The limits of predicting maladaptation to future 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

11 January 2024

Running Title: Limits of genomic offsets

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

***Corresponding Author**

Brandon M. Lind

Email: lind.brandon.m@gmail.com

Abstract

Anthropogenically driven changes in land use and climate patterns pose unprecedented challenges to species persistence. To understand the extent of these impacts, genomic offset methods have been used to forecast maladaptation of natural populations to future environmental change. However, while their use has become increasingly common, little is known regarding their predictive performance across a wide array of realistic and challenging scenarios. Here, we evaluate four offset methods (Gradient Forests, the Risk-Of-Non-Adaptedness, redundancy analysis, and LFMM2) using an extensive set of simulated datasets that vary demography, adaptive architecture, and the number and spatial patterns of adaptive environments. For each dataset, we train models using either *all*, *adaptive*, or *neutral* marker sets and evaluate performance using *in silico* common gardens by correlating known fitness with projected offset. Using over 4,850,000 of such evaluations, we find that 1) method performance is largely due to the degree of local adaptation across the metapopulation ($LA_{\Delta\mathrm{SA}}$), 2) *adaptive* marker sets provide minimal performance advantages, 3) within-landscape performance is variable across gardens and declines when offset models are trained using additional non-adaptive environments, and 4) despite (1), performance declines more rapidly in novel climates for metapopulations with higher $LA_{\Delta\mathrm{SA}}$ than lower $LA_{\Delta\mathrm{SA}}$. We discuss the implications of these results for management, assisted gene flow, and assisted migration.

1 | Introduction

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

In the context of population viability in the face of environmental change, many of these models rely on theoretical expectations that the limits of species' distributions are primarily determined by the distribution of environmental conditions (e.g., Good 1931), and that occupancy of highly suitable habitat enables increased abundance through greater survival and reproduction (i.e., fitness) of individuals (Brown, 1984). Such methods, termed species distribution models or ecological niche models (see Elith & Leathwick, 2009 for a discussion on terminology) are correlative approaches that are often used to predict (relative) habitat suitability for a single species (Lee-Yaw et al., 2022). This information is used to understand potential impacts on the species from future climate change. However, these methods often ignore aspects of the species' evolutionary history that could be important for predicting long-term population persistence, such as the environmental drivers of local adaptation or spatial patterns of adaptive genetic variation (Waldvogel et al., 2020).

Subsequent methods, termed genomic offsets (reviewed in Capblancq et al., 2020; Rellstab et al., 2021), have attempted to address these shortcomings by modeling relationships between environmental and genetic variation to predict maladaptation of natural populations to either future climates *in situ*, or to predict the relative suitability of these populations for the specific environment of a restoration site. Empirical attempts to confirm predictions from genomic offset models are rare and, compared to attempts *in silico* (Láruson et al., 2022), have found relatively weaker relationships between predicted maladaptation to common garden climates and the measurement of phenotypic proxies for fitness from individuals grown in these same environments (e.g., Capblancq & Forester, 2021; Fitzpatrick et al., 2021; Lind et al., 2024). Even so, these empirical results have consistently shown the expected negative relationship between predicted offset and common garden performance. Further, many of these studies found that genomic offsets often perform better than climate or geographic distance alone (e.g., Capblancq & Forester, 2021; Fitzpatrick et al., 2021; Láruson et al., 2022; Lind et al., 2024).

Across empirical and *in silico* studies, little difference in performance was found between

models trained using only adaptive markers (i.e., known *in silico*, or candidates from empirical genotype-environment [GEA] associations) and those chosen at random, suggesting that genome-wide data may be sufficient to capture signals relevant to environmental adaptation.

Together, these results suggest that genomic offset methods may provide valuable insight for management. Little is known, however, about how robust these methods are across a wide array of realistic empirical scenarios, nor the extent to which independent methods will arrive at similar conclusions when analyzing the same data. Indeed, concerns regarding the accuracy of ecological forecasting models present a primary limitation towards incorporating inferences from these models into management (Clark et al., 2001; Schmolke et al., 2010) and genomic offset models are no exception. Major questions still remain about how performance is affected by aspects of the evolutionary history of sampled populations, the type of signals in putatively ideal datasets that may mislead offset inference, the importance of identifying environmental drivers of local adaptation *a priori*, and the consistency of predictive performance across contemporary environmental space. Finally, because novel climates with no recent analog are expected to increase in the future (Lotterhos et al., 2021; Mahony et al., 2017) there is also uncertainty regarding the performance of forecasting models when predictions are made to environments that drastically differ from those used to train and build the models themselves (Fitzpatrick et al., 2018; Lind et al., 2024).

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

Here, we use a wide array of previously published biologically realistic, contrasting, and adversarial simulations from Lotterhos (2023) in an attempt to more precisely define the limits of predictive performance of five genomic offset methods (Table 1): Gradient Forests (GF_{offset}; *sensu* Fitzpatrick & Keller, 2015), the Risk Of Non-Adaptedness (RONA, Rellstab et al., 2016), Latent Factor Mixed

Models ($LFMM2_{offset}$, *sensu* Gain & François, 2021, and redundancy analysis (RDA_{offset}, *sensu* Capblancq & Forester, 2021). The main goal of this study was to understand how the evolutionary and experimental parameters used in the training and evaluation of offset methods affect the accuracy of the methods' projections of maladaptation under ideal empirical scenarios (i.e., using data with no inherent error). Using these scenarios, we ask the following six questions: 1) Which aspects of the past evolutionary history affect performance of offset methods? 2) How is offset performance affected by the proportion of loci with clinal alleles in the data? 3) Is method performance driven by causal loci or by genome-wide patterns of isolation-by-environment? 4) What is the variation of model performance across the landscape? 5) How does the addition of non-adaptive nuisance environments in training affect performance? 6) How well do offset models extrapolate to novel environments outside the range of environmental values used in training?

2 | Methods

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

2.1 | Explanation of Simulations and Training Data

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

The adaptive trait(s) were under selection by a different environmental variable, where the optimum trait value was given by the local environment on the landscape. The adaptive trait(s) undergoing selection responded to either a latitudinal temperature gradient (*temp*; $n_{\text{traits}} = 1$), or to both *temp* and a longitudinal "*Env2*" gradient ($n_{\text{traits}} =$

2). *Env2* represented distinct biological analogies depending on the context: in the *Stepping Stone - Mountain* landscape *Env2* was analogous to elevation (e.g., as with tree species), whereas in the *Estuary - Clines* landscape the *Env2* environment was analogous to gradients of salinity within coastal inlets connected only by the outer marine (ocean) environment (e.g., as with stickleback or oyster species).

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

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

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

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

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

2.2 | Evaluation of Offset Methods

We investigated the performance of five implementations of four genomic offset methods (Table 1): GF_{offset}, RDA_{offset},, $LFMM2_{offset}$, and RONA. While GF_{offset}, RDA_{offset}, and $LFMM2_{offset}$ can use multivariate environmental data to train models, RONA can only account for a single environment at one time (Table 1). Additionally, while GF_{offset} and RONA do not apply correction for population genetic structure, $LFMM2_{offset}$ does by default, and structure correction with RDA_{offset} is optional. We thus evaluate RDA_{offset} with (RDA-corrected) and without (RDA-uncorrected) population genetic structure correction (Table 1). For additional specifics related to the implementation of each offset method, see Supplemental Note S1.1-S1.4 and Fig. S1, Fig. S2, Fig. S3.

We varied construction of genomic offset training datasets for each replicate of the 1-, 2-, and 6-trait simulations by varying the marker set used in model training (Fig. 1A, Table 2; see *Q3* below). Each model was trained using genetic and environmental data from all 100 populations. The environmental variables used were only those imposing selection pressure. We predict offset from each model for each population to all 100 within-landscape common gardens from a full factorial *in silico* reciprocal transplant design (Fig. 1B). For each common garden, we quantified offset model performance as the rank correlation (Kendall's τ) between the population mean fitness (averaged over sampled individuals, Equation 3 in Lotterhos 2023) and projected population offset (Fig. 1C). Strong negative relationships between fitness and predicted offset indicate higher performance of the method (note y-axes of Kendall's τ are inverted within figures to show more intuitive performance relationships, Fig. 1C-11). We refer to the preceding processing of data as the *Adaptive Environment* workflow (Fig. 1, Table 2).

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

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

2.3 | Study Questions

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

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

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

where Y_{ij} is the within-landscape performance (Kendall's $\boldsymbol{\tau}$) of a single method for garden j in simulation i, with factors for landscape (L), demography (D), genic level (GL), presence of pleiotropy (P), proportion of QTN or neutral alleles with *temp* clines (respectively $p_{cQTN,t,i}$ and $p_{cNeut,t,i}$), proportion of QTN or neutral alleles with *Env2* clines (respectively $p_{cQTN,Env2,i}$ and $p_{cNeut,Env2,i}$), degree of local adaptation ($LA_{\Delta SA}$), and garden ID (G). The first four factors are illustrated in Fig. 1 of Lotterhos (2023).

Q2 - How is offset performance affected by the proportion of clinal alleles in the data? (Q1B)

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

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

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

Q3 - Is method performance driven by causal loci or by genome-wide patterns of Isolation By Environment? (Q2A)

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

markers, as well as non-QTN markers on the same linkage groups as QTNs). Only loci that passed MAF filtering were included in marker sets. If offset performance is determined solely by adaptive signals in genetic data, offsets trained using *adaptive* markers should have better performance than *all* or *neutral* markers, and *all* markers should have better performance than *neutral* markers.

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

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

Q4 - What is the variation of model performance across the landscape? (Q3a)

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

Q5 - How does the addition of non-adaptive nuisance environments in training affect performance? (Q2B)

In practice, the environments imposing selection are rarely known *a priori*. Additionally, the inclusion of environmental measures that are not correlated with the main axes of selection may reduce model performance compared to models trained using only causal environments. To

investigate the sensitivity of offset methods to environmental input we compared *Adaptive Environment* workflow models from 1-, 2-, and 6-trait simulations – where only the adaptive environment(s) are used in training (*0-nuisance*) – to models from the *Nuisance Environment* workflow trained with the same data but with the addition of nuisance environments (*N-nuisance*, where $N > 0$; Table 2).

We use nuisance environmental variables from Lotterhos (2023) that were real BioClim variables (*TSsd*, *PSsd*, and *ISO*) taken from British Columbia and Alberta, Canada, which have minimal correlation with causal environments and each other (Fig. S4). These three nuisance environments differ from previous implementations of such variables (Láurson et al. 2022) in that they are spatially autocorrelated whereas nuisance environments in Láruson et al. (2022) were not. For 1-trait scenarios, *Env2* was also used as a nuisance environmental variable.

If offset methods are unaffected by the addition of nuisance environmental variables, performance should not differ between *0-nuisance* and *N-nuisance* implementations. Finally, in empirical settings the set of adaptive environments are not known *a priori*. We also explored whether GF would rank adaptive environments higher than nuisance environments using weighted importance output from GF.

Q6 - How well do offset models extrapolate to novel environments outside the range of environments used in training? (Q2C)

Even if offset methods have high within-landscape performance, this does not directly address situations where future environmental conditions are vastly different from the environmental conditions used for training (i.e., novel environments). If performance decreases with increasing environmental novelty relative to training data, this raises questions about the utility of genomic offsets for predicting 1) relative *in situ* vulnerability of populations to future climate change, and 2) the relative suitability of populations to restoration sites that differ drastically than those used in training.

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

3 | Results

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

The ANOVA model (Eq. 1) indicated that the degree of local adaptation of the metapopulation ($LA_{\Delta SA}$) was the primary factor influencing offset performance, followed by common garden location, demography, and landscape (Table S1; Fig. S6). Within the simulations, $LA_{\Delta SA}$ was impacted by pleiotropy, the relative strength of selection, and landscape, (Fig. S7; see also Figs. S2A, S2B in Lotterhos, 2023), so there may be some confounding among these factors.

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

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

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

Q2 - How is offset performance affected by the proportion of clinal alleles in the data? (Q1B)

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

Overall, relationships between performance and p_{cNeut} (second column, Fig. S11) had stronger relationships than between performance and p_{cQTN} (first column, Fig.

S11). However, sometimes performance increased with p_{cNeut} and sometimes it decreased, depending on the method (Fig. S11), indicating that each method is differentially sensitive to clinal alleles in the data. Ultimately, strong population genetic structure along environmental clines in 2-trait simulations (Fig. S12) drove relationships with p_{cNeut} (Fig. S13) which in turn drove relationships with performance (Fig. S14, Fig. S11).

Q3 - Is method performance driven by causal loci or by genome-wide patterns of Isolation-By-Environment? (Q2A)

Overall, 1- and 2-trait *Adaptive Environment* models had relatively similar performance among marker sets. For instance, models trained using *all* or *neutral* markers had similar performance while models trained using *adaptive* markers performed slightly higher than the 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). A*daptive* markers from 6-trait evaluations provided varied performance advantages across methods (Fig. 4).

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).

Together these results indicate that while higher degrees of local adaptation may lead to increased levels of *IBE* in the genome, the signal of *IBE* of input markers generally has minimal and varied impact on performance differences for the scenarios evaluated here. Alternatively, the levels of *IBE* present in the simulated genomes may exceed a minimum threshold of *IBE*, beyond which differences in

performance between marker sets are minimized.

Q4 - What is the variation of model performance across the landscape? (Q3a)

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

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

Q5 - How does the addition of non-adaptive nuisance environments in training affect performance? (Q2B)

Training offset models with the addition of non-adaptive nuisance environmental variables generally reduced offset method performance (Fig. 5). This decline was most dramatic for offset trained on 1-trait simulations (Fig. 5A) compared to the decline observed for 2-trait (Fig. 5B) and 6-trait (Fig. 5C) simulations. The only instances for which median performance did not decrease monotonically with nuisance level were for 2-trait simulations evaluated with GF_{offset} (Fig. S30).

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

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

Q6 - How well do offset models extrapolate to novel environments outside the range of environments used in training? (Q2C)

The datasets that had the greatest within-landscape performance (i.e., those with higher levels of $LA_{\Delta SA}$) were also those that experienced the steepest decline in

performance with increasing climate novelty (red shade, Fig. 6). Importantly, declines in performance for datasets with greater $LA_{\Delta SA}$ were not due to instances where all populations had zero fitness (and thus performance was undefined and manually set to 0; Supplemental Note S4, Fig. S33). Despite little change in the median performance for datasets with low levels of LA, most performance scores from these datasets were below Kendall's τ=0.5, and therefore had little predictive value in novelty scenarios.

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

4 | Discussion

Solutions are needed to mitigate the negative impacts of global change on biodiversity. In the last decade, genomic offset methods have been identified as a complement to other ecological forecasting models because they incorporate intraspecific variation (Keller & Fitzpatrick, 2015; Capblancq et al., 2020; Rellstab et al., 2021). Our evaluations show that offset methods are differentially impacted by both the evolutionary history of sampled populations as well as the decisions made during model training. Our analyses emphasize the importance of sampling locally adapted populations, identifying the drivers underlying environmental selection pressures *a priori*, and restricting offset projections to climates similar to those used in training. Below, we discuss the implications of these findings towards restoration, conservation, and the management of biodiversity.

4.1 | The importance of local adaptation

A basic assumption of genomic offset methods is that the sampled populations are adapted to their local environment (Rellstab et al., 2016, 2021), but this assumption has not been formally tested. Our analyses show that indeed the degree of local adaptation ($LA_{\Delta SA}$) is one of the primary factors that determine model performance for most methods. A value of $LA_{\Delta SA}$ ~ 0.5 indicates that fitness in demes is on average 50% higher in sympatry than allopatry. Values of $LA_{\Delta SA}$ represent the average deme-level magnitudes of $LA_{\Delta HA}$ and $LA_{\Delta LF}$ across the metapopulation (Blanquart et al., 2013). Previous metaanalyses of studies measuring local adaptation of natural populations have used different measures of *LA* from the ones we calculate here, but do show that some species evolve large fitness differences among populations (Hereford, 2009; Leimu & Fischer, 2008). Given the prevalence of *LA* found previously (Hereford, 2009; Leimu & Fisher, 2010), we may therefore expect some genomic offset methods to do reasonably well when local adaptation in the metapopulation is high (i.e., when $LA_{\Delta SA} > 0.5$).

4.2 | The importance of the signals within genomic marker sets

Because of the assumption that locally adapted populations will be necessary for

satisfactory model performance, initial implementations of genomic offset models focussed on putatively adaptive markers where this signal may be strongest (Keller & Fitzpatrick, 2015; Rellstab et al., 2016). More recently, investigators have varied the set of markers used to train models but have found little influence on performance (Fitzpatrick et al., 2021; Lachmuth, Capblancq, Keller, et al., 2023; Láruson et al., 2022; Lind et al., 2024). Our results are similar to previous investigations, finding that the *adaptive* marker sets provide minimal advantage over *all* or *neutral* marker sets, but not universally or by great margins.

One hypothesis put forth as to why adaptive marker sets perform similarly to all markers is that genome-wide data captures sufficient signatures of IBE (Lachmuth, Capblancq, Keller, et al., 2023; Lind et al., 2024). Our analysis found weak positive relationships between performance and levels of *IBE* within marker sets. Even so, and except for RONA, there were no universal relationships within methods between the difference in *IBE* of marker sets and the difference in performance of the models trained with these markers.

While we found little impact of levels of *IBE* on overall performance, the way in which we measured IBE may have masked causative relationships. For instance, in our measure of IBE we correlated environmental distance with pairwise F_{ST}. In doing so, our measure of IBE distills genetic distance down to a single value from a large number of loci, and gives less weight to loci with rare alleles. In future studies, creating a marker set by ranking loci by single-locus measures of IBE offers another opportunity to understand the impact of IBE on performance. Such marker sets could be used to compare to performance from putatively adaptive marker sets or marker sets composed of all or random loci. Empirical datasets will also be able to specifically address geographical distances while quantifying IBE (e.g., Bradburd et al., 2013).

While measures of IBE are one signal remaining to be explored in future analyses, the proportion of clinal neutral loci within marker sets was shown to impact performance, sometimes being positively related to performance and sometimes negatively depending on the context. These and other signals within data that could improve or mislead offset models also warrant further investigation.

4.3 / The importance of adaptive environmental variables

In empirical settings, the environ- mental drivers of local adaptation are rarely known *a priori*. Even so, our results emphasize the importance of identifying these variables before training offset models, as there were often declines in performance between models trained using only adaptive environmental variables (*0-nuisance*) and those trained using additional non-adaptive nuisance environmental variables (*N-nuisance*).

The importance of identifying these selective environments may be particularly germane to two general empirical scenarios. In the first empirical scenario,, sparsely sampling an environmentally heterogeneous range may enrich genetic signals (e.g., coincident

population structure) most correlated to environmental variables that maintain a gradient across this extent, and miss signals relevant to more local scales. In the second empirical scenario, identifying the environmental variables underlying selection is particularly important when a specific genomic offset method is ill-suited to differentiate importance among input variables. For instance, RDA (and therefore RDA_{offset}) assumes that the environmental variables used to build models are not collinear; (as implemented here; Capblancq & Forester, 2021; Legendre & Legendre, 2012). Because of this, empirical datasets must be limited to a subset of available environmental measures. The process of excluding environmental variables in this way may omit signals of adaptive drivers (particularly when true drivers are not well measured), or perhaps incorporate environmental variables that do not coincide with drivers of selection. In these cases, performance is likely to decline. As such, this may indicate that methods such as RDA_{offset} are likely to perform worse in, or less uniformly across, realistic empirical settings than what our current findings suggest.

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

There are some differences between the nuisance environmental variables implemented here and those that have been implemented previously. For instance, Láruson et al. (2022) created nuisance variables by randomly sampling a multivariate normal distribution. In contrast to findings here, Láruson et al. (2022) found that model performance was relatively unaffected with the addition of nuisance variables. The minimal influence of nuisance variables on performance found by Láruson et al. (2022) may differ from the performance declines reported here because the nuisance variables we used were spatially autocorrelated, while those from Láruson et al. (2022) were not. Inclusion of nuisance variables that are spatially autocorrelated may mislead offset models more generally than variables with little spatial autocorrelation because of the spurious relationship between environmental structure and genetic structure.

4.4 / The effect of environmental novelty

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

Thus, genomic offset models are likely ill-suited for estimating fitness ranks of populations in environments that differ drastically from those used to train the models themselves. This is particularly relevant for applications of offset methods that attempt to estimate *in situ* risk of climate change to years or climate scenarios where the environment is expected to be increasingly novel. While climate novelty is often measured with respect to historical variability (e.g., Lotterhos et al., 2021; Mahony et al., 2017; Williams et al., 2007), indices of local climate change indicate that local environments in terrestrial systems could experience change in excess of three standard deviations relative to historic values (Williams et al., 2007). Similar indices in marine systems indicate potential for even greater novelty (Lotterhos et al., 2021). We observed performance declines below the analogous z_E=3.13 *Climate Novelty* scenario, indicating offset predictions will likely be inaccurate in many real-world climate change predictions. These issues are also germane to measures derived from offset values (Gougherty et al., 2021; Lachmuth, Capblancq, Keller, et al., 2023; Lachmuth, Capblancq, Prakash, et al., 2023), which currently do not consider the degree of climate novelty in the prediction (but see DeSaix et al., 2022).

Our results present a best-case scenario for predicting performance in novel environments, as in many cases there will be biological reasons as to why climate-fitness relationships will differ in future environments from relationships measured within the contemporary climate space (see Fig. 5 in Capblancq et al., 2020). For the simulations here, the relationship between contemporary and novel environments with fitness was the same.

4.5 / Genomic offsets in practice

Our evaluations show that genomic offset methods hold promise for predicting maladaptation to environmental change within a historical baseline, in metapopulations that evolve strong local adaptation. However, our analyses also emphasize the limits of these methods in some

systems or scenarios. In practice, species that are locally adapted to measurable environmental variables will be best suited for offset methods when predicting the relative performance of populations in a contemporary common garden, but paradoxically these species may be least suited to using these methods to predict their vulnerability to novel climates.

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

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

Acknowledgements

This research was funded by NSF-2043905 (KEL) and Northeastern University. The funding bodies did not have any role in the design of the study, analysis, interpretation of results, or in writing of the manuscript.

Author Contributions

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

Conflict of Interests

The authors declare no conflicts of interest.

Data Availability

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

All directories, notebooks, and scripts can be found on GitHub, and will be archived on Zenodo.
https://github.com/ModelValidationProgram/MVP-offsets

References

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters*, *16*(9), 1195–1205. https://doi.org/10.1111/ele.12150

Bonan, G. B. (2008). Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, *320*(5882), 1444–1449. https://doi.org/10.1126/science.1155121

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the Effects of Geographic and Ecological Isolation on Genetic Differentiation. *Evolution*, *67*(11), 3258–3273. https://doi.org/10.1111/evo.12193

Brown, J. H. (1984). On the Relationship between Abundance and Distribution of Species. *The American Naturalist*, *124*(2), 255–279. https://doi.org/10.1086/284267

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), 245–269. https://doi.org/10.1146/annurev-ecolsys-020720-042553

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.13722

Clark, J. S., Carpenter, S. R., Barber, M., Collins, S., Dobson, A., Foley, J. A., Lodge, D. M., Pascual, M., Jr., R. P., Pizer, W., Pringle, C., Reid, W. V., Rose, K. A., Sala, O., Schlesinger, W. H., Wall, D. H., & Wear, D. (2001). Ecological Forecasts: An Emerging Imperative. Science, *293*(5530), 657–660. https://doi.org/10.1126/science.293.5530.657

DeSaix, M. G., George, T. L., Seglund, A. E., Spellman, G. M., Zavaleta, E. S., & Ruegg, K. C. (2022). Forecasting climate change response in an alpine specialist songbird reveals the importance of considering novel climate. *Diversity and Distributions*, *28*(10), 2239–2254. https://doi.org/10.1111/ddi.13628

Doney, S. C., Ruckelshaus, M., Duffy, J. E., Barry, J. P., Chan, F., English, C. A., Galindo, H. M., Grebmeier, J. M., Hollowed, A. B., Knowlton, N., Polovina, J., Rabalais, N. N., Sydeman, W. J., & Talley, L. D. (2012). Climate Change Impacts on Marine Ecosystems. *Annual Review of Marine Science*, *4*(1), 11–37. https://doi.org/10.1146/annurev-marine-041911-111611

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fitzpatrick, M. C., Blois, J. L., Williams, J. W., Nieto-Lugilde, D., Maguire, K. C., & Lorenz, D. J. (20*18*). How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Global Change Biology*, *24*(8), 3575–3586. https://doi.org/10.1111/gcb.141*38*

Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources*. https://doi.org/10.1111/1755-0998.13374

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecology Letters, 18(1), 1–16. https://doi.org/10.1111/ele.1*237*6

Gain, C., & François, O. (20*21*). LEA 3: Factor models in population genetics and ecological genomics with R. Molecular Ecology Resources, 21(8), 2738–2748. https://doi.org/10.1111/1755-0998.13366

Good, R. D. (1931). A THEORY OF PLANT GEOGRAPHY. *New Phytologist*, *30*(3), 149–149. https://doi.org/10.1111/j.1469-8137.1931.tb07414.x

Gougherty, A. V., Keller, S. R., & Fitzpatrick, M. C. (2021). Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change*, 1–15. https://doi.org/10.1038/s41558-020-00968-6

Hereford, J. (2009). A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. The American Naturalist, *173*(5), 579–588. https://doi.org/10.1086/597611

Hoegh-Guldberg, O., & Bruno, J. F. (2010). The Impact of Climate Change on the World's Marine Ecosystems. Science, *328*(5985), 1523–1528. https://doi.org/10.1126/science.1189930

Lachmuth, S., Capblancq, T., Keller, S. R., & Fitzpatrick, M. C. (2023). Assessing uncertainty in genomic offset forecasts from landscape genomic models (and implications for restoration and assisted migration). *Frontiers in Ecology and Evolution*, *11*, 1155783. https://doi.org/10.3389/fevo.2023.1155783

Lachmuth, S., Capblancq, T., Prakash, A., Keller, S. R., & Fitzpatrick, M. C. (2023). Novel genomic offset metrics account for local adaptation in climate suitability forecasts and inform assisted migration. *BioRxiv*, 2023.06.05.541958. https://doi.org/10.1101/2023.06.05.541958

Láruson, Á. J., Fitzpatrick, M. C., Keller, S. R., Haller, B. C., & Lotterhos, K. E. (2022). Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. *Evolutionary Applications*, *15*(3), 403–416. https://doi.org/10.1111/eva.13354

Lee-Yaw, J. A., McCune, J. L., Pironon, S., & Sheth, S. N. (2022). Species distribution models rarely predict the biology of real populations. *Ecography*, *2022*(6). https://doi.org/10.1111/ecog.05877

Legendre, P., & Legendre, L. (2012). *Numerical Ecology* (Vol. 24). Elsevier.
Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*(12), e4010. https://doi.org/10.1371/journal.pone.0004010.s001

Lind, B. M., Candido-Ribeiro, R., Singh, P., Lu, M., Vidakovic, D. O., Booker, T. R., Whitlock, M., Isabel, N., Yeaman, S., & Aitken, S. N. (2024). How useful is genomic data for predicting maladaptation to future climate? *In Press at Global Change Biology*. https://doi.org/10.1101/2023.02.10.528022

Lotterhos, K. E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. *Proceedings of the National Academy of Sciences*, *120*(12). https://doi.org/10.1073/pnas.2220313120

Lotterhos, K. E., Fitzpatrick, M. C., & Blackmon, H. (2022). Simulation Tests of Methods in Evolution, Ecology, and Systematics: Pitfalls, Progress, and Principles. Annual Review of Ecology, Evolution, and Systematics, 53(1), 113–

136. https://doi.org/10.1*146*/annurev-ecolsys-102320-093722

Lotterhos, K. E., Láruson, Á. J., & Jiang, L.-Q. (2021). Novel and disappearing climates in the global surface ocean from 1800 to 2100. *Scientific Reports*, *11*(1), 15535. https://doi.org/10.1038/s41598-021-94872-4

Mahony, C. R., Cannon, A. J., Wang, T., & Aitken, S. N. (2017). A closer look at novel climates: new methods and insights at continental to landscape scales. Global Change Biology, 23(9), 3934–3955. https://doi.org/10.1111/gcb.13645

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. Evolutionary Applications, 14(5), 1202–1212. https://doi.org/10.1111/eva.13205

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., Bodénès, C., Sperisen, C., Kremer, A., & Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. Molecular Ecology, *25*(23), 5907–5924. https://doi.org/10.1111/mec.13889

Schmolke, A., Thorbek, P., DeAngelis, D. L., & Grimm, V. (2010). Ecological models supporting environmental decision making: a strategy for the future. *Trends in Ecology & Evolution*, *25*(8), 479–486. https://doi.org/10.1016/j.tree.2010.05.001

Waldvogel, A.-M., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R., Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O., Bergland, A., Flatt, T., Guillaume, F., & Pfenninger, M. (2020). Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, *4*(1). https://doi.org/10.1002/evl3.154

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. Molecular Ecology, 23(23), 5649–5662. https://doi.org/10.1111/mec.12938

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution, 38(6), 1358–1370.

Williams, J. W., Jackson, S. T., & Kutzbach, J. E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. Proceedings of the National Academy of Sciences, *104*(14), 5738–5742. https://doi.org/10.1073/pnas.0606292104

1405

Figures for:

The limits of predicting maladaptation to future 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

11 January 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?**
Gradient Forests[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)	[$AE_{1\text{-trait}}$ environments + Env2] +/- [ISO + PSsd] +/- [TSsd]	Yes	Yes	175,500
	2-trait	180 (1)	[$AE_{2\text{-trait}}$ environments + ISO + PSsd] +/- [TSsd]			432,000*
	6-trait	1 (1)	[$AE_{6\text{-trait}}$ environments + ISO + PSsd + TSsd]			1,200*
Climate Novelty (CN)†	1-trait	45 (10)	$AE_{1\text{-trait}}$ environments	Prediction only (using AE training models)	No	64,800*
	2-trait	180 (10)	$AE_{2\text{-trait}}$ environments			259,200*
	6-trait	1 (1)	$AE_{6\text{-trait}}$ 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*" = [$AE_{1\text{-trait}}$ environments + Env2] and "*1-trait 4-nuisance*" = [$AE_{1\text{-trait}}$ 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. Counts of evaluations were tabulated in SC 02.10.01.

