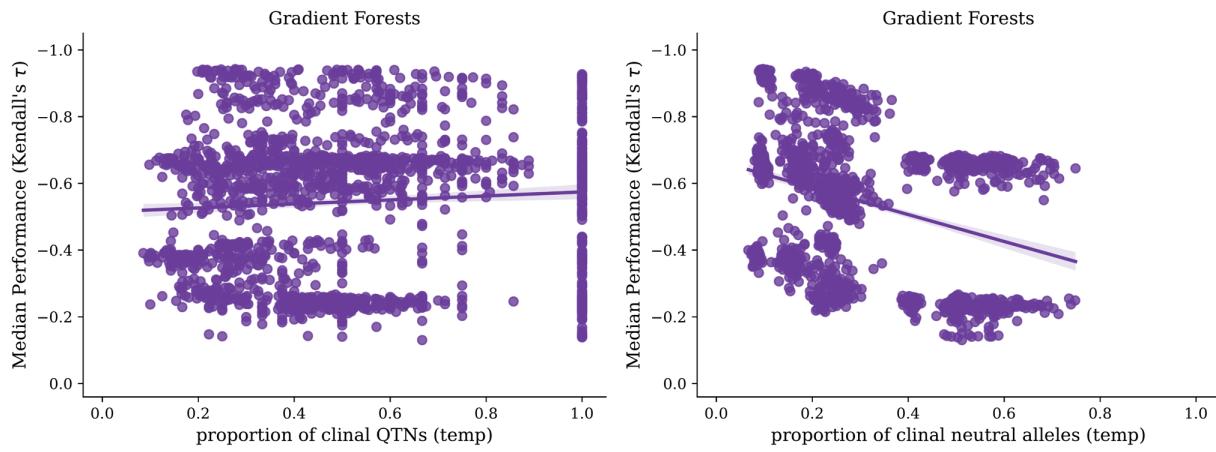
475



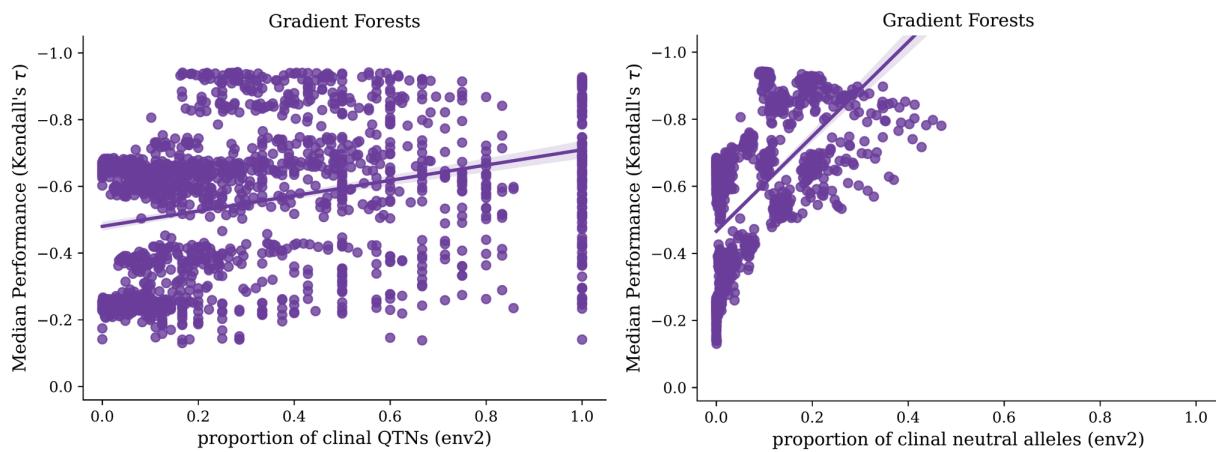
476



477

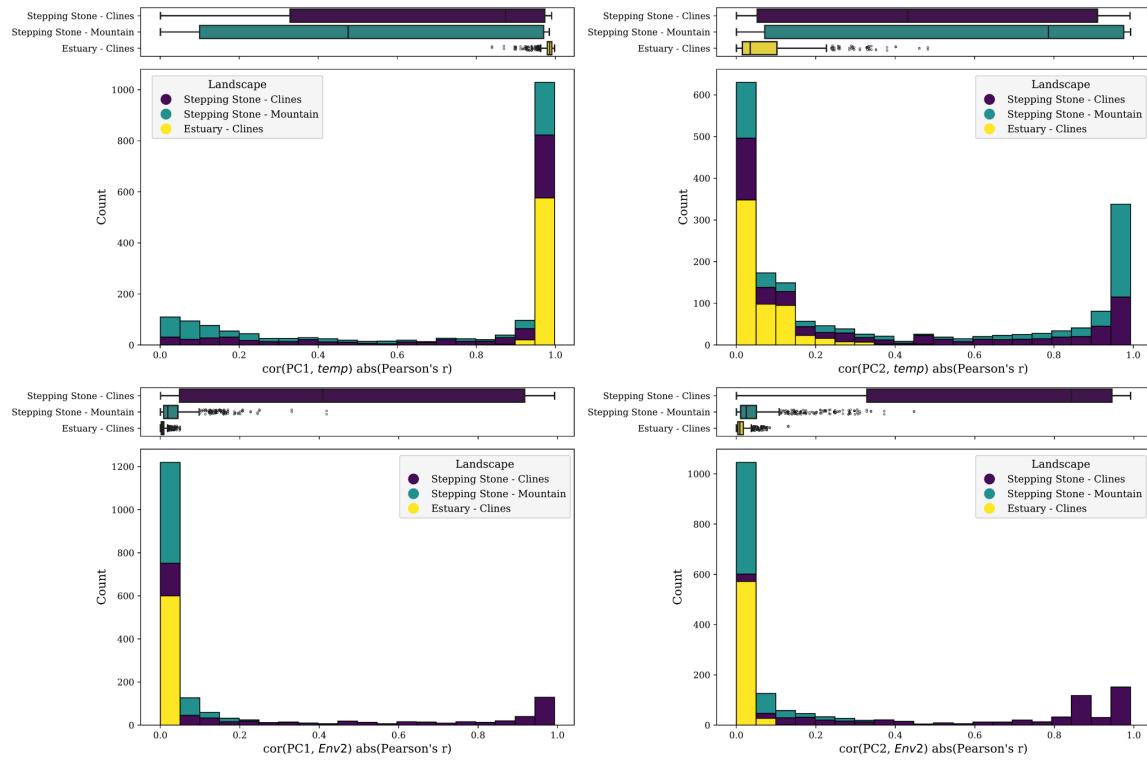


478



479 **Fig S11** Impact on method performance (y-axes) from the proportion of QTNs with clinal
 480 relationships with temp (first column) or Env2 (second column). Model performance is quantified
 481 as Kendall's rank correlation between offset and fitness; shown are median values from scores
 482 from 10 replicates per seed (100 common gardens for each replicate). Data included in this figure
 483 is from evaluation of 2-trait simulations using *all* markers. Code to create these figures can be
 484 found in 02.01.03.

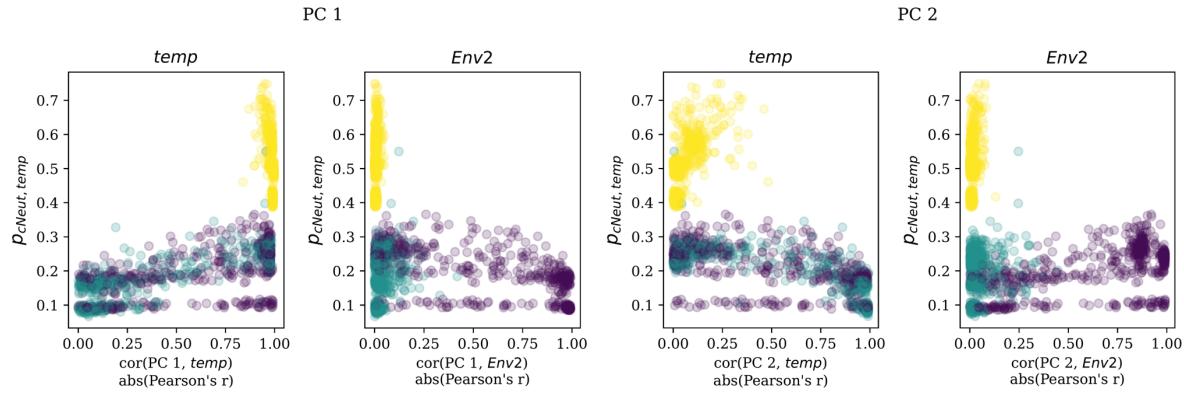
485



486

487 **Fig S12** Stacked bar plot showing correlation between environmental variables (rows) and axes
488 of population genetic structure (Principal Component Analysis axes [PC axes]; columns). Data
489 included in this figure is from all 2-trait simulations. Code to create this figure can be found in SC
490 02.10.03.

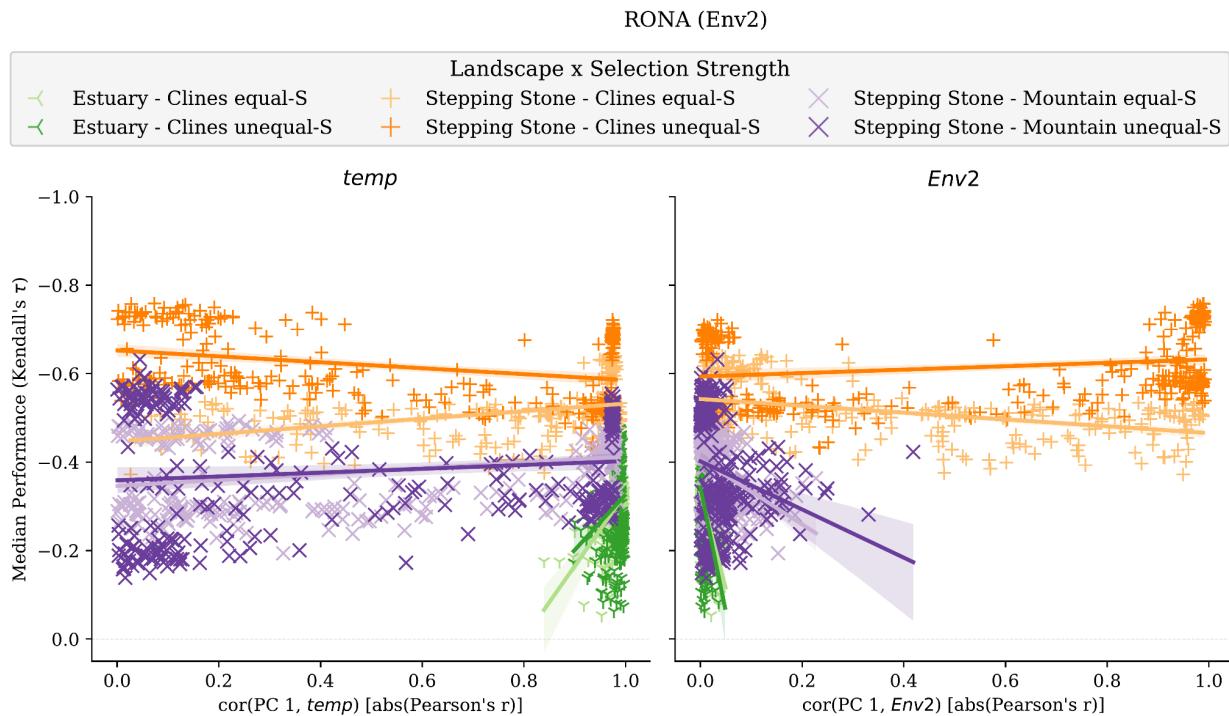
491



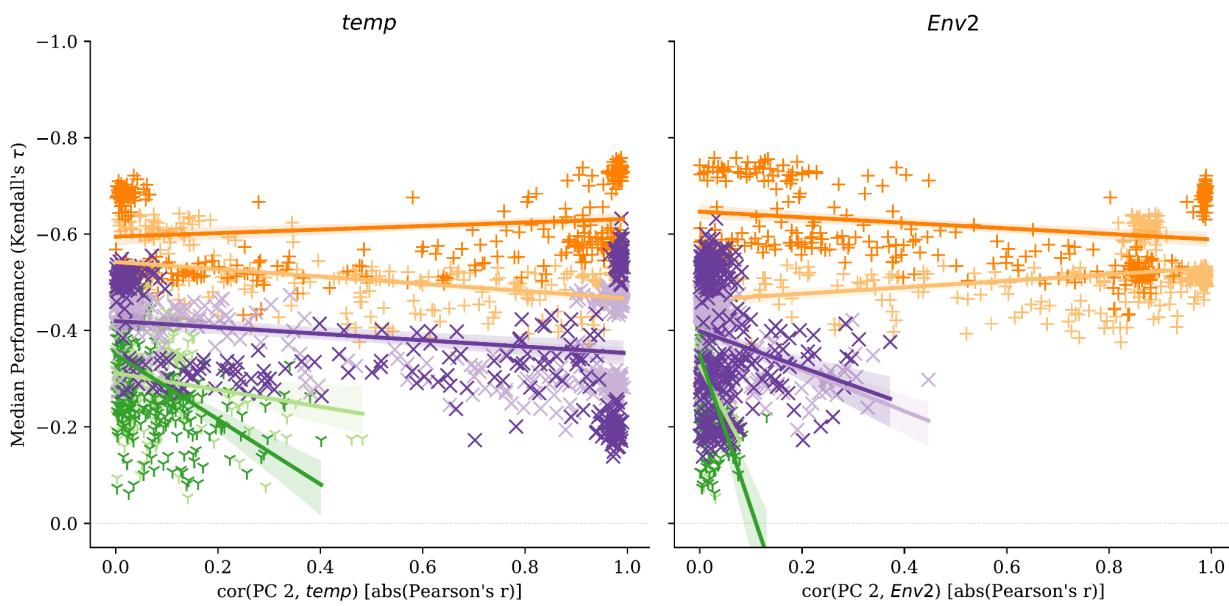
492

493 **Fig S13** Relationship between the proportion of clinal neutral loci for *temp* (y-axes, first row) or
 494 *Env2* (y-axes, second row) with the strength of the relationship between environmental variables
 495 and axes of population genetic structure. Purple = *Stepping Stone - Clines*; teal = *Stepping Stone*
 496 - *Clines*; yellow = *Estuary - Clines*. Data included in this figure is from all 2-trait simulations. Code
 497 to create this figure can be found in 02.10.03.

498 (Fig S14)

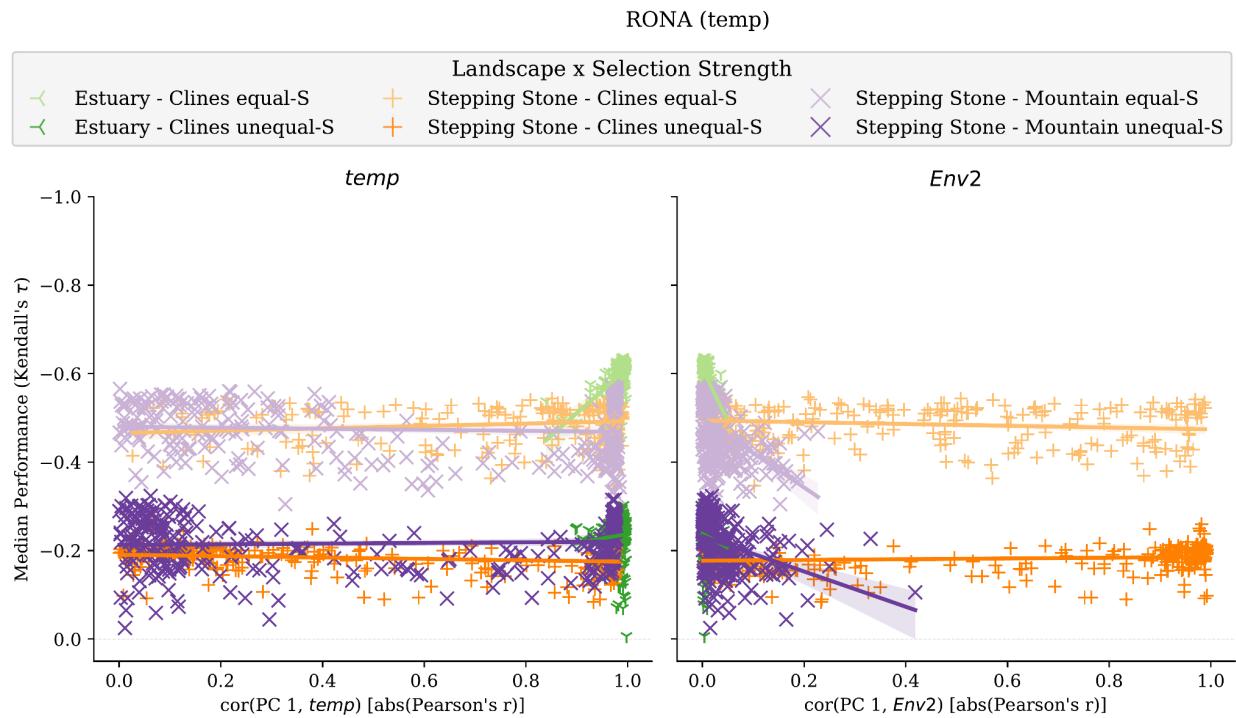


499

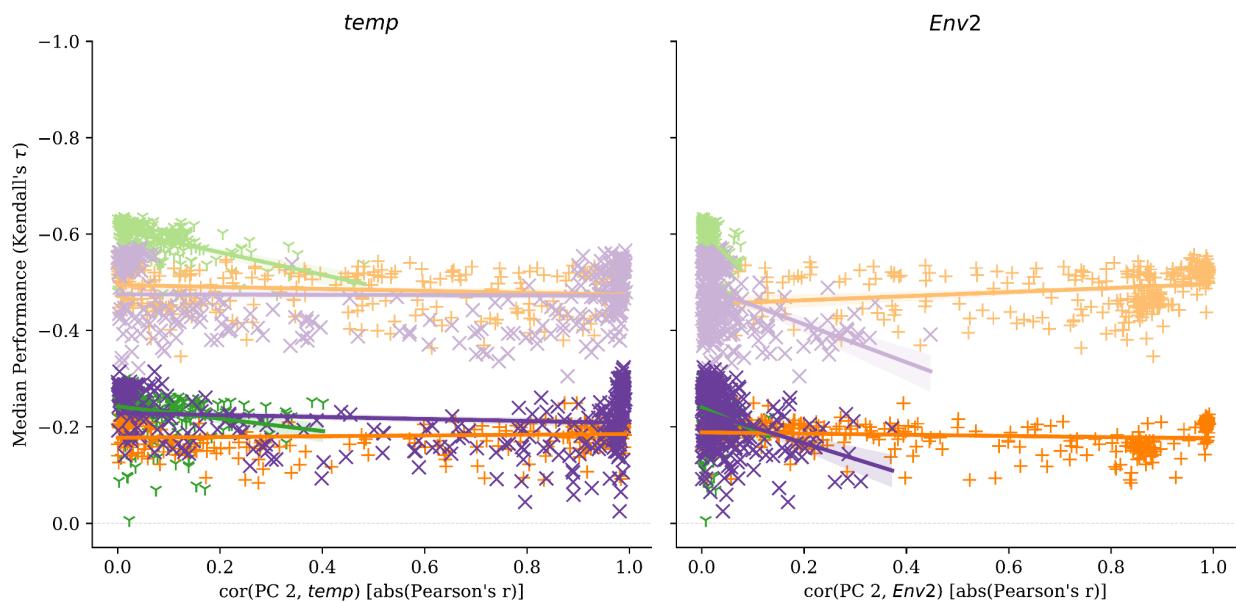


500

501 (Fig S14 continued)