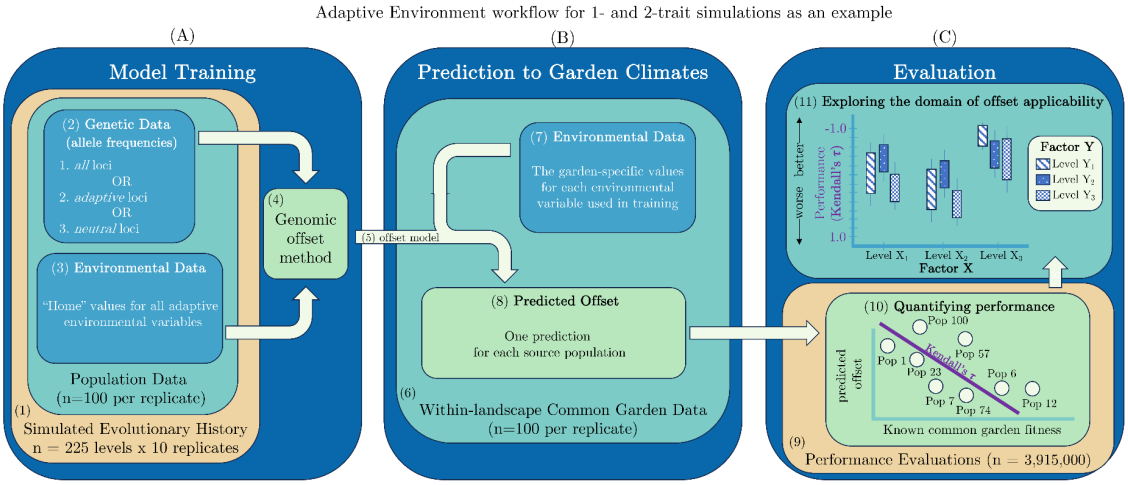
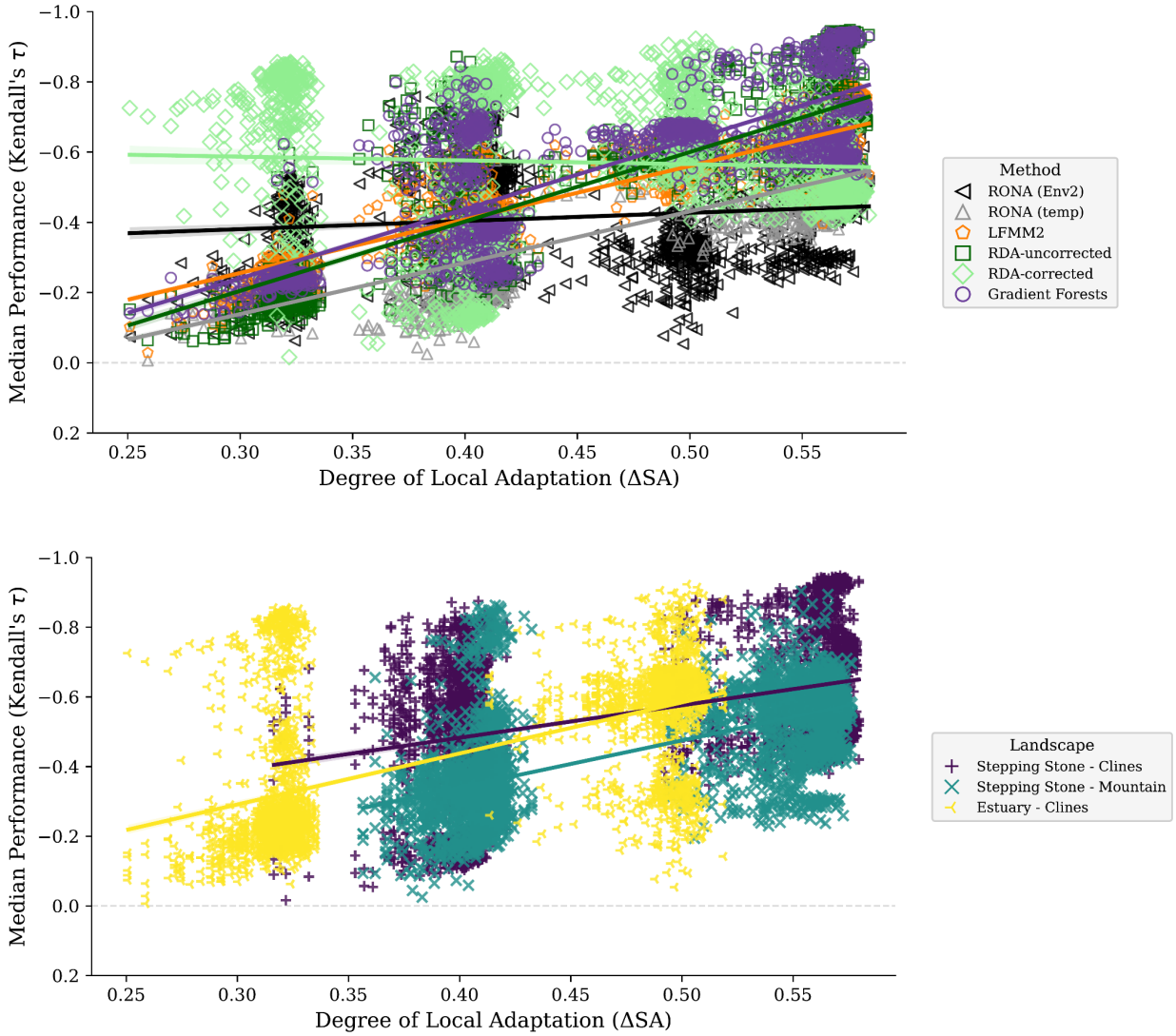


Figure 1 Analysis of 1-, 2-, and 6-trait simulations included three main phases: A) model training, B) model prediction, and C) evaluation of models. The *Adaptive Environment* workflow is shown as an example of the processing of 1- and 2-trait simulation data for genomic offset evaluation. In total, 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)


Median Performance (Kendall's τ)
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.25
0.30
0.35
0.40
0.45
0.50
0.55
Degree of Local Adaptation (ΔSA)
Method
RONA (Env2)
RONA (temp)
LFMM2
RDA-uncorrected
RDA-corrected
Gradient Forests
Median Performance (Kendall's τ)
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.25
0.30
0.35
0.40
0.45
0.50
0.55
Degree of Local Adaptation (ΔSA)
Landscape
Stepping Stone - Clines
Stepping Stone - Mountain
Estuary - Clines

(Fig. 2 continued)

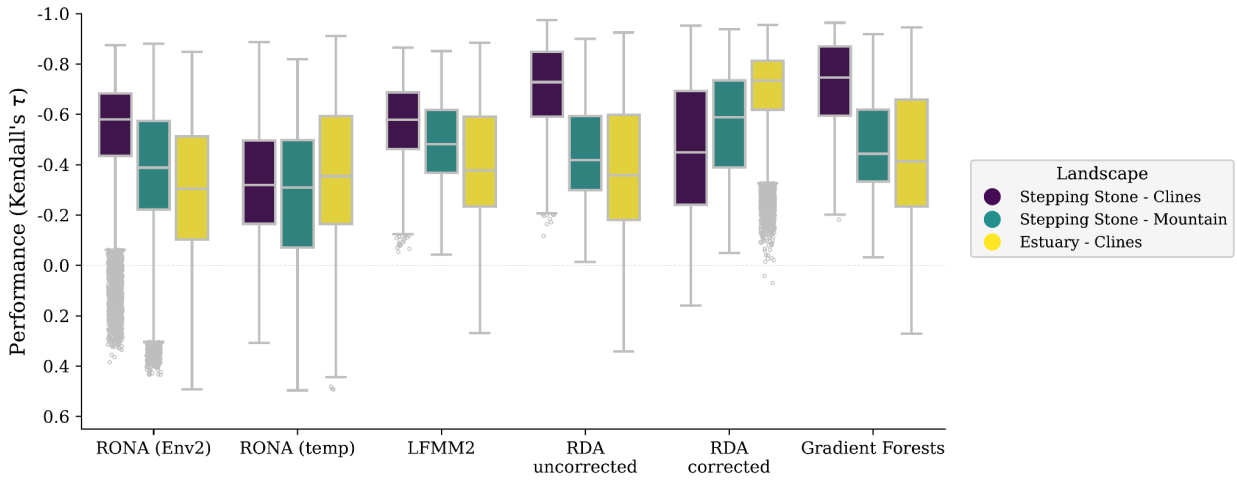


Figure 2 Predictive performance of genomic offset models (y-axes) is driven by the degree of local adaptation (A) and the spatial patterns of adaptive environments across the landscape (B, C). For each model, a median value from performance scores from 100 common gardens is shown for A and B; C shows 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)

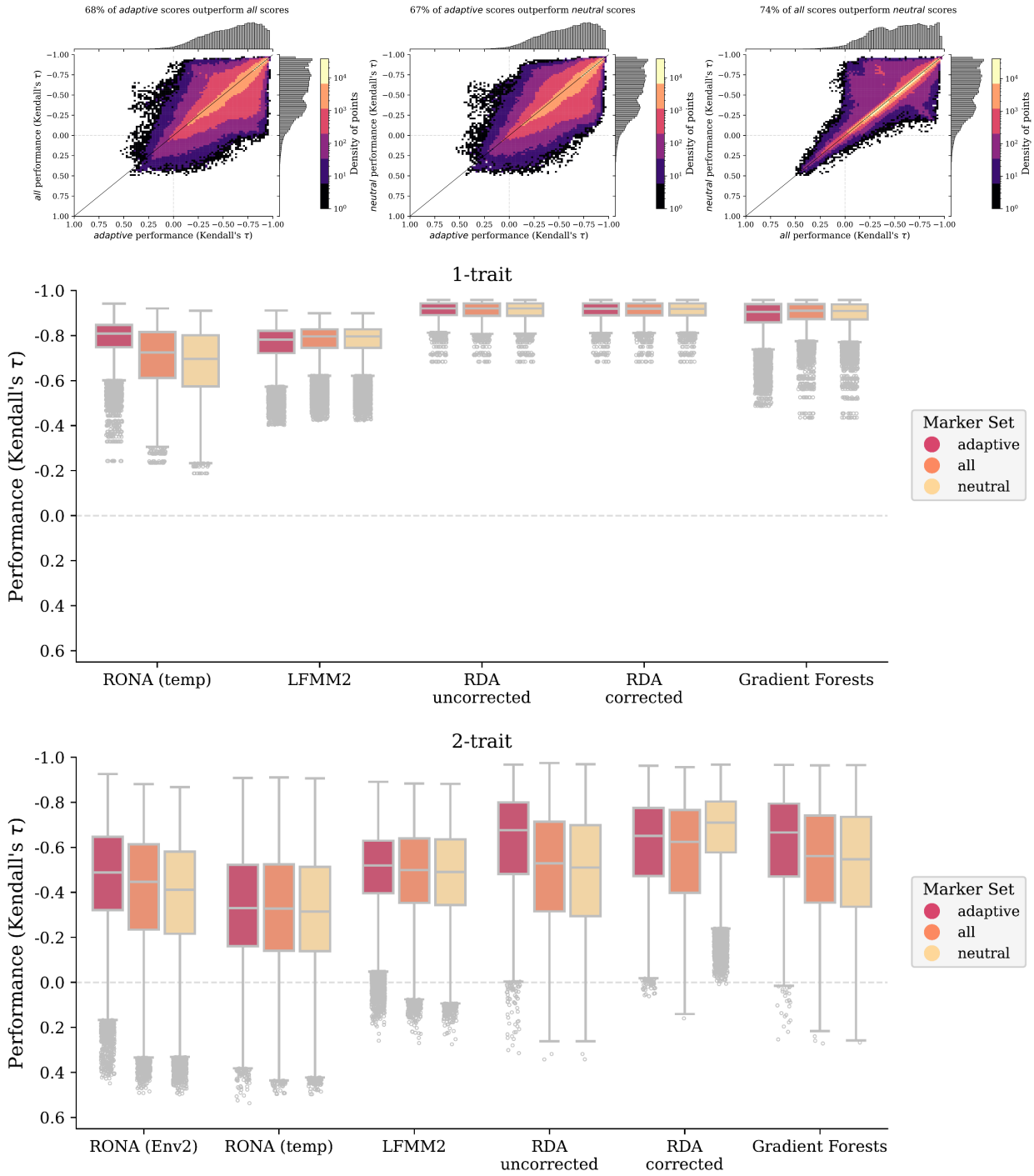

68% of adaptive scores outperform all scores
67% of adaptive scores outperform neutral scores
74% of all scores outperform neutral scores
all performance (Kendall's τ)
neutral performance (Kendall's τ)
adaptive performance (Kendall's τ)
all performance (Kendall's τ)
Density of points
1-trait
2-trait
Performance (Kendall's τ)
Marker Set
adaptive
all
neutral
RONA (Env2)
RONA (temp)
LFMM2
RDA uncorrected
RDA corrected
Gradient Forests

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.

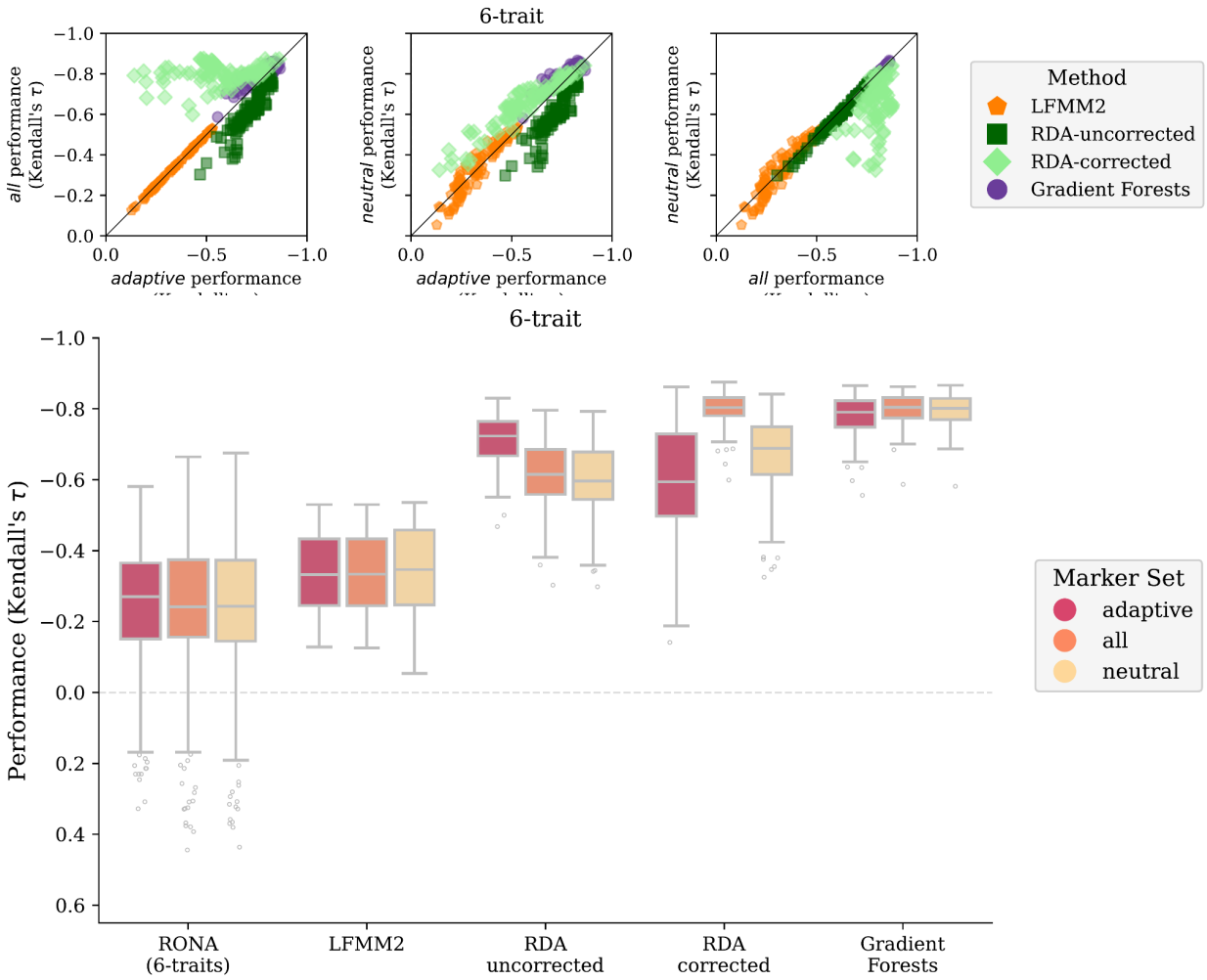


Figure 4 Comparison of marker choice across genomic offset methods for the 6-trait simulation. A-C are scatterplots of pairwise comparisons of performance between marker sets (RONA is not shown, except in SN 02.05.10). D are boxplots from the same data in A-C ($RONA_{6\text{-traits}}$ is the combined 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)

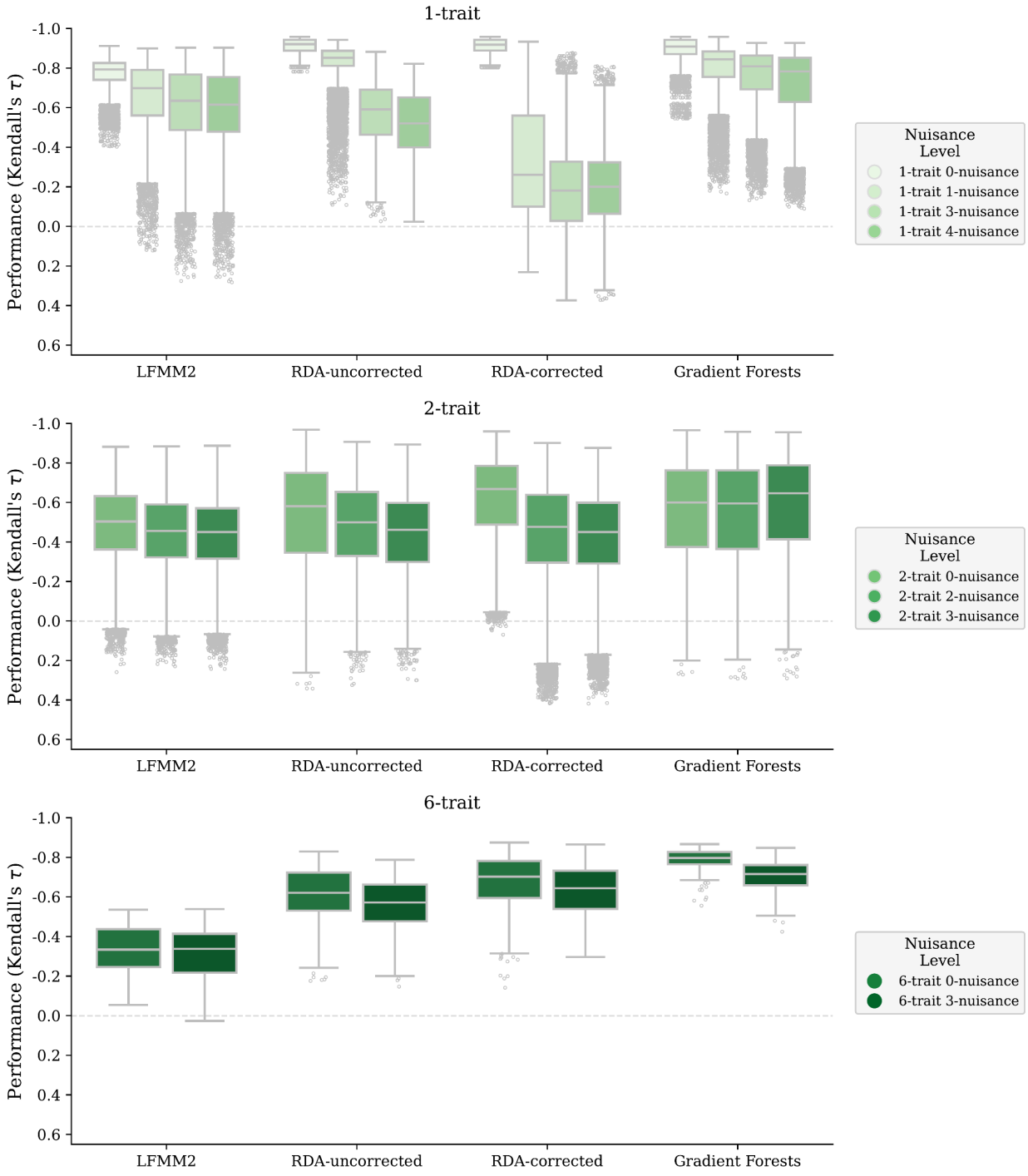

1-trait
2-trait
6-trait
Performance (Kendall's τ)
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
LFMM2
RDA-uncorrected
RDA-corrected
Gradient Forests
Nuisance Level
1-trait 0-nuisance
1-trait 1-nuisance
1-trait 3-nuisance
1-trait 4-nuisance
2-trait 0-nuisance
2-trait 2-nuisance
2-trait 3-nuisance
6-trait 0-nuisance
6-trait 3-nuisance

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.

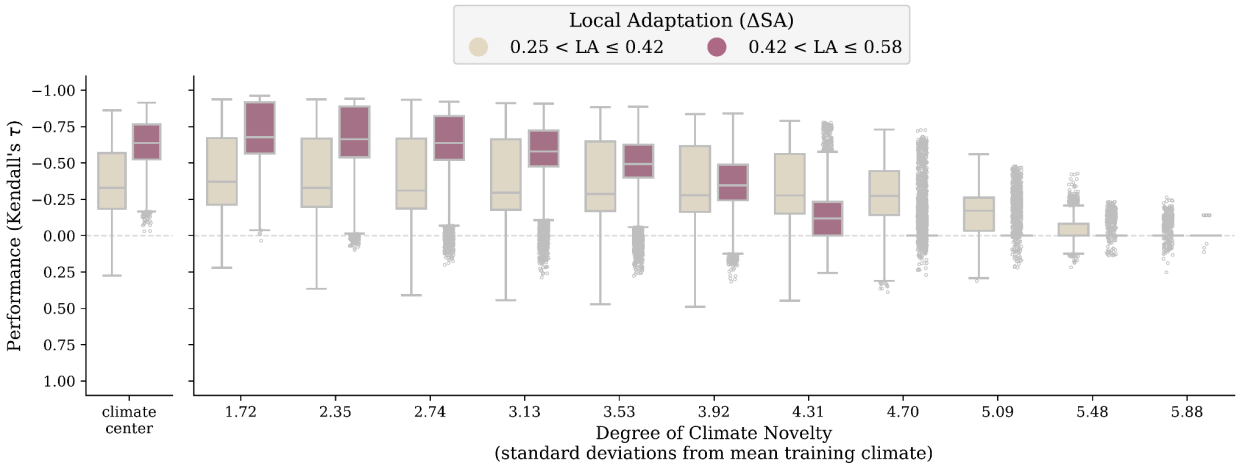


Figure 6 Performance decays with climate novelty relative to training data. Shown is model performance (y-axes) across methods at climate center (A) and across common gardens each representing increasing degrees of climate novelty relative to training data (x-axis of B) where all 100 populations have 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 deviations 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.

Supplemental Information

The limits of predicting maladaptation to future 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

11 January 2024

Running Title: *The 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

Table of Contents

Supplemental Notes

S1 - Implementation of Offset Methods

See Supplemental Note S2 for specific citations of code.

1.1 | Gradient Forests

For a given set of input loci (*all*, *adaptive*, or *neutral*; see Q3 in Methods), and for all workflows, Gradient Forests (GF_{Offset}) is trained using `ntree=500`, `corr.threshold=0.5`, and `maxLevel=`$(0.368 * \frac{N}{2})$, where N is the number of populations. Using default linear extrapolation, the trained model is projected onto the landscape using the `predict` function and the same environmental values used in training. This creates the "current" projection used to calculate offset below.

The trained models are then fit to the climate of each of 100 common gardens on the landscape for the *Adaptive Environment* and *Nuisance Environment* scenarios, or to each of the 11 *Climate Novelty* scenarios. Specifically, for each garden, the `predict` function is used to take the trained model and the garden's climate to create a projection similar to that using current climate data (previous paragraph). Then the Euclidean distance is taken between the current and future projections to calculate offset.

1.2 | The Risk Of Non-Adaptedness

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

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

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

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

183 Of note, in some instances, particularly *Adaptive Environment* datasets
184 simulated with oligogenic architectures, there were no loci with significant linear
185 relationships with environmental variables and these instances were given NA
186 performance values (i.e., excluded from analyses).

187 1.3 | Landscape and Ecological Association (LEA) Studies R package

188 We used the `genetic.offset` function in the LEA package to estimate
189 $LFMM2_{offset}$ for each workflow (Fig. 1). The `genetic.offset` function was used
190 with default settings, except for *K*, the number of subdivisions within the data. To
191 determine *K* needed for the `genetic.offset` function for 1- and 2-trait
192 simulations, we first used filtered SNP data (see Section 2.1) to estimate 21 principal
193 components (PCs) using principal component analysis (PCA). Then we equated *K*
194 to the number of PC axes that explain greater than 1.3x the variation of the next
195 subsequent axis (see line 677-697 of c-AnalyzeSimOutput.R from Lotterhos, 2023).
196 This resulted in varied *K* across simulation levels and replicates (Fig S1). For the 6-
197 trait simulation, it was never the case that a PC axis explained >1.3x the variation
198 explained by the previous axis, so we used the elbow rule to estimate *K*=7 (Fig S2).

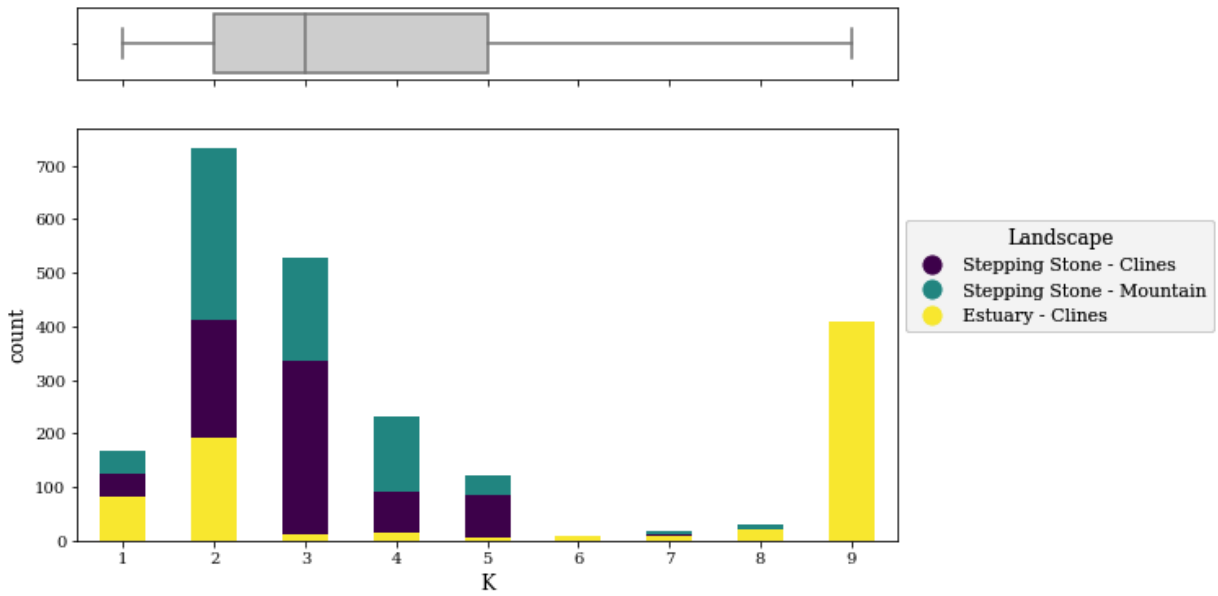


Fig S1 Distribution of K used for the lfmm2 `genetic.offset` function for 1- and 2-trait simulations. K was estimated by determining the number of principal component axes that explain at least 1.3x times the amount of variation of the subsequent axis. Code used to create this figure can be found in SC 02.09.01.

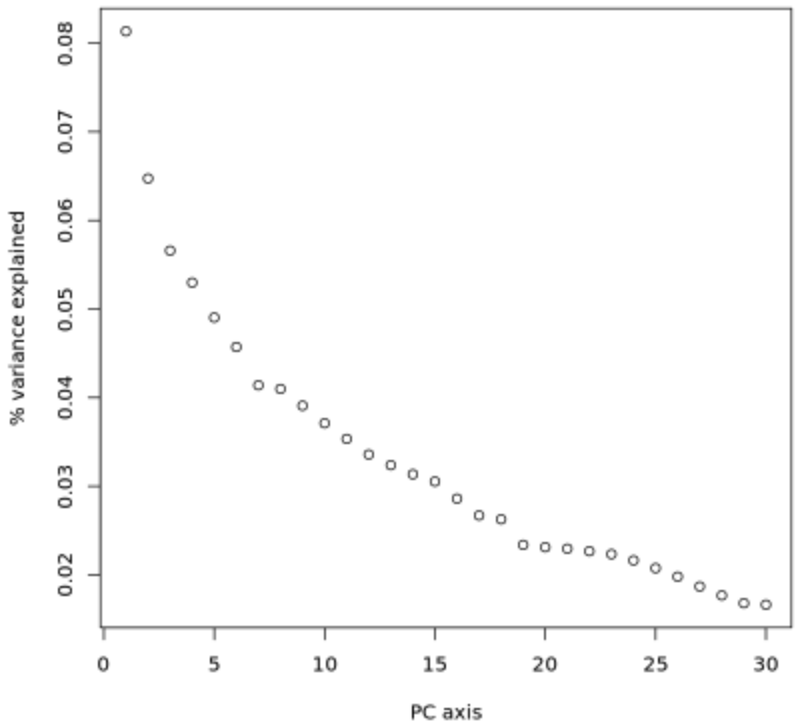
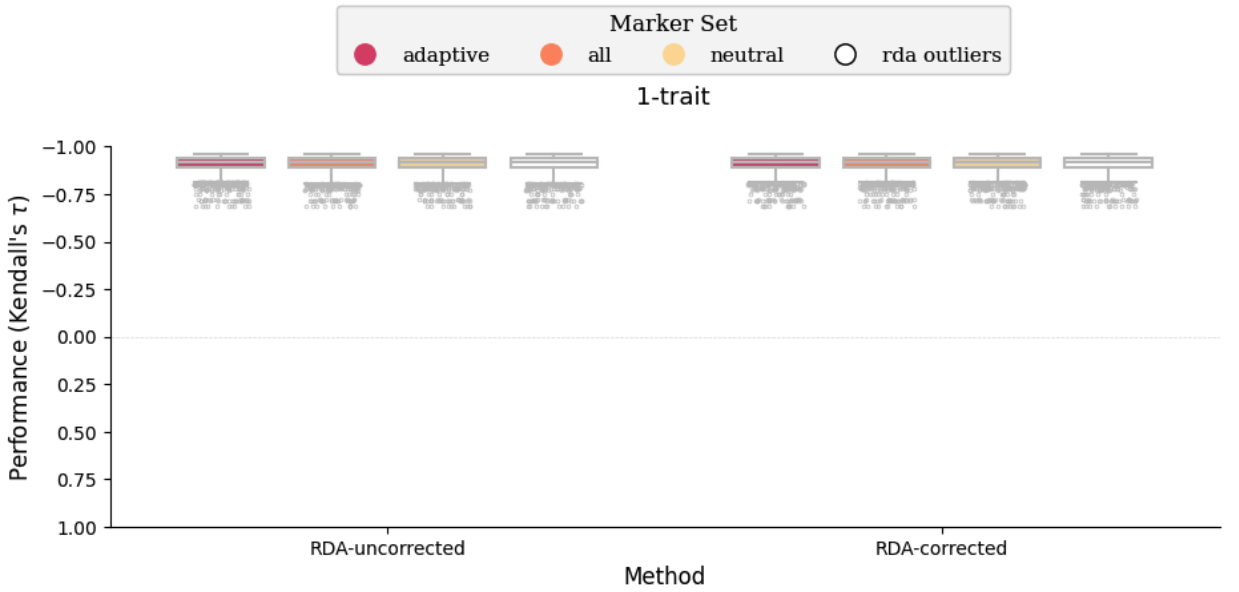


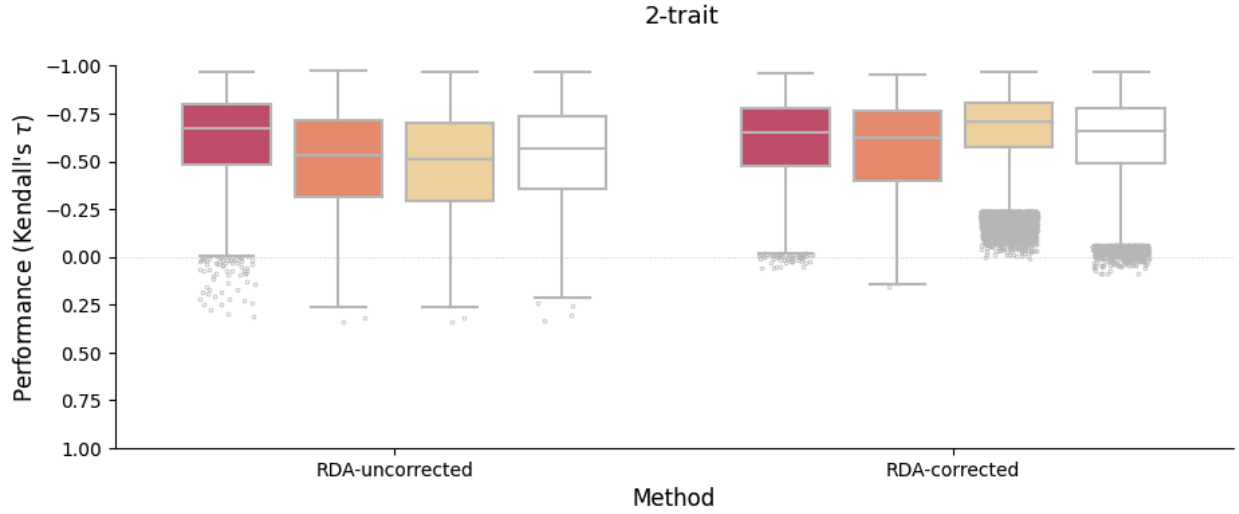
Fig S2 Percent variance explained from principal component (PC) axes from principal component analysis of SNP data from the 6-trait simulation. The "elbow rule" was used to estimate K=7 for this simulation. Code used to create this figure can be found in SC 02.05.11.