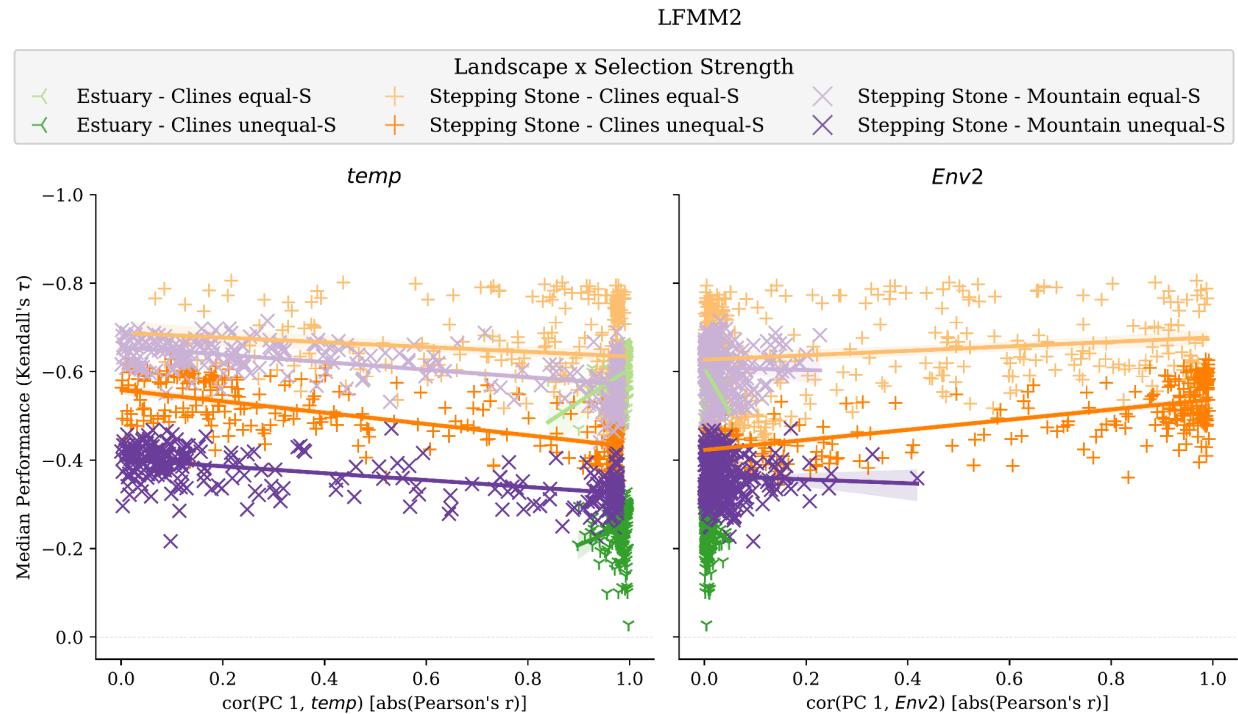


502

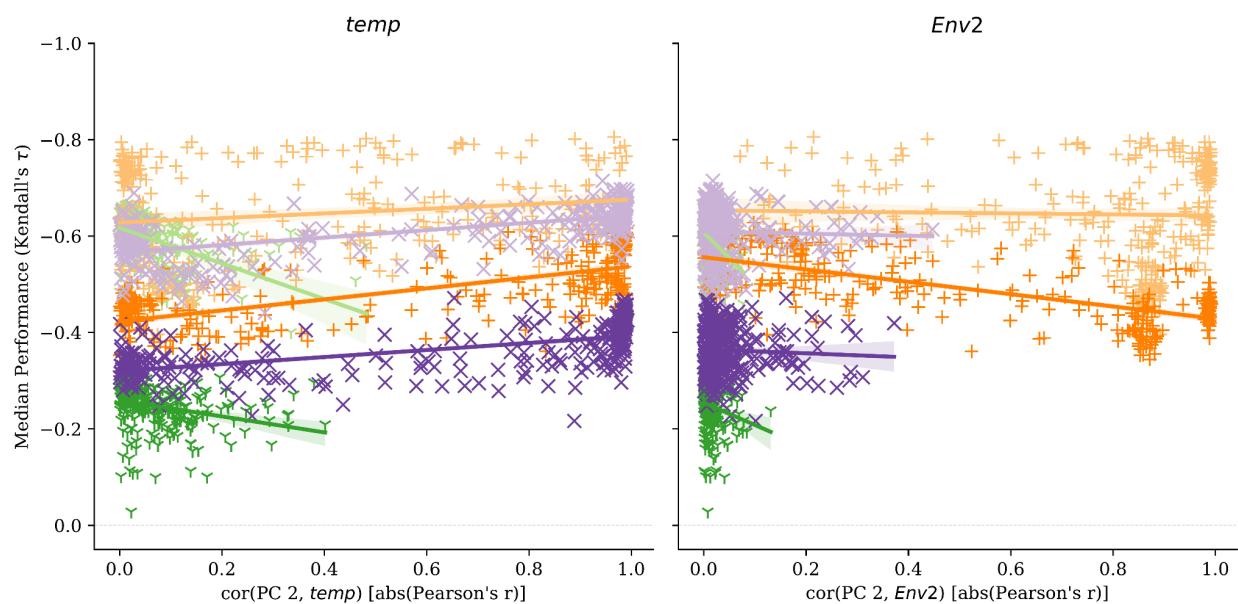


503

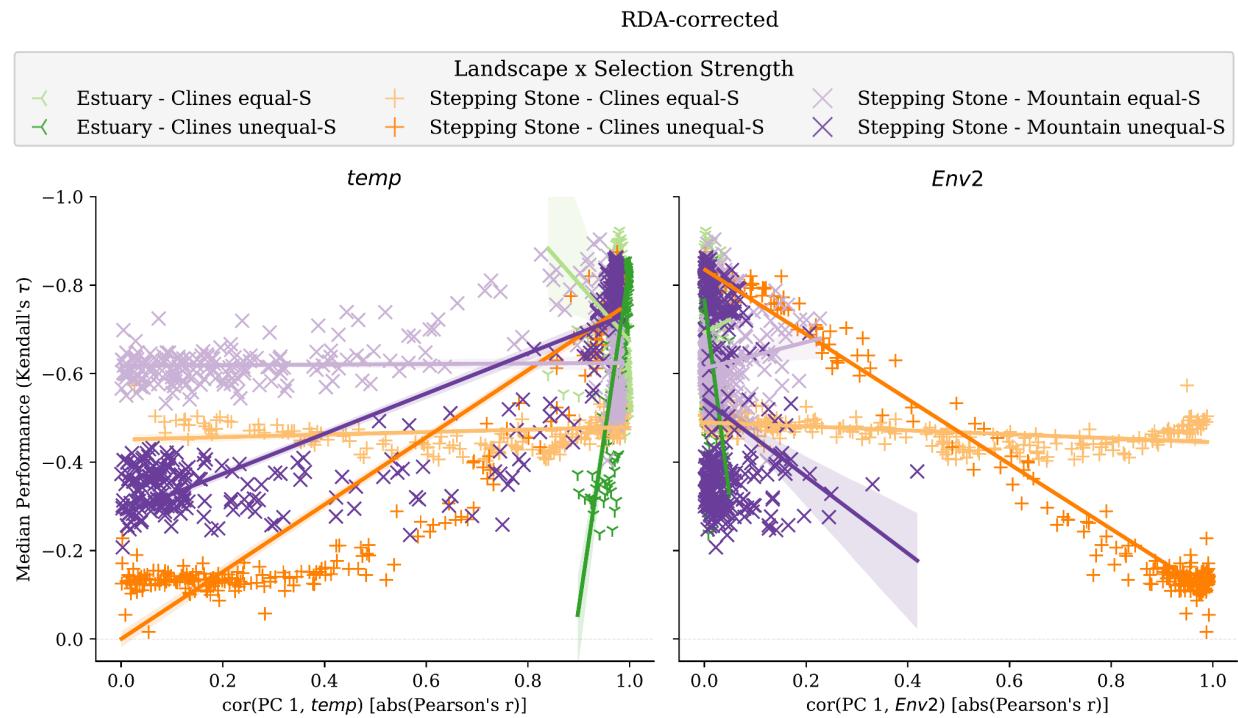
504 (Fig S14 continued)



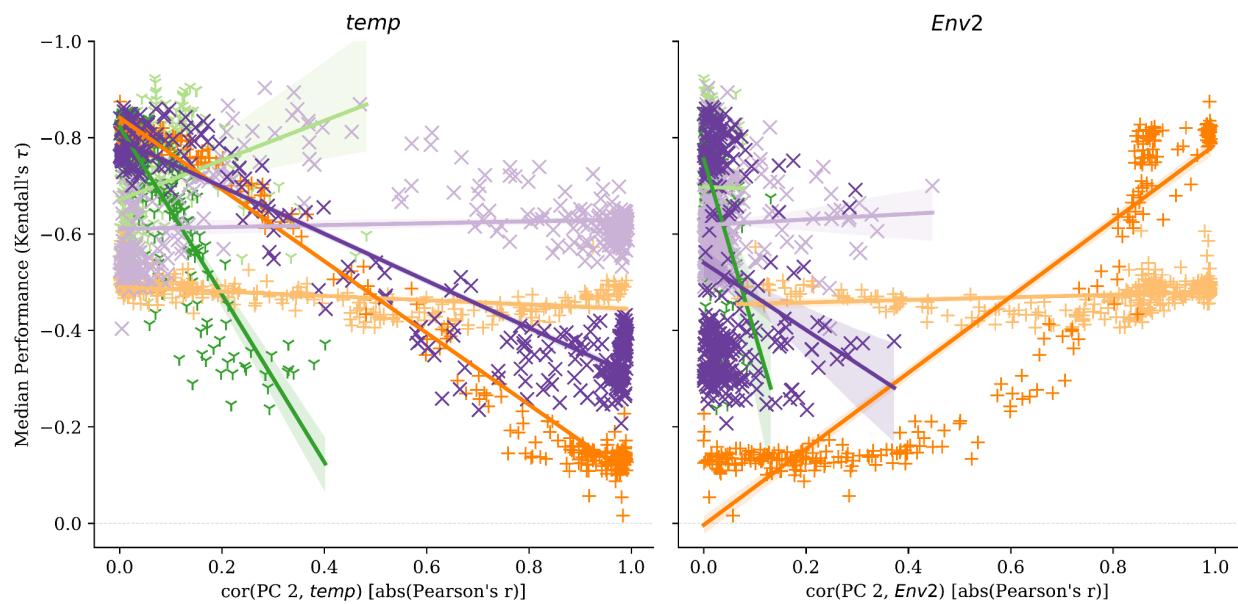
506



507 (Fig S14 continued)

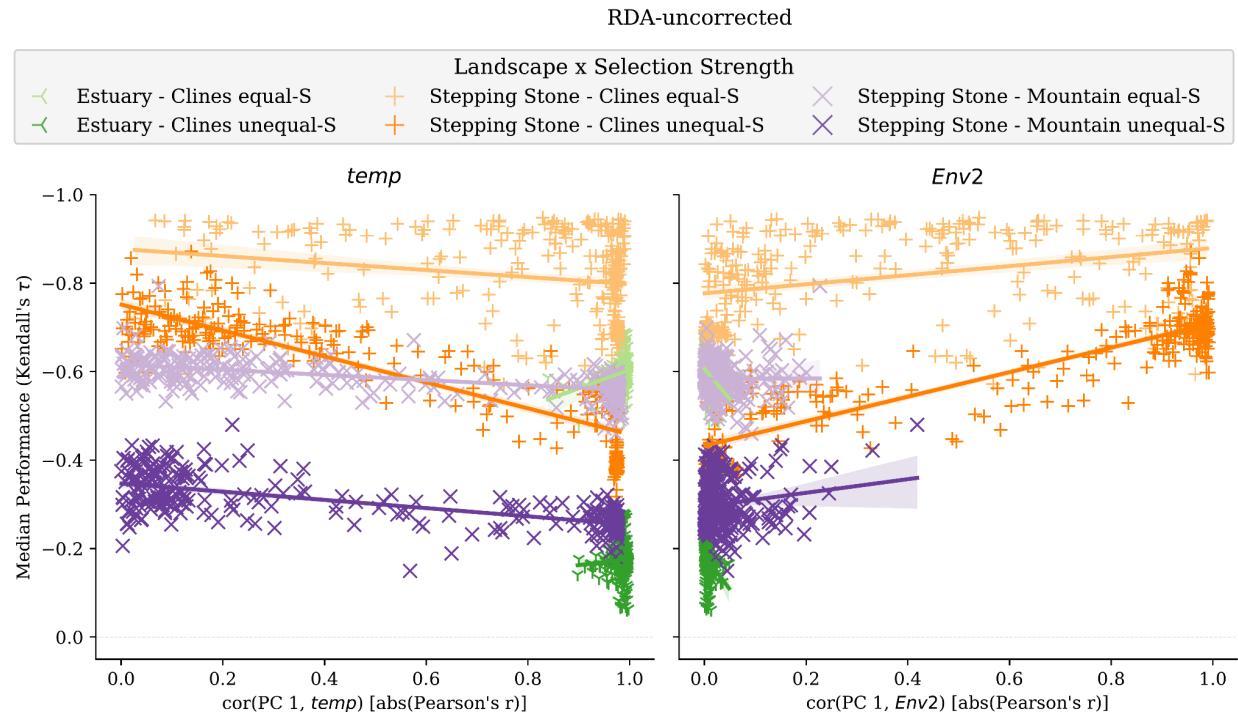


508

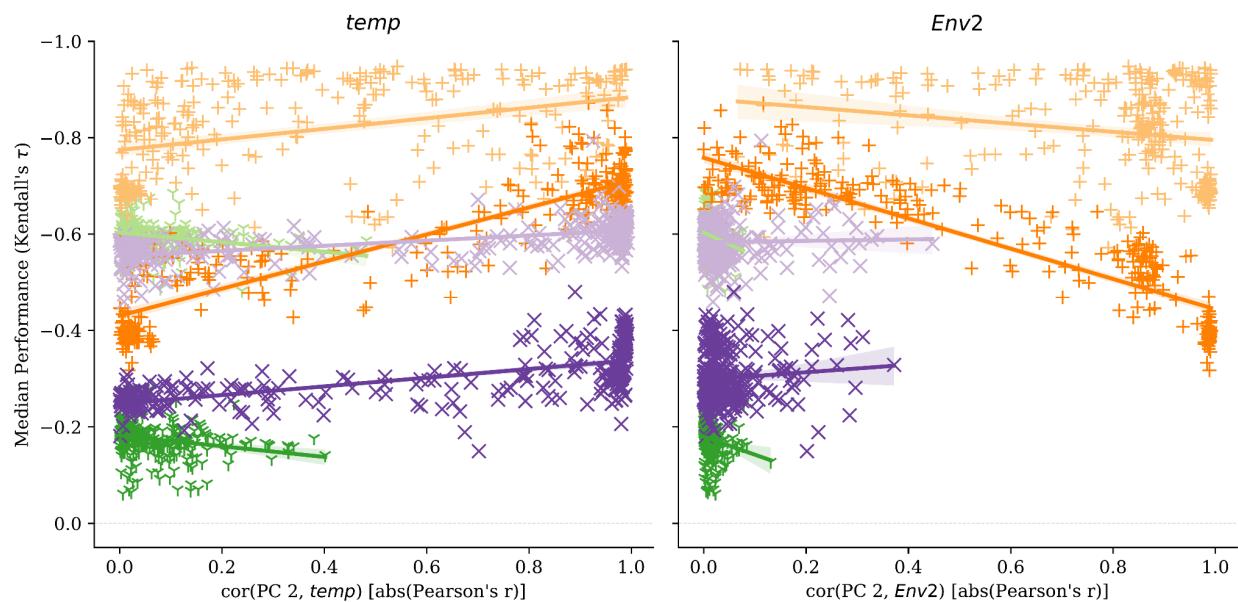


509

510 (Fig S14 continued)

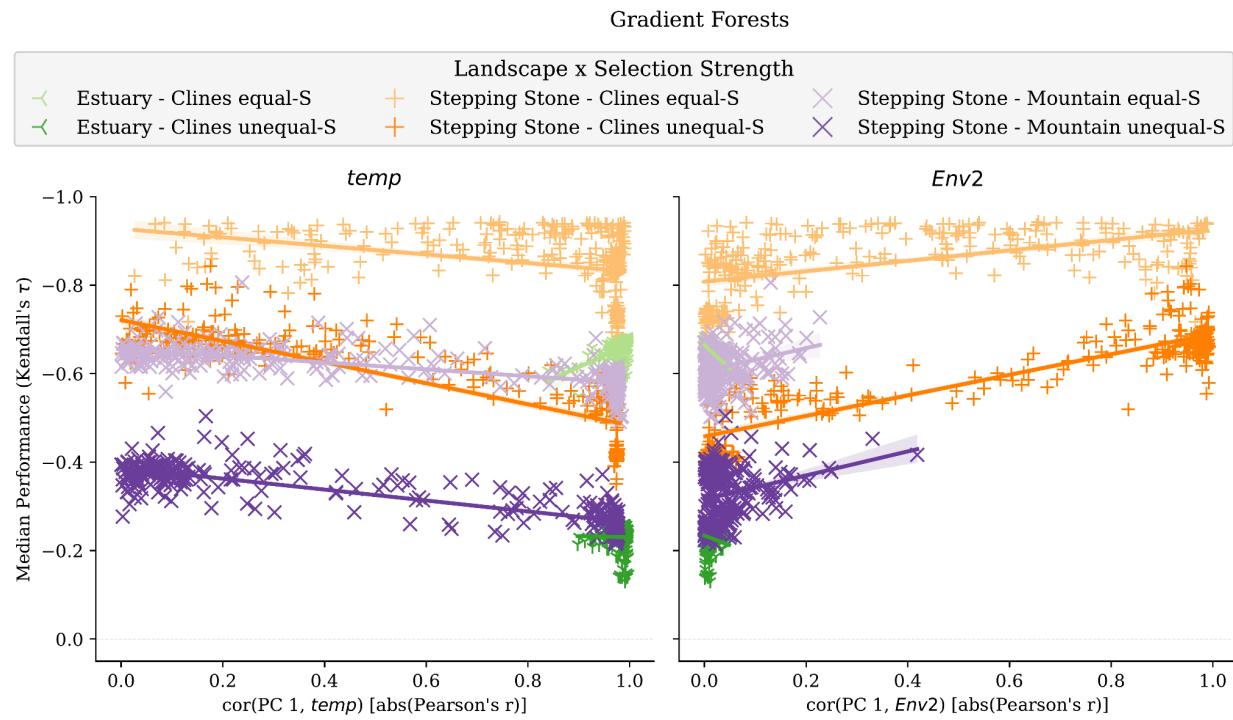


511

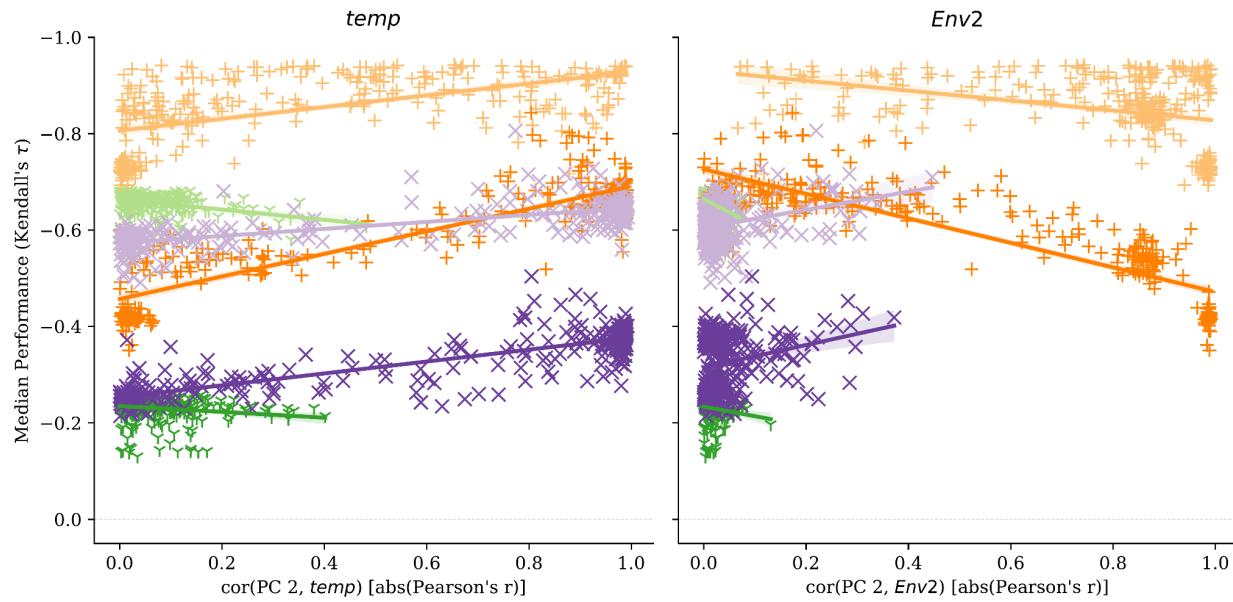


512

513 (Fig S14 continued)



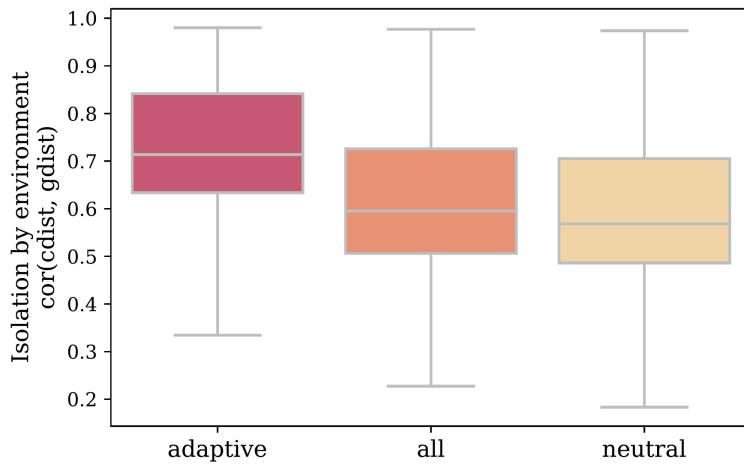
514



515

516 **Fig S14** Relationship between median performance and absolute correlation (Pearson's r)
 517 between environmental variables and axes of population genetic structure (principal component
 518 analysis axes). Each subfigure is for a different method (see panel titles). Data used in this figure
 519 is from 2-trait simulations. Code to create this figure can be found in SC 02.10.03.