1.4 | Redundancy Analysis

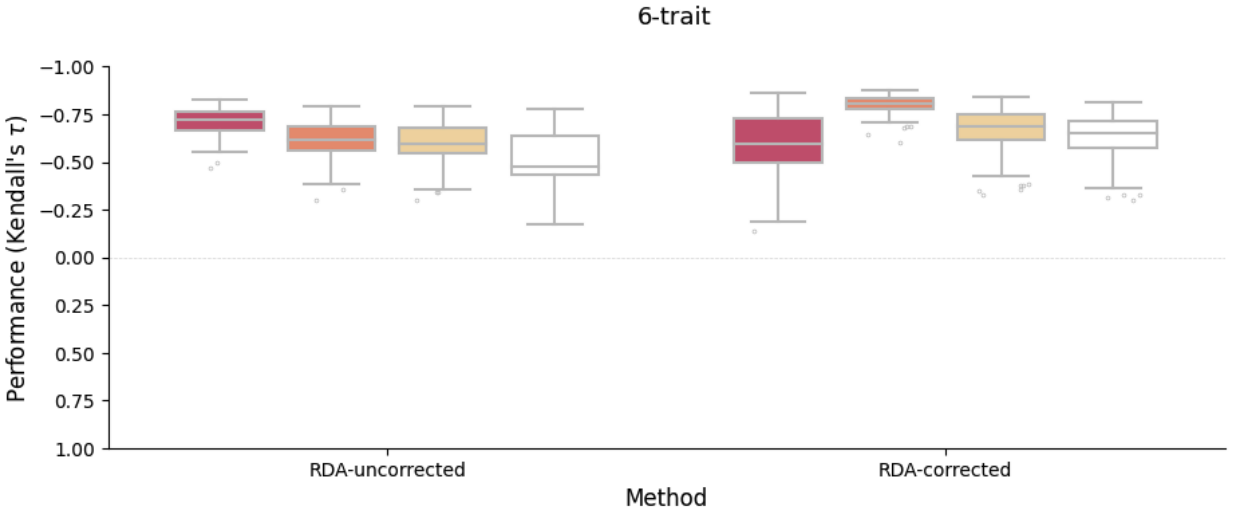
RDA_{offset} was implemented as in Capblancq & Forester (2021). Note that the environmental variables used here across workflows had minimal correlation, as required by RDA (Fig. S4). In addition to the three marker sets used as input (*all*, *adaptive*, or *neutral*; see Q3 in Methods), we also used *RDA-outliers* as input to RDA offset estimation. *RDA-outlier* loci were those from separate RDA models trained using *all* loci and adaptive environments, and were included in this study because of their use in the original implementation of RDA_{offset} by Capblancq & Forester (2021). *RDA-outliers* were identified as in Capblancq et al. (2018) for loci with q-values < 0.05. For each 1-, 2-, and 6-trait simulation replicate, RDA_{offset} was estimated with (RDA-corrected) and without (RDA-uncorrected) correction for population genetic structure. When correcting for structure, the loadings for the first two PCs from PCA estimated with *all* loci were used. Because *RDA-outliers* performed on par with or worse than other marker sets in 1-trait (Fig S2A), 2-trait (Fig S3B), and 6-trait (Fig S3C) evaluations from the *Adaptive Environment* workflow (Fig. 1) we focus on *all*, *neutral*, and *adaptive* marker sets for the main text.


Marker Set
adaptive
all
neutral
rda outliers
1-trait
Performance (Kendall's τ)
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
RDA-uncorrected
RDA-corrected
Method

225


2-trait
Performance (Kendall's τ)
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
RDA-uncorrected
RDA-corrected
Method

226


6-trait
Performance (Kendall's τ)
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
RDA-uncorrected
RDA-corrected
Method

227

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. Data in this figure is from the *Adaptive Environment* workflow. Code to create this figure can be found in SC 02.06.02.

S2 - Coding workflows

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

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

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

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

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

Simulation data used below within scripts and notebooks has been processed from SLiM output separately by Lotterhos (2023) into more user-friendly forms - see here for more information: https://github.com/ModelValidationProgram/MVP-NonClinalAF/tree/main/sim_ouput_20220428_metadata

1.1 | The *Adaptive Environment* coding workflow

The *Adaptive Environment* workflow represents the general pipeline for processing simulations and running genomic offset methods, most other

processing code is built on top of this main pipeline (i.e., scripts and notebooks source/import functions from these scripts to avoid code redundancy).

1.1.1 | 1- and 2-trait simulations

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

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

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

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

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

Nested dictionaries containing validation results from each method are reformatted and combined into a single object in notebooks within the SC 02.01.00

directory. These combined objects are used throughout remaining analyses in jupyter notebooks found in subdirectories of SC 02.

1.1.2 | 6-trait simulation

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

GF was trained using the same script as 1- and 2-trait simulations (`MVP_gf_training_script.R`). GF offset was predicted manually in SC 02.05.02, and validated manually in SC 02.05.03. In SC 02.05.04 - 02.05.05 LFMM was trained and validated manually. Similarly, RDA was trained and validated in SC 02.05.06 - SC 02.05.07, and RONA trained and validated in SC 02.05.08 - SC 02.05.09.

1.2 | The *Climate Novelty* coding workflow

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

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

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

1.3 | The *Nuisance Environment* coding workflow

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

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

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

1.4 | Misc

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

S3 - Defining *Climate Novelty* scenarios

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

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

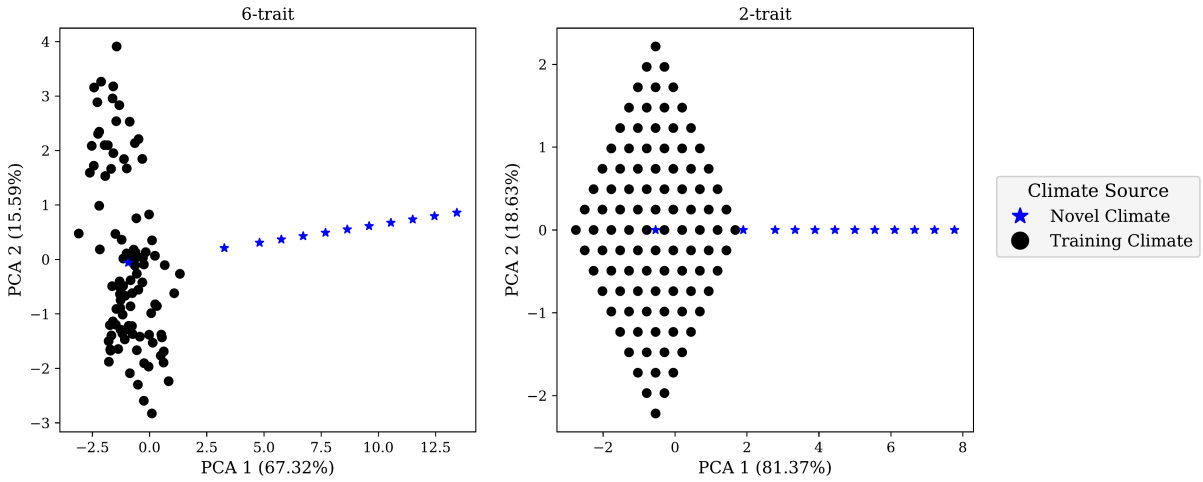


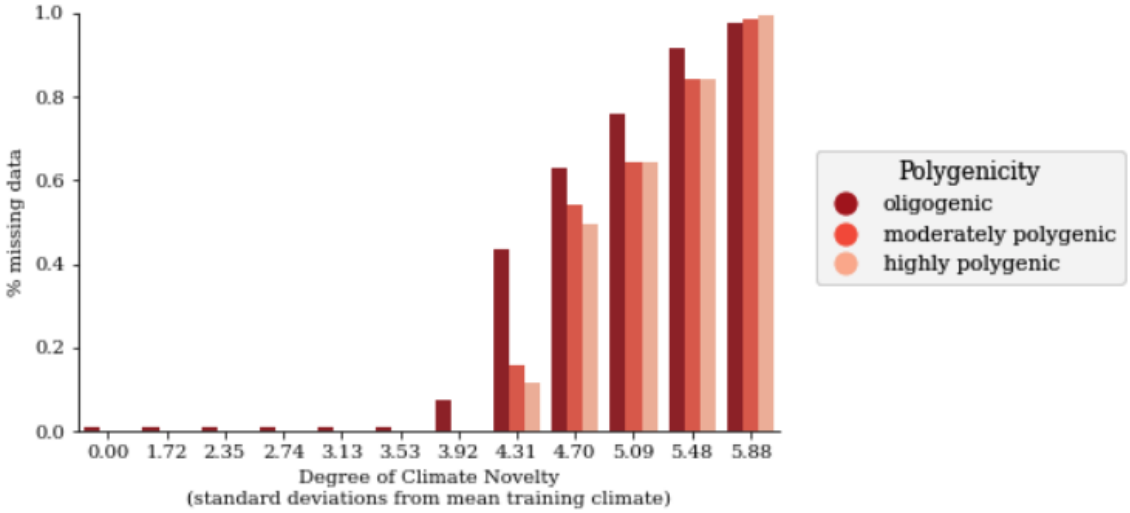
Fig S5 Differentiation of *Climate Novelty* environments (blue stars, including climate center) from within-landscape environments (black circles) using Principal Component Analysis (PCA) of environmental data. Environmental data is centered and standardized relative to the within-landscape environmental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

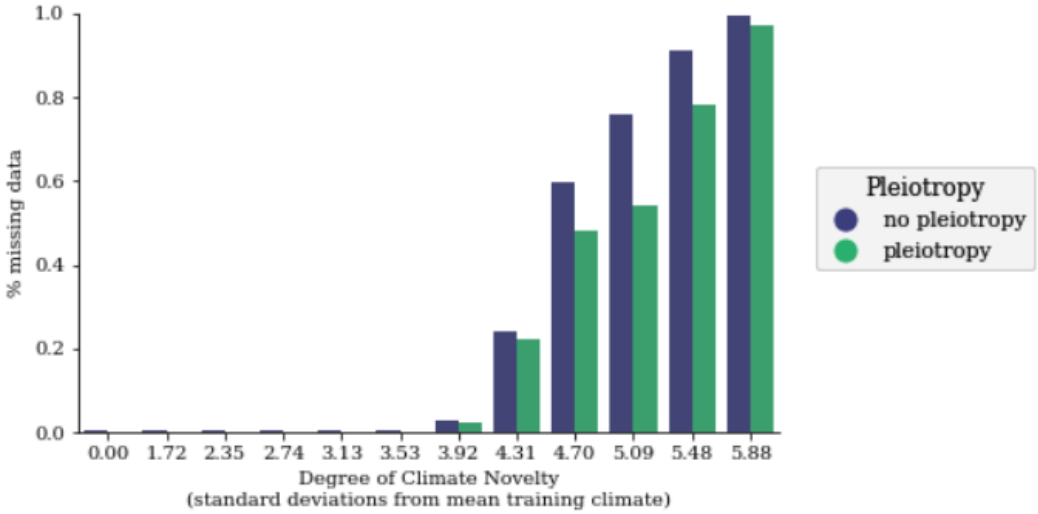
S4 - Missing data in *Climate Novelty* evaluations

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

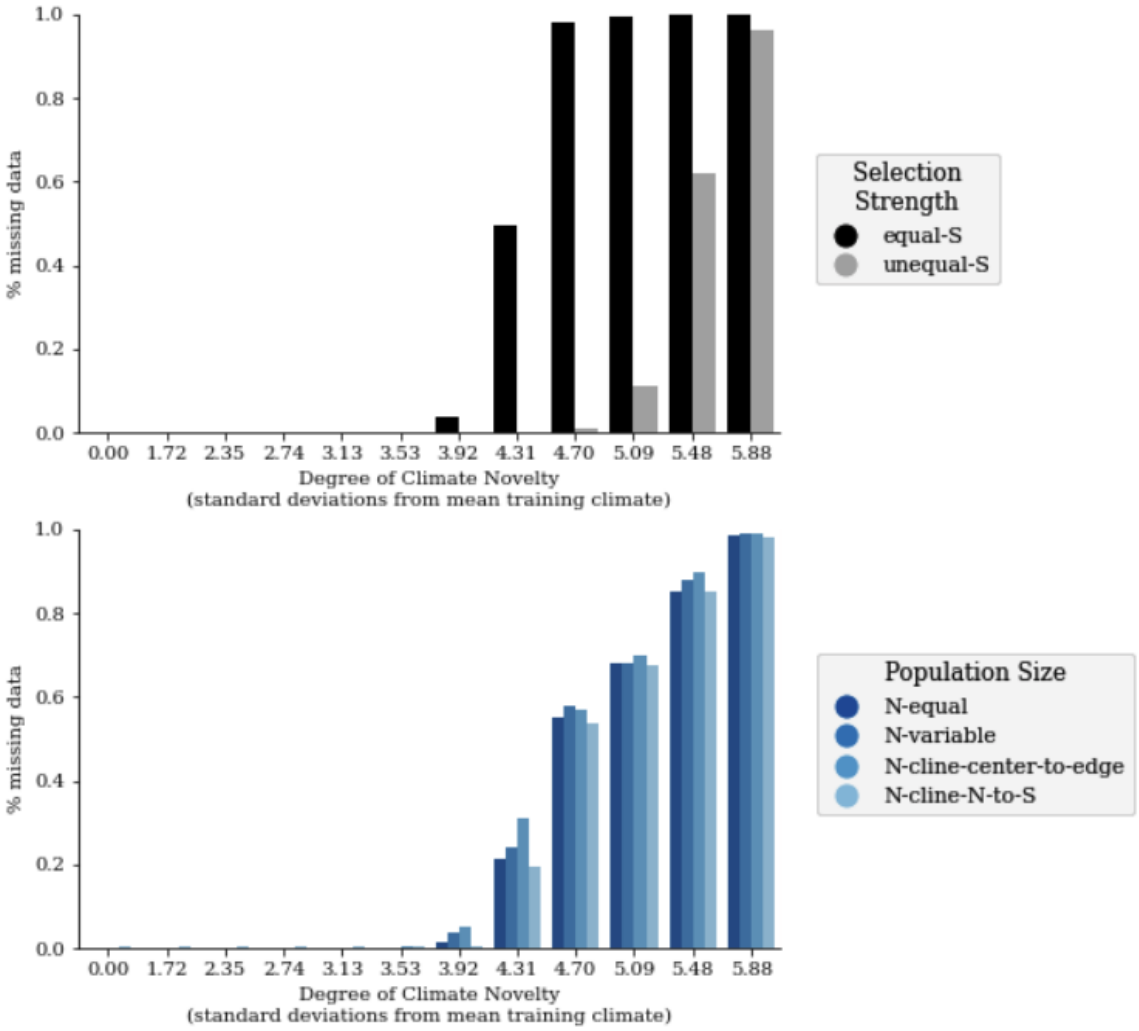
To understand impacts of missing data, we created figures that grouped simulation and experimental levels across novelty scenarios (Fig. S33). We also printed out specific scenarios in the code (SC 02.04.05). Importantly, missing data is not substantial until *Climate Novelty (CN) Scenario 4.31*, which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

(Fig S33)

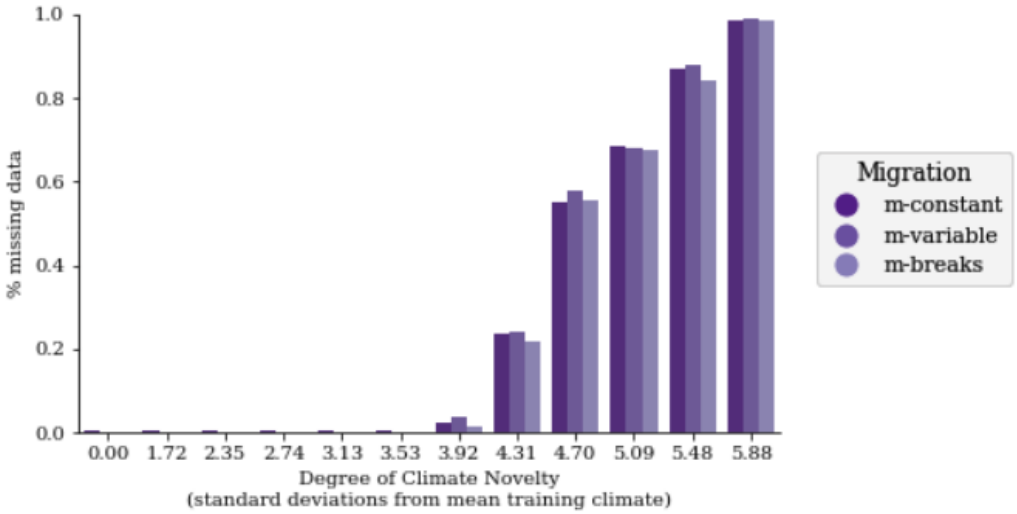

1.0
0.8
0.6
0.4
0.2
0.0
% missing data
0.00 1.72 2.35 2.74 3.13 3.53 3.92 4.31 4.70 5.09 5.48 5.88
Degree of Climate Novelty
(standard deviations from mean training climate)
Polygenicity
oligogenic
moderately polygenic
highly polygenic

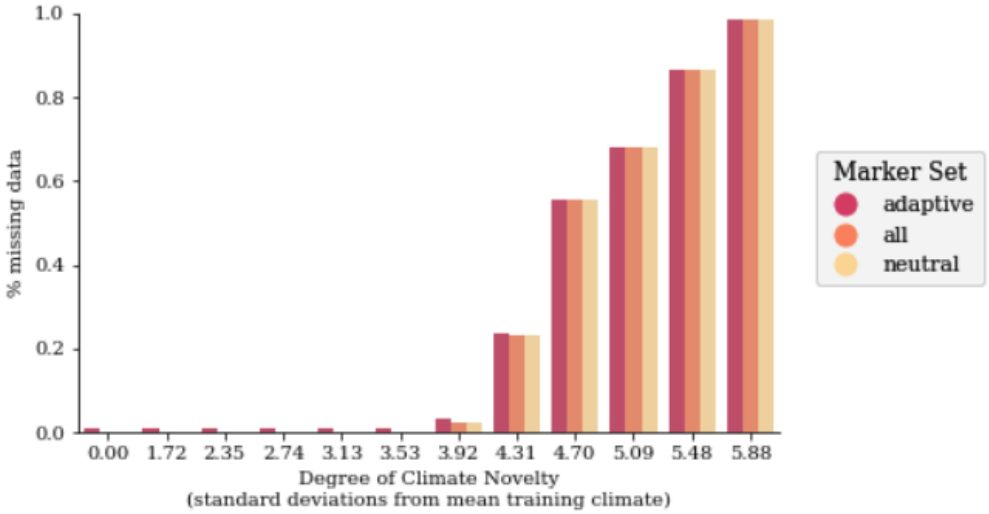

1.0
0.8
0.6
0.4
0.2
0.0
% missing data
0.00 1.72 2.35 2.74 3.13 3.53 3.92 4.31 4.70 5.09 5.48 5.88
Degree of Climate Novelty
(standard deviations from mean training climate)
Pleiotropy
no pleiotropy
pleiotropy

(Fig S33 continued)

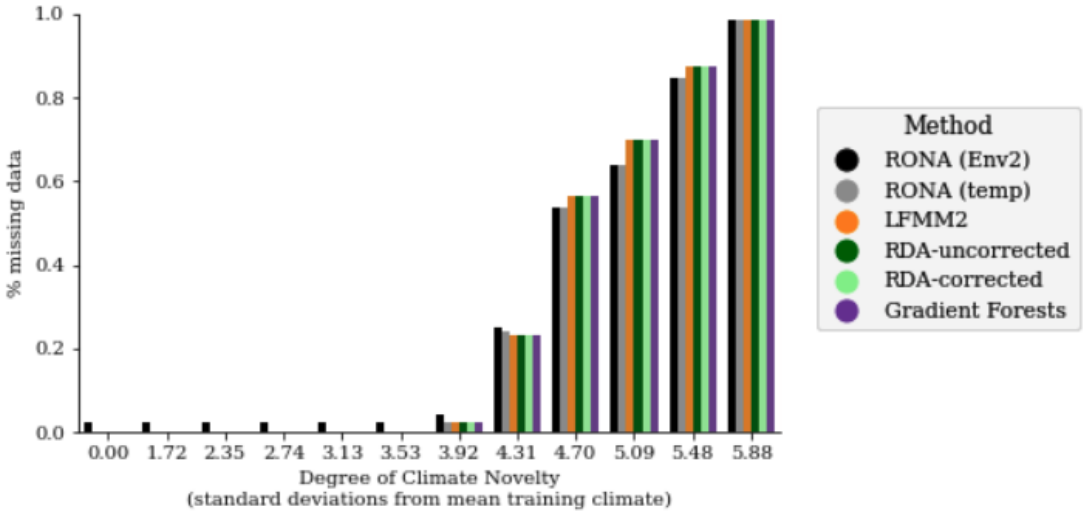


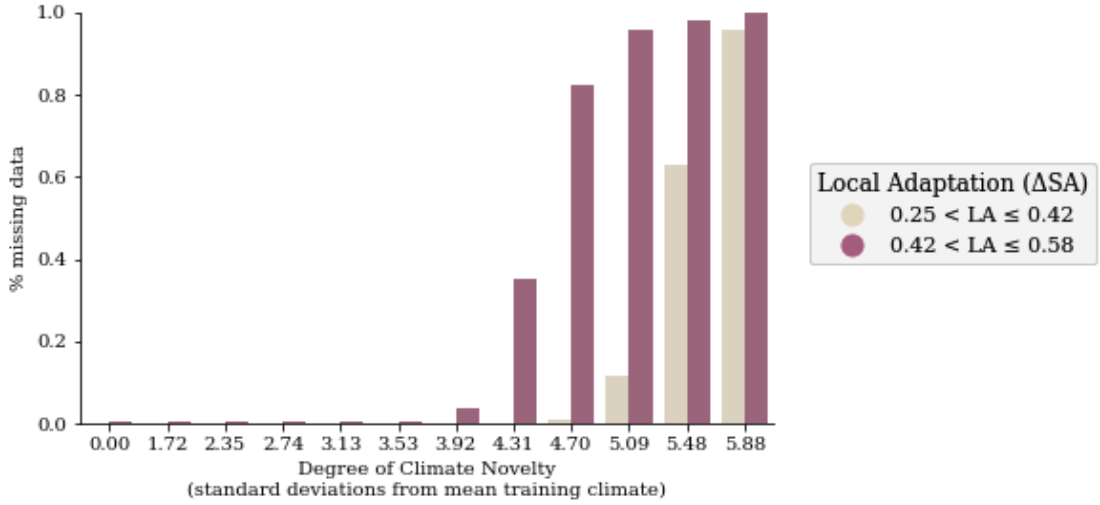
(Fig S33 continued)





(Fig S33 continued)





(Fig S33 continued)

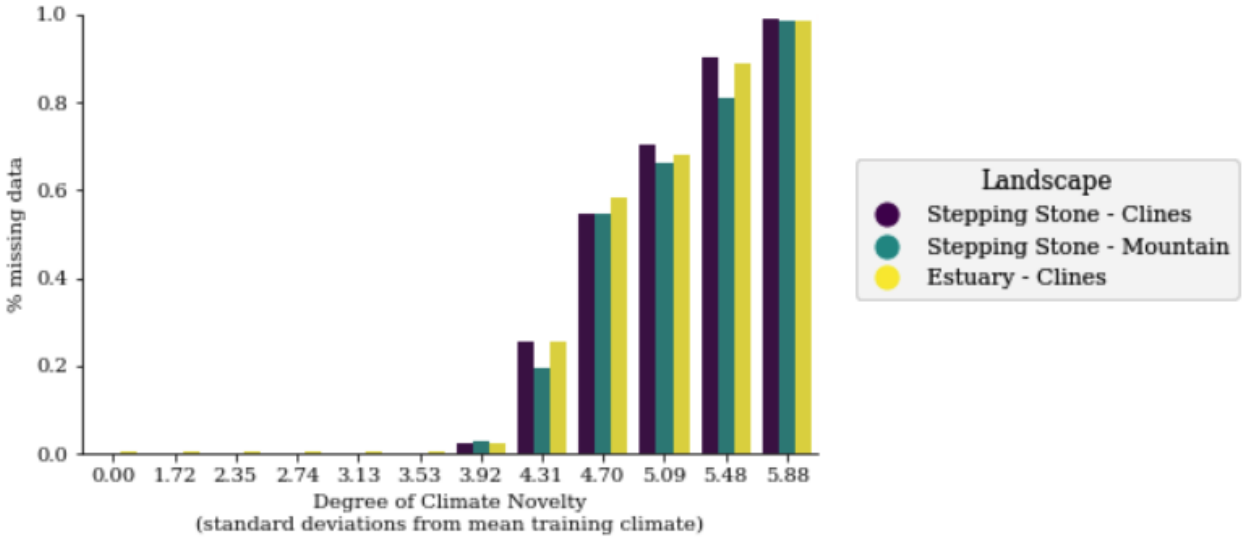


Fig S33 The effect of simulation parameters on missing data for *Climate Novelty* scenarios. Shown are the percent missing data (y-axes) due to experimental and simulation parameters (legends). Missing data is when all populations have zero fitness in a given novelty scenario, and thus performance cannot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

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

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

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

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

RONA (Env2)

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	11.076279	1.0	203.959434	3.027995e-46	0.11
$p_{cQTN,\ Env2}$	7.130868	1.0	131.308340	2.171984e-30	0.07
$p_{cNeut,\ temp}$	350.186388	1.0	6448.358576	0.000000e+00	3.40
$p_{cNeut,\ Env2}$	165.769820	1.0	3052.497975	0.000000e+00	1.61
Residual	9774.859470	179995.0	NaN	NaN	94.82

RONA (temp)

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	461.435097	1.0	6854.545092	0.000000e+00	3.54
$p_{cQTN,\ Env2}$	321.669143	1.0	4778.344045	0.000000e+00	2.47
$p_{cNeut,\ temp}$	21.729532	1.0	322.788745	4.136405e-72	0.17
$p_{cNeut,\ Env2}$	119.950781	1.0	1781.849799	0.000000e+00	0.92
Residual	12116.925221	179995.0	NaN	NaN	92.91

RDA-uncorrected

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	326.478563	1.0	8168.479995	0.0	3.29
$p_{cQTN,\ Env2}$	171.117641	1.0	4281.356235	0.0	1.72
$p_{cNeut,\ temp}$	589.314088	1.0	14744.613849	0.0	5.93
$p_{cNeut,\ Env2}$	1649.284886	1.0	41265.038898	0.0	16.61
Residual	7194.056785	179995.0	NaN	NaN	72.45

RDA-corrected

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	36.345449	1.0	970.987128	1.344472e-212	0.48
$p_{cQTN,\ Env2}$	26.680765	1.0	712.790185	1.002191e-156	0.35
$p_{cNeut,\ temp}$	281.267846	1.0	7514.213388	0.000000e+00	3.70
$p_{cNeut,\ Env2}$	512.076667	1.0	13680.388309	0.000000e+00	6.74
Residual	6737.472479	179995.0	NaN	NaN	88.72

Gradient Forests

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	350.833368	1.0	9184.348121	0.0	3.69
$p_{cQTN,\ Env2}$	210.540691	1.0	5511.673549	0.0	2.21
$p_{cNeut,\ temp}$	375.213154	1.0	9822.578289	0.0	3.94
$p_{cNeut,\ Env2}$	1702.524610	1.0	44569.816059	0.0	17.89
Residual	6875.637914	179995.0	NaN	NaN	72.26

LFMM2

	sum_sq	df	F	PR(>F)	perc_sum_sq
$p_{cQTN,\ temp}$	256.979852	1.0	8357.260709	0.0	3.72
$p_{cQTN,\ Env2}$	173.645491	1.0	5647.137802	0.0	2.51
$p_{cNeut,\ temp}$	718.499204	1.0	23366.365591	0.0	10.40
$p_{cNeut,\ Env2}$	227.994551	1.0	7414.627593	0.0	3.30
Residual	5534.718859	179995.0	NaN	NaN	80.08

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). Code to create these tables can be found in SC 02.02.05.

RONA-sal_opt

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	13.066056	1.0	939.445081	2.543708e-166	34.04
final_LA	0.321152	1.0	23.090730	1.673759e-06	0.84
Residual	24.993162	1797.0	NaN	NaN	65.12

RONA-temp_opt

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	0.086501	1.0	8.324025	0.003959	0.18
final_LA	30.251825	1.0	2911.137332	0.000000	61.72
Residual	18.673983	1797.0	NaN	NaN	38.10

lfmm2

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	3.673454	1.0	873.091702	9.874623e-157	8.64
final_LA	31.290136	1.0	7436.913517	0.000000e+00	73.58
Residual	7.560714	1797.0	NaN	NaN	17.78

GF

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	8.505224	1.0	749.172746	3.638160e-138	10.54
final_LA	51.787680	1.0	4561.657568	0.000000e+00	64.18
Residual	20.401019	1797.0	NaN	NaN	25.28

rda-nocorr

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	10.763364	1.0	827.118396	6.020460e-150	12.54
final_LA	51.668736	1.0	3970.521018	0.000000e+00	60.21
Residual	23.384518	1797.0	NaN	NaN	27.25

rda-structcorr

	sum_sq	df	F	PR(>F)	perc_sum_sq
all	3.192007	1.0	78.914320	1.537690e-18	4.20
final_LA	0.062105	1.0	1.535395	2.154664e-01	0.08
Residual	72.686890	1797.0	NaN	NaN	95.71

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.

Nuisance Level	*Adaptive* models	*All* models	*Neutral* 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

Table S4 Gradient Forests (GF) sometimes incorrectly identifies the environments driving adaptation. Shown are the proportions of simulation levels (N 1-trait = 45 levels; N2-trait = 180 levels; one replicate each) where weighted feature importance output from GF correctly identified the adaptive environments in the top-most ranks. If at least one nuisance environment was ranked above an adaptive environment this was counted as incorrect. Data used to create this table is from the GF models output from the *Nuisance Environment* workflow. Code used to create this table can be found in SC 02.10.02.

Supplemental Figures

Figs. S1-S3 are in Supplemental Note S1.

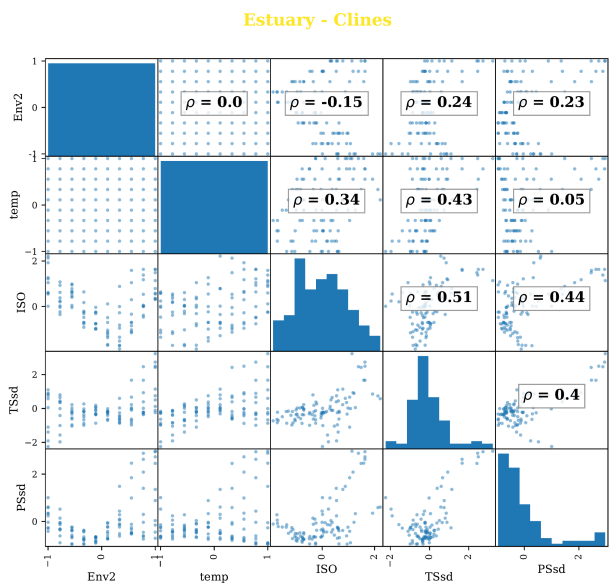


Fig S4 Correlation (Spearman's rho) among environmental variables faceted by landscape. On-diagonal entries are histograms of environmental values, off-diagonal entries are scatter plots between pairwise variables. Included are environmental variables from 1-trait (*temp*), 2-trait (*temp*, *Env2*), 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.

(Fig S4 continued)

6-trait

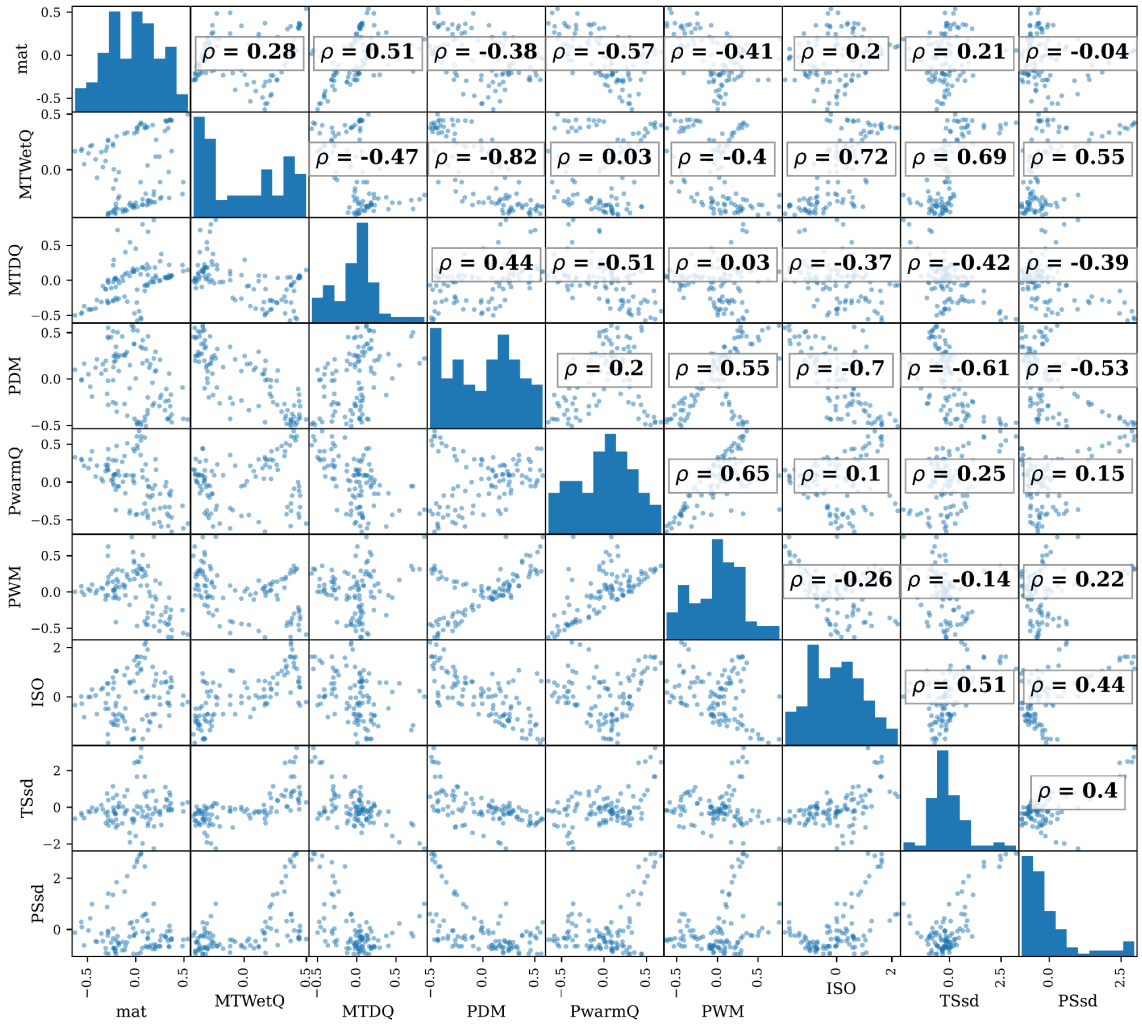


Fig S5 is in Supplemental Note S3

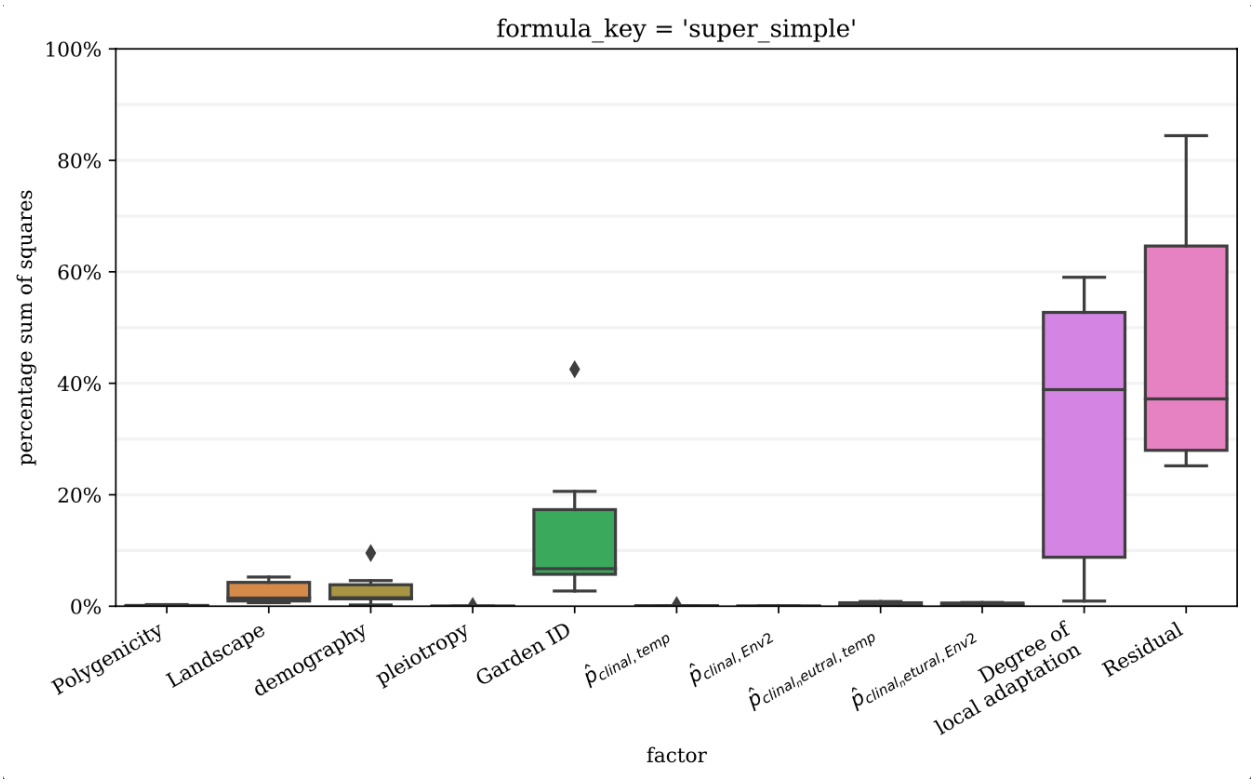


Fig S6 Percent sum of squares of the various factors from the ANOVA model in Table S1. Boxplots are created from the percent sum of squares from each method's individual ANOVA model. Data included in this figure are from models trained using all markers and simulations with two selective environments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

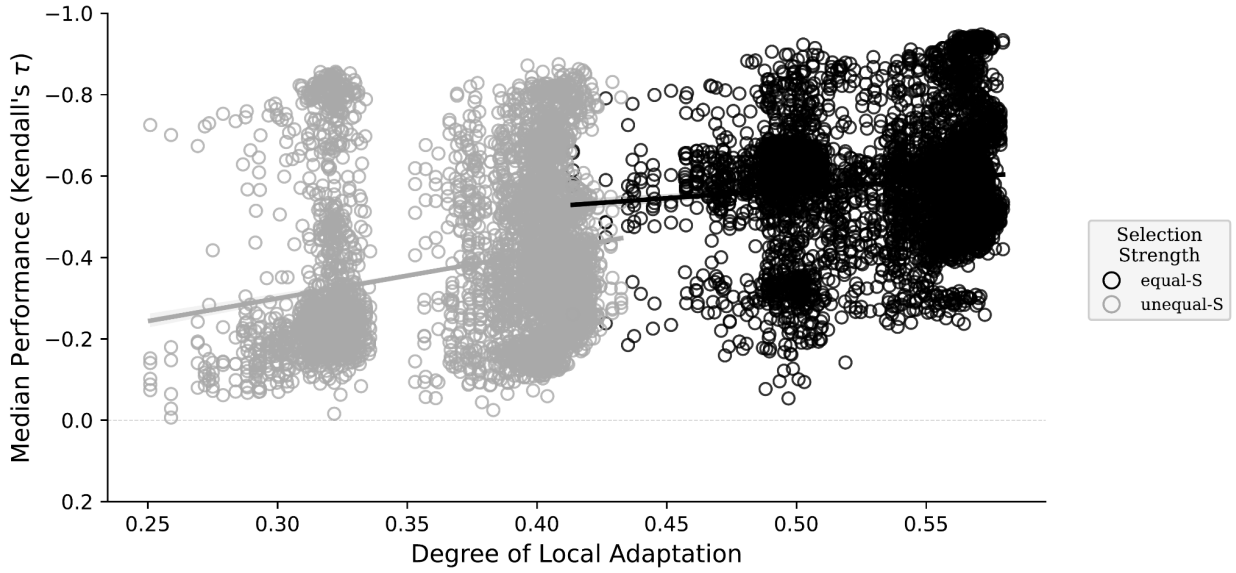


Fig S7 Effect of the degree of local adaptation (x-axes) on method performance (y-axes) colored by the relative strength of selection on the two traits. Shown are the linear relationships between the median validation scores (circles, taken from validation scores across all 100 common gardens on the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

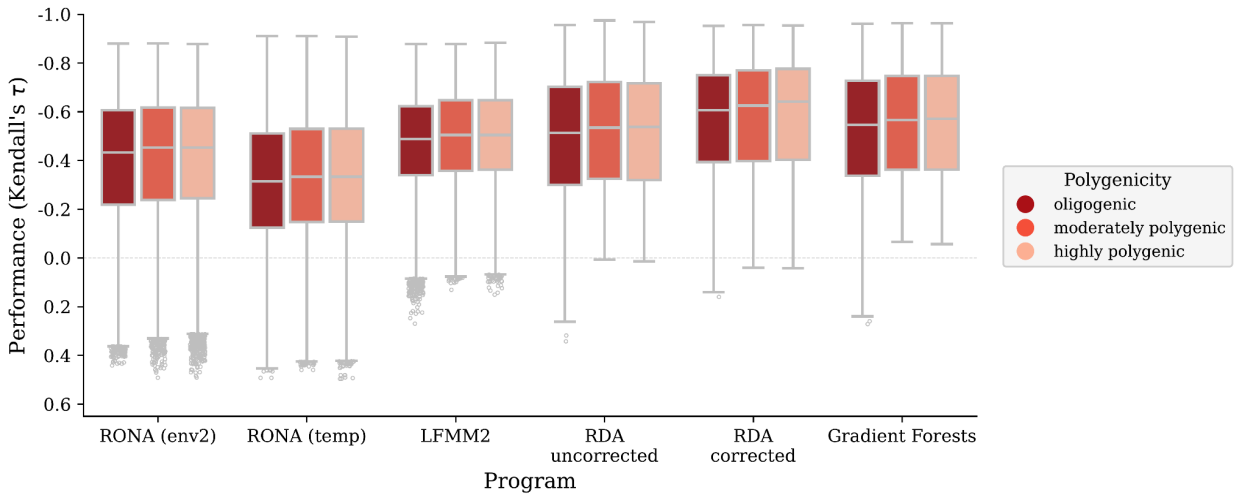


Fig S8 Effect of polygenicity on performance of offset methods trained using all markers on simulations with two adaptive traits. Code to create this figure can be found in SC 02.02.01.

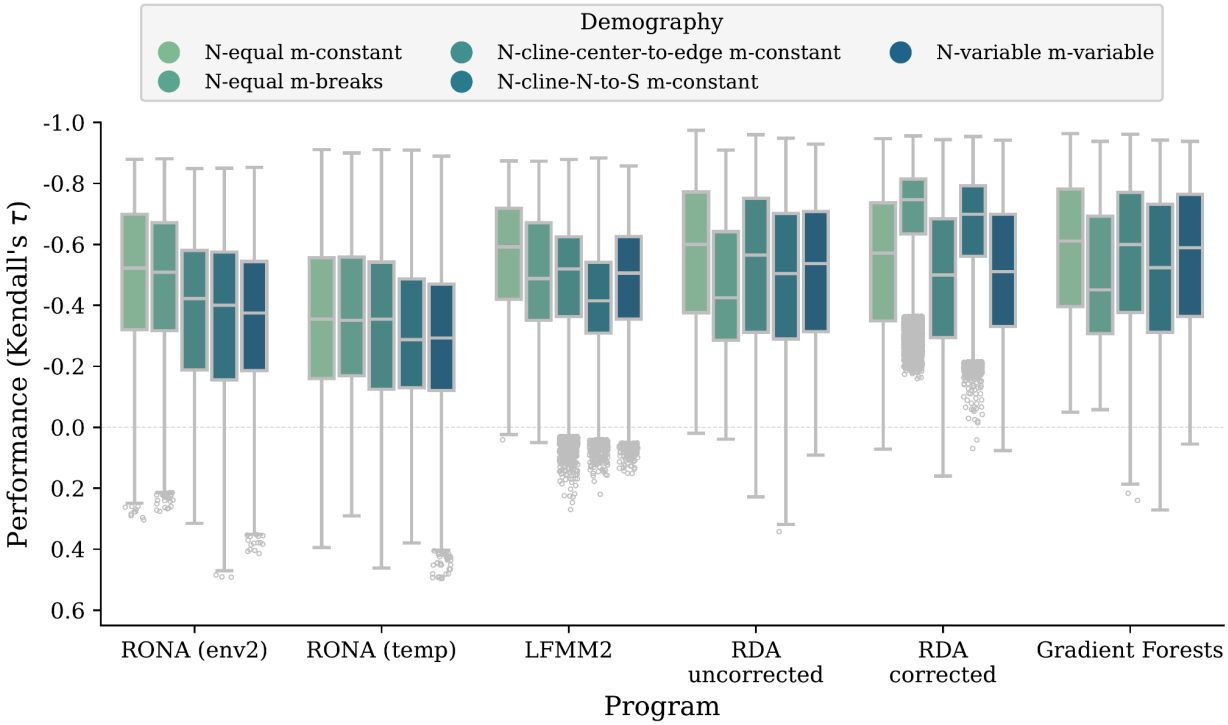


Fig S9 Effect of demography on performance of offset methods trained using all markers on simulations with two adaptive traits. Code to create this figure can be found in SC 02.02.01.

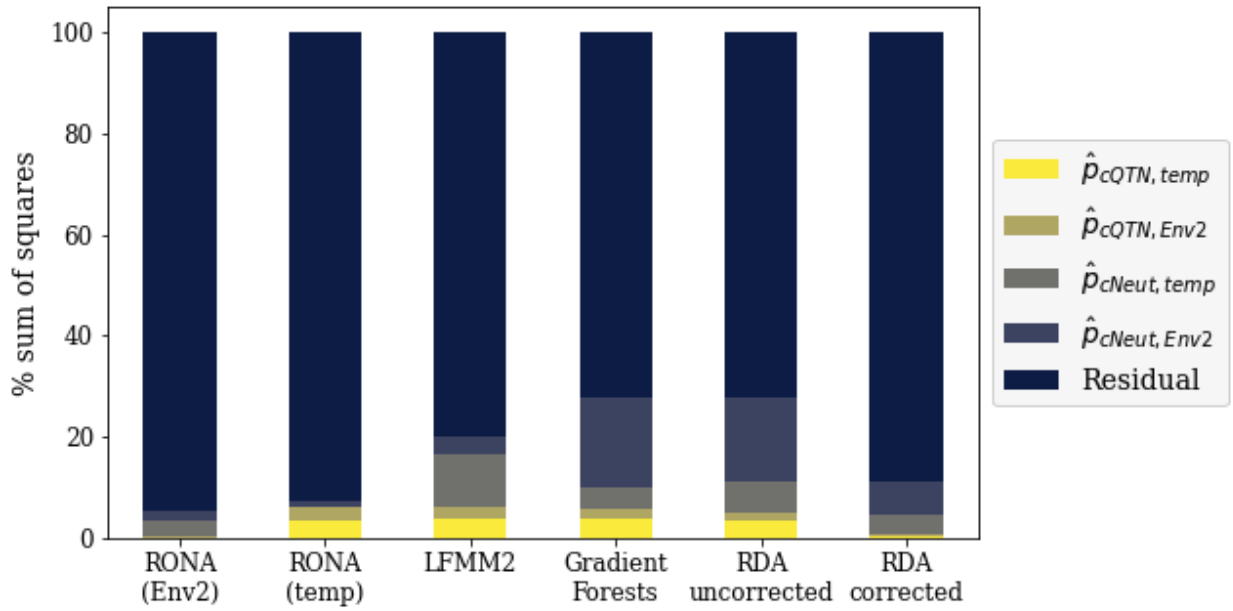
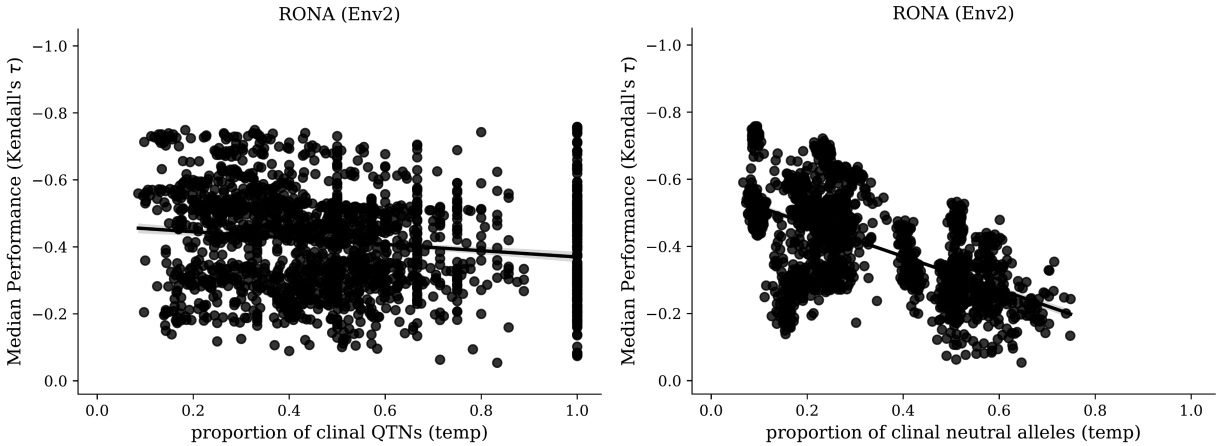
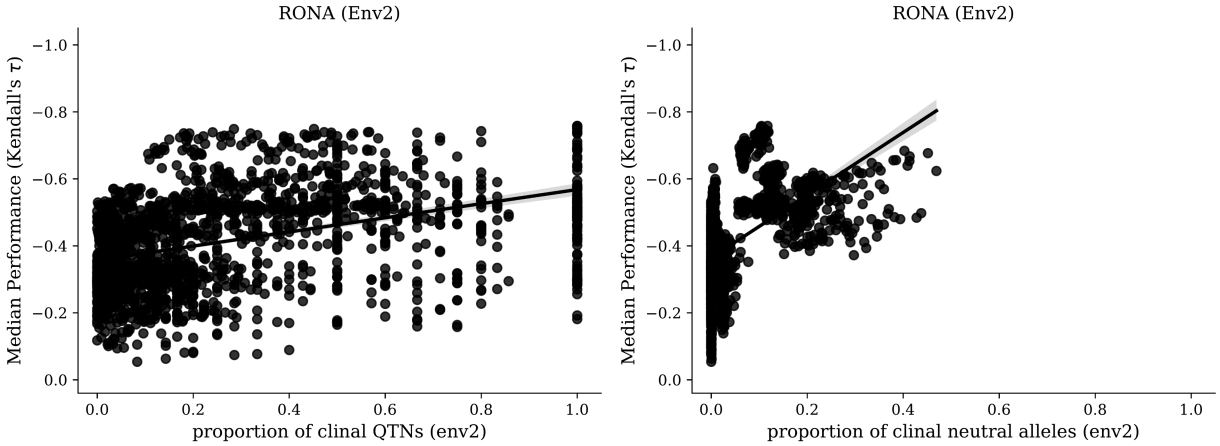


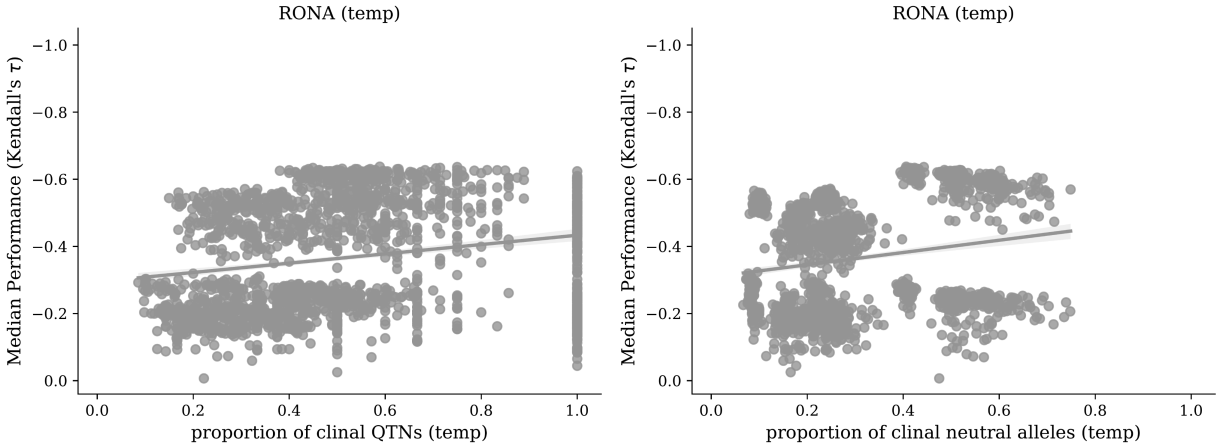
Fig S10 Stacked bar plot of the percent sum of squares from Type II ANOVAs from regressing the proportion of clinal QTNs and clinal neutral alleles on offset performance (see Equation 2 of the main text). Code to create this table is in 02.02.05.


RONA (Env2)
Median Performance (Kendall's τ)
proportion of clinal QTNs (temp)
RONA (Env2)
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (temp)

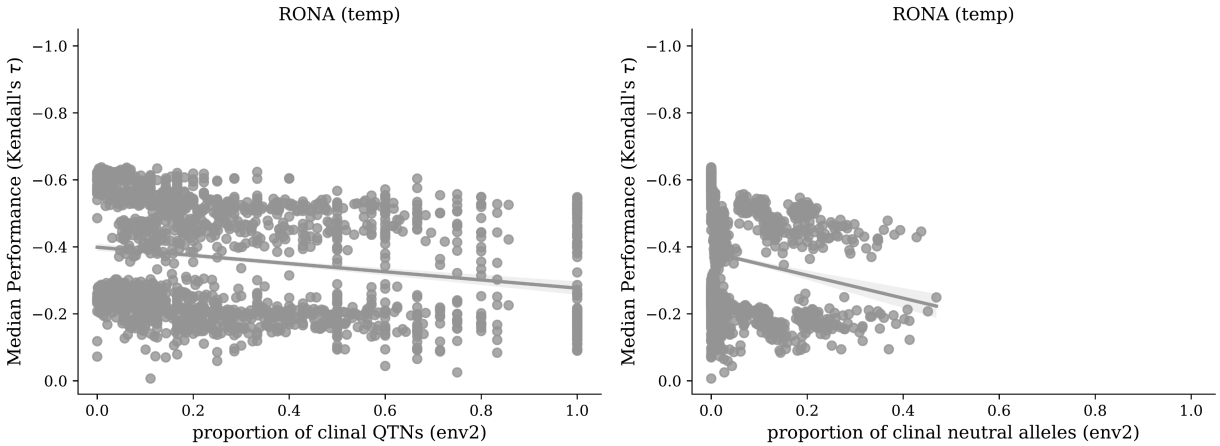
517


RONA (Env2)
Median Performance (Kendall's τ)
proportion of clinal QTNs (env2)
RONA (Env2)
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (env2)

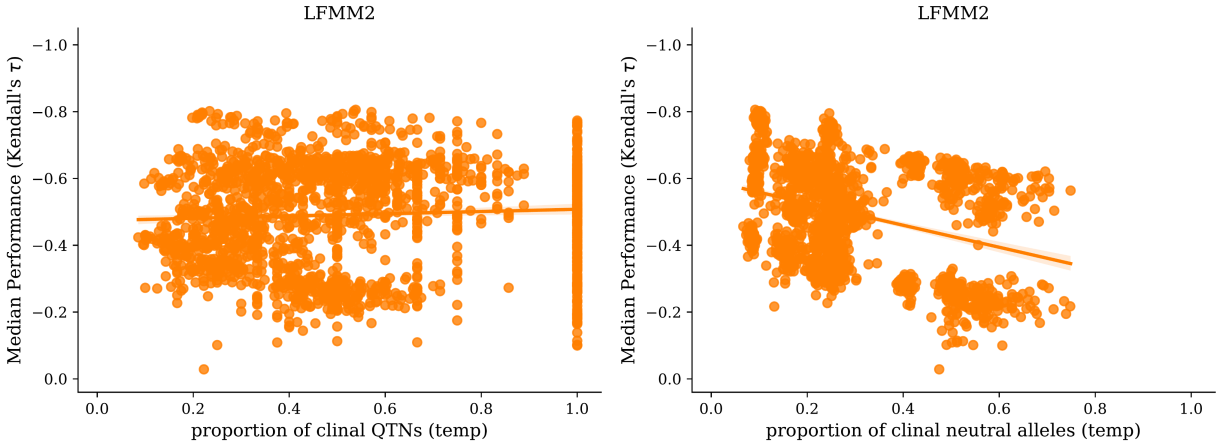
518


RONA (temp)
Median Performance (Kendall's τ)
proportion of clinal QTNs (temp)
RONA (temp)
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (temp)

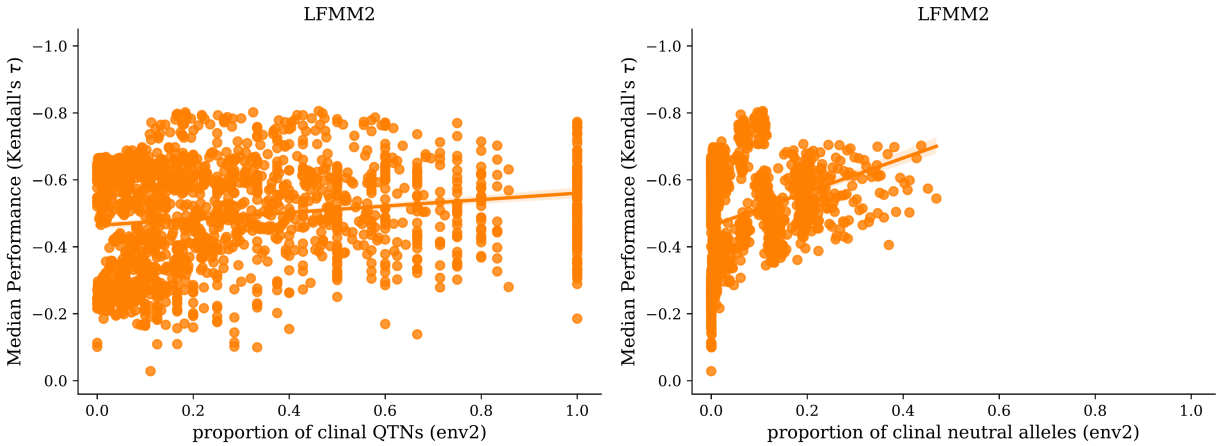
519


RONA (temp)
Median Performance (Kendall's τ)
proportion of clinal QTNs (env2)
RONA (temp)
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (env2)