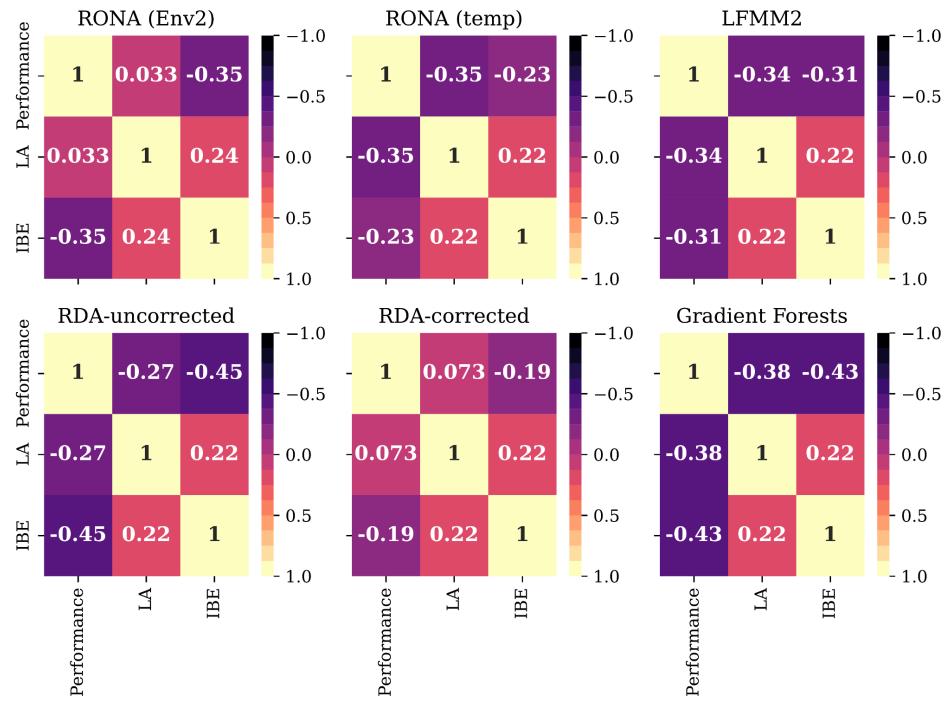
520



521 **Fig S15** Adaptive markers contain greater levels of isolation-by-environment (IBE) than other
522 marker sets. *IBE* is quantified as Spearman's rank correlation between population pairwise F_{ST}
523 and Euclidean distance of adaptive environments. Code to create this figure can be found in SC
524 02.02.10.

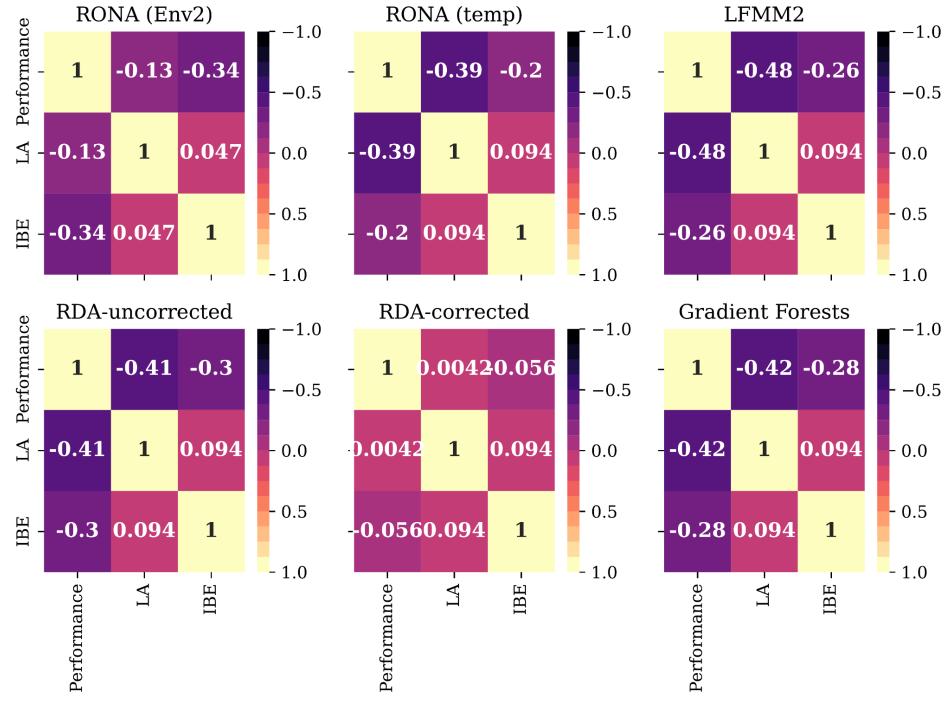
525 (Fig S16)

526

A) adaptive markers

527

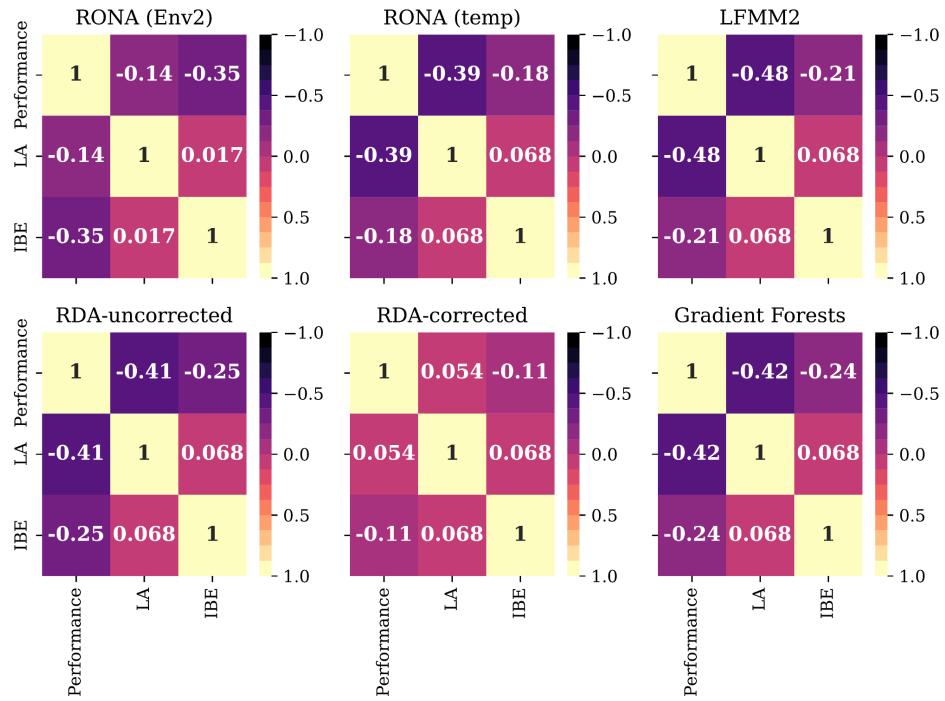
528

B) all markers

529

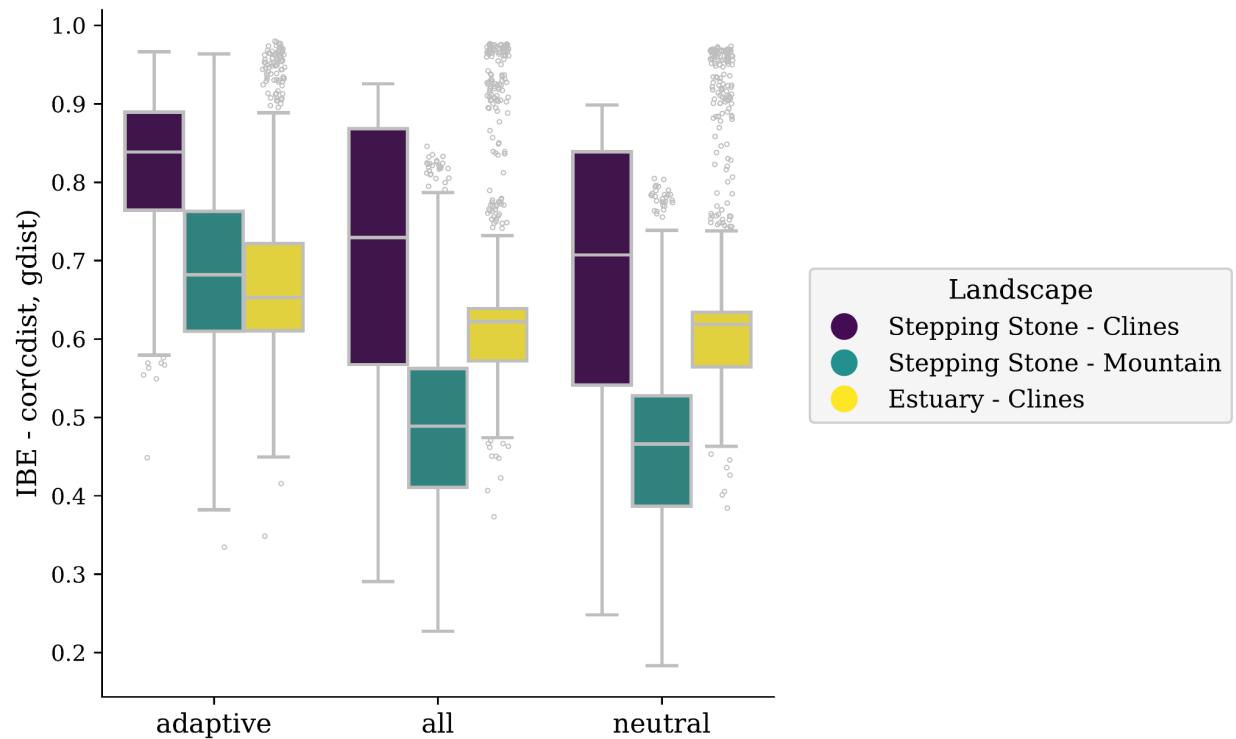
530

C) neutral markers

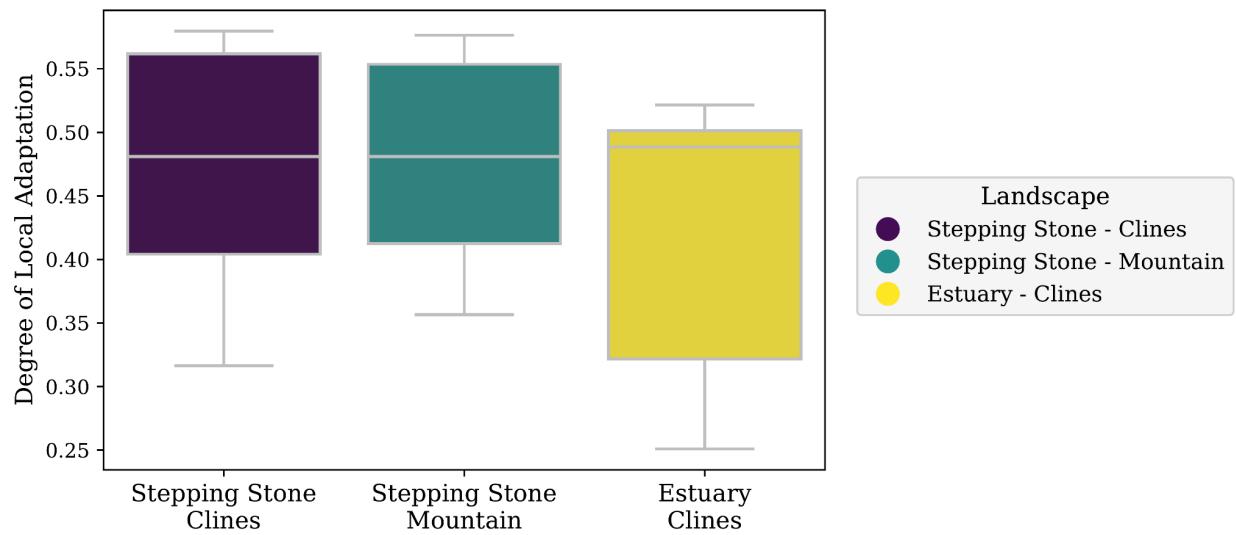


531

532 **Fig S16** The relationship between the degree of local adaptation (LA_{ASA}), levels of IBE within
 533 marker sets, and median performance of models trained with one of the three marker sets: (A)
 534 *adaptive*, (B) *all*, and (c) *neutral* marker sets. IBE is quantified as Spearman's rank correlation
 535 between population pairwise F_{ST} and Euclidean distance of adaptive environments. Data included
 536 in these figures are from 1- and 2-trait simulations. Code to create these figures can be found in
 537 SC 02.02.10.



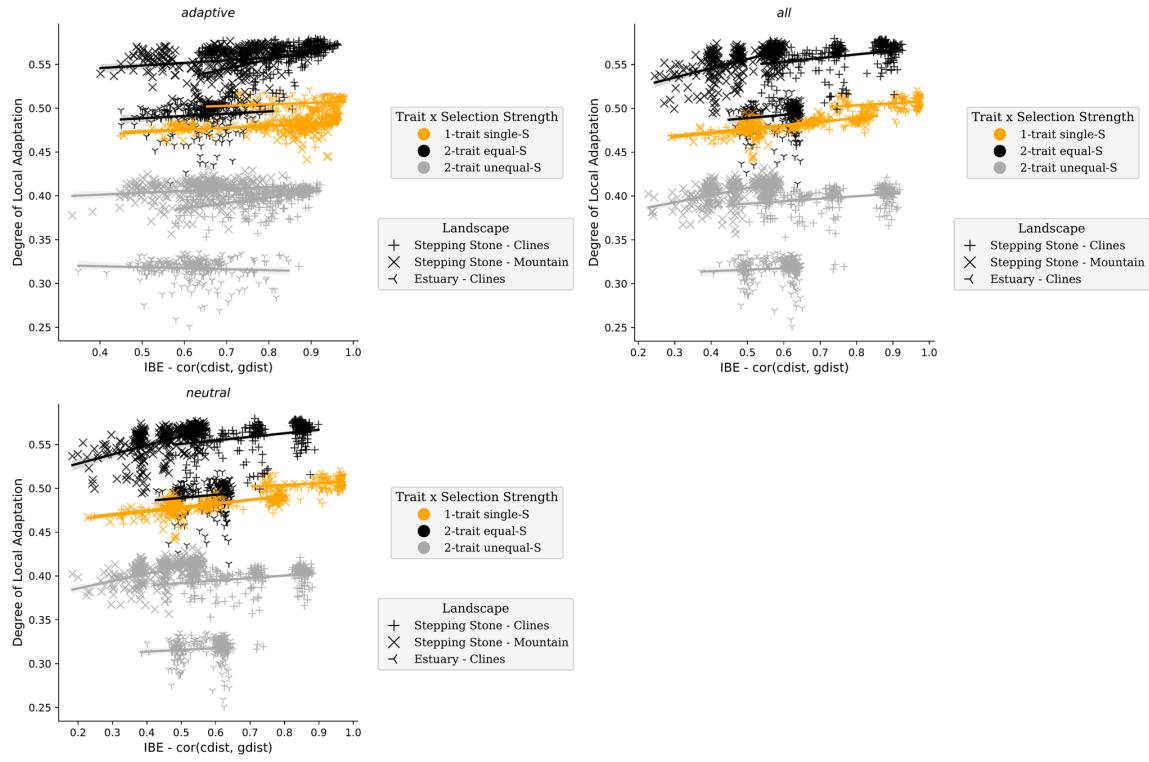
538



539

540 **Fig S17** Levels of isolation-by-environment in marker sets vary across landscapes (A) and the
 541 degree of local adaptation reached by metapopulations on these landscapes (B). The pattern in
 542 (A) given (B) is in contrast to patterns between levels of IBE and the degree of local adaptation
 543 (Fig. S29). I IBE is quantified as Spearman's rank correlation between population pairwise F_{ST}
 544 (gdist) and Euclidean distance of adaptive environments (cdist). Data in this figure is from all 1-
 545 and 2-trait simulations. Code to create this figure can be found in SC 02.02.10.

546

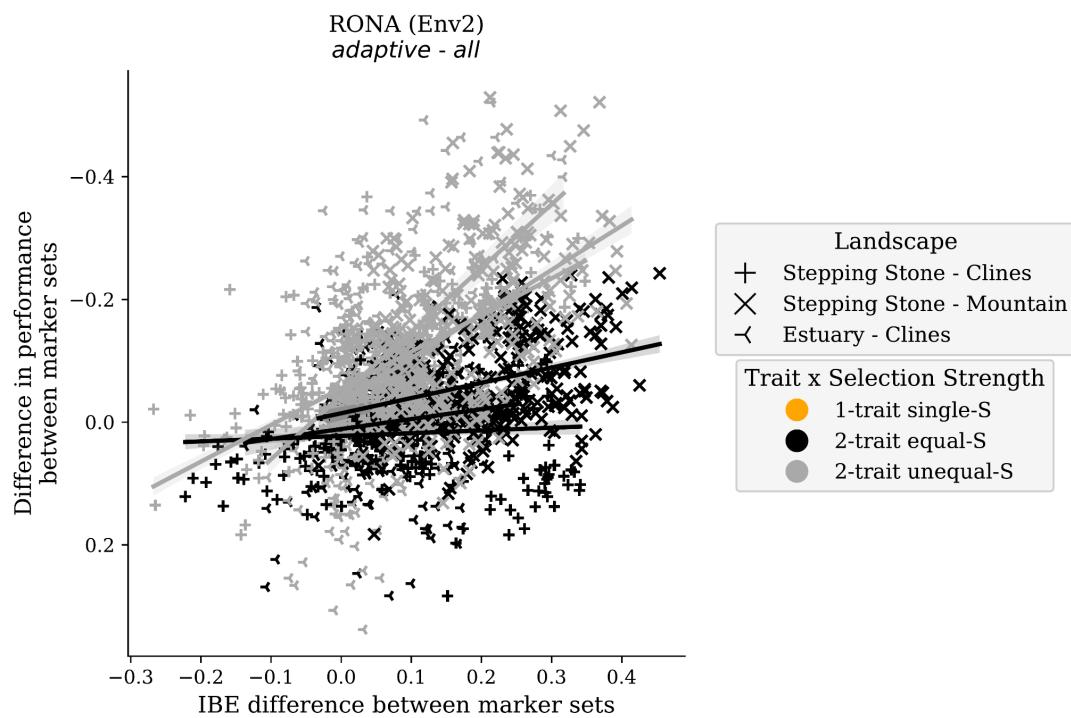


547

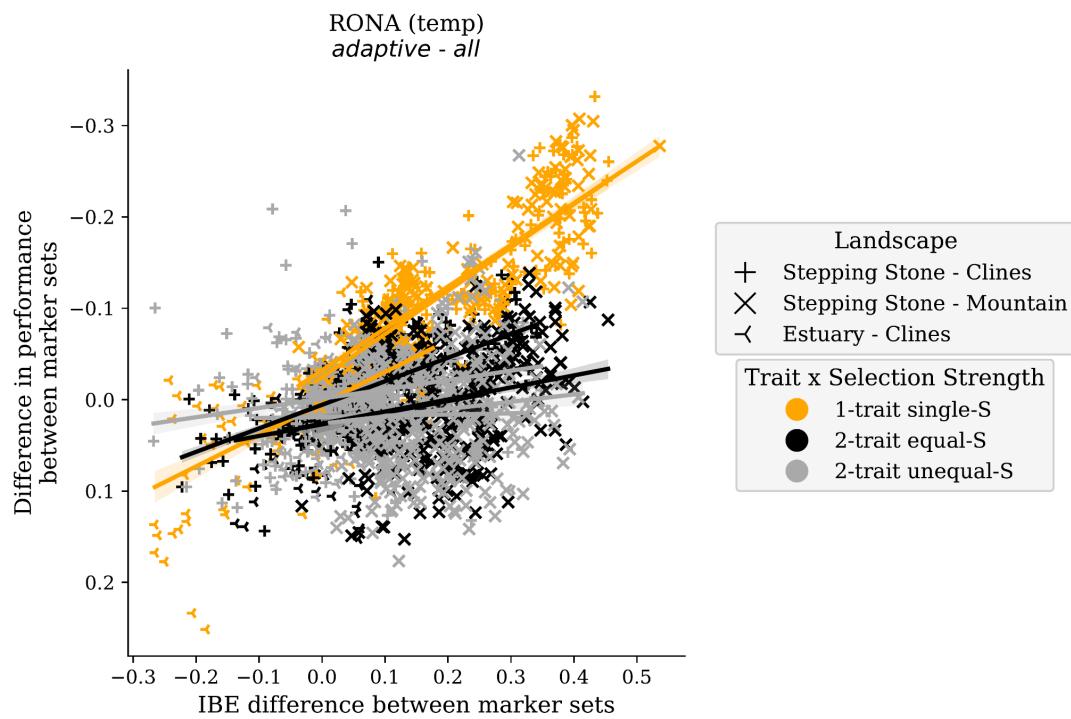
548 **Fig S18** The levels of isolation-by-distance in marker sets (panels) are weakly correlated with
 549 the degree of local adaptation ($LA_{\Delta SA}$) within simulation levels. IBE is quantified as Spearman's
 550 rank correlation between population pairwise F_{ST} ($gdist$) and Euclidean distance of adaptive
 551 environments ($cdist$). Data included in this figure is from all marker sets from 1- and 2-trait
 552 simulations. Code to create this figure can be found in 02.02.10.