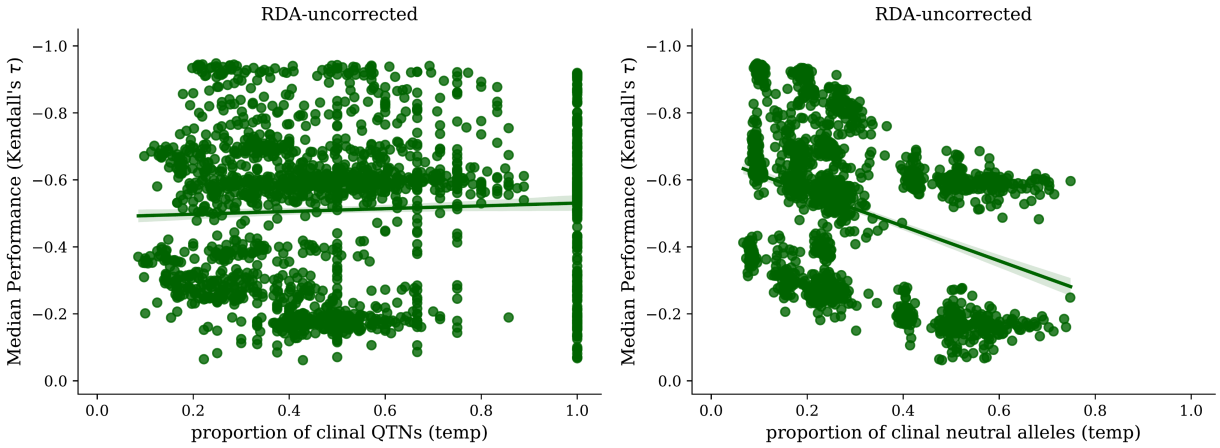
520


LFMM2
Median Performance (Kendall's τ)
proportion of clinal QTNs (temp)
LFMM2
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (temp)

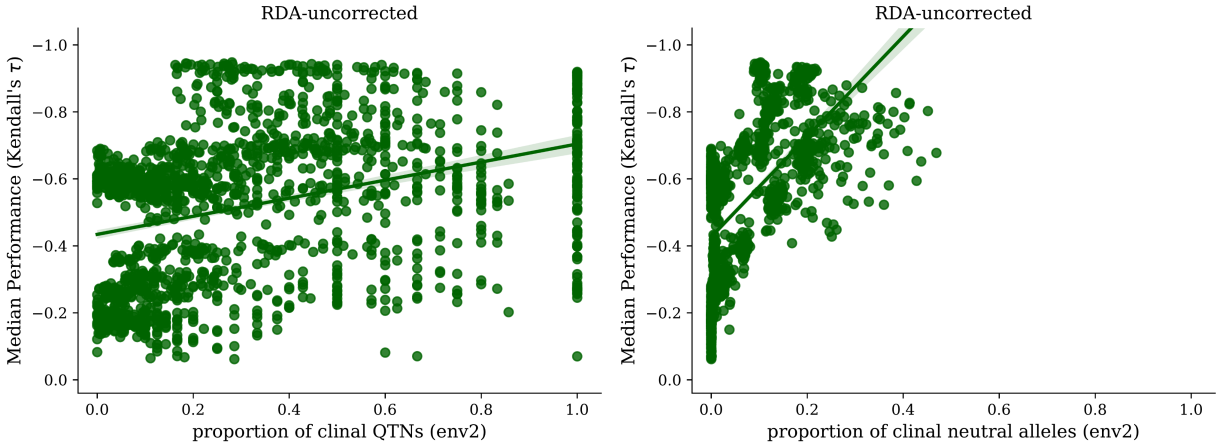
521


LFMM2
Median Performance (Kendall's τ)
proportion of clinal QTNs (env2)
LFMM2
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (env2)

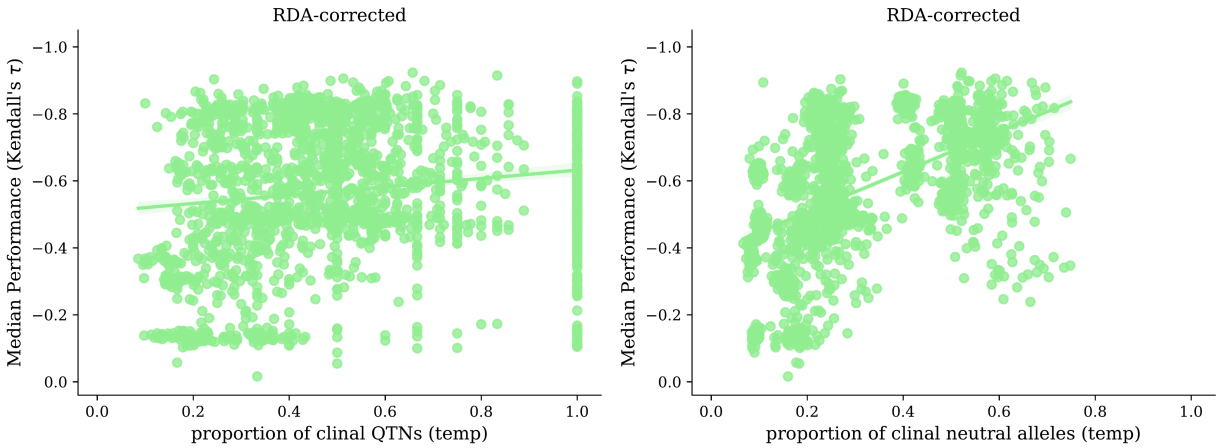
522


RDA-uncorrected
Median Performance (Kendall's τ)
proportion of clinal QTNs (temp)
RDA-uncorrected
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (temp)

523


RDA-uncorrected
Median Performance (Kendall's τ)
proportion of clinal QTNs (env2)
RDA-uncorrected
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (env2)

524


RDA-corrected
Median Performance (Kendall's τ)
proportion of clinal QTNs (temp)
RDA-corrected
Median Performance (Kendall's τ)
proportion of clinal neutral alleles (temp)

525

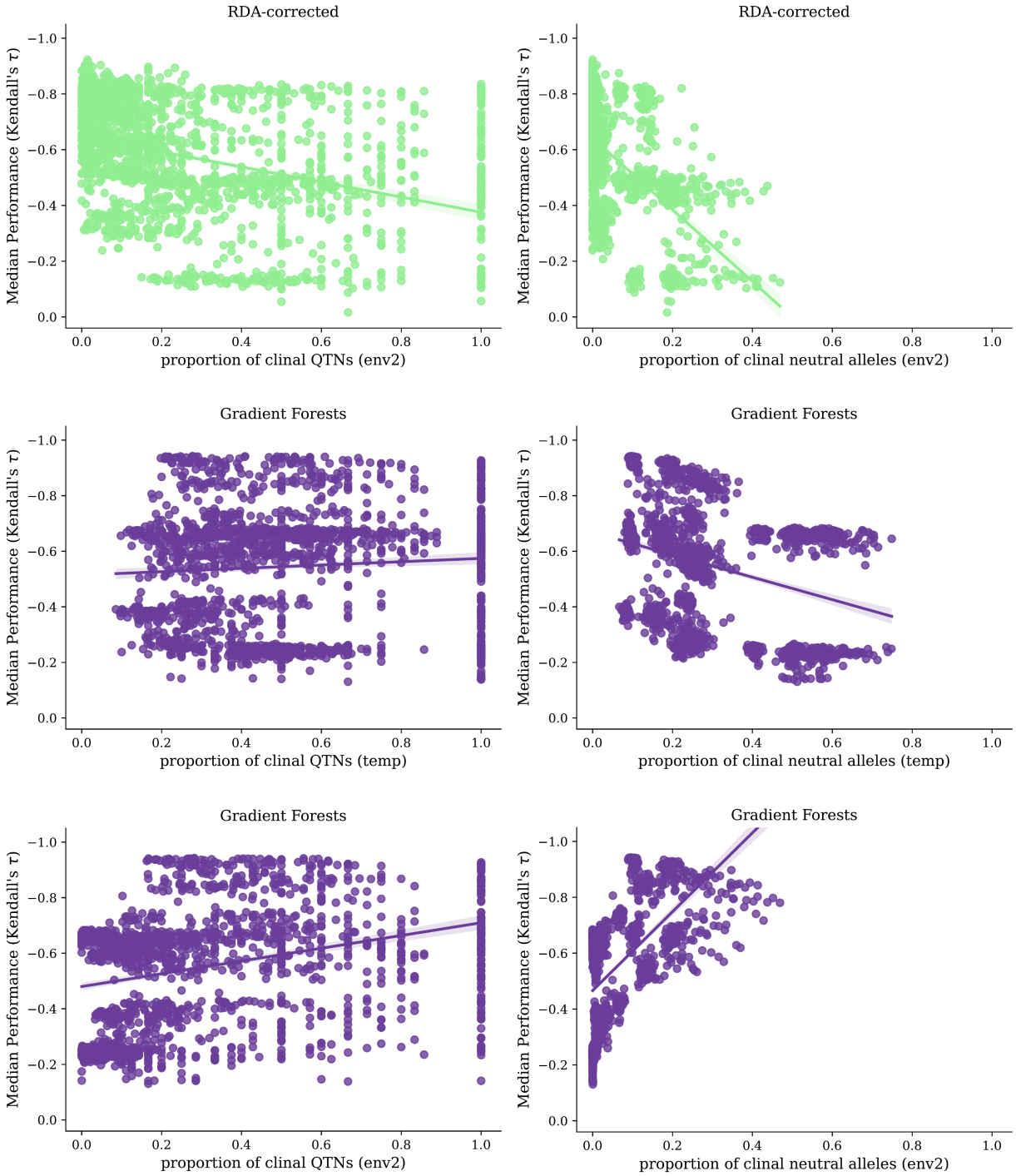


Fig S11 Impact on method performance (y-axes) from the proportion of QTNs with clinal relationships with *temp* (first column) or *Env2* (second column). Model performance is quantified as Kendall's rank correlation between offset and fitness; shown are median values from scores from 10 replicates per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

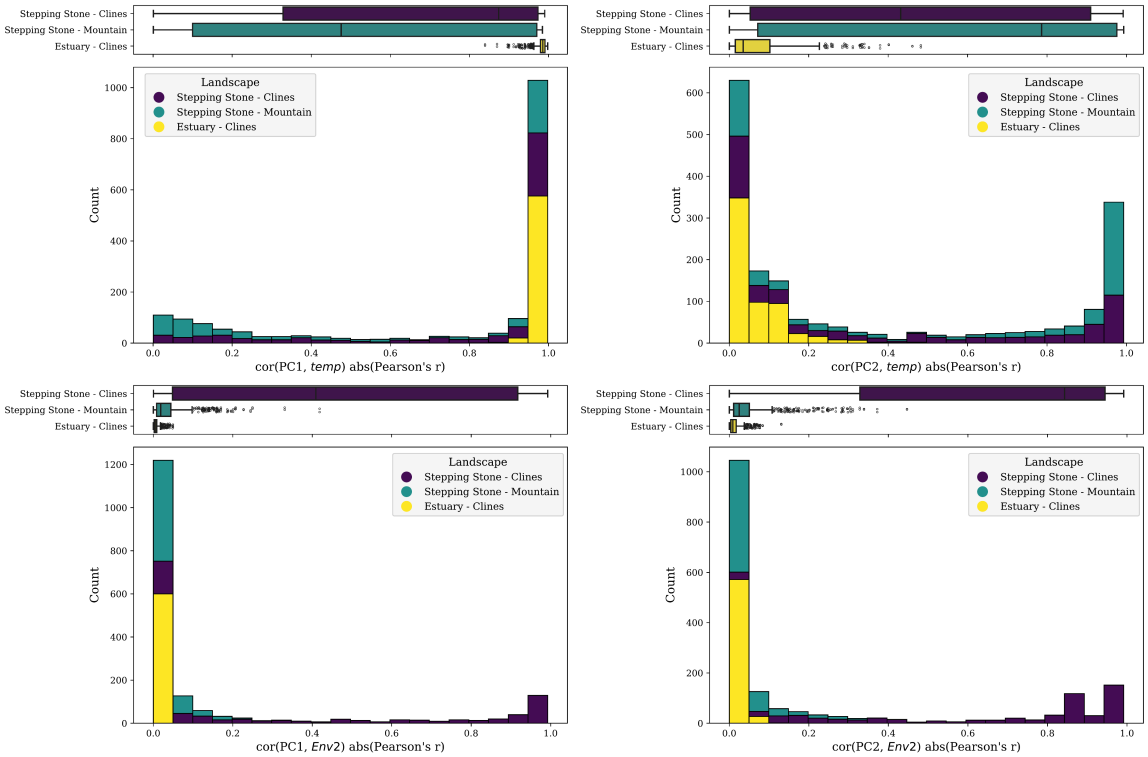


Fig S12 Stacked bar plot showing correlation between environmental variables (rows) and axes of population genetic structure (Principal Component Analysis axes [PC axes]; columns). Data included in this figure is from all 2-trait simulations. Code to create this figure can be found in SC 02.10.03.

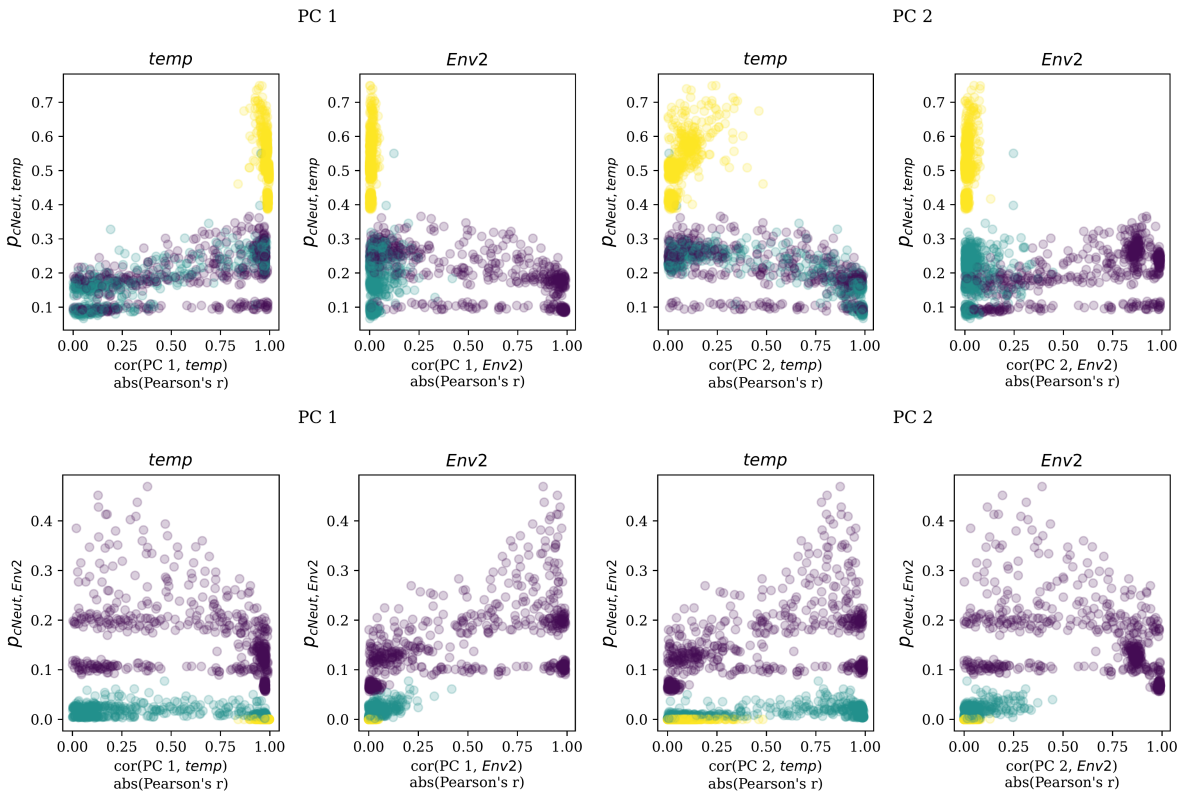
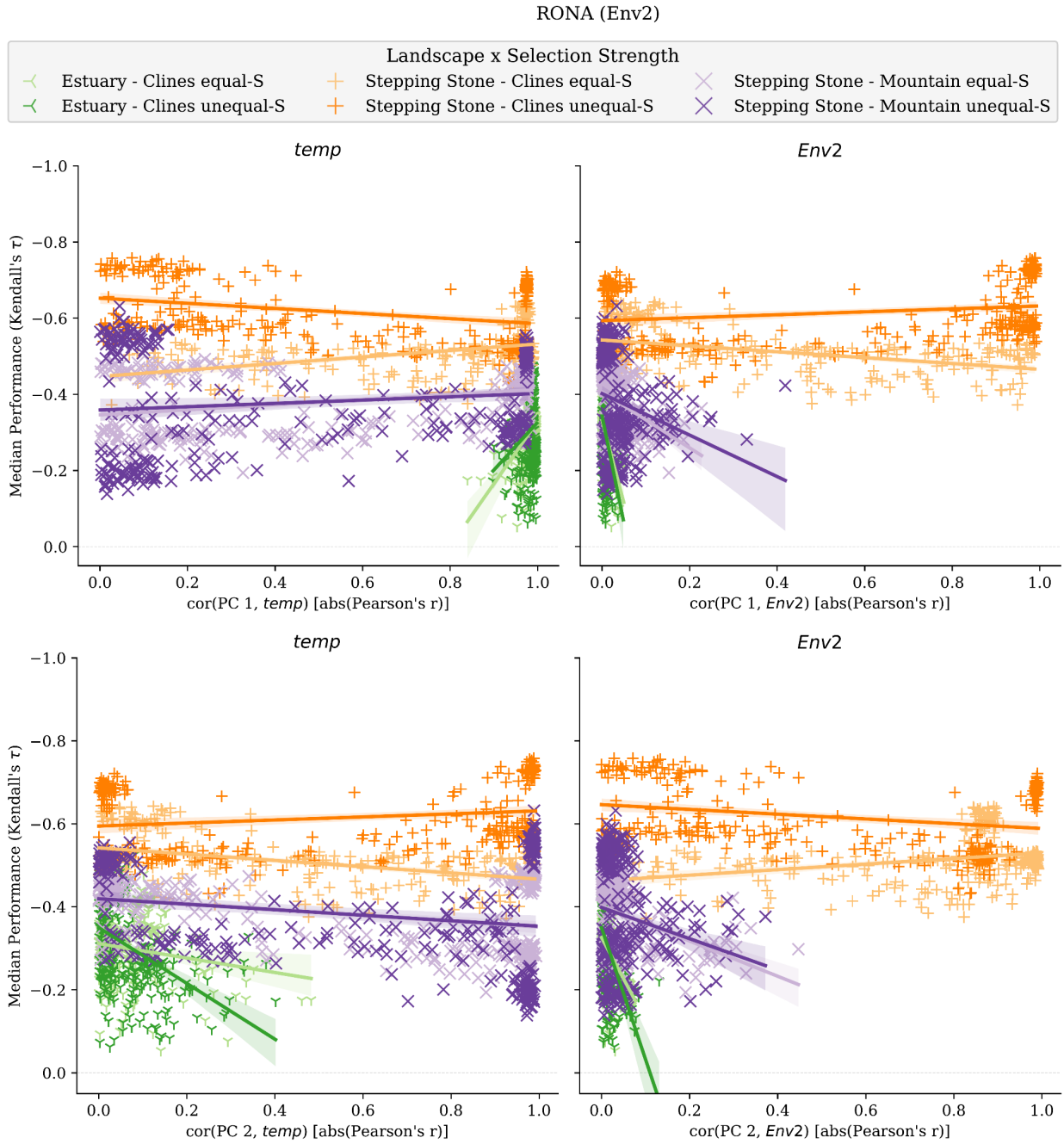
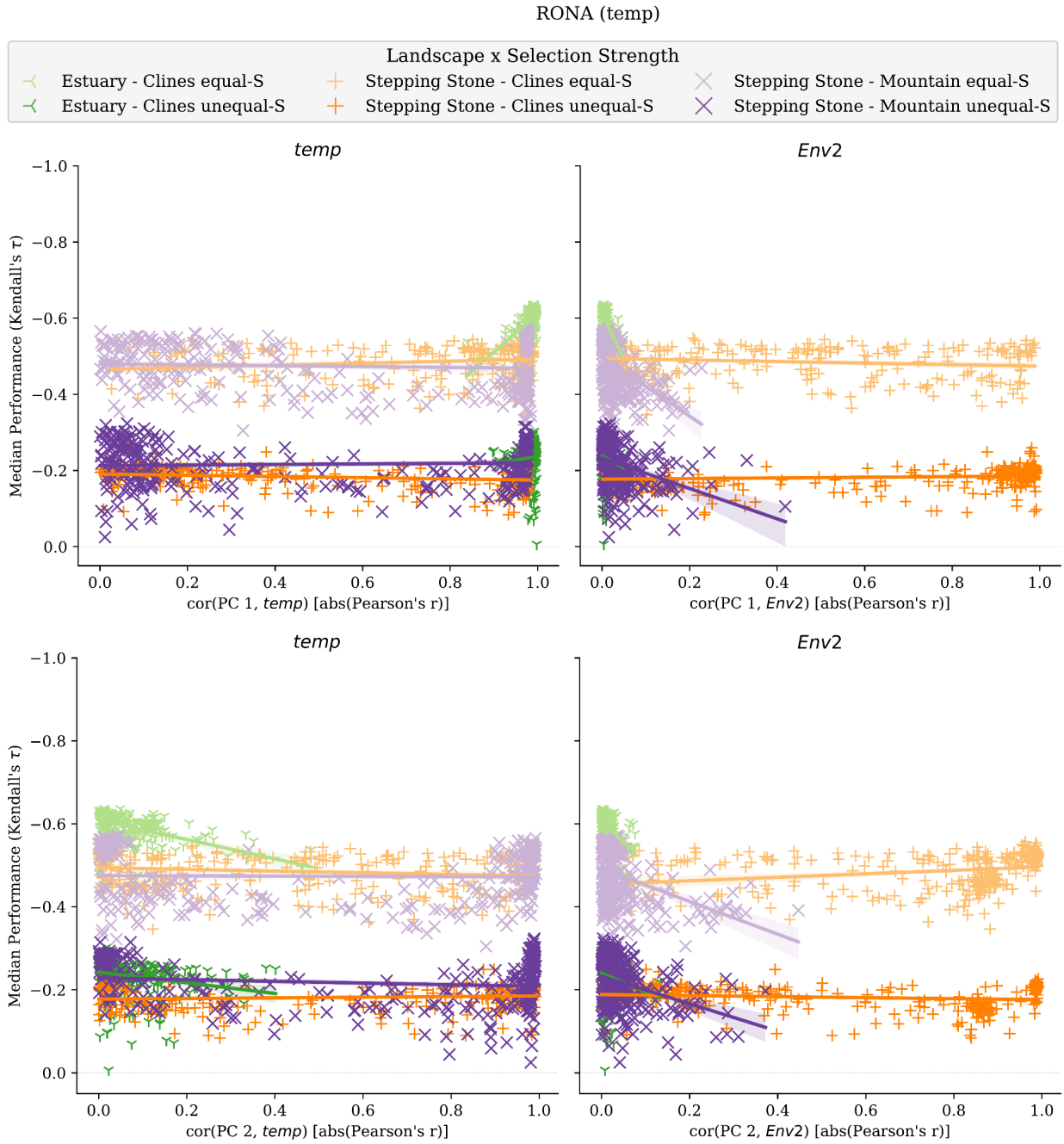


Fig S13 Relationship between the proportion of clinal neutral loci for *temp* (y-axes, first row) or *Env2* (y-axes, second row) with the strength of the relationship between environmental variables and axes of population genetic structure. Purple = *Stepping Stone - Clines*; teal = *Stepping Stone -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

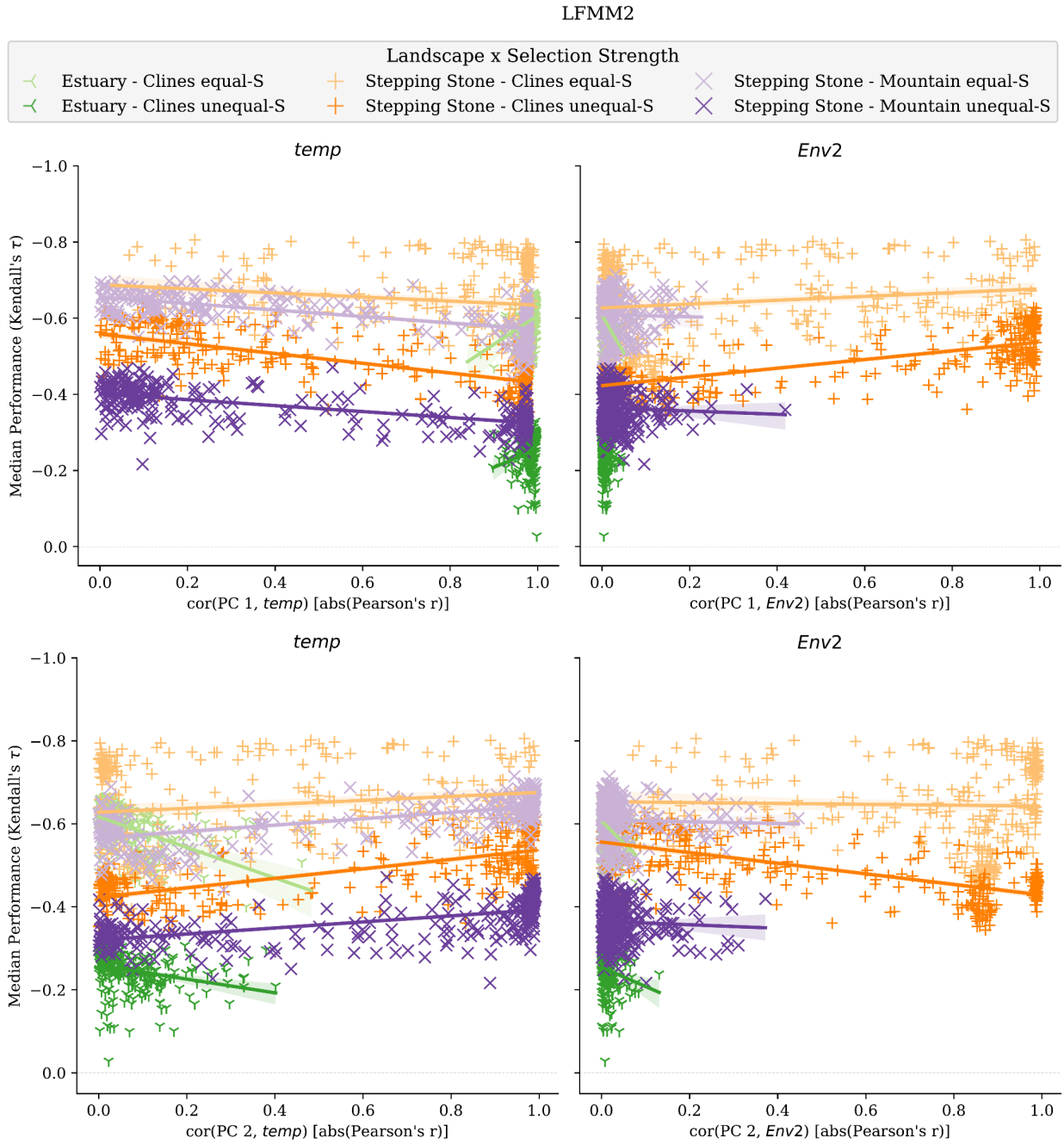
(Fig S14)


RONA (Env2)
Landscape x Selection Strength
Estuary - Clines equal-S
Estuary - Clines unequal-S
Stepping Stone - Clines equal-S
Stepping Stone - Clines unequal-S
Stepping Stone - Mountain equal-S
Stepping Stone - Mountain unequal-S
temp
Env2
Median Performance (Kendall's τ)
cor(PC 1, temp) [abs(Pearson's r)]
cor(PC 1, Env2) [abs(Pearson's r)]
cor(PC 2, temp) [abs(Pearson's r)]
cor(PC 2, Env2) [abs(Pearson's r)]

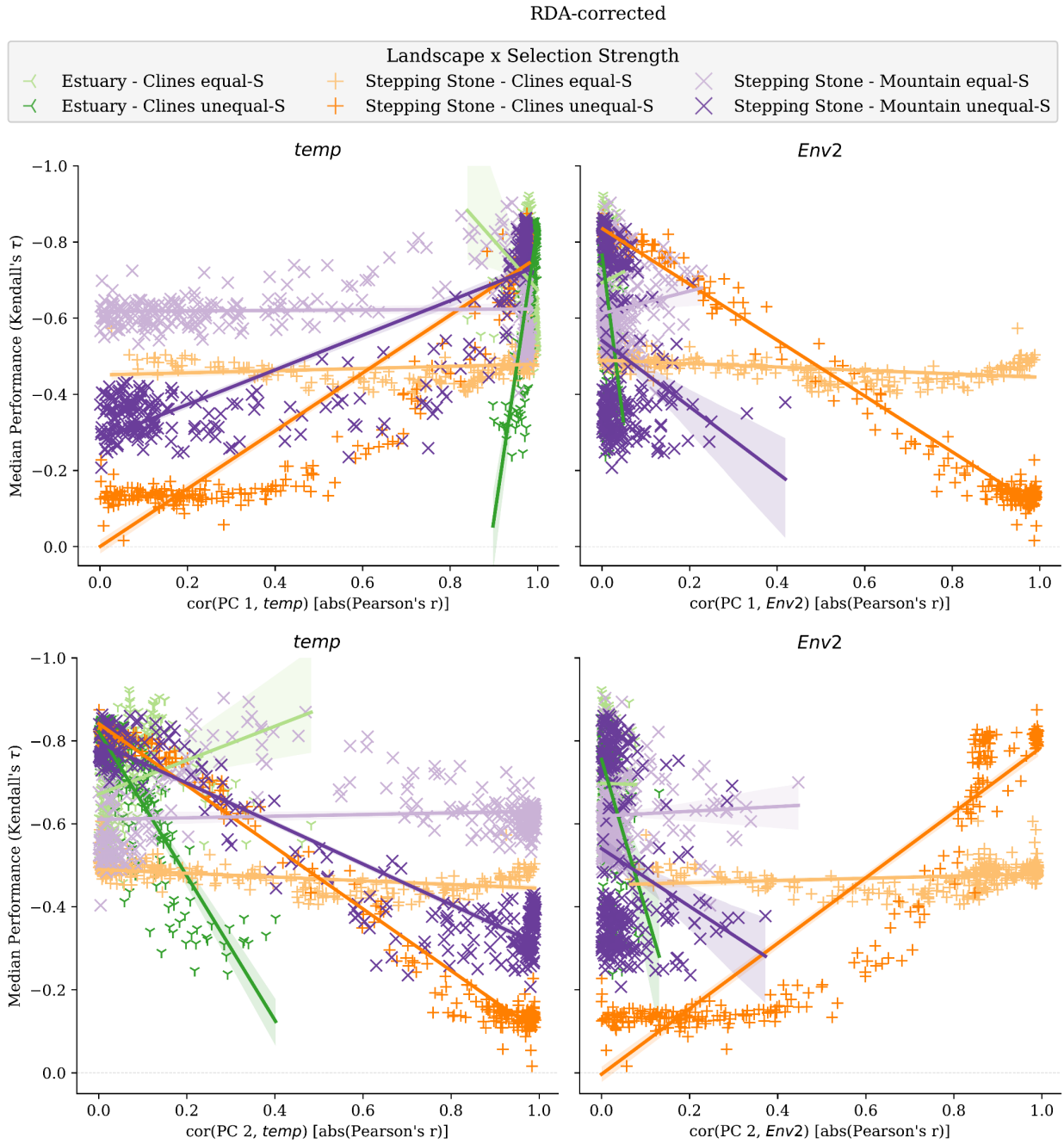
(Fig S14 continued)



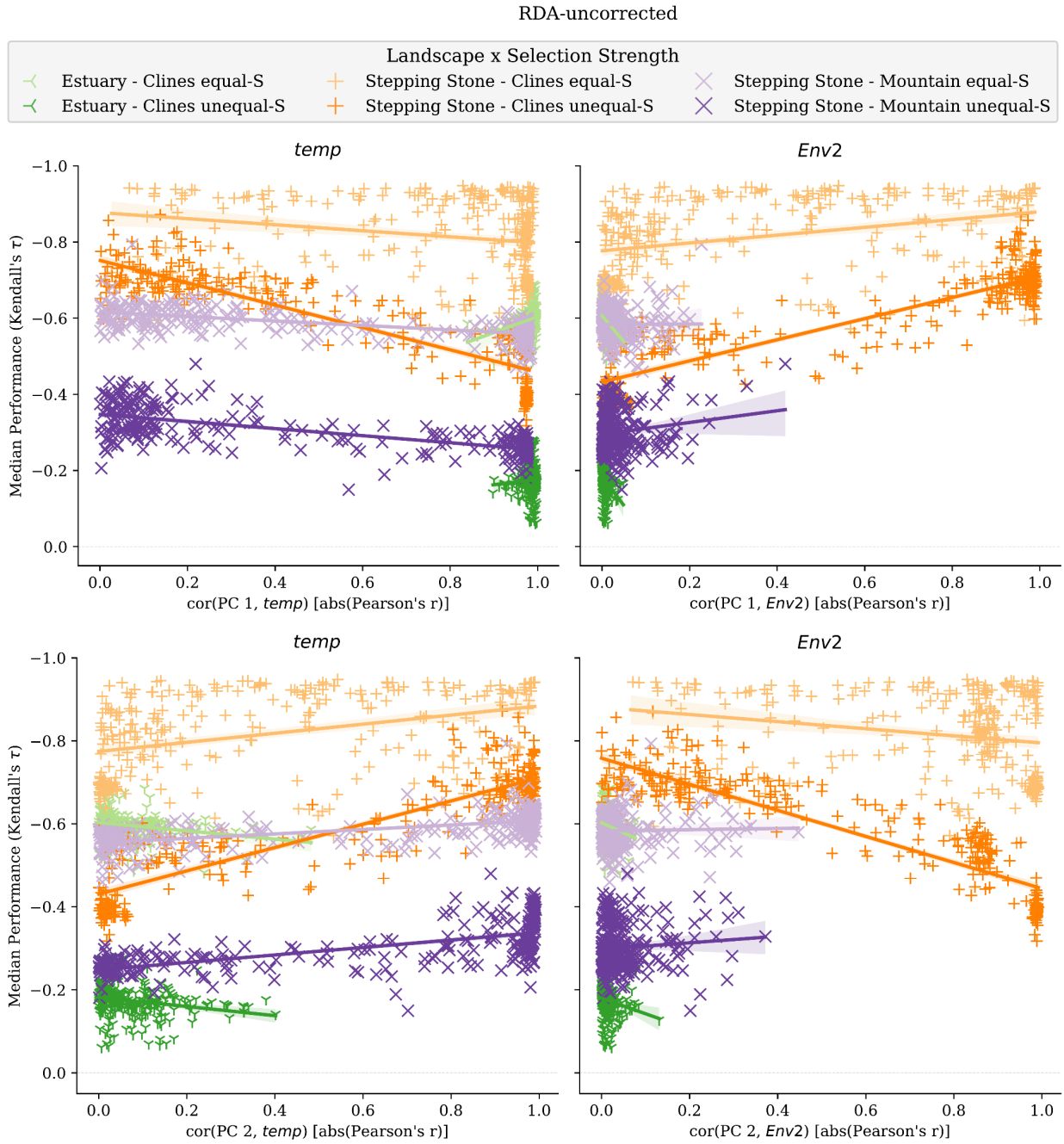
(Fig S14 continued)



(Fig S14 continued)



(Fig S14 continued)



(Fig S14 continued)

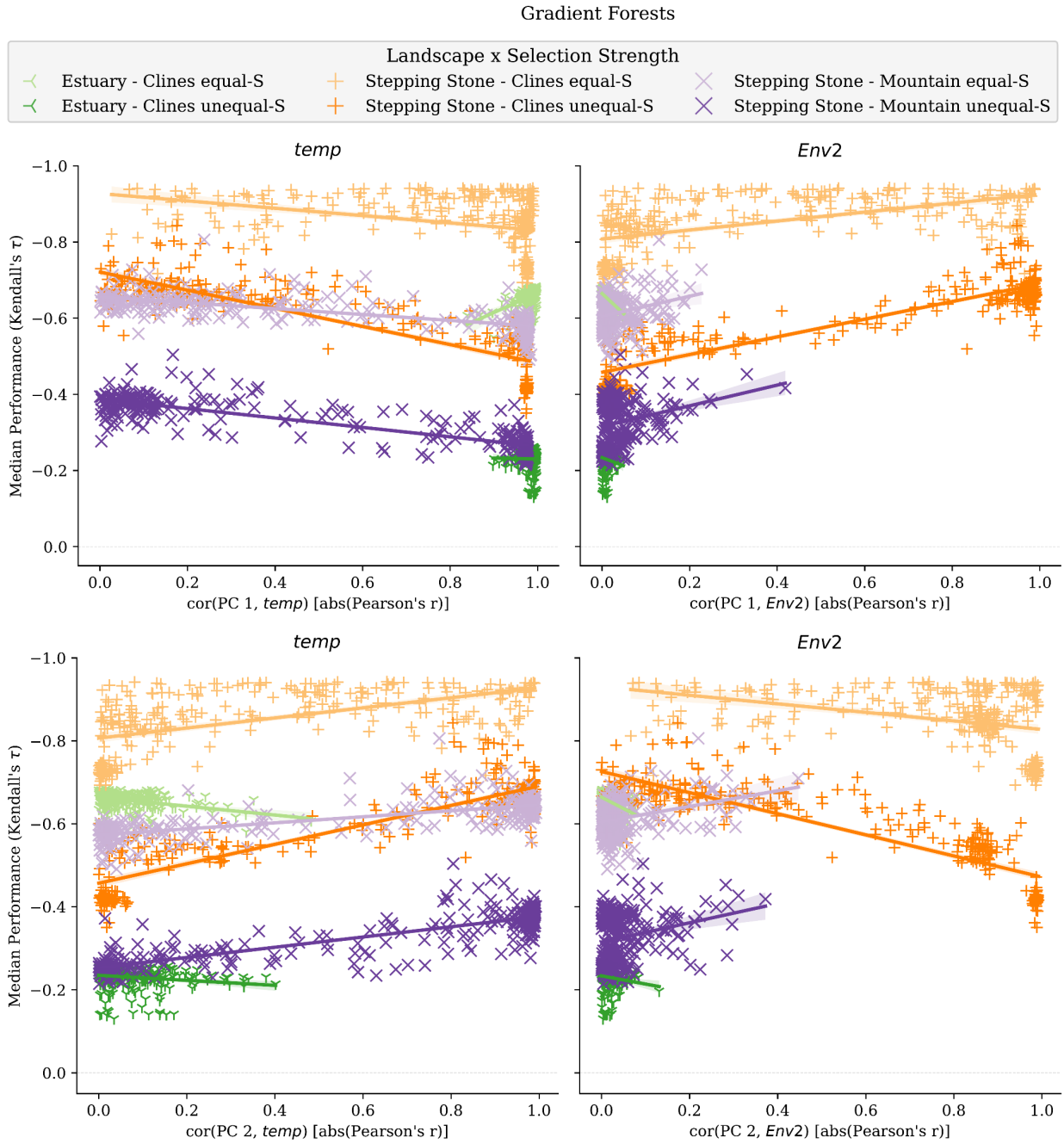


Fig S14 Relationship between median performance and absolute correlation (Pearson's r) between environmental variables and axes of population genetic structure (principal component analysis axes). Each subfigure is for a different method (see panel titles). Data used in this figure is from 2-trait simulations.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

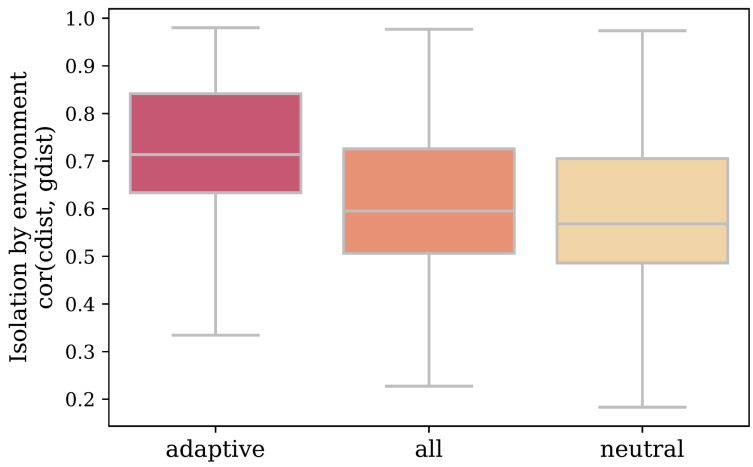
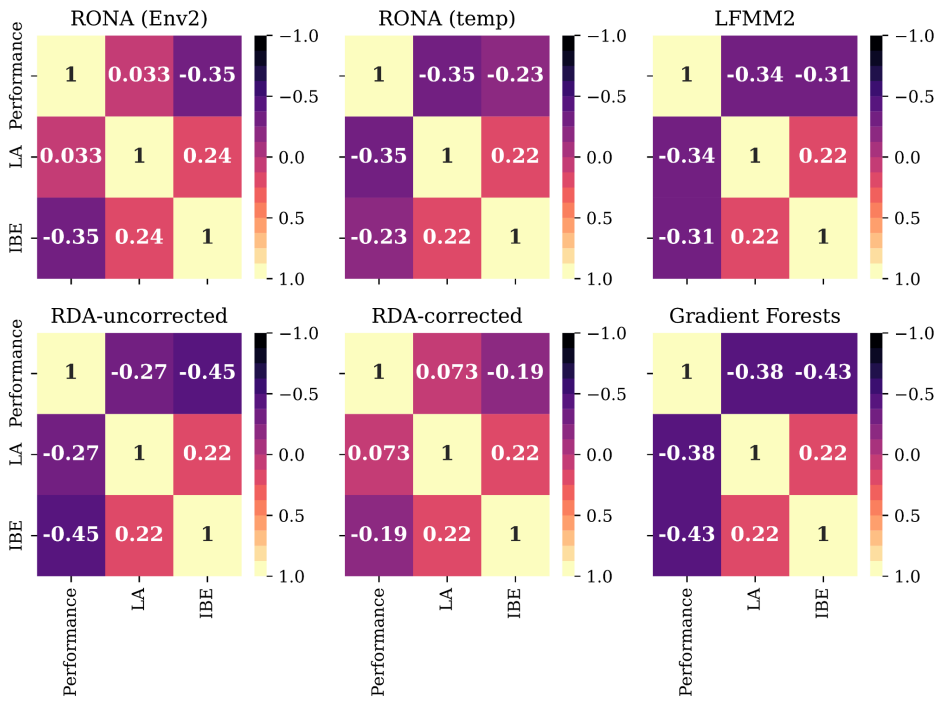


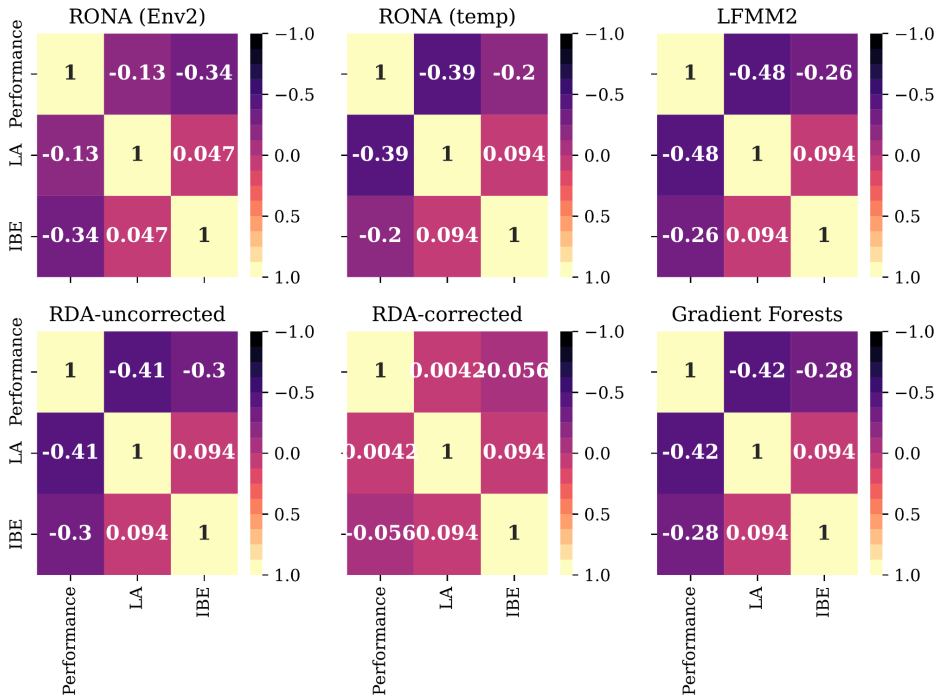
Fig S15 *Adaptive* markers contain greater levels of isolation-by-environment (IBE) than other marker sets. *IBE* is quantified as Spearman's rank correlation between population pairwise F_{ST} and Euclidean distance of adaptive environments. Code to create this figure can be found in SC 02.02.10.

(Fig S16)

A) *adaptive* markers



B) all markers



C) neutral markers

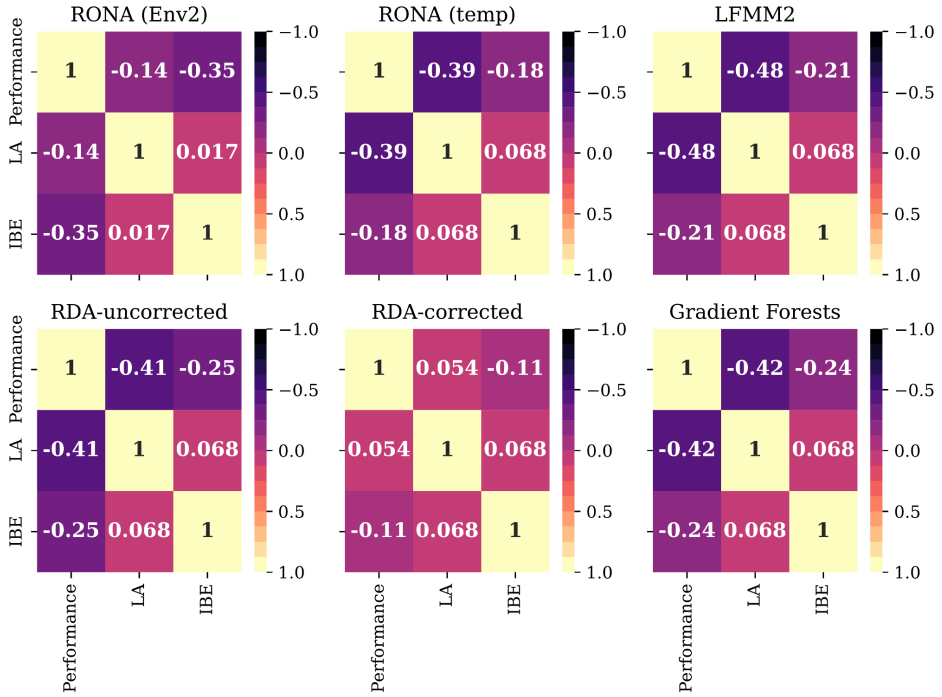


Fig S16 The relationship between the degree of local adaptation ($LA_{\Delta SA}$), levels of *IBE* within marker sets, and median performance of models trained with one of the three marker sets: (A) *adaptive*, (B) *all*, and (c) *neutral* marker sets. *IBE* is quantified as Spearman's rank correlation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

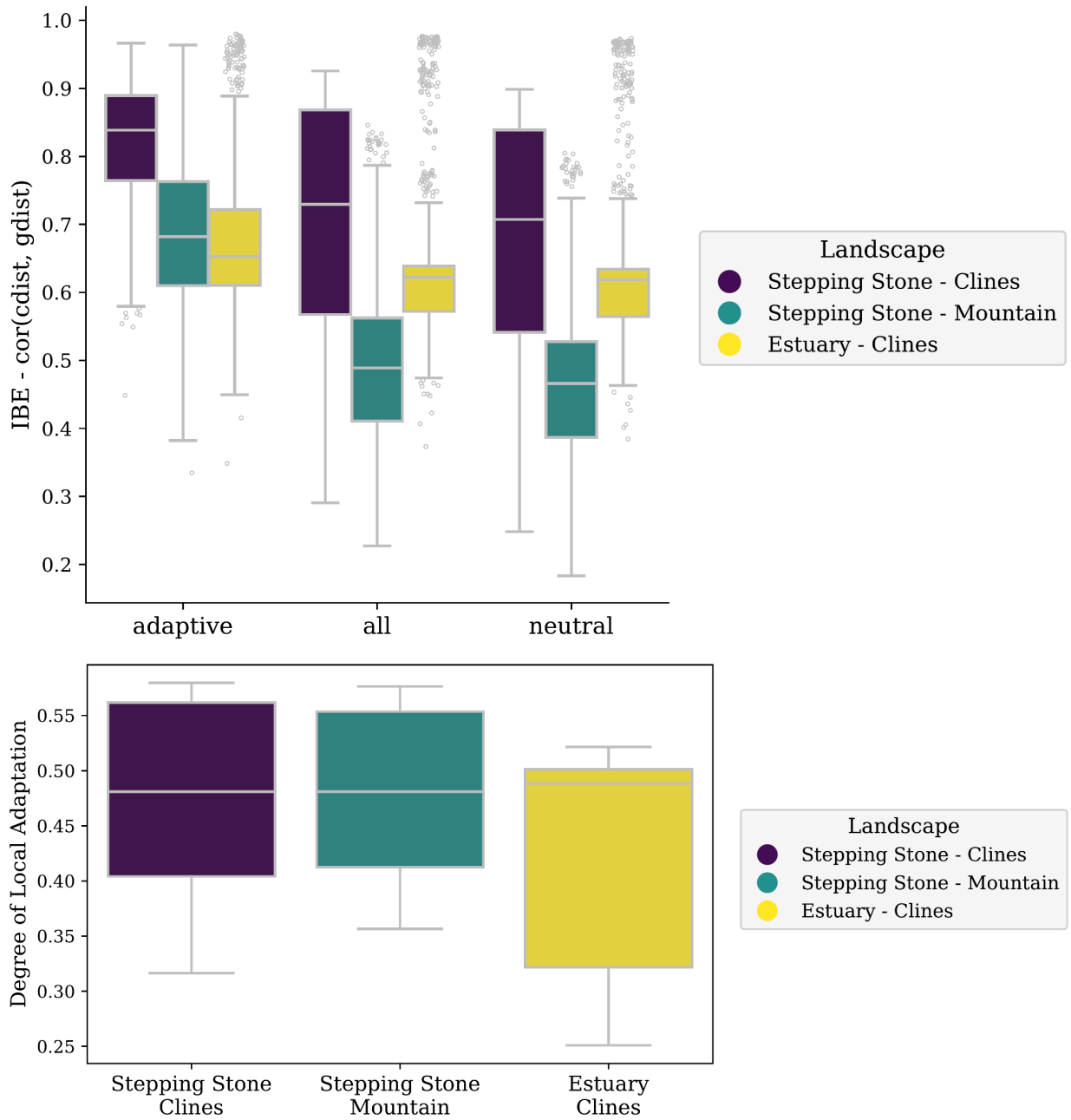


Fig S17 Levels of isolation-by-environment in marker sets vary across landscapes (A) and the degree of local adaptation reached by metapopulations on these landscapes (B). The pattern in (A) given (B) is in contrast to patterns between levels of IBE and the degree of local adaptation (Fig. S29). *I IBE*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

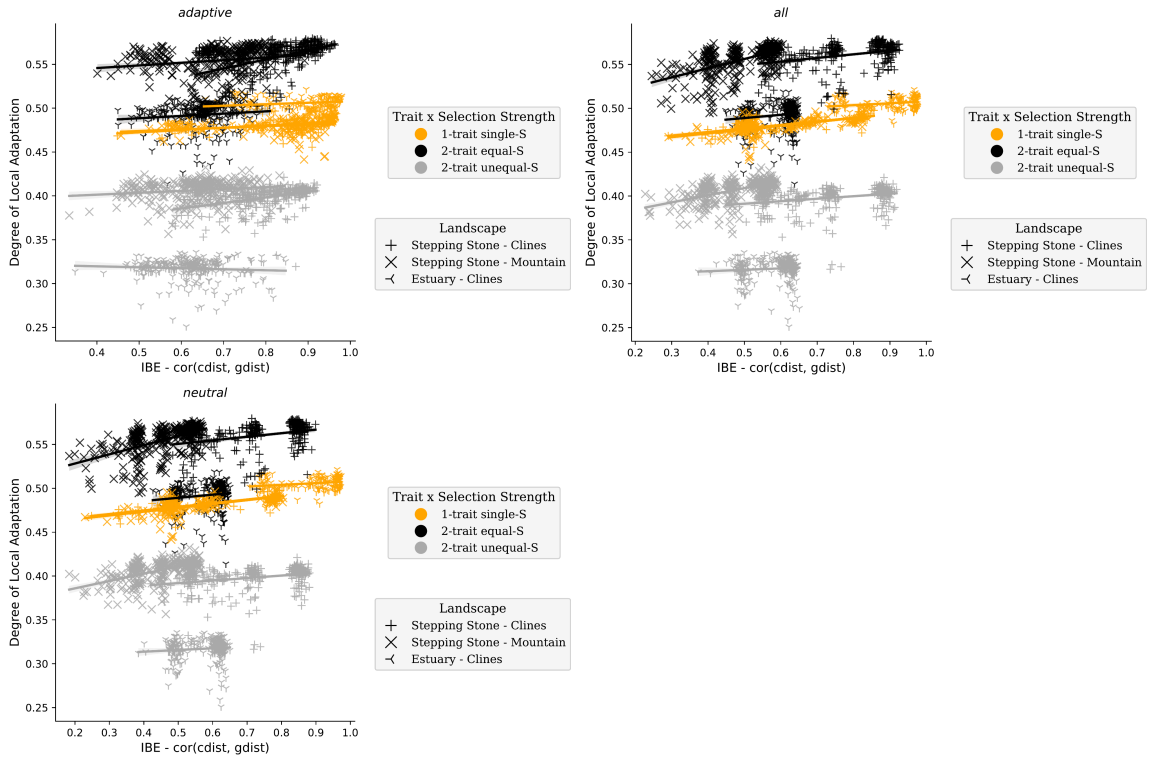
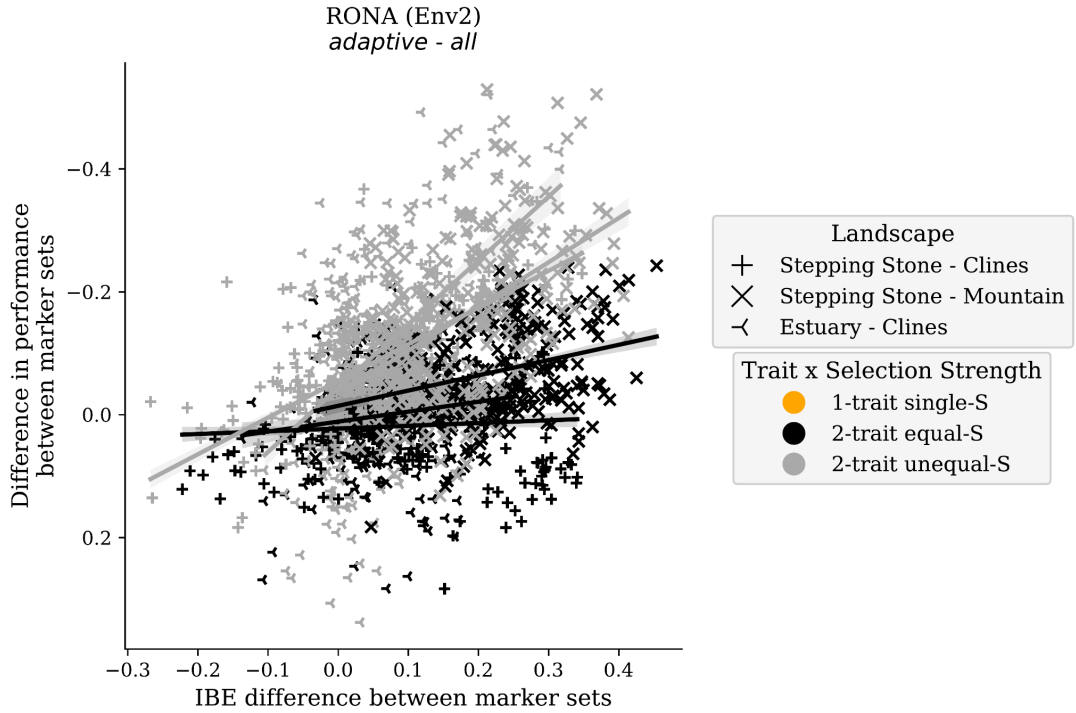
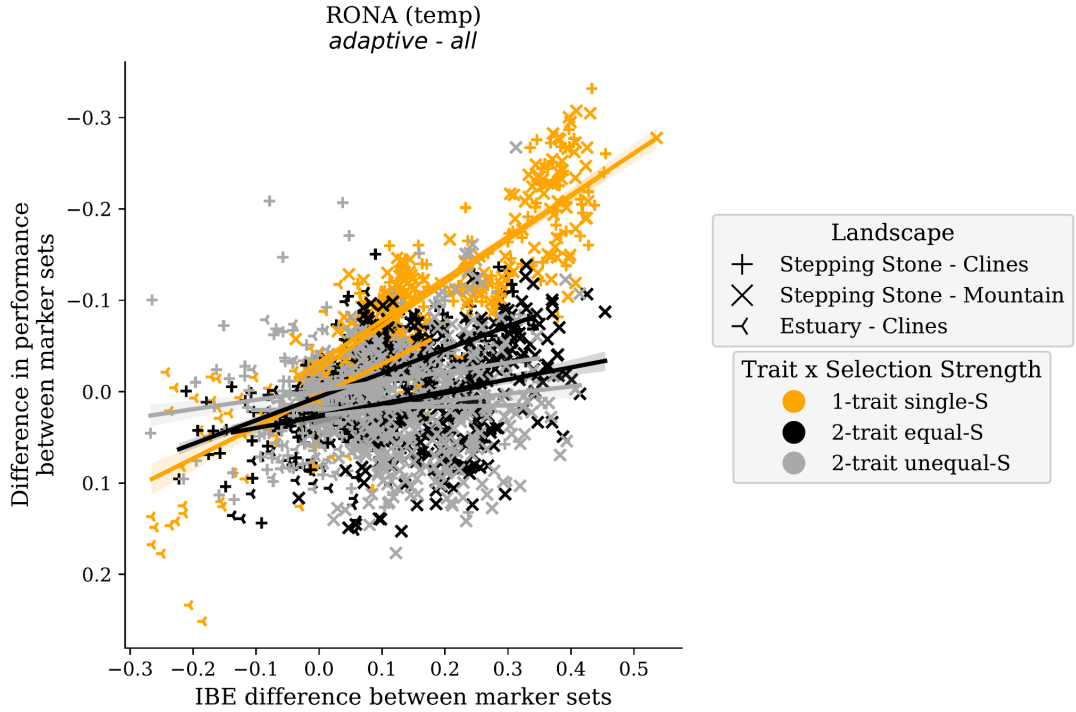


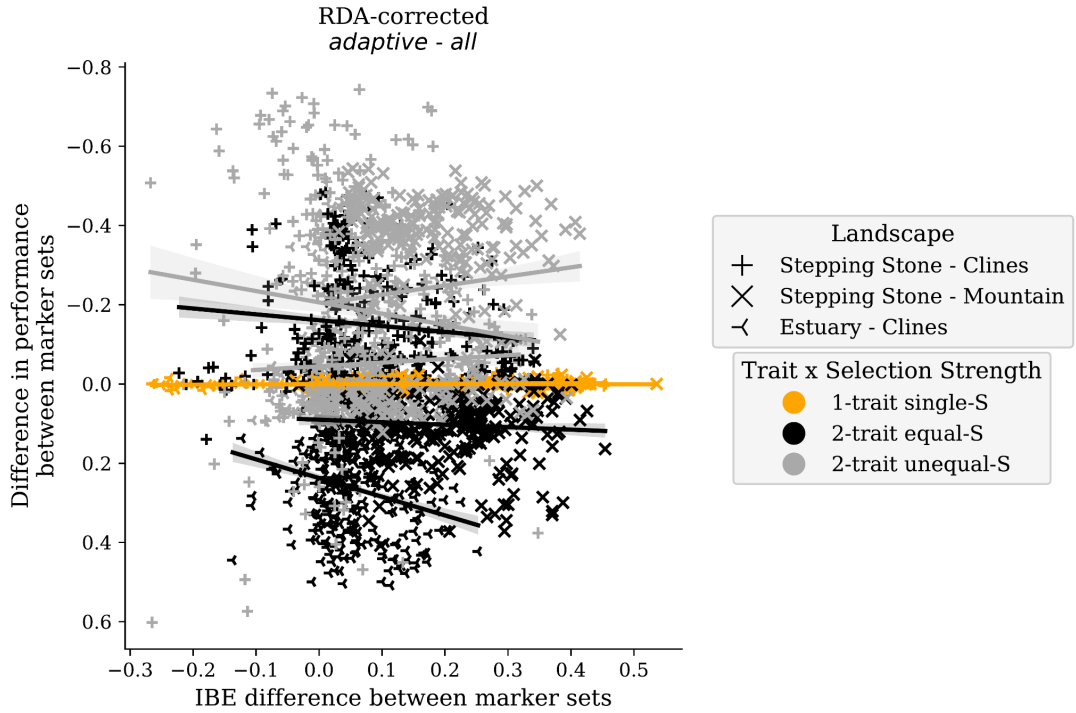
Fig S18 The levels of isolation-by-distance in marker sets (panels) are weakly correlated with the degree of local adaptation ($LA_{\Delta SA}$) within simulation levels. *IBE* is quantified as Spearman's rank correlation between population pairwise F_{ST} (gdist) and Euclidean distance of adaptive environments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