553 (Fig

S19)



554

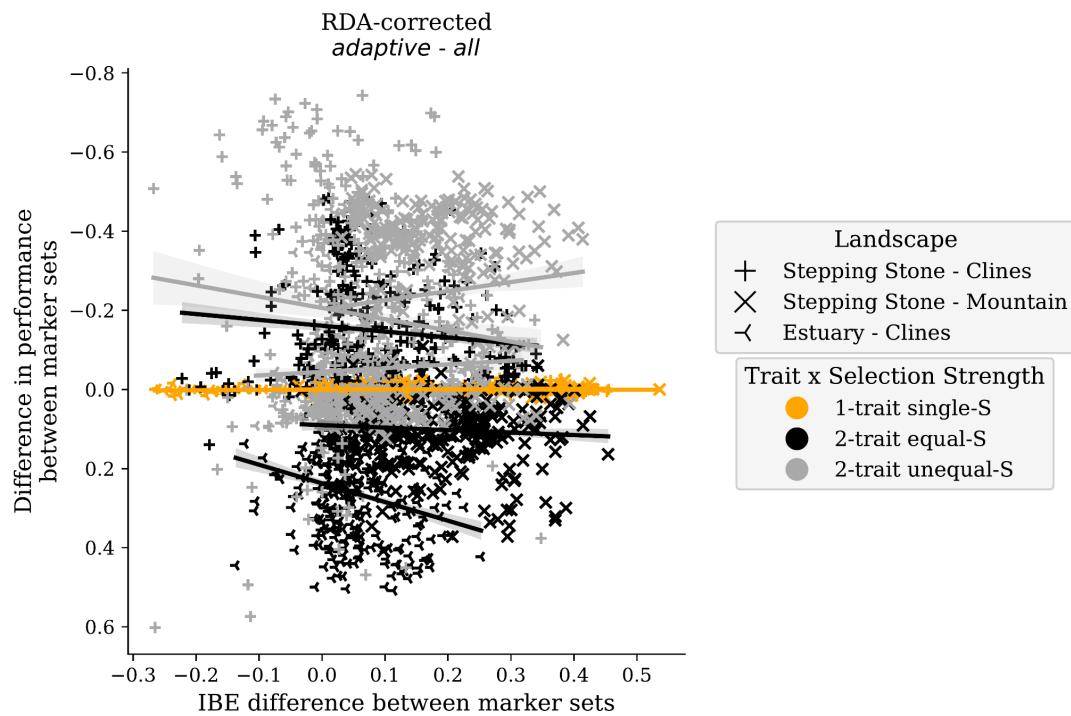


555

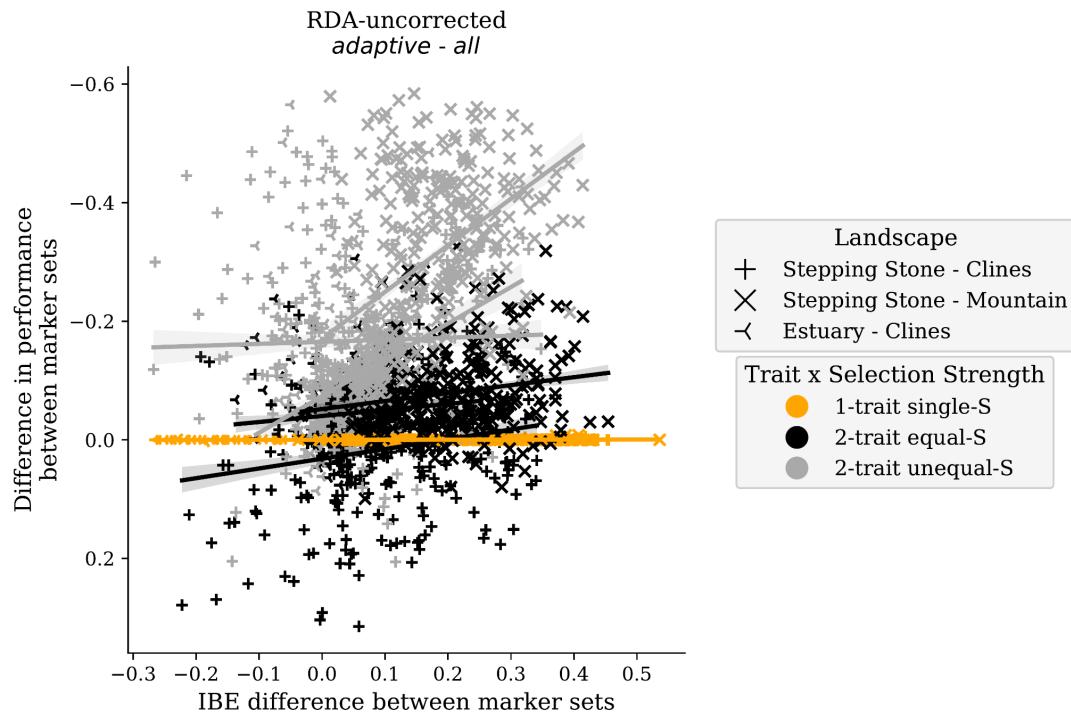
556 (Fig

S19

continued)



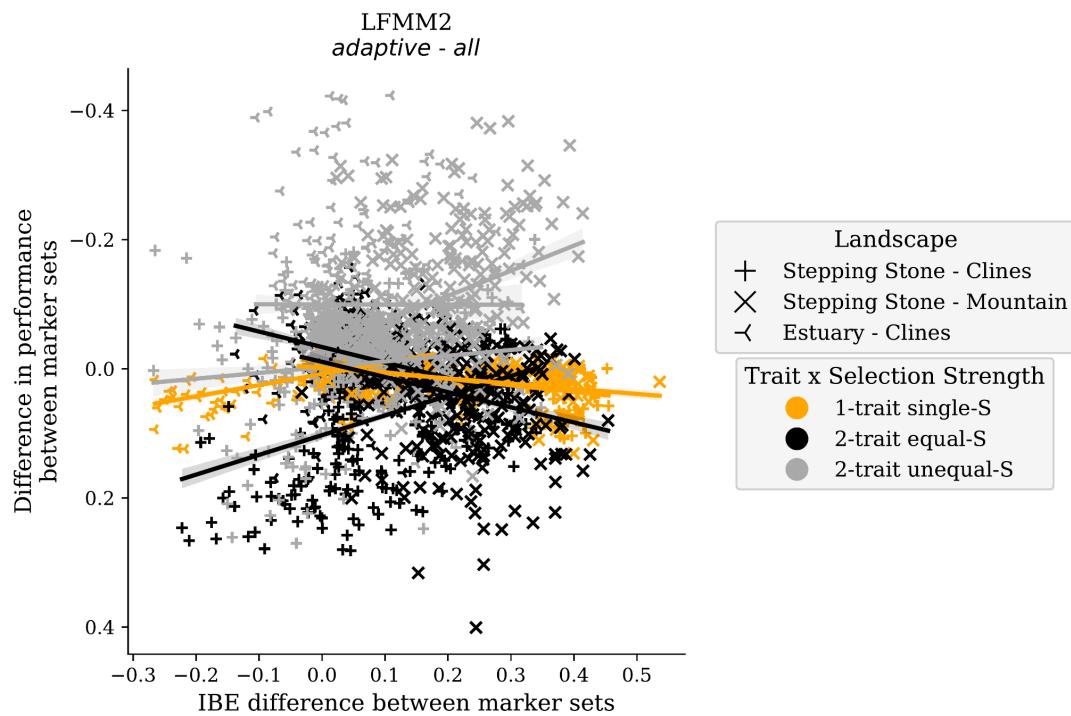
557

558
559

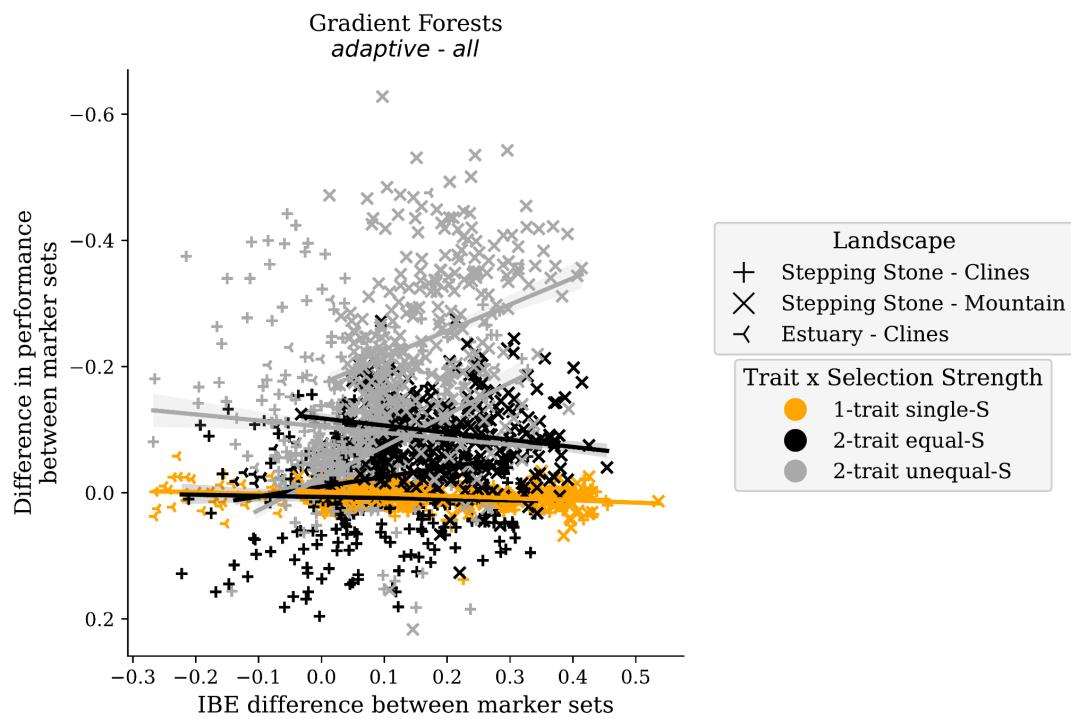
560 (Fig

S19

continued)



561



562

563 **Fig S19** Differences in levels of IBE between marker sets used to train models is generally
564 unrelated to differences in model performances. Shown is the difference in median performance
565 between *adaptive* and *all* marker sets and the difference in *IBE* between these marker sets. *IBE*
566 is quantified as Spearman's rank correlation between population pairwise F_{ST} and Euclidean
567 distance of adaptive environments. Data in this figure is from 1- and 2-trait simulations. Code to
568 create these figures can be found in SC 02.02.12.

569

	1	2	3	4	5	6	7	8	9	10	
Latitude	10	91.0	92.0	93.0	94.0	95.0	96.0	97.0	98.0	99.0	100.0
	9	81.0	82.0	83.0	84.0	85.0	86.0	87.0	88.0	89.0	90.0
	8	71.0	72.0	73.0	74.0	75.0	76.0	77.0	78.0	79.0	80.0
	7	61.0	62.0	63.0	64.0	65.0	66.0	67.0	68.0	69.0	70.0
	6	51.0	52.0	53.0	54.0	55.0	56.0	57.0	58.0	59.0	60.0
	5	41.0	42.0	43.0	44.0	45.0	46.0	47.0	48.0	49.0	50.0
	4	31.0	32.0	33.0	34.0	35.0	36.0	37.0	38.0	39.0	40.0
	3	21.0	22.0	23.0	24.0	25.0	26.0	27.0	28.0	29.0	30.0
	2	11.0	12.0	13.0	14.0	15.0	16.0	17.0	18.0	19.0	20.0
570	1	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0

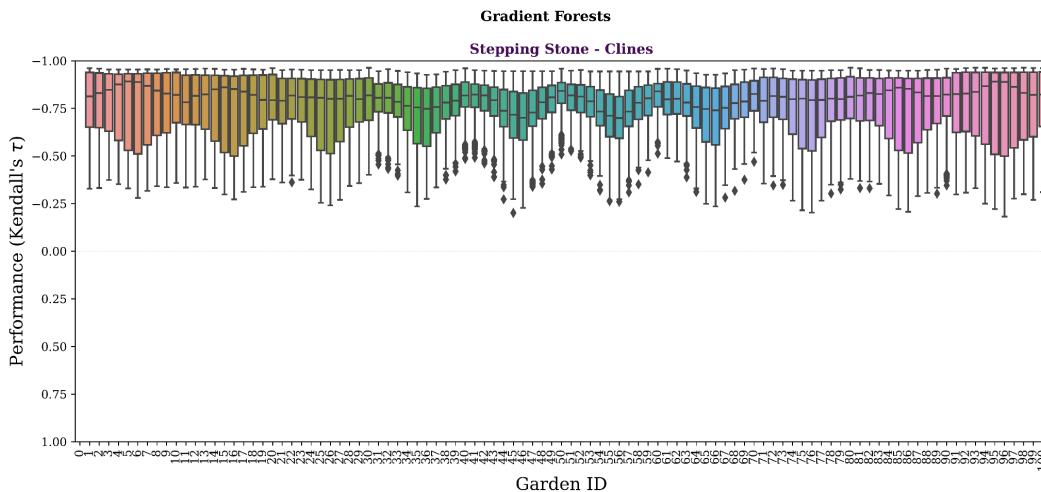
571

Longitude

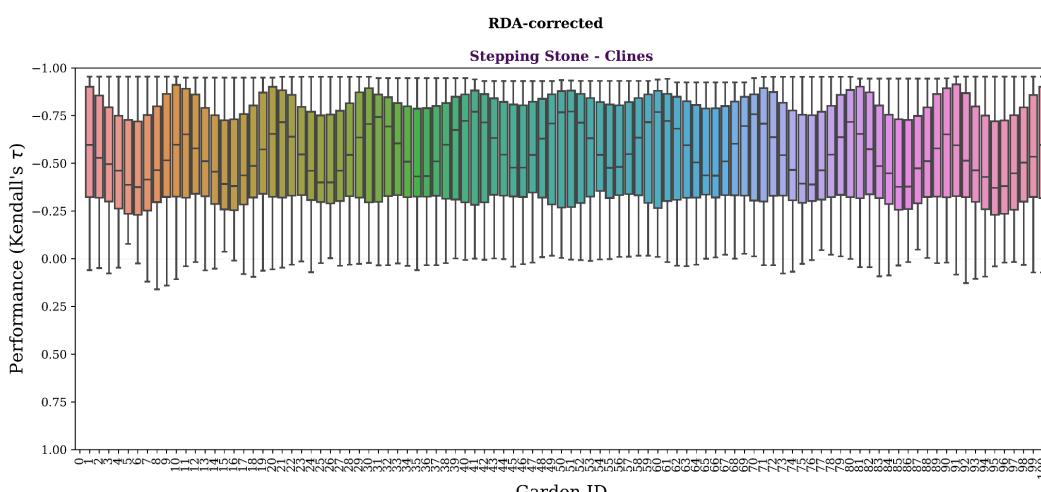
572

573 **Fig S20** A map of Garden ID (unbolded entries) across each landscape for 1-, 2- and 6-trait
 574 simulations (latitudinal and longitudinal grids are bolded). This map can be used to interpret the
 575 ordering of gardens along x-axes of Figs. S21 S22 and S23. Code used to create this figure can be
 576 found in SC 02.02.04.

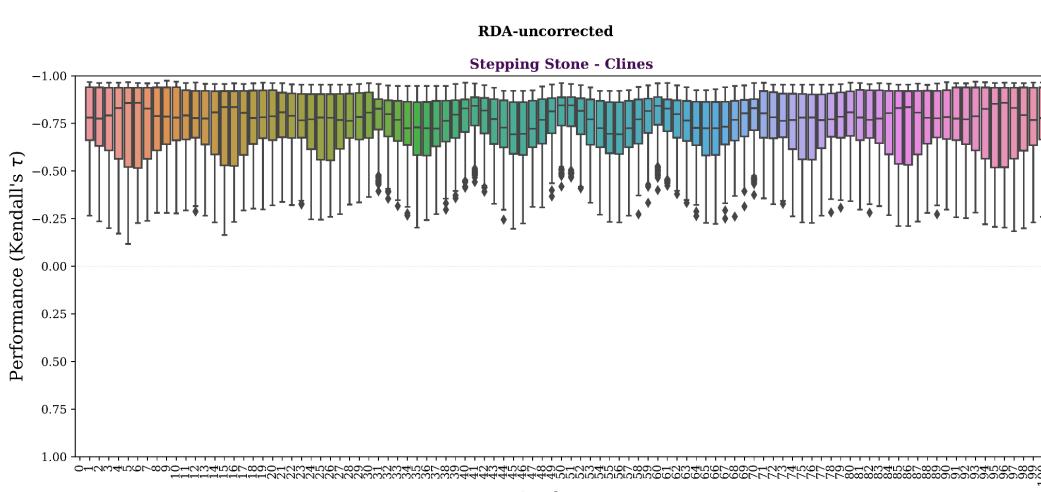
577 (Fig. S21)



578

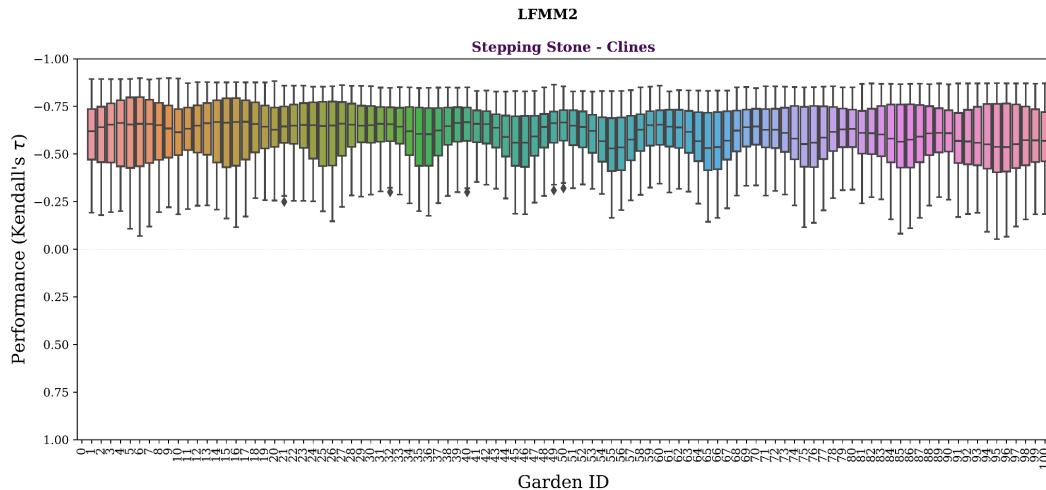


579

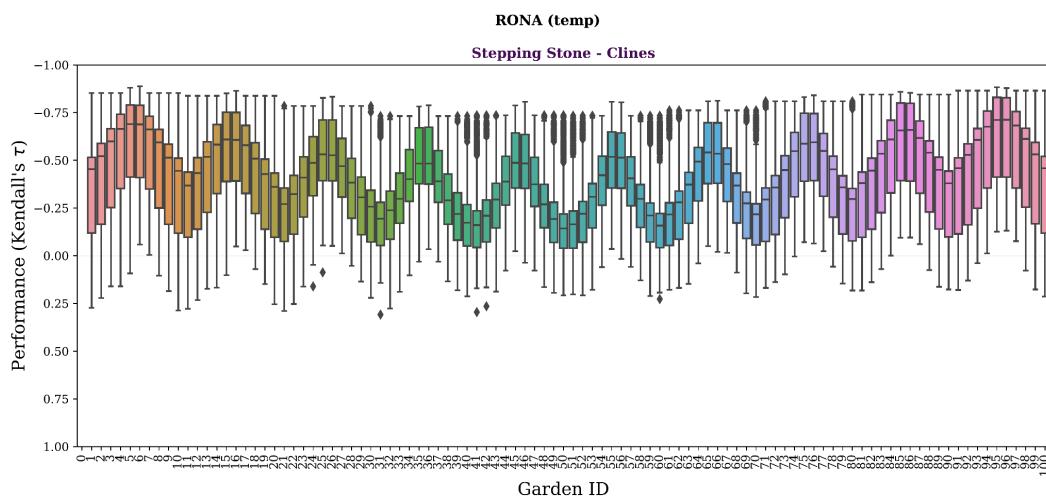


580

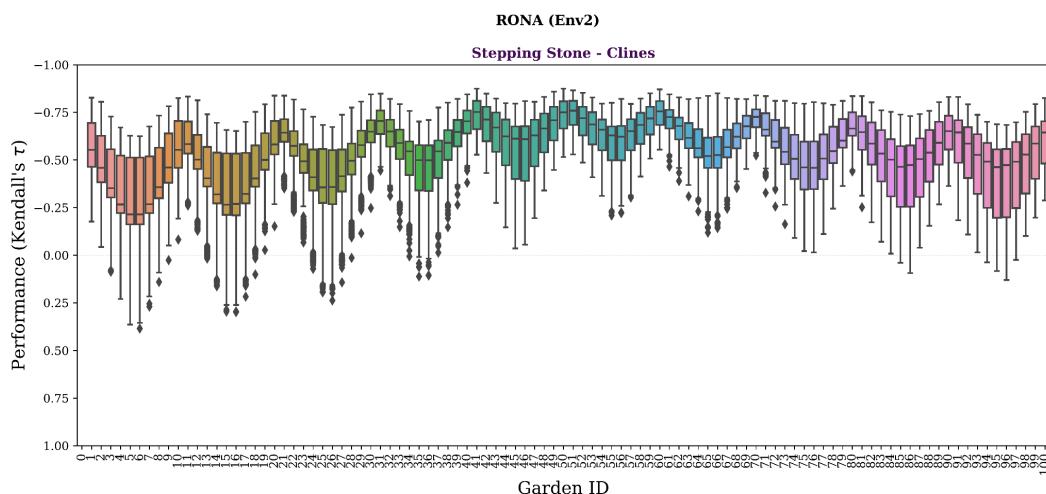
581 (Fig S21 continued)



582



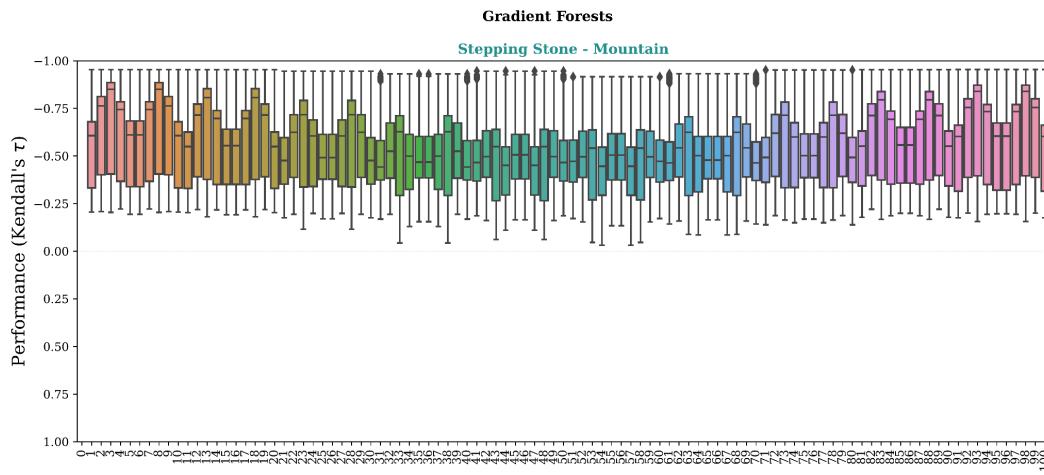
583



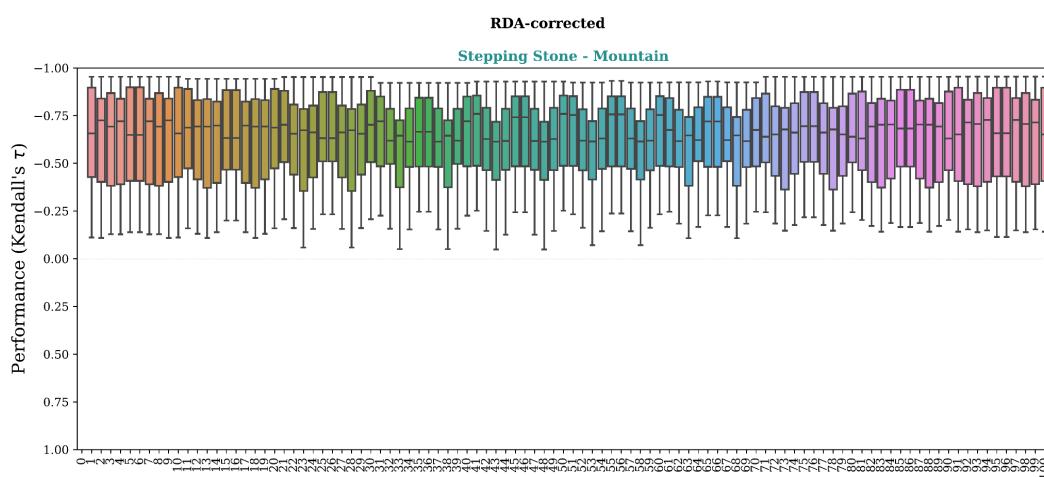
584

585 **Fig S21** Genomic offset methods have variable performance across the *Stepping-Stone Clines*
586 landscape. Shown is the variability of each offset method performance (y-axes) across the 100
587 common gardens (x-axes). Gardens are ordered from left to right by garden ID. This ordering of
588 gardens is equivalent to the southwestern-most garden first and northeastern-most garden last
589 (see Fig. S20 for a map of garden ID across each landscape). Similar figures for *Stepping-Stone*
590 *Mountain* and *Estuary-Clines* landscapes can be found in Fig S22 and Fig S23, respectively. Data
591 included in this figure is from evaluation of 1- and 2-trait simulations using *all* markers. Code used
592 to create these figures can be found in SC 02.02.04.