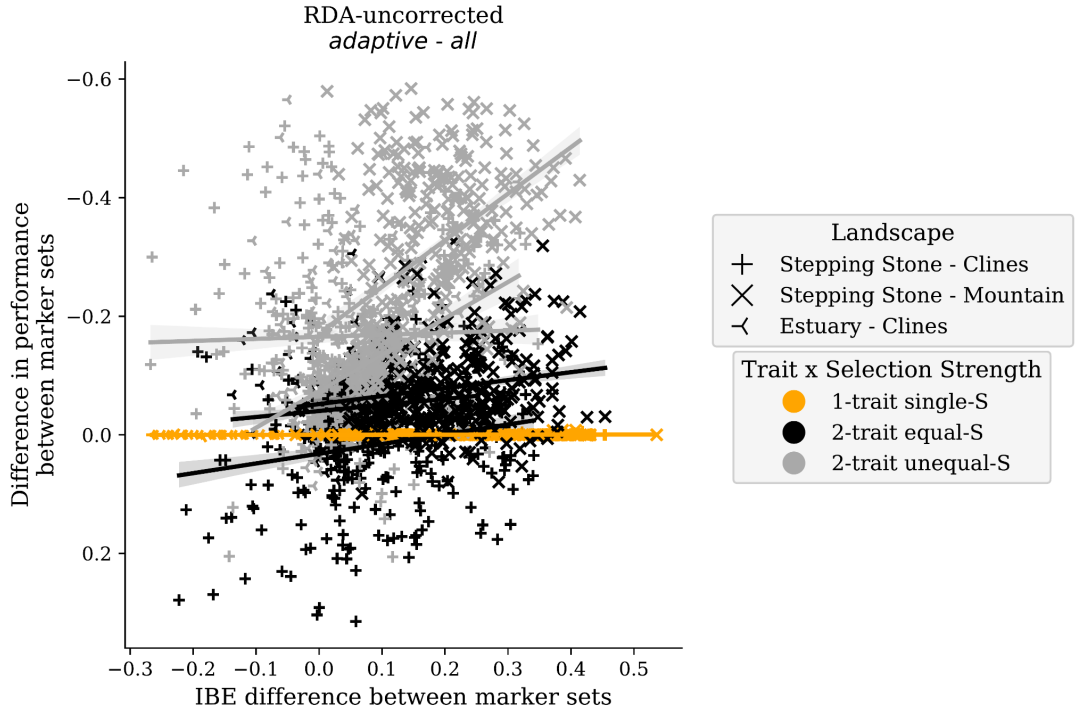
(Fig S19)


RONA (Env2)
adaptive - all
Difference in performance between marker sets
IBE difference between marker sets
Landscape
Stepping Stone - Clines
Stepping Stone - Mountain
Estuary - Clines
Trait x Selection Strength
1-trait single-S
2-trait equal-S
2-trait unequal-S

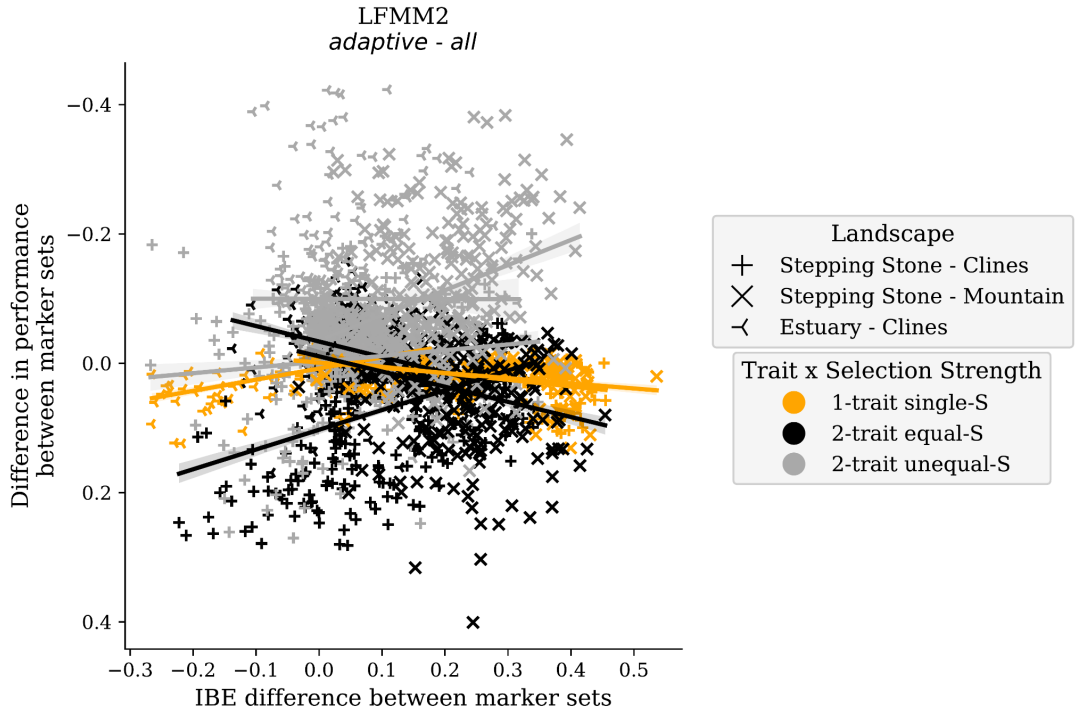

RONA (temp)
adaptive - all
Difference in performance between marker sets
IBE difference between marker sets
Landscape
Stepping Stone - Clines
Stepping Stone - Mountain
Estuary - Clines
Trait x Selection Strength
1-trait single-S
2-trait equal-S
2-trait unequal-S

(Fig S19 continued)





(Fig S19 continued)



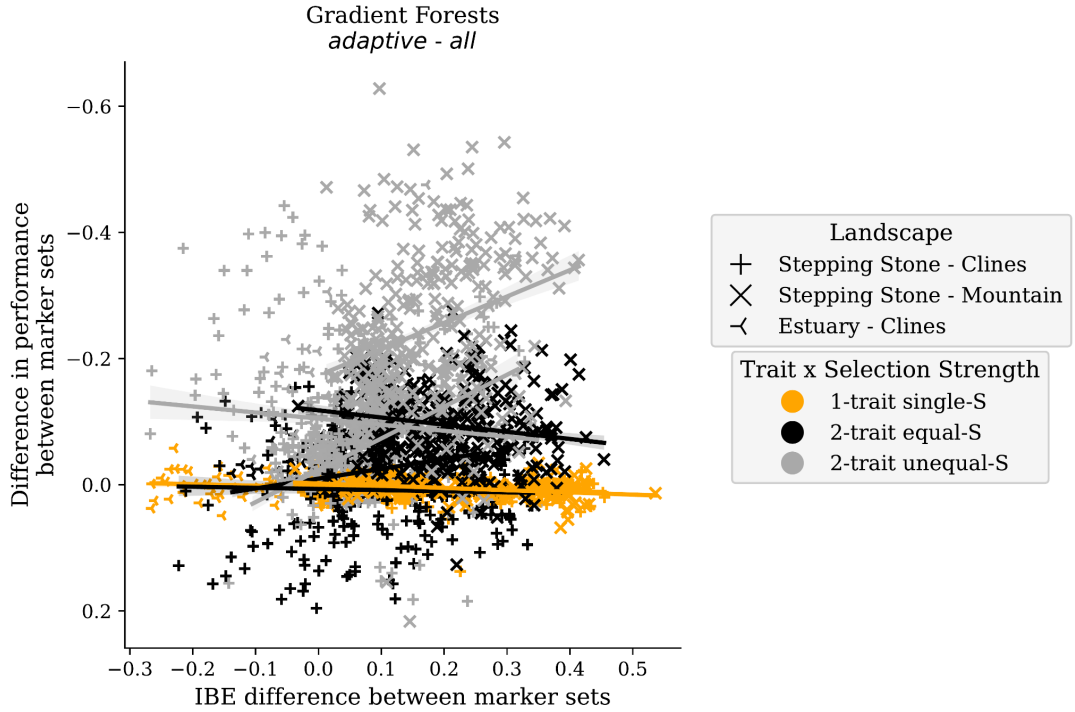


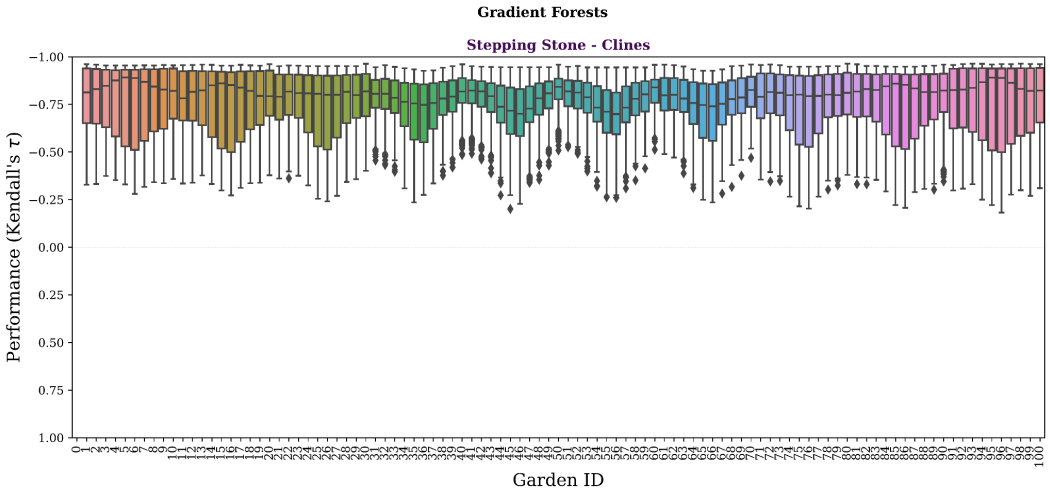
Fig S19 Differences in levels of IBE between marker sets used to train models is generally unrelated to differences in model performances. Shown is the difference in median performance between *adaptive* and *all* marker sets and the difference in *IBE* between these marker sets. *IBE* is quantified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

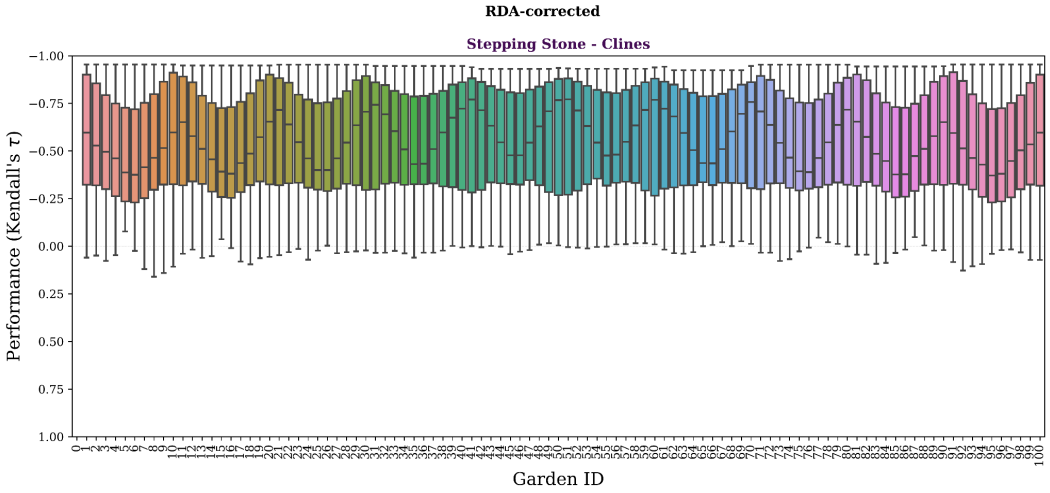
Latitude	1	2	3	4	5	6	7	8	9	10
10	91.0	92.0	93.0	94.0	95.0	96.0	97.0	98.0	99.0	100.0
9	81.0	82.0	83.0	84.0	85.0	86.0	87.0	88.0	89.0	90.0
8	71.0	72.0	73.0	74.0	75.0	76.0	77.0	78.0	79.0	80.0
7	61.0	62.0	63.0	64.0	65.0	66.0	67.0	68.0	69.0	70.0
6	51.0	52.0	53.0	54.0	55.0	56.0	57.0	58.0	59.0	60.0
5	41.0	42.0	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
1	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0

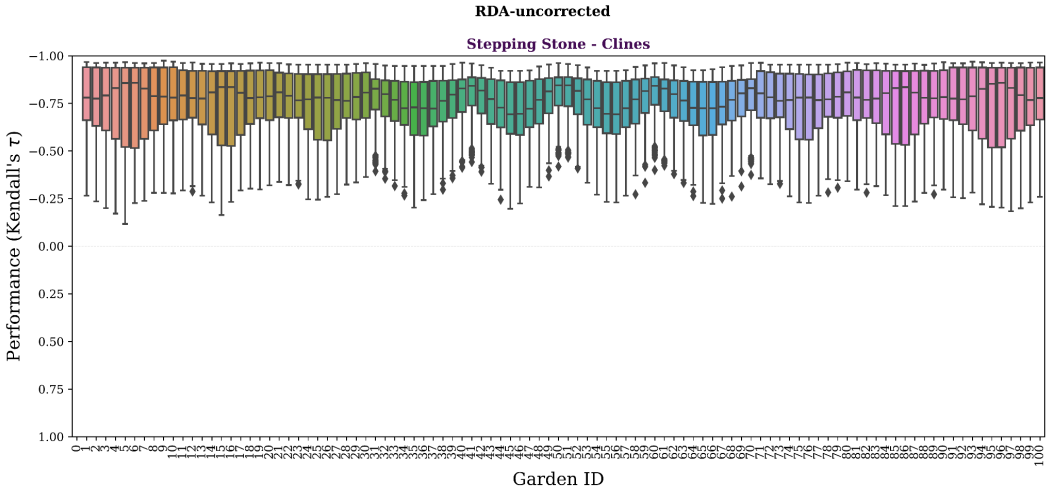
Longitude

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). This map can be used to interpret the ordering of gardens along x-axes of Figs. S21 S22 and S23. Code used to create this figure can be found in SC 02.02.04.

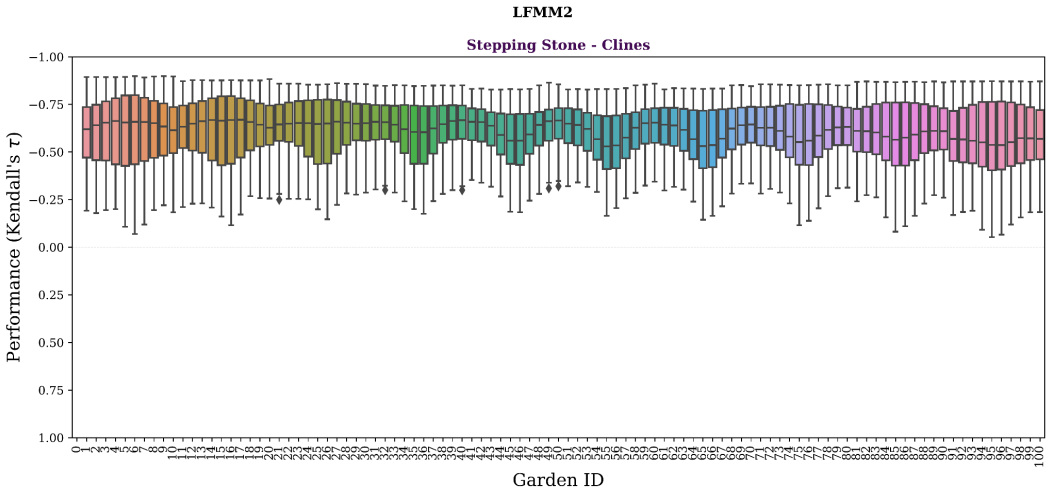
(Fig. S21)

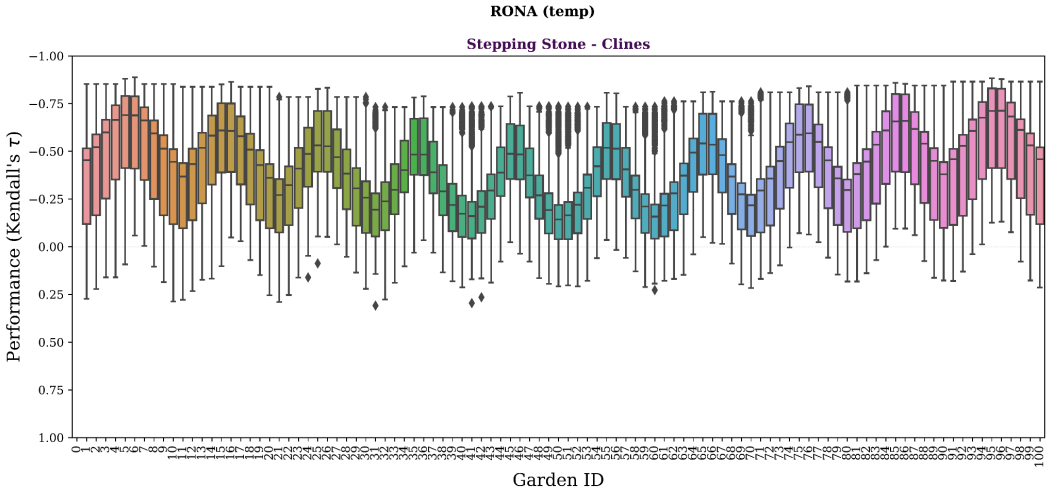






(Fig S21 continued)


LFMM2
Stepping Stone - Clines
Performance (Kendall's τ)
Garden ID


RONA (temp)
Stepping Stone - Clines
Performance (Kendall's τ)
Garden ID

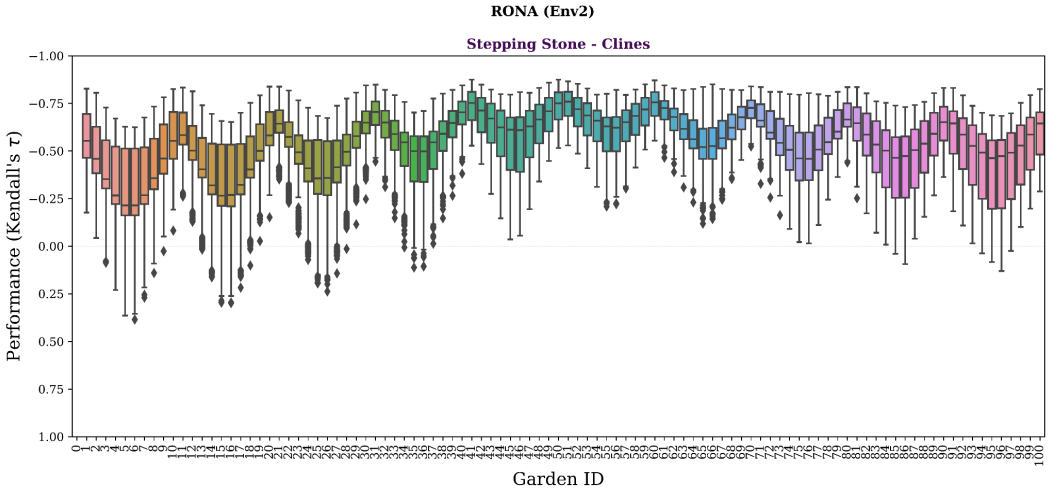
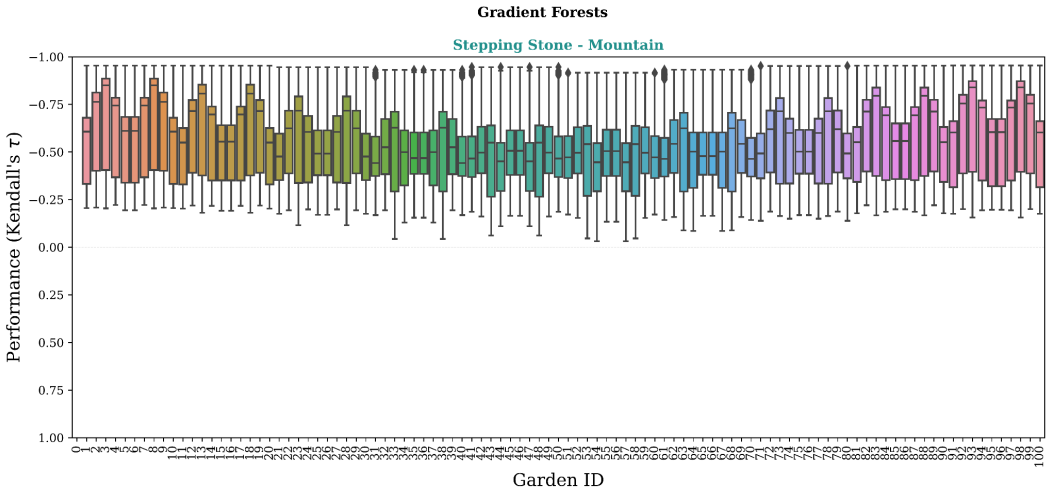
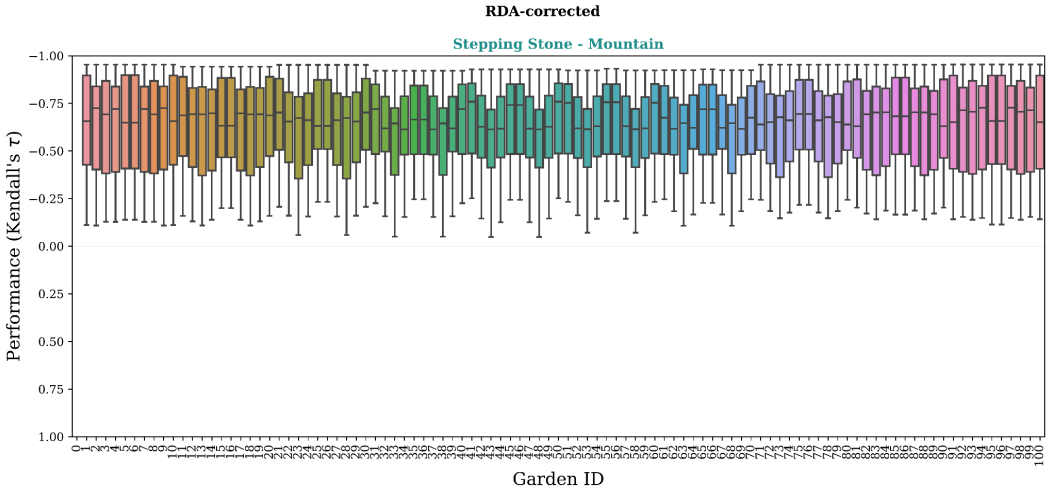
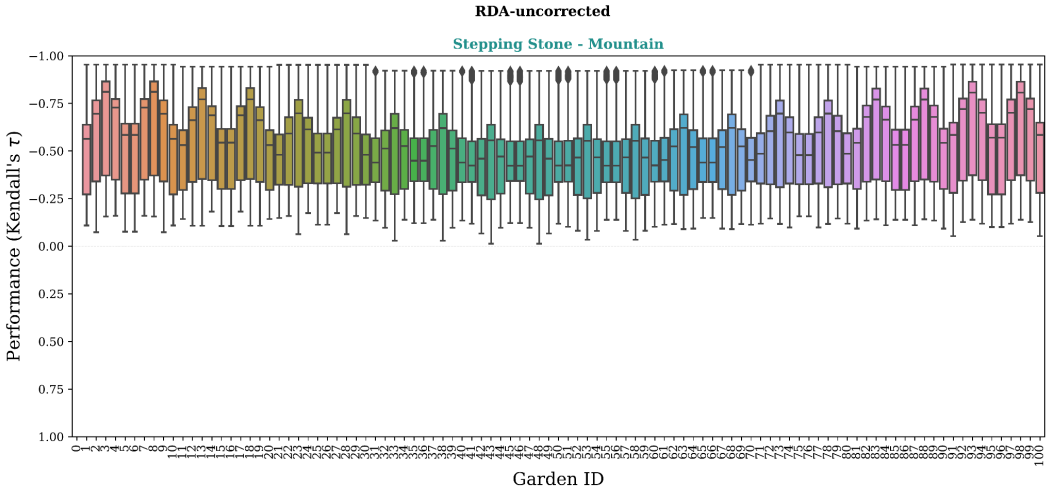

RONA (Env2)
Stepping Stone - Clines
Performance (Kendall's τ)
Garden ID

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

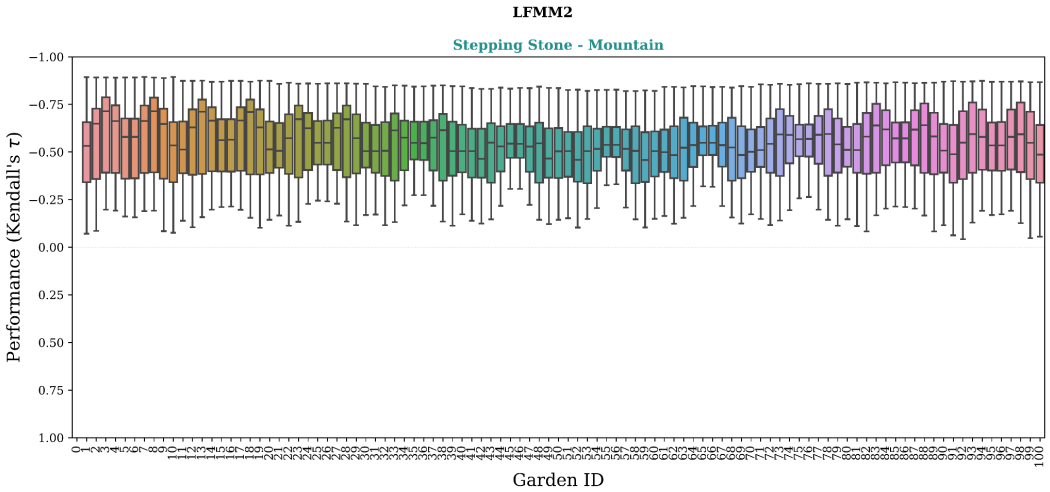
(Fig S22)

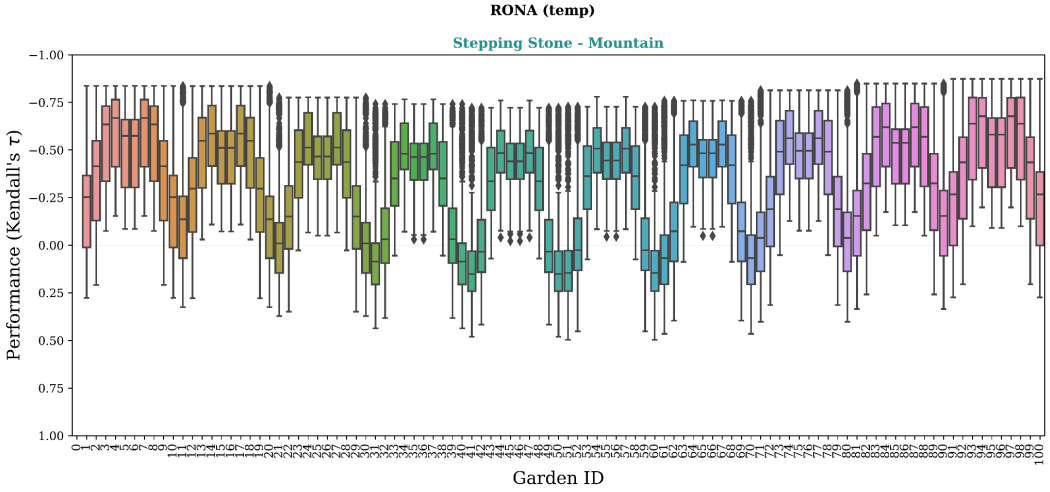

Gradient Forests
Stepping Stone - Mountain
Performance (Kendall's τ)
Garden ID


RDA-corrected
Stepping Stone - Mountain
Performance (Kendall's τ)
Garden ID


RDA-uncorrected
Stepping Stone - Mountain
Performance (Kendall's τ)
Garden ID

(Fig S22 continued)