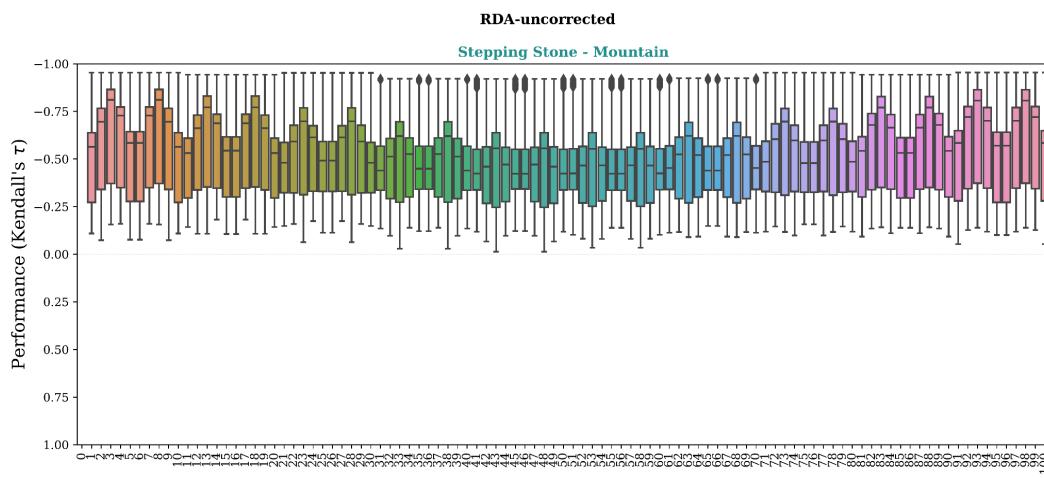
593 (Fig S22)



594



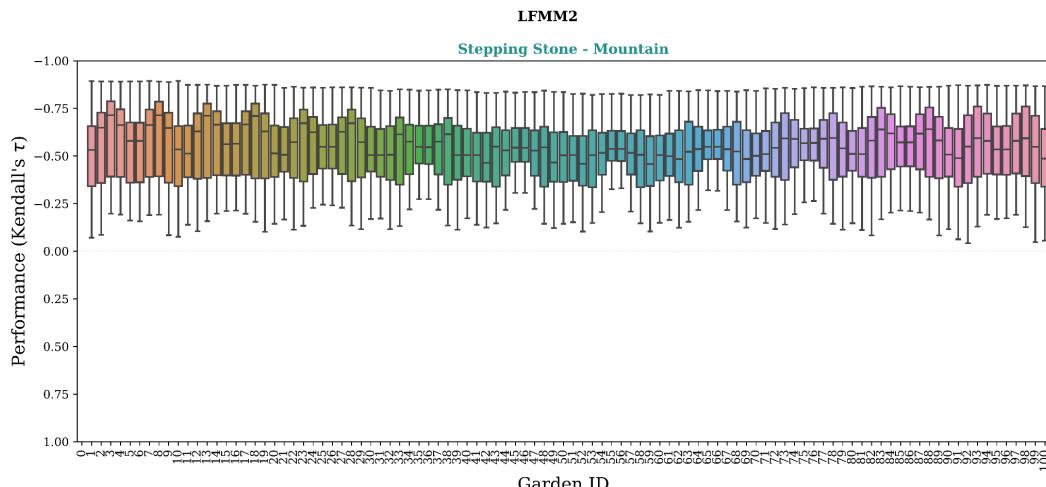
595



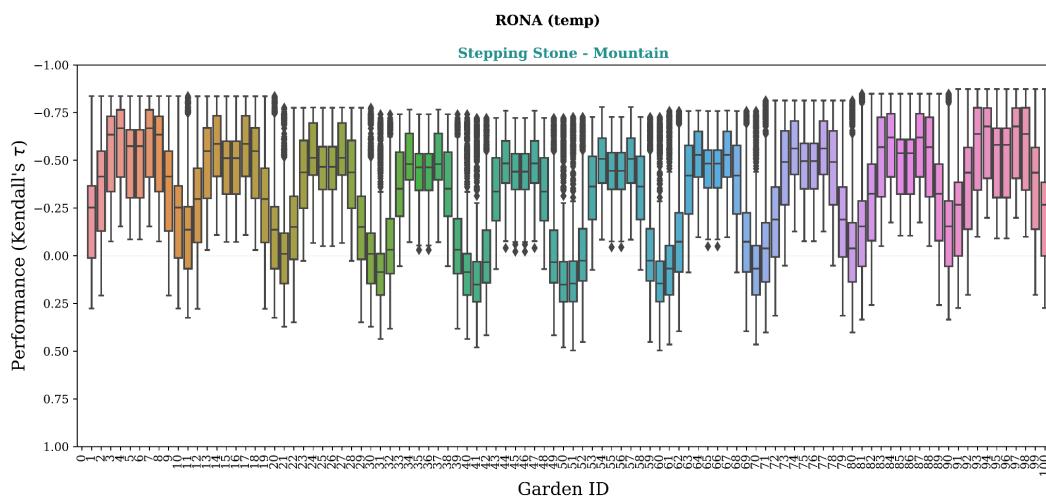
596

597

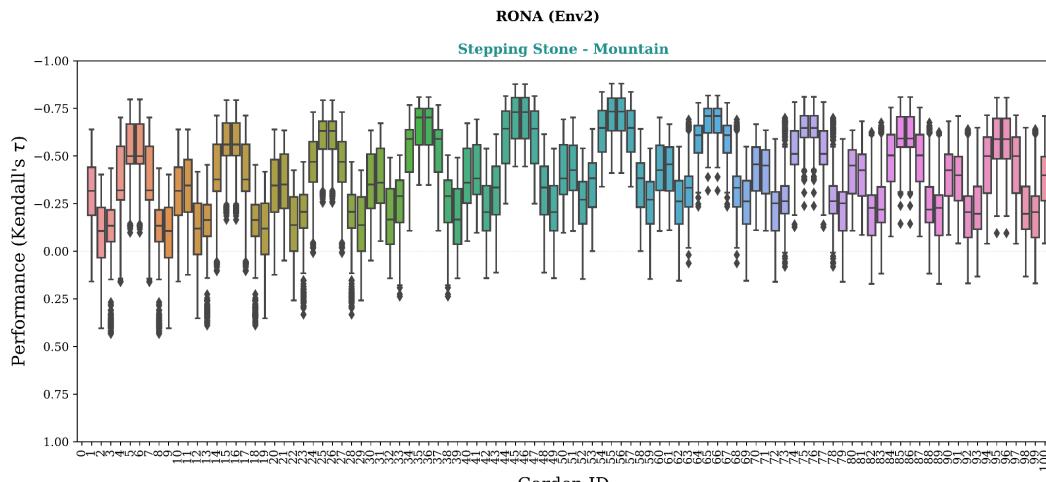
598 (Fig S22 continued)



599



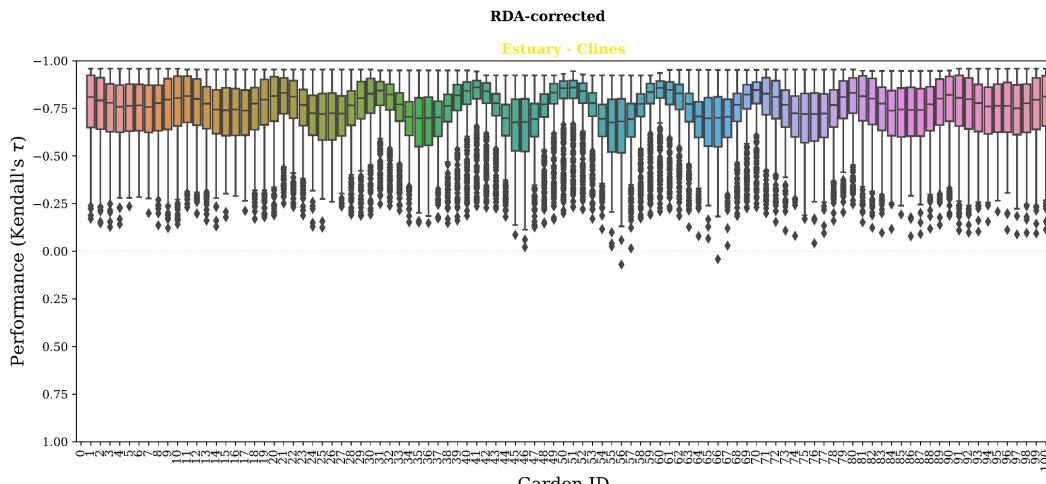
600



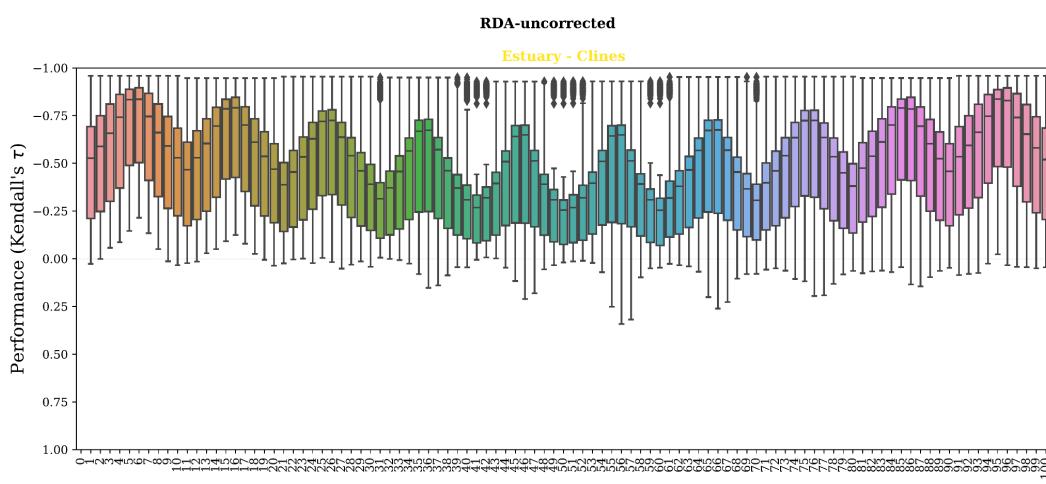
601

602 **Fig S22** Genomic offset methods have variable performance across the *Stepping-Stone* -
603 *Mountain* landscape. Shown is the variability of each offset method performance (y-axes) across
604 the 100 common gardens (x-axes). Gardens are ordered from left to right by garden ID. This
605 ordering of gardens is equivalent to the southwestern-most garden first and northeastern-most
606 garden last (see Fig. S20 for a map of garden ID across each landscape). Similar figures for
607 *Stepping-Stone* - *Clines* and *Estuary* - *Clines* landscapes can be found in Fig S21 and Fig S23,
608 respectively. Data included in this figure is from evaluation of 1- and 2-trait simulations using *all*
609 markers. Code used to create this figure can be found in SC 02.02.04.

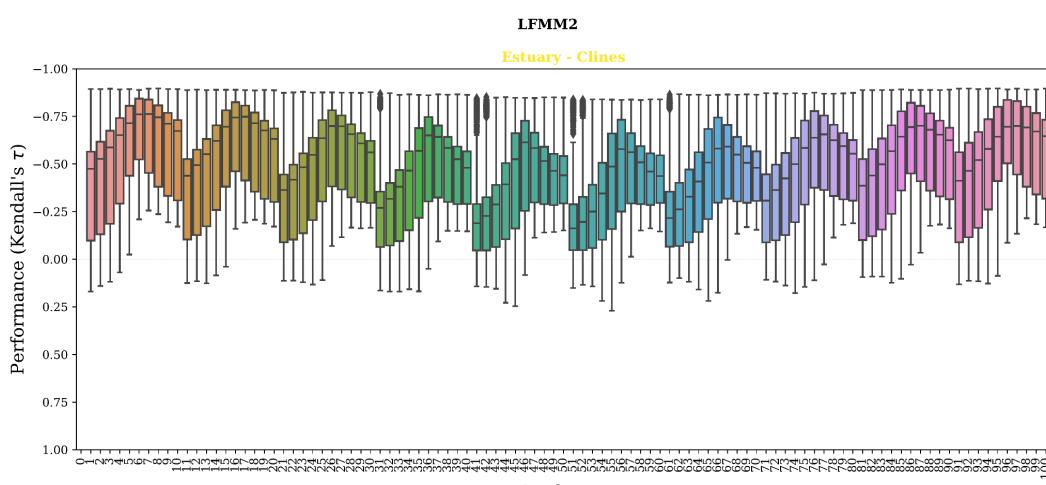
610 (Fig S23)



611

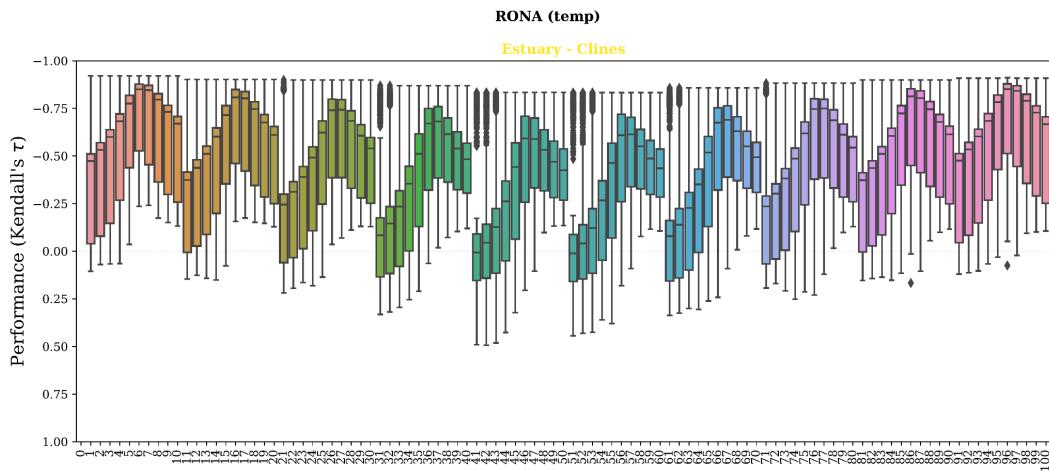


612

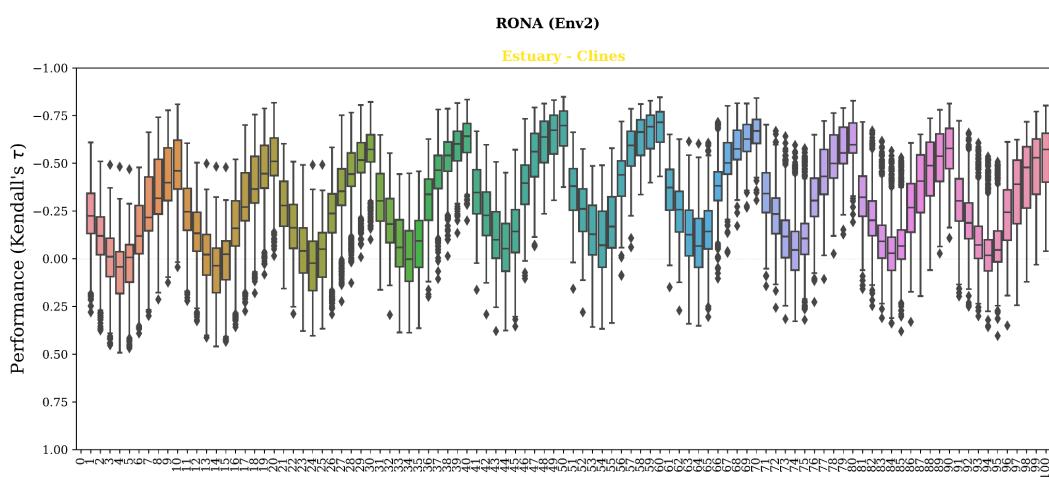


613

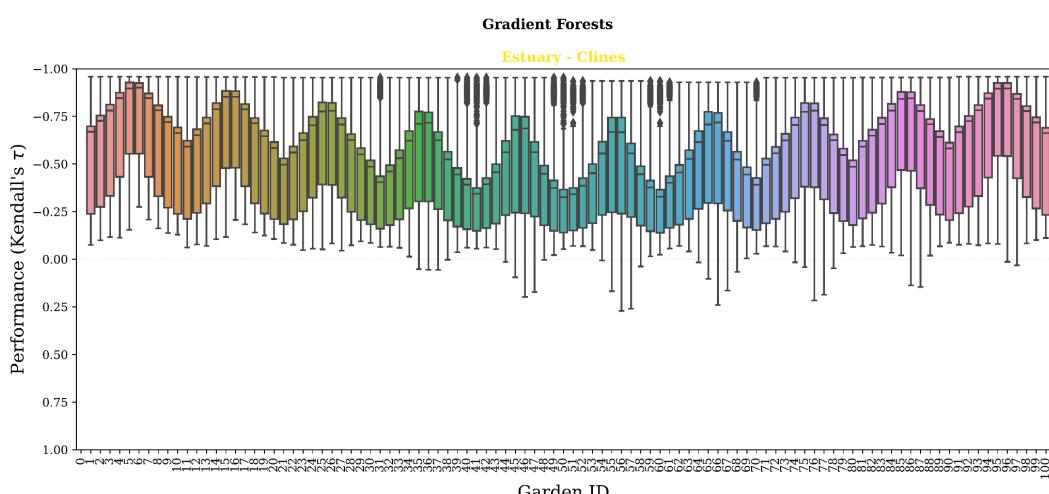
614 (Fig S23 continued)



615



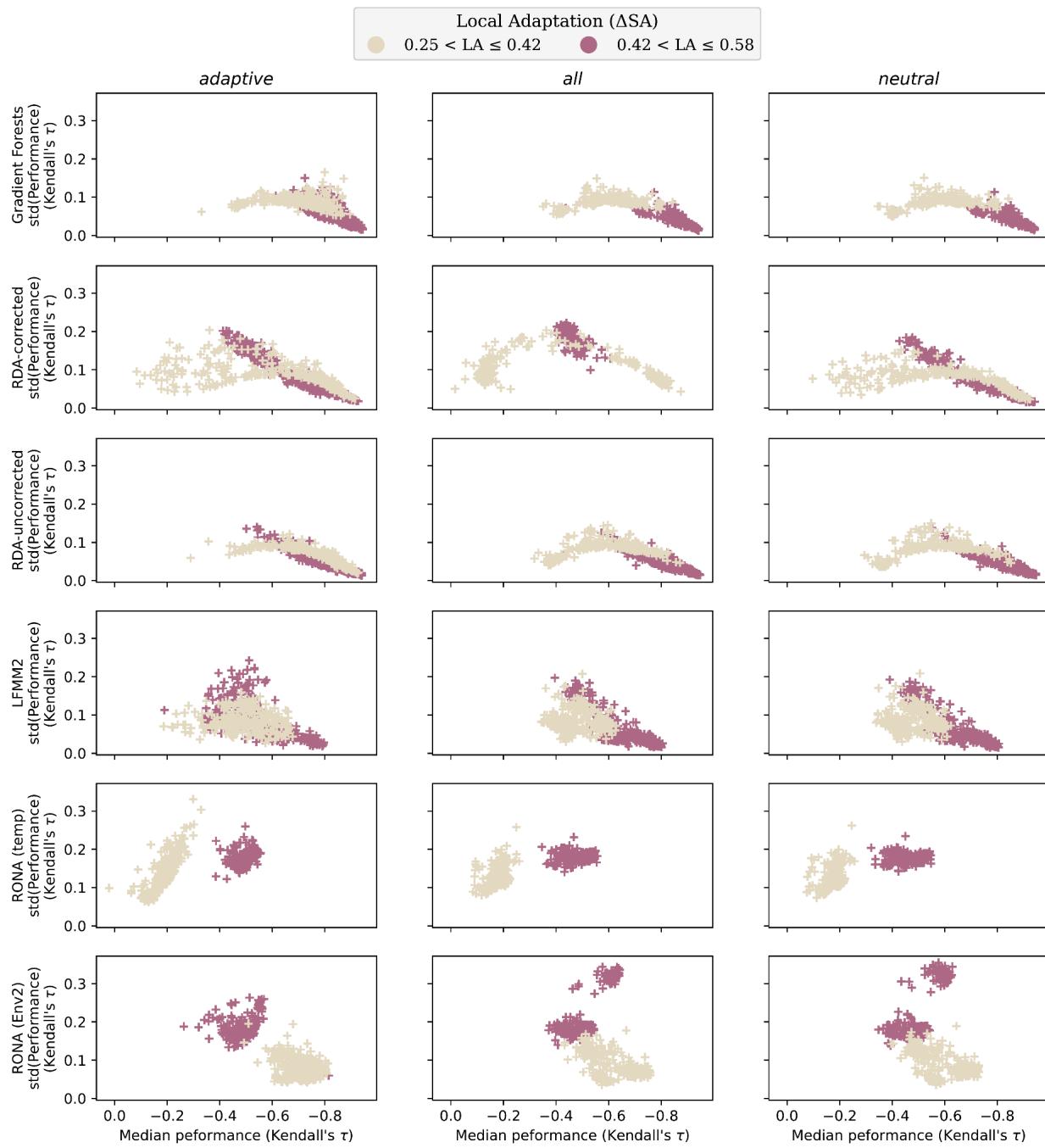
616



617

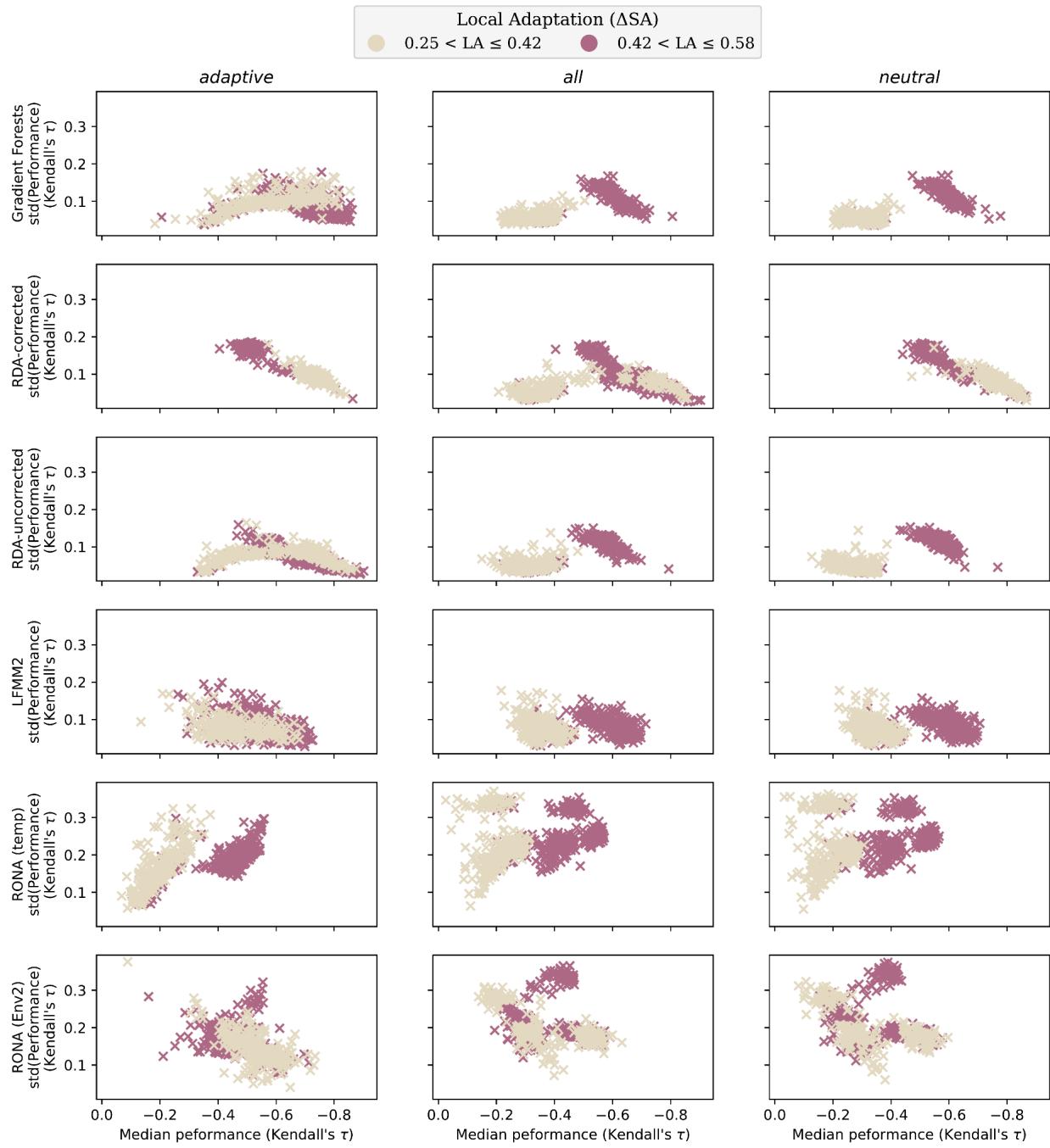
618 **Fig S23** Genomic offset methods have variable performance across the *Estuary - Clines*
619 landscape. Shown is the variability of each offset method performance (y-axes) across the 100
620 common gardens (x-axes). Gardens are ordered from left to right by garden ID. This ordering of
621 gardens is equivalent to the southwestern-most garden first and northeastern-most garden last
622 (see Fig. S20 for a map of garden ID across each landscape). Similar figures for *Stepping-Stone -*
623 *Clines* and *Stepping-Stone - Mountain* landscapes can be found in Fig S21 and Fig S22,
624 respectively. Data included in this figure is from evaluation of 1- and 2-trait simulations using *all*
625 markers. Code used to create this figure can be found in SC 02.02.04.

626 (Fig. S24)



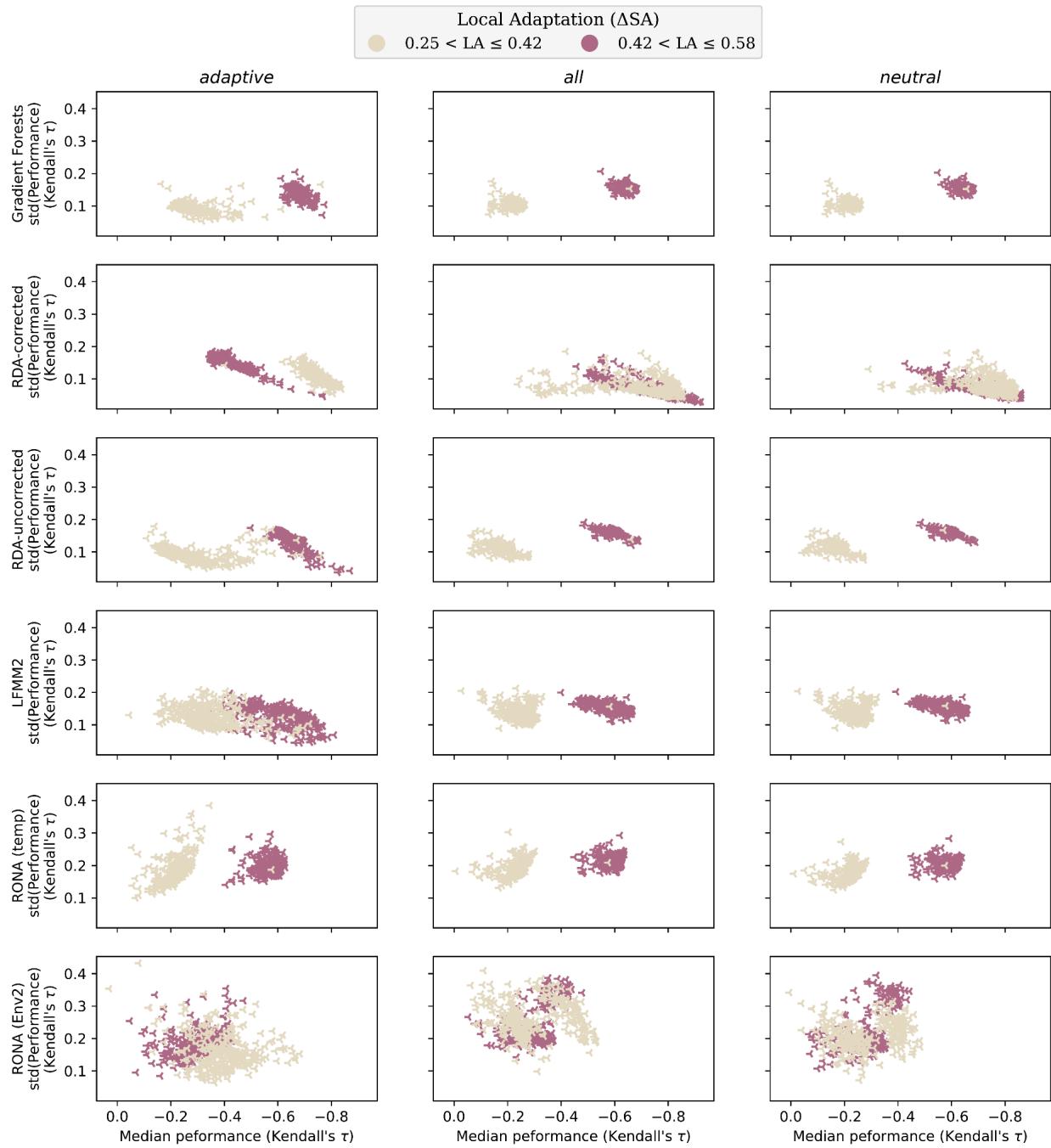
627

628 **Fig S24** Variability of genomic offset performance (y-axes) for a given model (+) often decreases
 629 with increasing median performance (x-axes). Shown are patterns from each offset method
 630 (rows) for each marker set (columns) used in training. Data included in this figure is from
 631 evaluation of 2-trait simulations from *Stepping-Stone - Clines* landscapes processed through the
 632 *Adaptive Environment* workflow. For similar figures for *Stepping-Stone - Mountain* and *Estuary -*
 633 *Clines* landscapes, see Figs. S25-S26, respectively. Code used to create these figures can be found
 634 in SC 02.02.07.



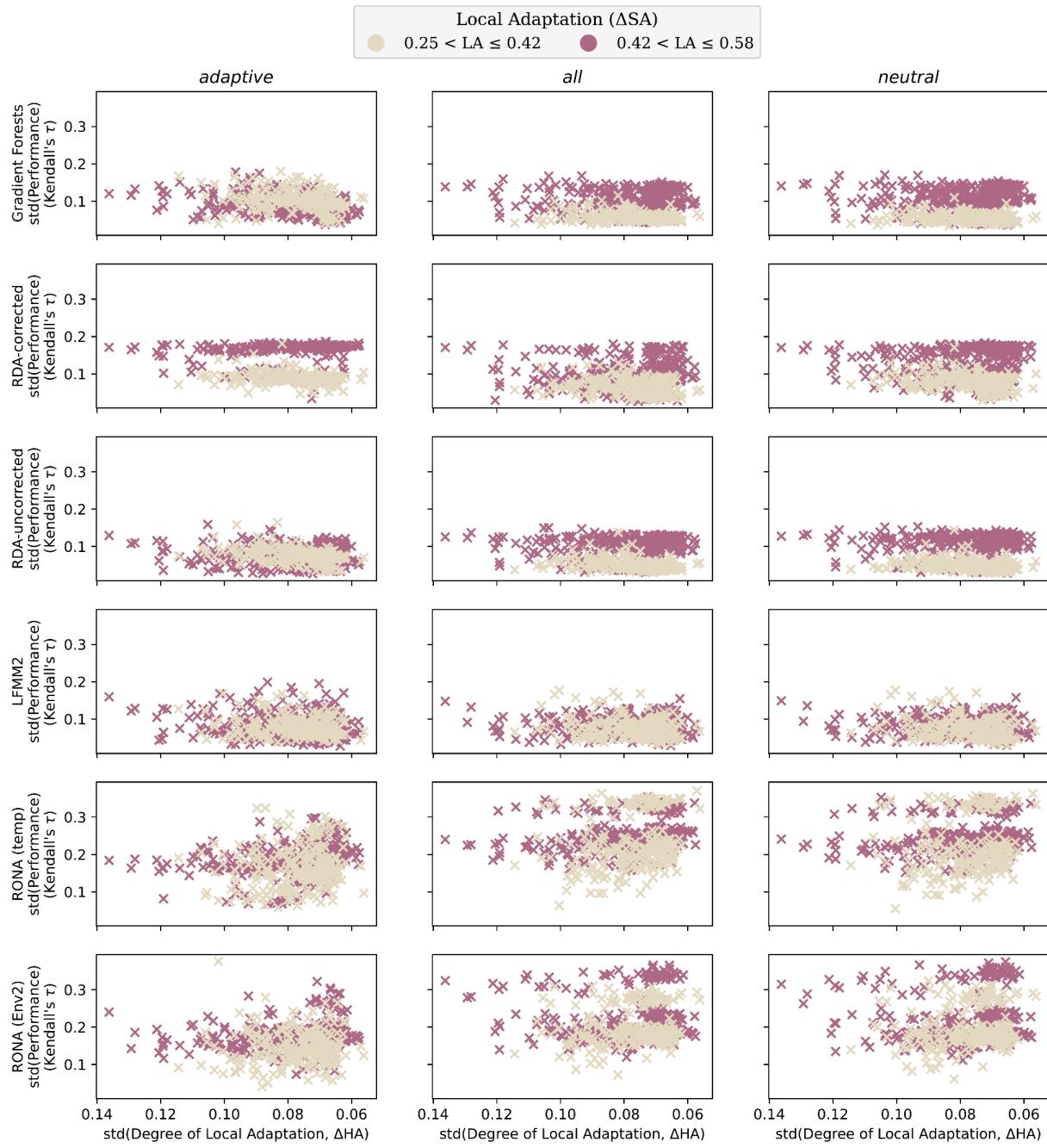
635

636 **Fig S25** Variability across evaluations of genomic offsets often decreases with increasing average
 637 performance across marker sets. Data included in this figure is from evaluation of 2-trait
 638 simulations from *Stepping-Stone - Mountain* landscapes. For similar figures for *Stepping-Stone -*
 639 *Clines* and *Estuary - Clines* landscapes, see Figs. 24 and S26, respectively. Code used to create
 640 these figures can be found in SC 02.02.07.



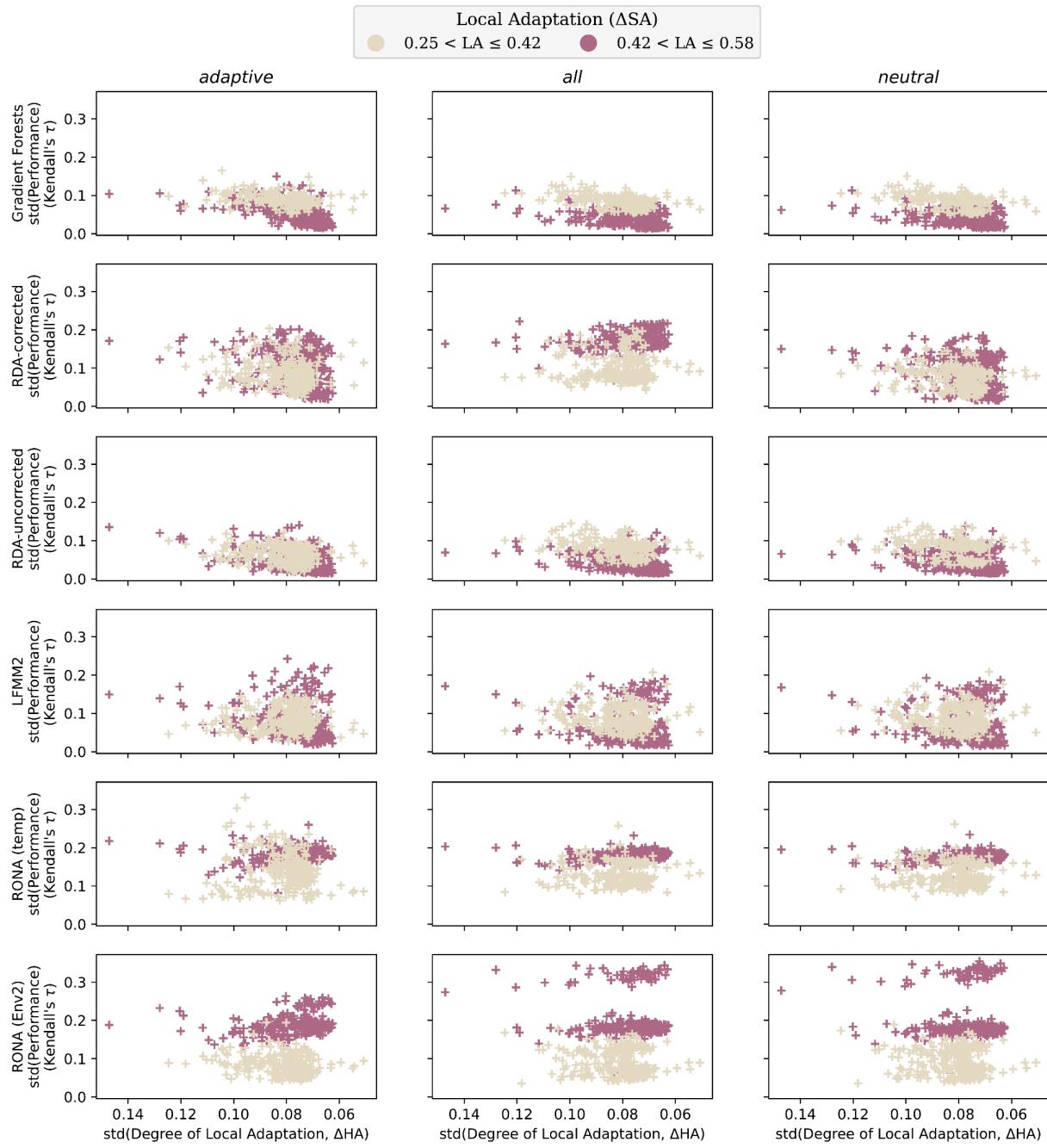
641

642 **Fig S26** Variability across evaluations of genomic offsets often decreases with increasing average
 643 performance across marker sets. Data included in this figure is from evaluation of 2-trait
 644 simulations from *Estuary - Clines* landscapes. For similar figures for *Stepping-Stone - Clines* and
 645 *Stepping-Stone - Mountain* landscapes, see Figs. S24 and S25, respectively. Code used to create
 646 these figures can be found in SC 02.02.07.



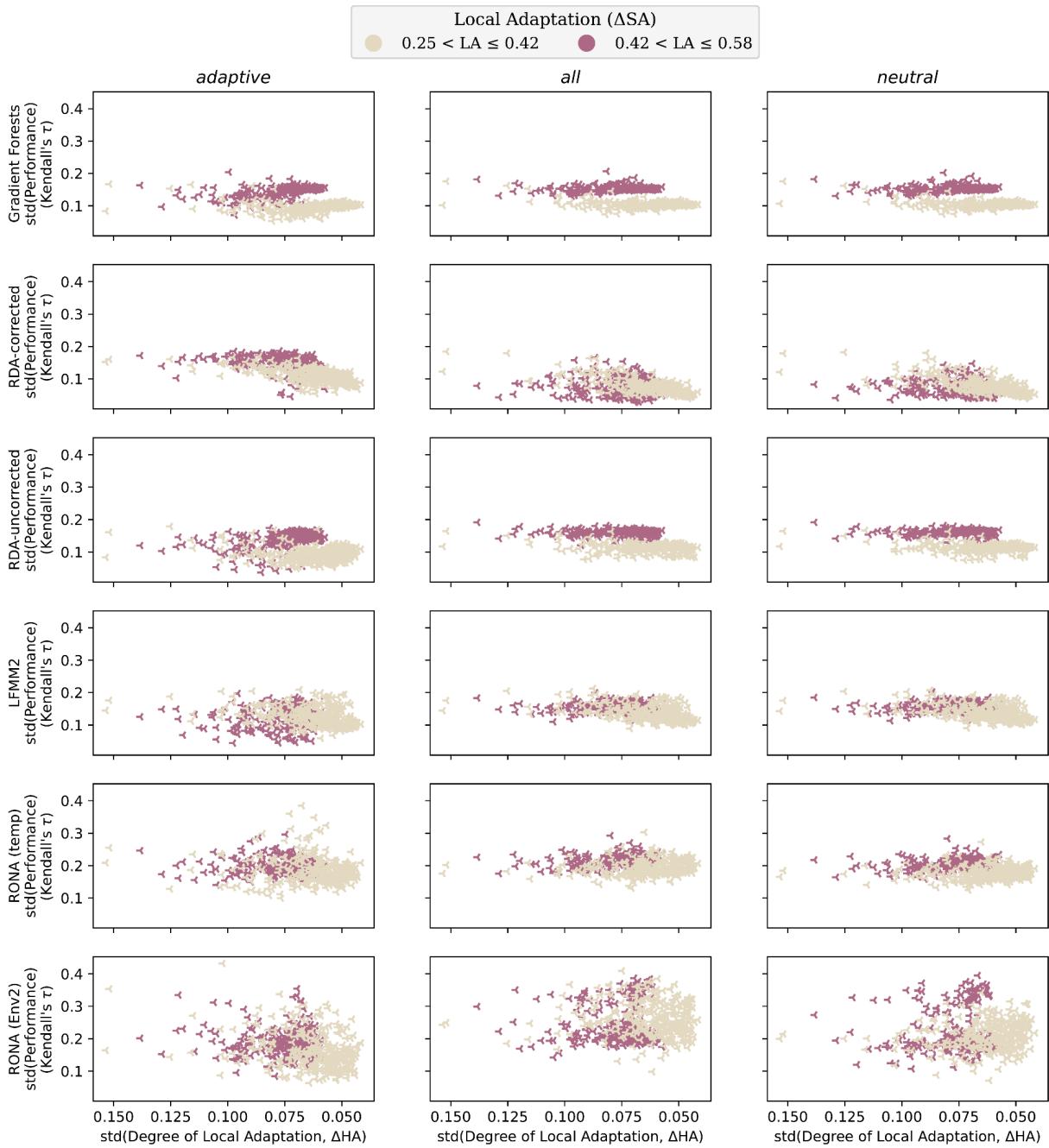
647

648 **Fig S27** Variability across evaluations of genomic offsets is often unrelated to the variability in
 649 the degree of local adaptation across populations. Data included in this figure is from evaluation
 650 of 2-trait simulations from *Stepping-Stone - Mountain* landscapes. For similar figures for
 651 *Stepping-Stone - Clines* and *Estuary - Clines* landscapes, see Figs. S27 and S28, respectively. Code
 652 used to create these figures can be found in SC 02.02.07.



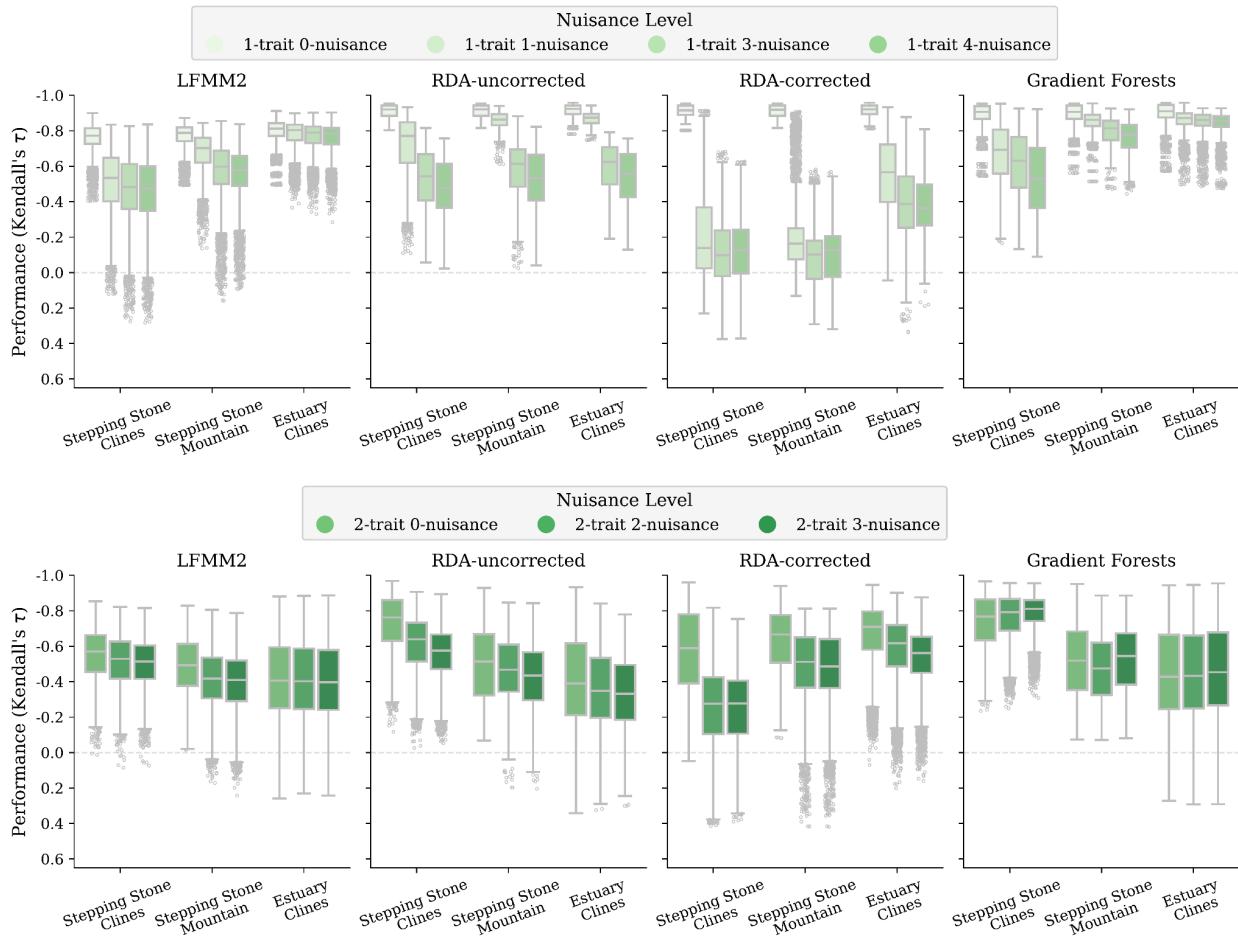
653

654 **Fig S28** Variability across evaluations of genomic offsets is often unrelated to the variability in
 655 the degree of local adaptation across populations. Data included in this figure is from evaluation
 656 of 2-trait simulations from *Stepping Stone - Clines* landscapes. For similar figures for *Estuary -*
 657 *Clines* and *Stepping-Stone - Mountain* landscapes, see Figs. S26 and S28, respectively. Code used
 658 to create these figures can be found in SC 02.02.07.



659

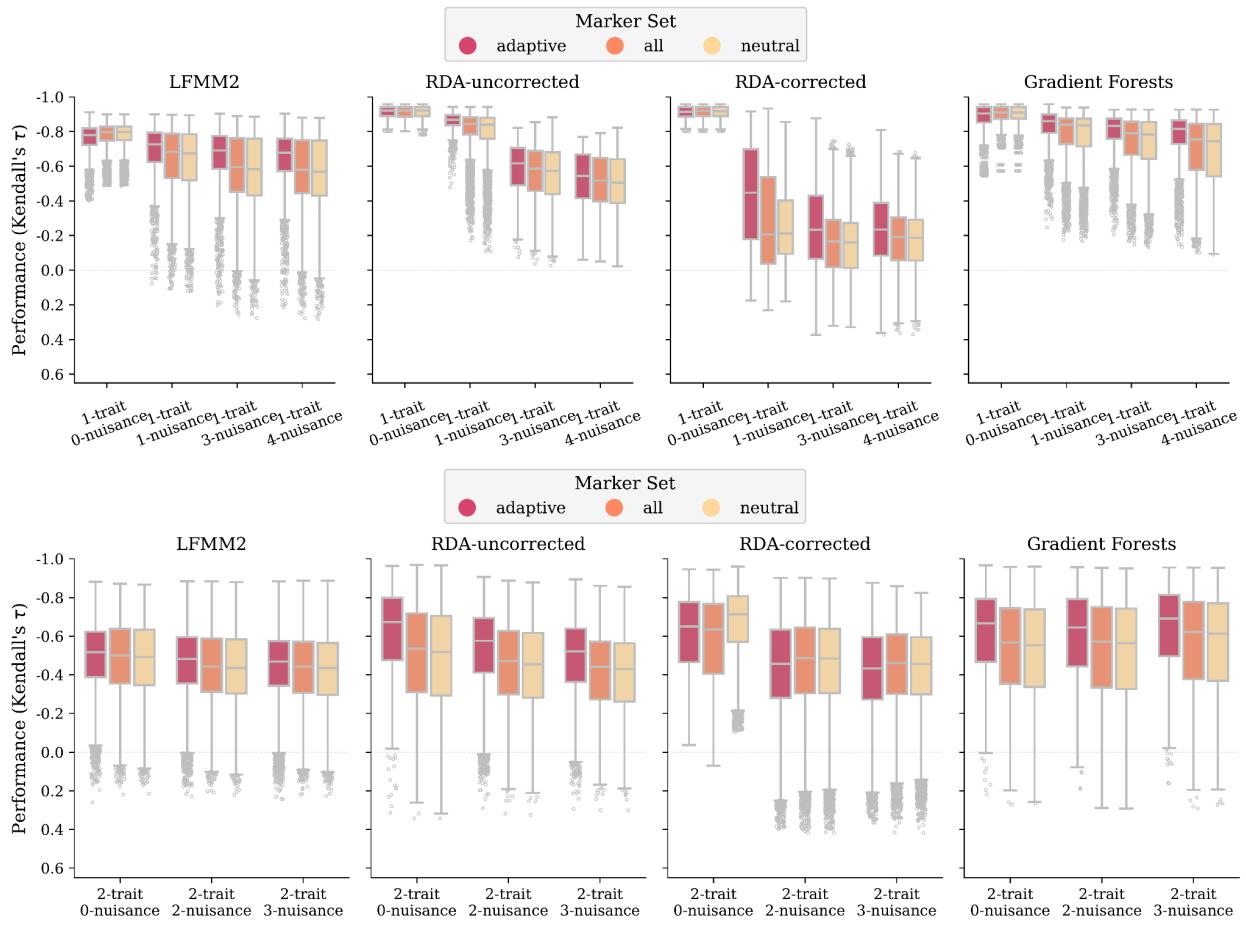
660 **Fig S29** Variability across evaluations of genomic offsets is often unrelated to the variability in
 661 the degree of local adaptation across populations. Data included in this figure is from evaluation
 662 of 2-trait simulations from *Estuary - Clines* landscapes. For similar figures for *Stepping-Stone -*
 663 *Clines* and *Stepping-Stone - Mountain* landscapes, see Figs. S26 and S27, respectively. Code used
 664 to create these figures can be found in SC 02.02.07.



665

666

667 **Fig S30** Effect of non-adaptive nuisance environmental variables on offset performance faceted
 668 by landscape. Shown are offsets from 1- and 2-trait simulations trained using only adaptive
 669 environments (0-nuisance) or with adaptive environments and the addition of $N > 0$ non-adaptive
 670 environmental variables (N -nuisance). RONA is not shown because it is univariate with respect to
 671 environmental variables. The nuisance variables for 1-trait simulations are: Env2, ISO, TSsd, PSsd;
 672 and for 2-trait simulations are ISO, TSsd, PSsd; see Table 2. The *Nuisance Environment* workflow
 673 was used to produce this data. Code to create these figures can be found in SC 02.02.06.



674

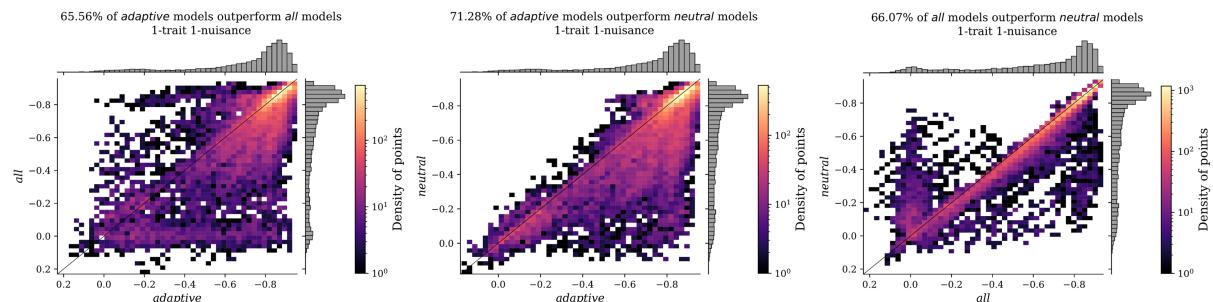
675

676 **Fig S31** Effect of non-adaptive nuisance environmental variables on offset performance faceted
 677 by marker set. Shown are offsets from 1- (A) and 2-trait (B) simulations trained using only
 678 adaptive environments (0-nuisance) or with adaptive environments and the addition of N>0 non-
 679 adaptive environmental variables (N-nuisance). RONA is not shown because it is univariate with
 680 respect to environmental variables. The nuisance variables for 1-trait simulations are: Env2, ISO,
 681 TSsd, PSsd; and for 2-trait simulations are ISO, TSsd, PSsd; see Table 2. Code to create these
 682 figures can be found in SC 02.02.06.

683 (Fig. S32)

684

1-trait 1-nuisance

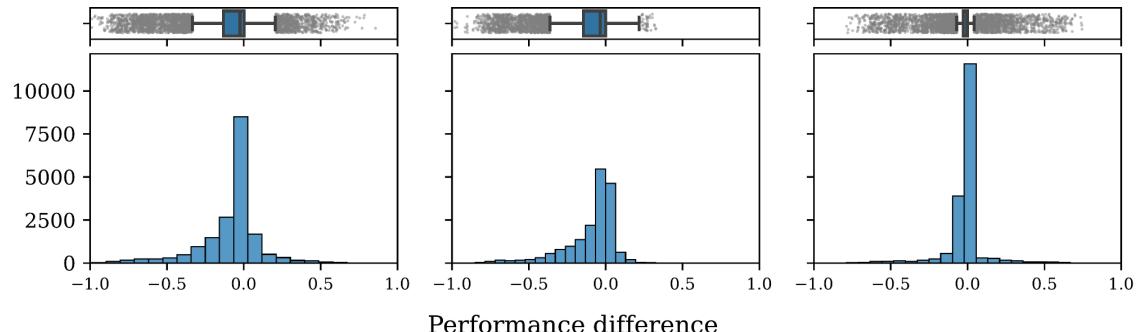


685

adaptive - all

adaptive - neutral

all - neutral

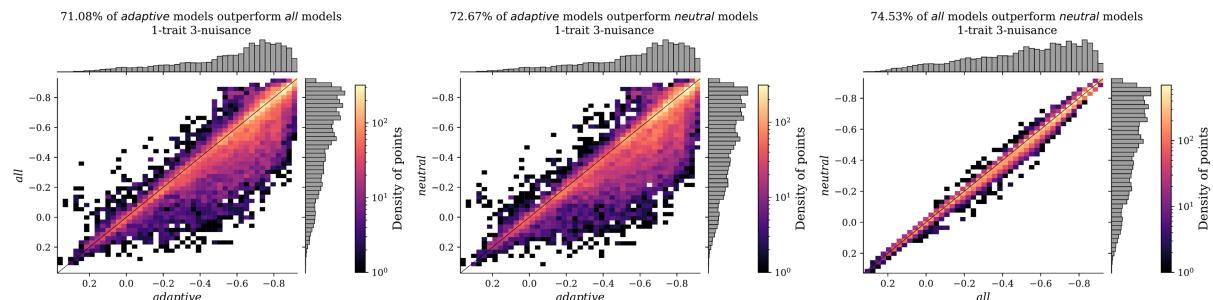


686

687 (Fig. S32 continued)

688

1-trait 3-nuisance

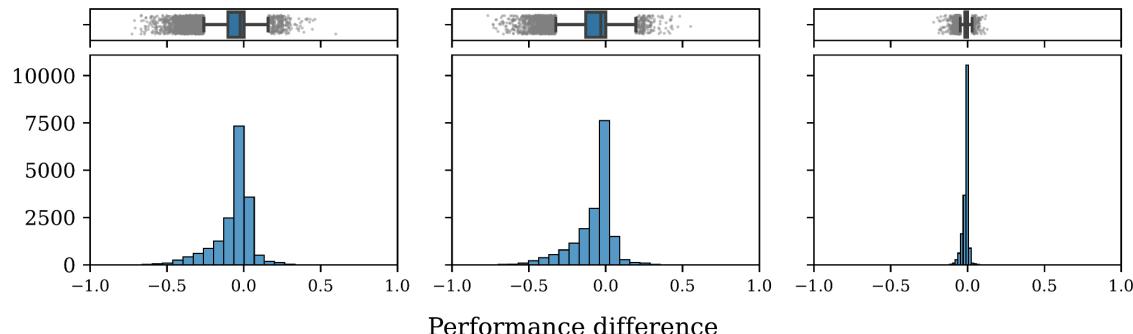


689

adaptive - all

adaptive - neutral

all - neutral

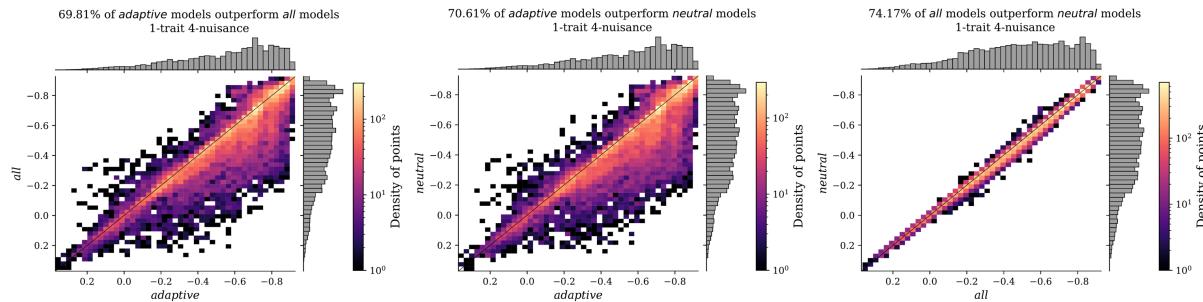


690

691 (Fig. S32 continued)

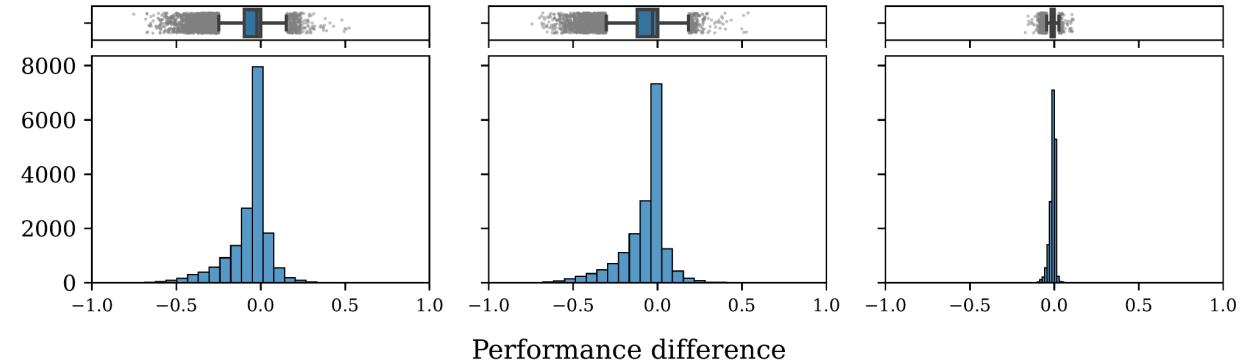
692

1-trait 4-nuisance



693

adaptive - all

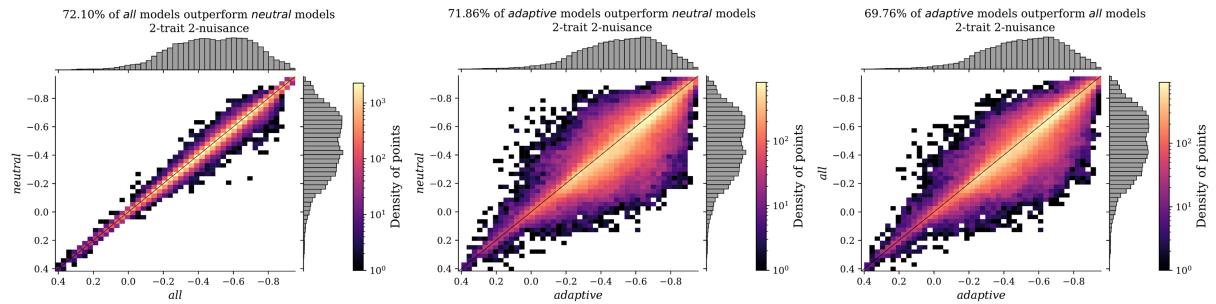


694

695 (Fig. S32 continued)

696

2-trait 2-nuisance

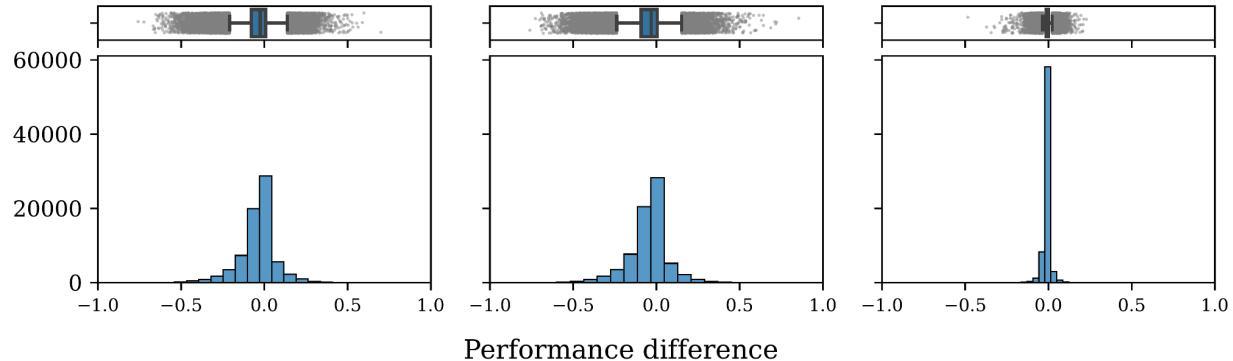


697

adaptive - *all*

adaptive - *neutral*

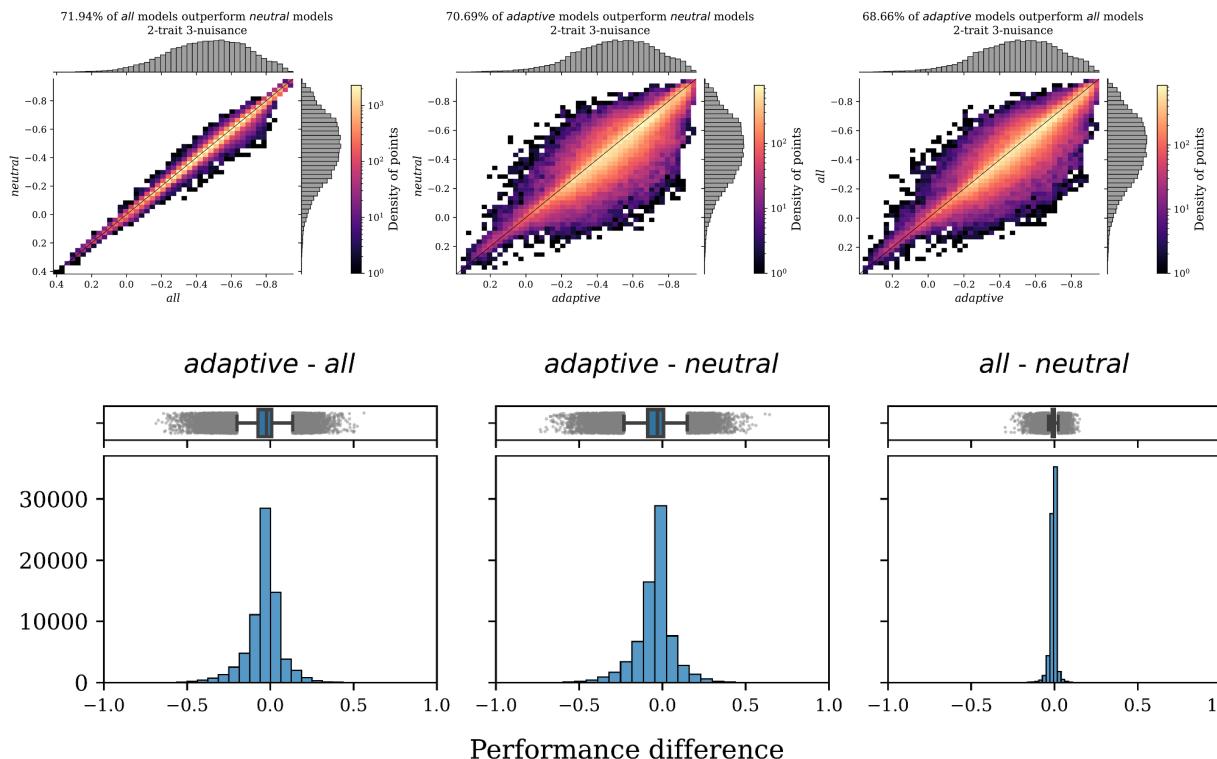
all - *neutral*



698

699 (Fig. S32 continued)

700

2-trait 3-nuisance

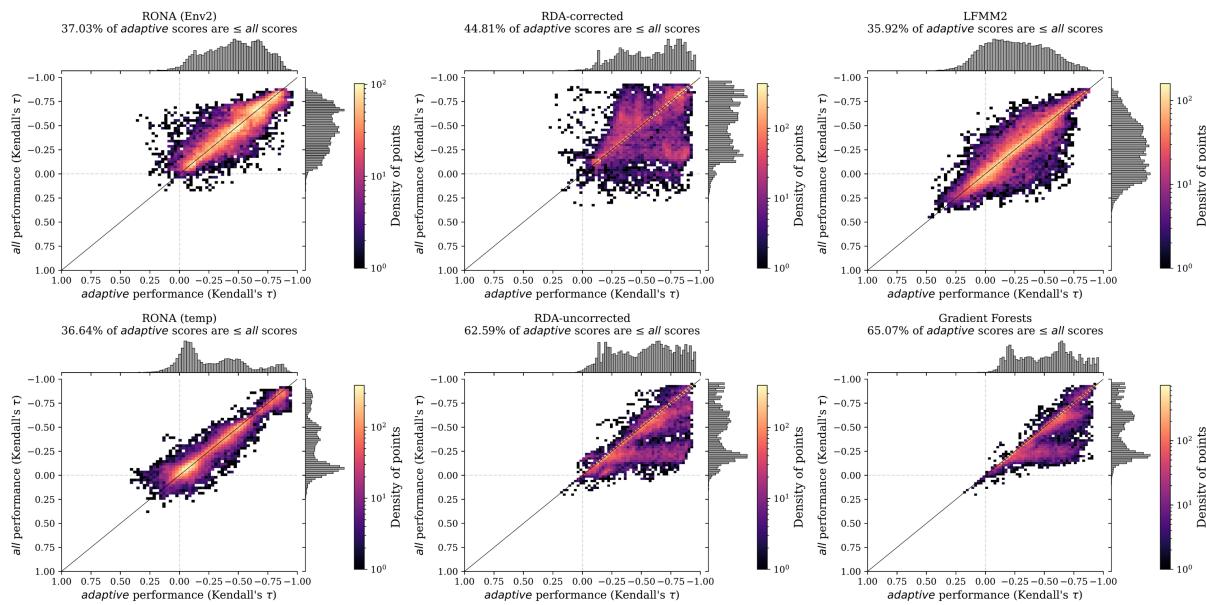
701

702

Fig S32 Pairwise comparison of performance differences between marker sets for *Nuisance Environment* scenarios. The first row for each nuisance level (*N-trait N-nuisance*) are scatterplots of pairwise comparisons of performance between marker sets (histograms in each margin) from both 1- and 2-trait models where density of points is indicated by color in legend (note color scale is different for each figure to accentuate patterns in data). The second row for each nuisance level are histograms for the difference in performance between marker sets for a given model. Method-specific figures are not shown except in SC 02.02.06. Data for these figures includes 1- and 2-trait *Nuisance Environment* evaluations. Code to create these figures can be found in SC 02.02.06.

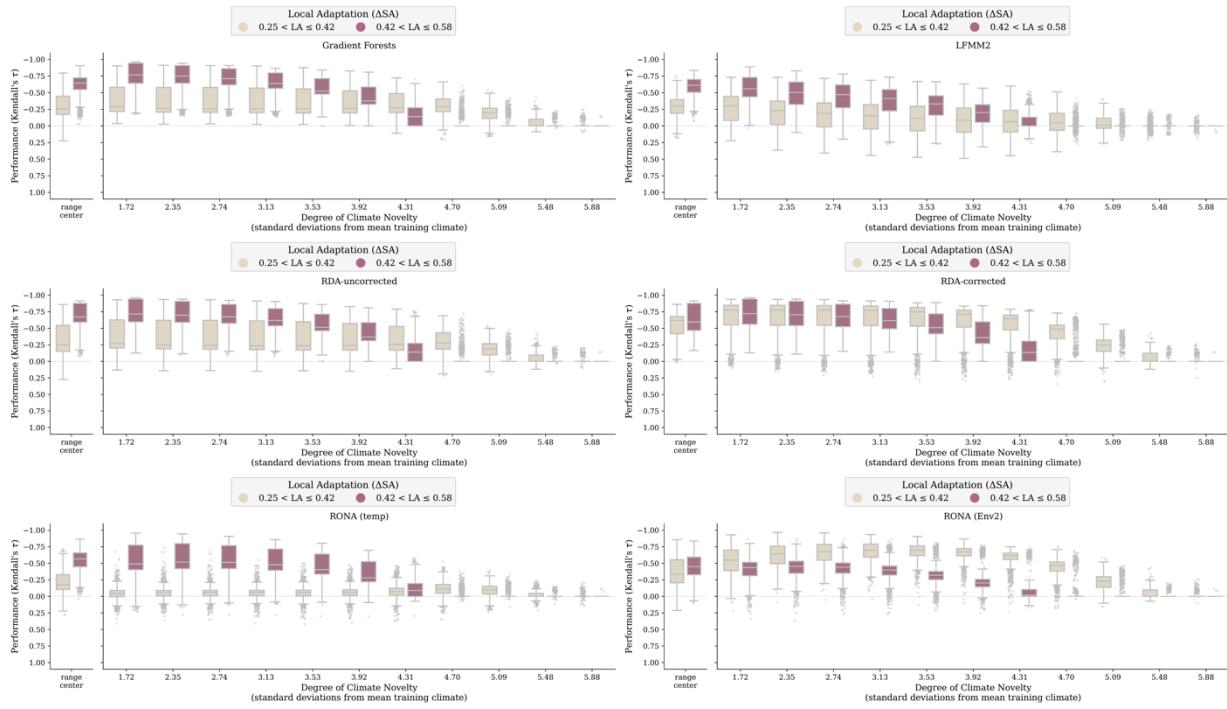
712 Fig S33 is in Supplemental Note S4

713



714

715 **Fig S34** Pairwise comparison of performance differences between marker sets for *Climate*
 716 *Novelty* scenarios. Shown are scatterplots of pairwise comparisons of performance between
 717 marker sets (histograms in each margin) from both 1- and 2-trait models where density of points
 718 is indicated by color in legend (note color scale is different for each figure to accentuate patterns
 719 in data). Data for these figures includes 1- and 2-trait *Climate Novelty* evaluations. Code to create
 720 these figures can be found in SC 02.04.05.



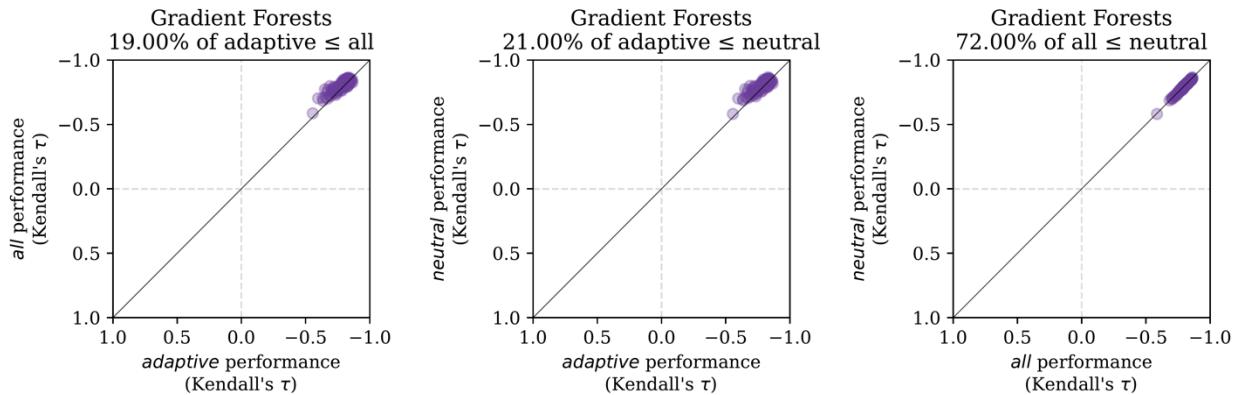
721

722

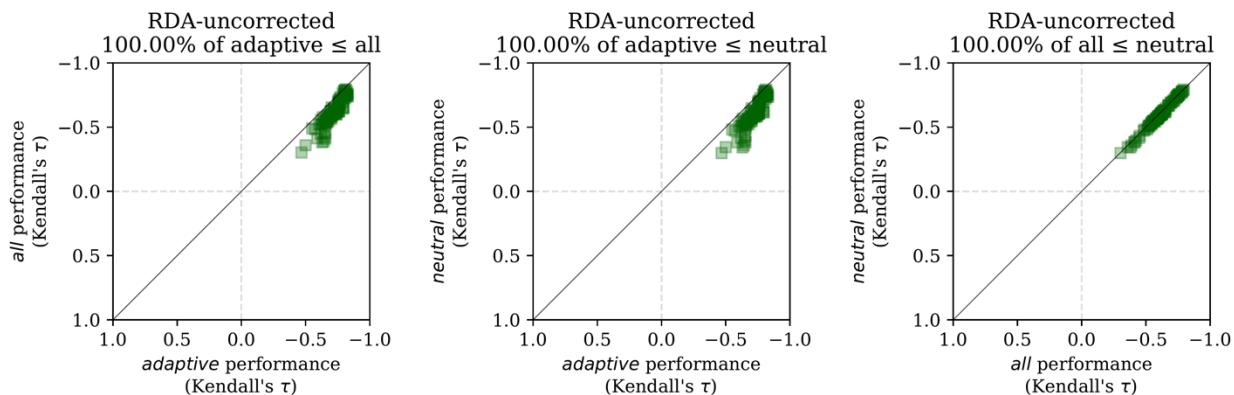
723

724 **Fig S35** Performance decays with climate novelty relative to training data. Shown is the same
 725 information as that presented in Fig. 6 of the main text, except the results are separated by
 726 method (titles within figures). Code to create these figures can be found in SC 02.04.05.

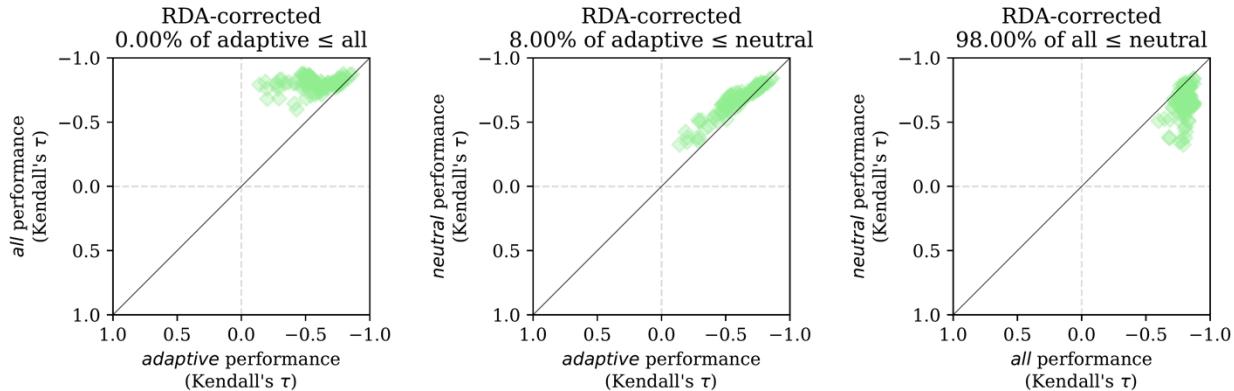
727 (Fig S36)



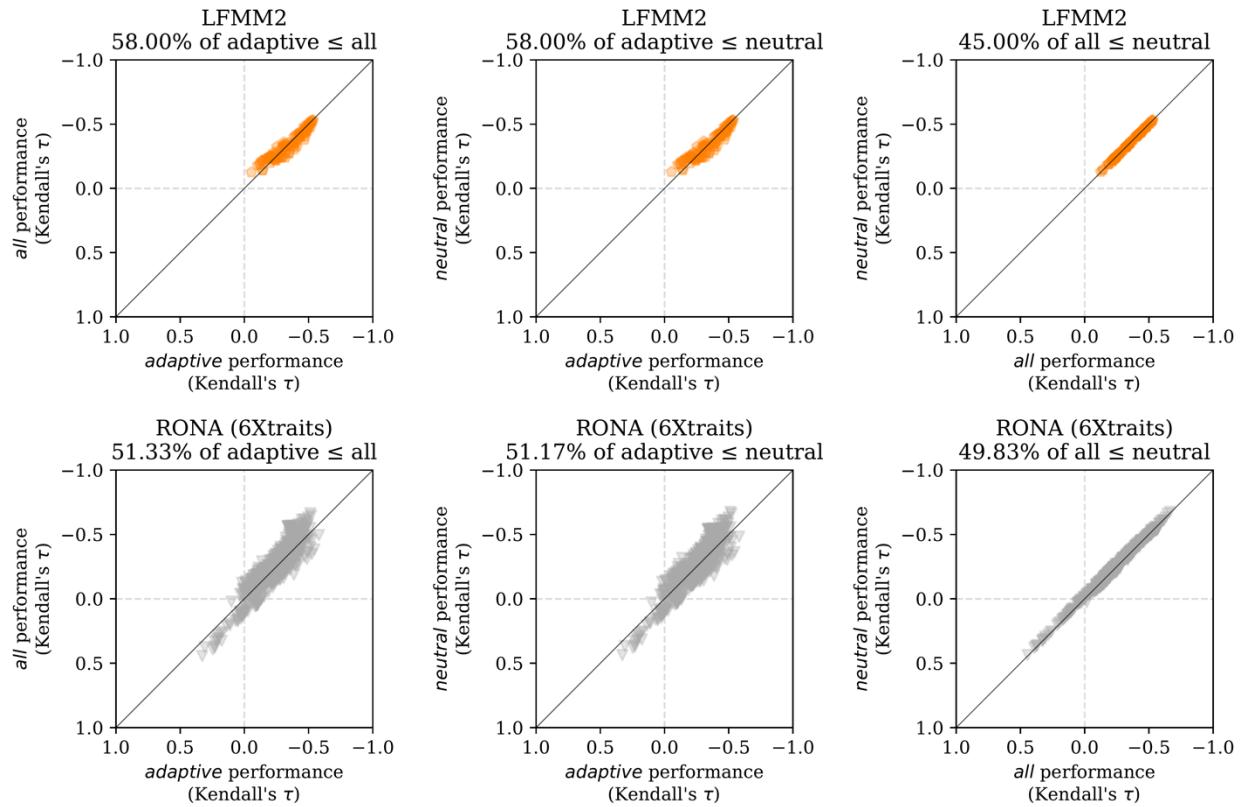
728



729



730

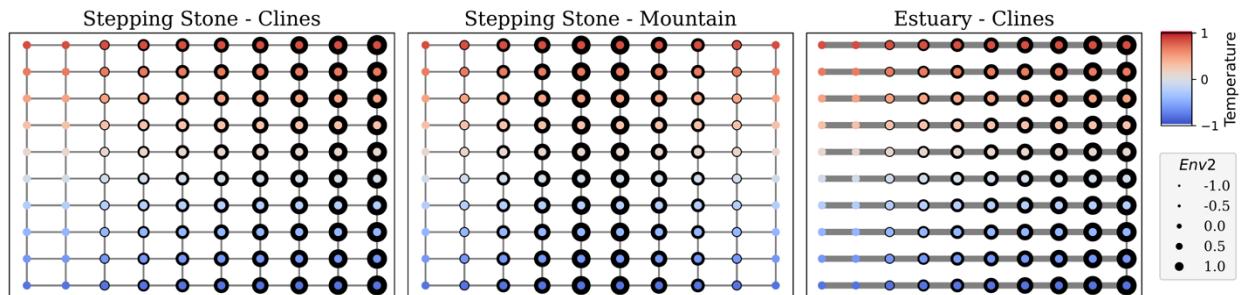


731

732

733 **Fig S36** Comparison of marker choice across genomic offset methods for the 6-trait simulation.
 734 Shown is the same information as Fig 4 of the main text, except the results are separated by
 735 method and all marker pairwise comparisons are shown. Code to create these figures can be
 736 found in SC 02.05.10.

737



738 **Fig S37** The spatial distribution of environmental variables from Wright-Fisher simulations,
 739 reproduced from Lotterhos (2023). Points on each landscape represent the multivariate
 740 environmental optima of sampled populations. The color of each point indicates the value along
 741 the latitudinal environmental gradient analogous to temperature (*temp*, used in 1- and 2-trait
 742 simulations), and the size of the black point is indicative of the value along the longitudinal
 743 environmental gradient (*Env2*; present only in 2-trait simulations). Gray lines connecting
 744 populations indicate potential for gene flow, the magnitude and direction of which was varied
 745 (thick gray lines in the *Estuary - Clines* landscape indicates greater migration rates compared to
 746 other landscapes). Code used to create this figure can be found in SC 02.10.04.

747 Supplemental References

- 748 Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape
749 genomics. *Methods in Ecology and Evolution*. <https://doi.org/10.1111/2041-210x.13722>
- 750 Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to
751 identify signatures of local adaptation. *Molecular Ecology Resources*, 18(6), 1223–1233.
752 <https://doi.org/10.1111/1755-0998.12906>
- 753 Lotterhos, K. E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the
754 underlying genes. *Proceedings of the National Academy of Sciences*, 120(12).
755 <https://doi.org/10.1073/pnas.2220313120>