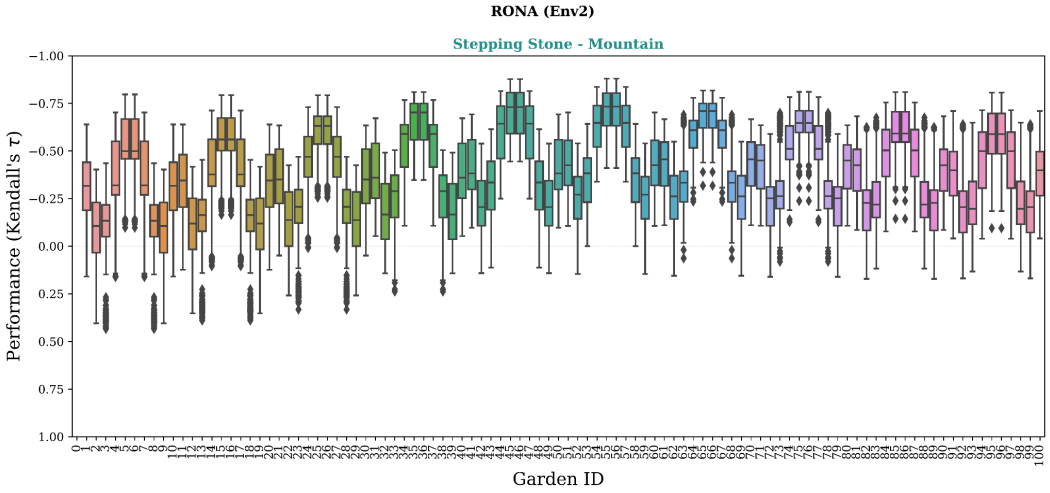
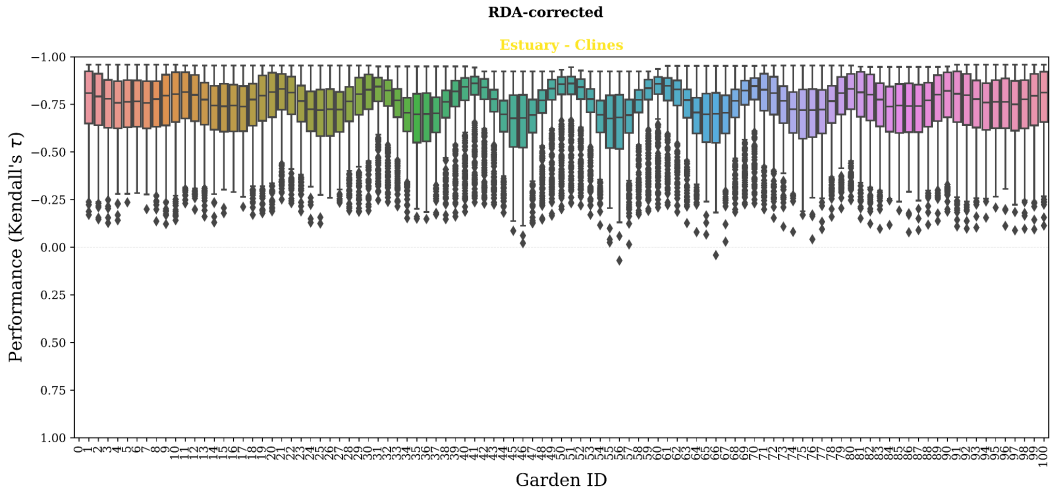
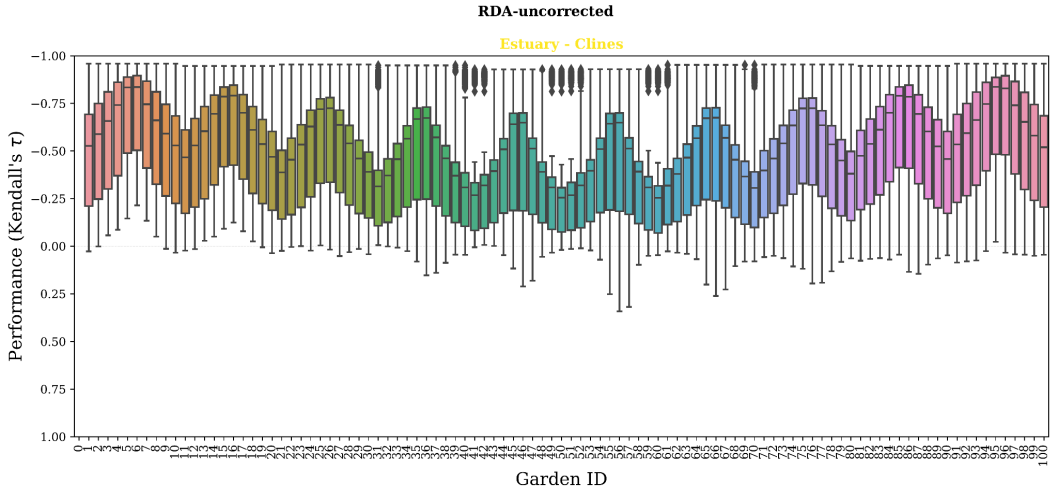
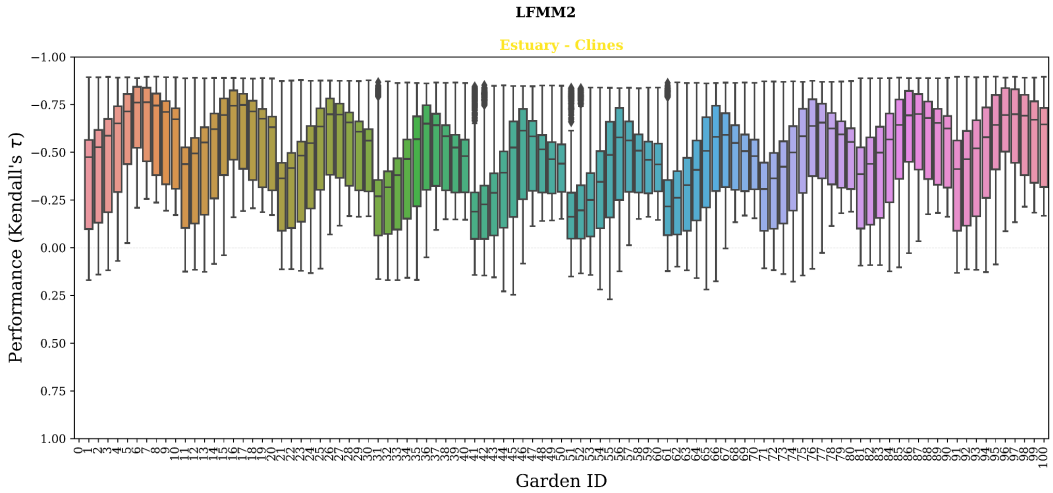


Fig S22 Genomic offset methods have variable performance across the *Stepping-Stone - Mountain* landscape. Shown is the variability of each offset method performance (y-axes) across the 100 common gardens (x-axes). Gardens are ordered from left to right by garden ID. This ordering of gardens is equivalent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

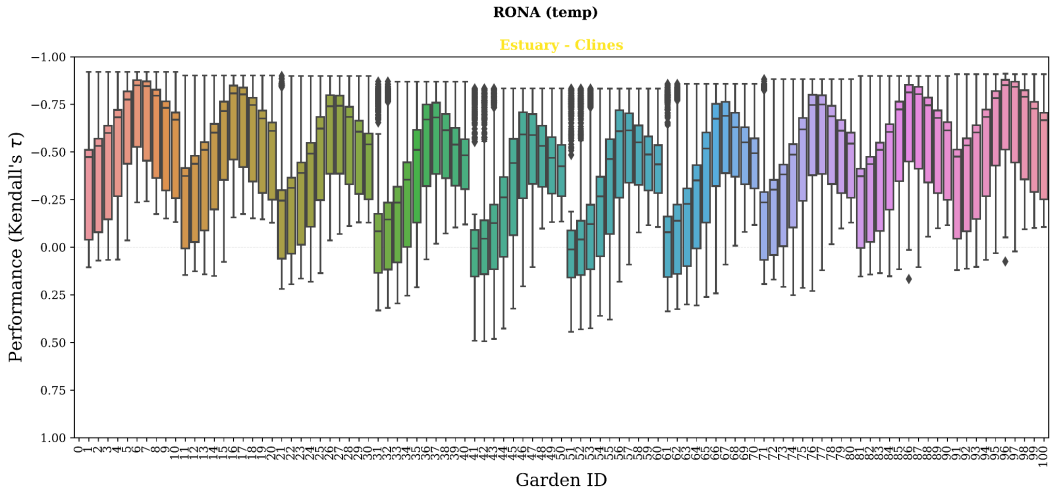
(Fig S23)

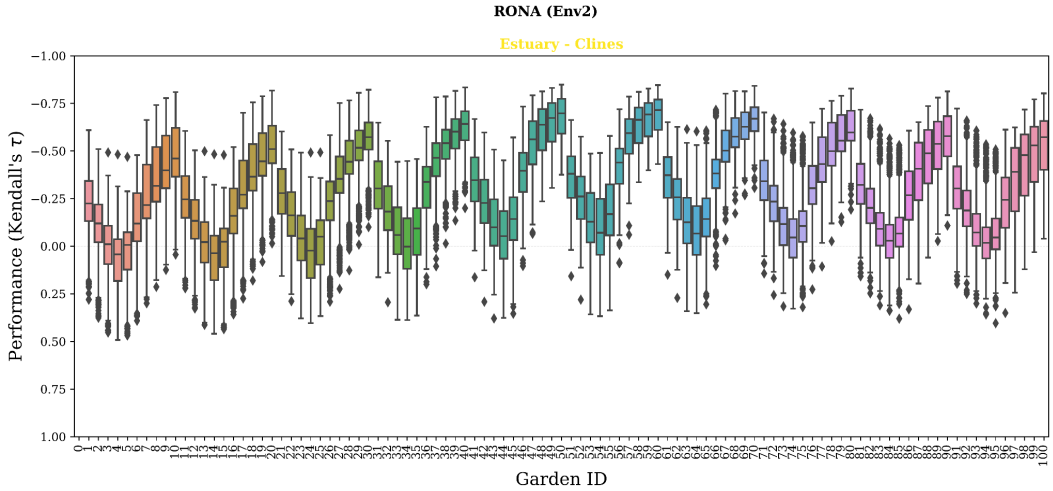

RDA-corrected
Estuary - Clines
Performance (Kendall's τ)
Garden ID


RDA-uncorrected
Estuary - Clines
Performance (Kendall's τ)
Garden ID


LFMM2
Estuary - Clines
Performance (Kendall's τ)
Garden ID

(Fig S23 continued)





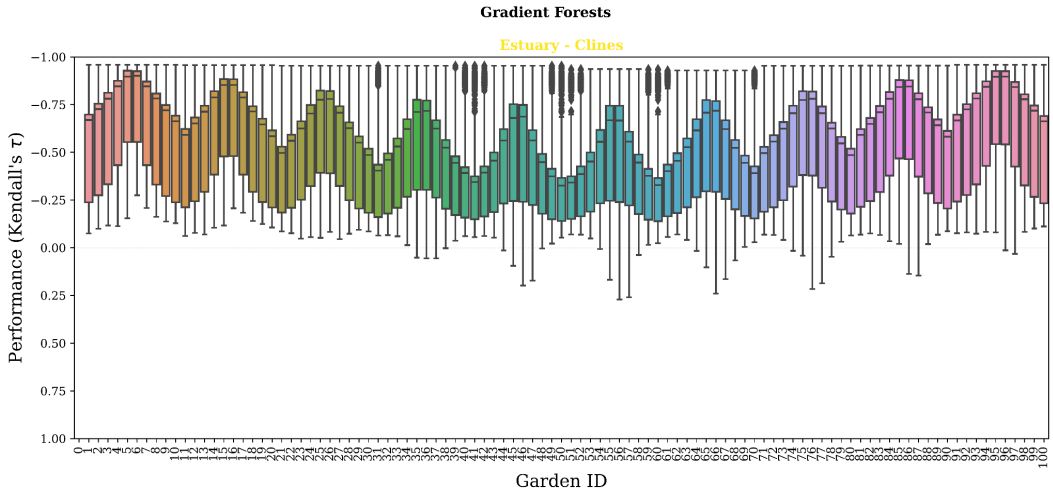


Fig S23 Genomic offset methods have variable performance across the *Estuary - Clines* landscape. Shown is the variability of each offset method performance (y-axes) across the 100 common gardens (x-axes). Gardens are ordered from left to right by garden ID. This ordering of gardens is equivalent to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

(Fig. S36)

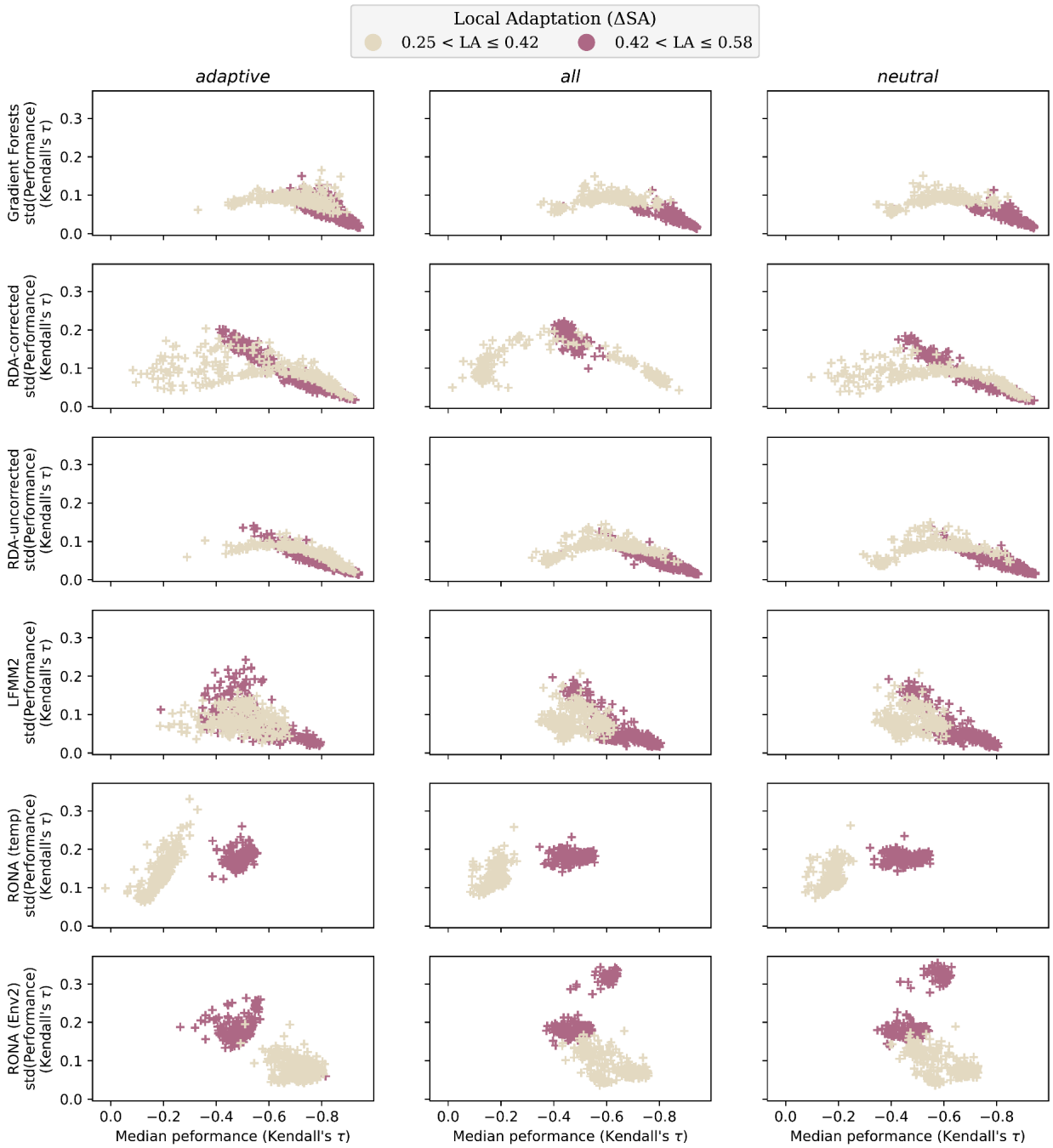

Local Adaptation (ΔSA)
0.25 < LA ≤ 0.42
0.42 < LA ≤ 0.58
adaptive
all
neutral
Gradient Forests std(Performance) (Kendall's τ)
RDA-corrected std(Performance) (Kendall's τ)
RDA-uncorrected std(Performance) (Kendall's τ)
LFMM2 std(Performance) (Kendall's τ)
RONA (temp) std(Performance) (Kendall's τ)
RONA (Env2) std(Performance) (Kendall's τ)
0.0
0.1
0.2
0.3
0.0
−0.2
−0.4
−0.6
−0.8
Median peformance (Kendall's τ)

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

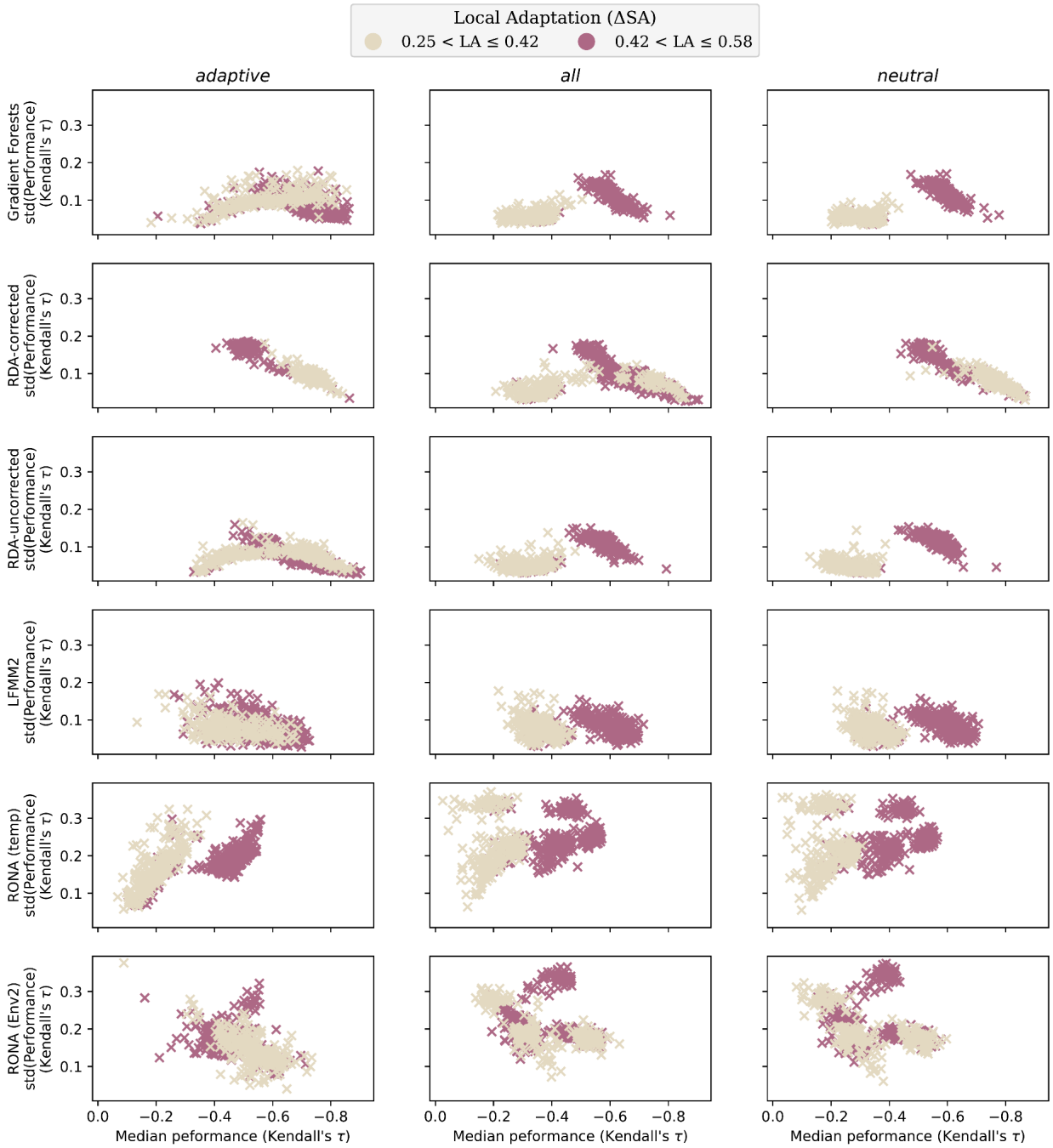


Fig S25 Variability across evaluations of genomic offsets often decreases with increasing average performance across marker sets. Data included in this figure is from evaluation of 2-trait simulations from *Stepping-Stone - Mountain* landscapes. For similar figures for *Stepping-Stone - Clines* and *Estuary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

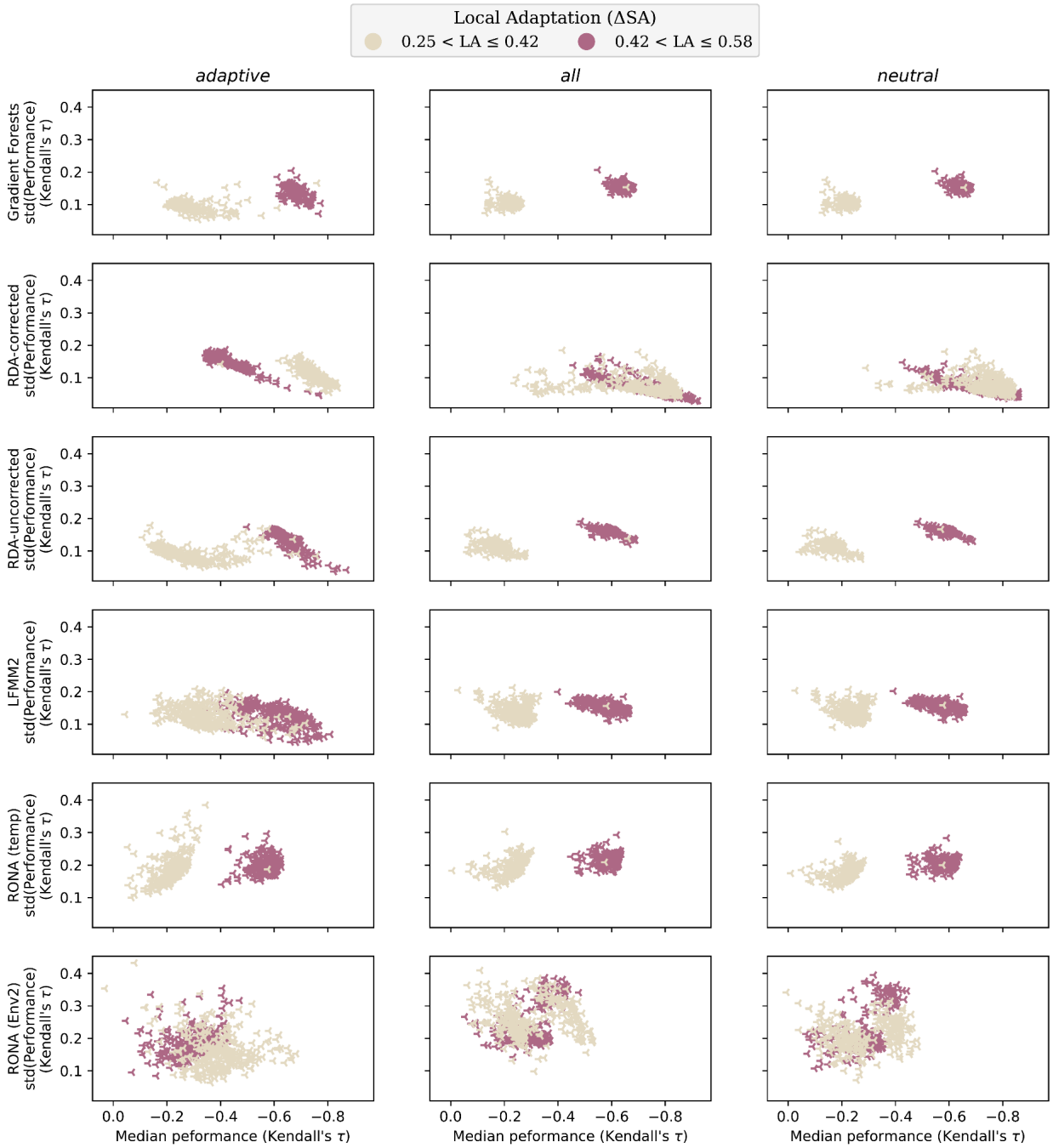


Fig S26 Variability across evaluations of genomic offsets often decreases with increasing average performance across marker sets. Data included in this figure is from evaluation of 2-trait simulations from *Estuary - Clines* landscapes. For similar figures for *Stepping-Stone - Clines* and *Stepping-Stone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

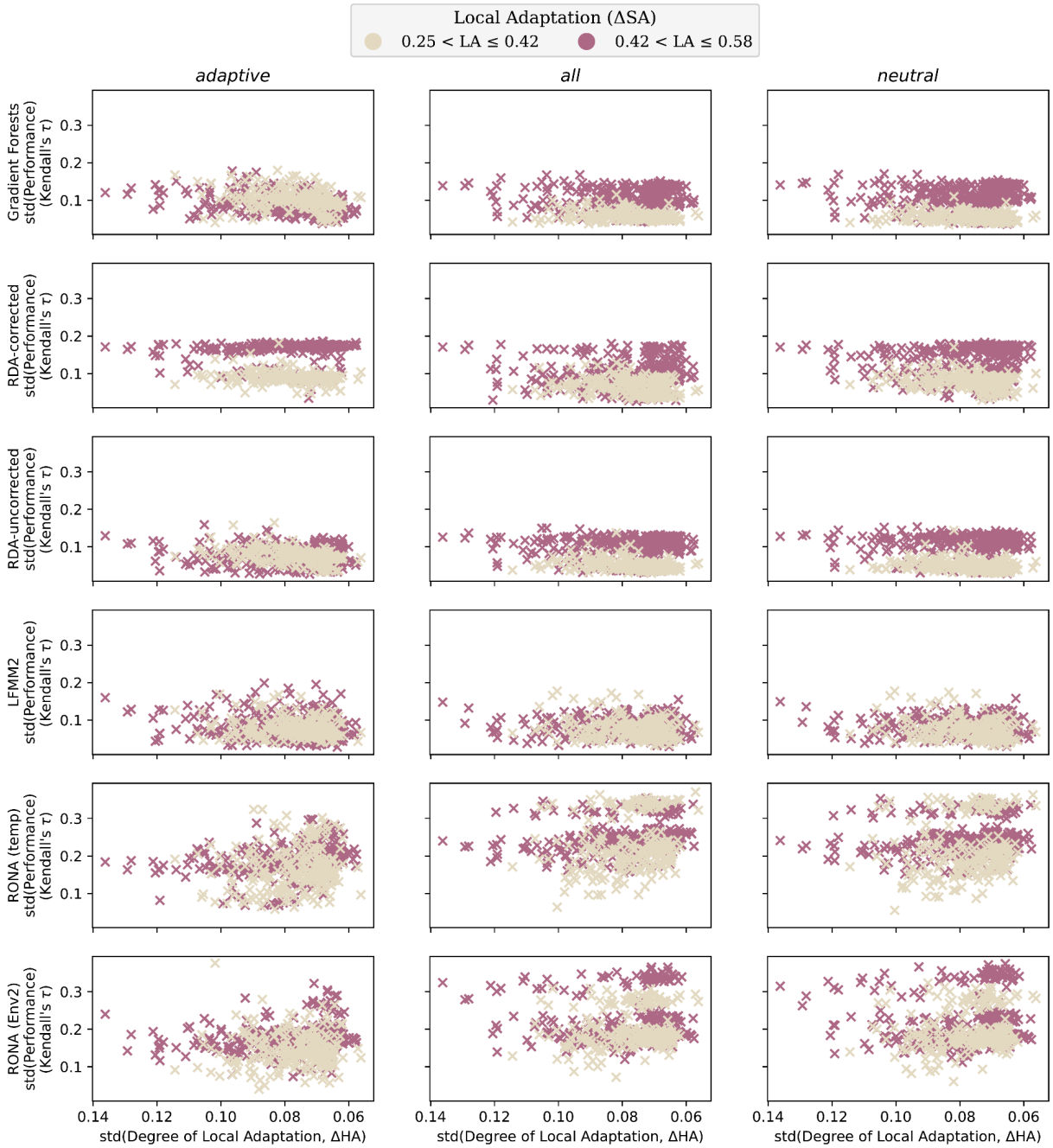


Fig S27 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations. Data included in this figure is from evaluation of 2-trait simulations from *Stepping-Stone - Mountain* landscapes. For similar figures for *Stepping-Stone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

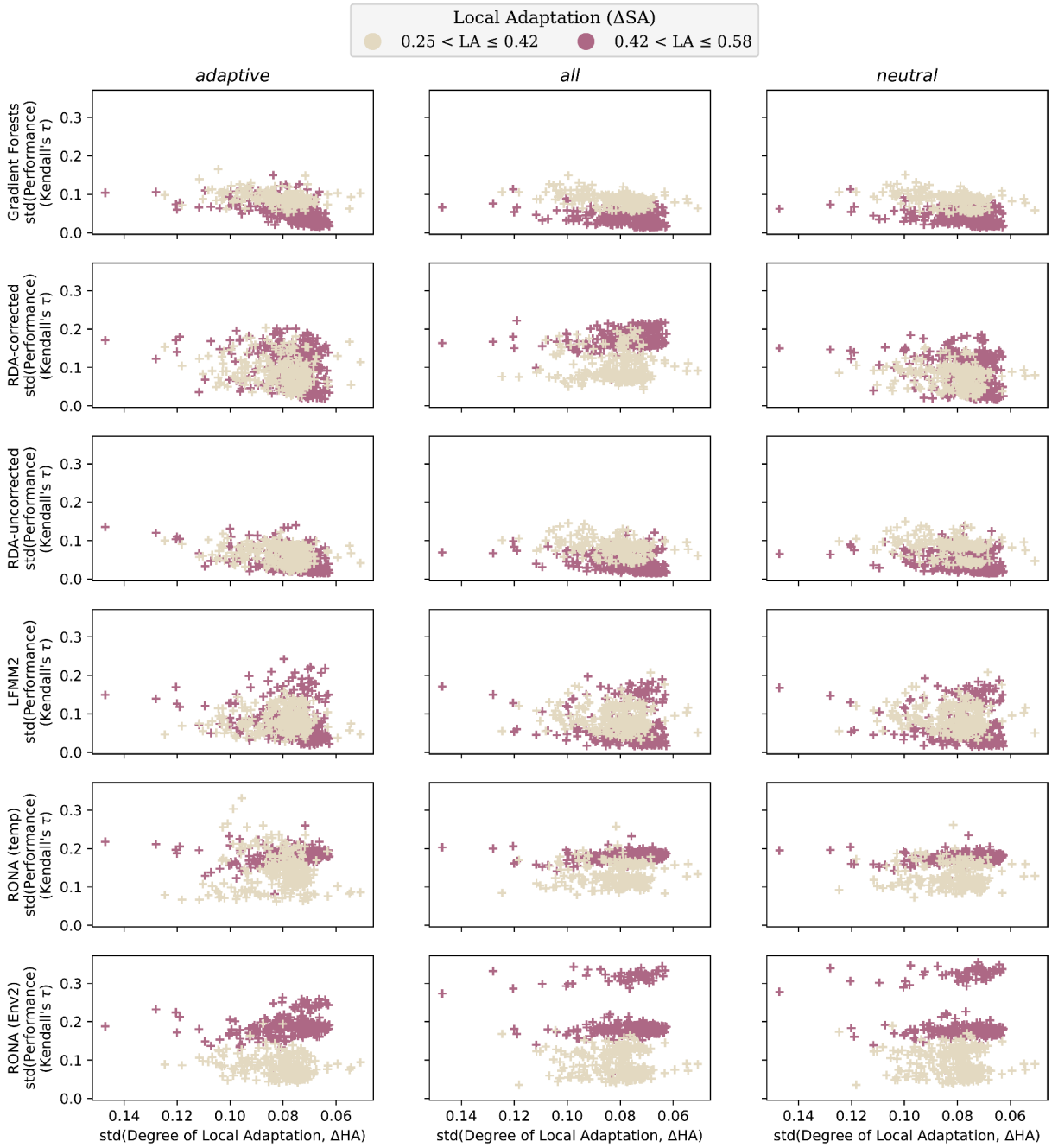


Fig S28 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations. Data included in this figure is from evaluation of 2-trait simulations from *Stepping Stone - Clines* landscapes. For similar figures for *Estuary - Clines*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

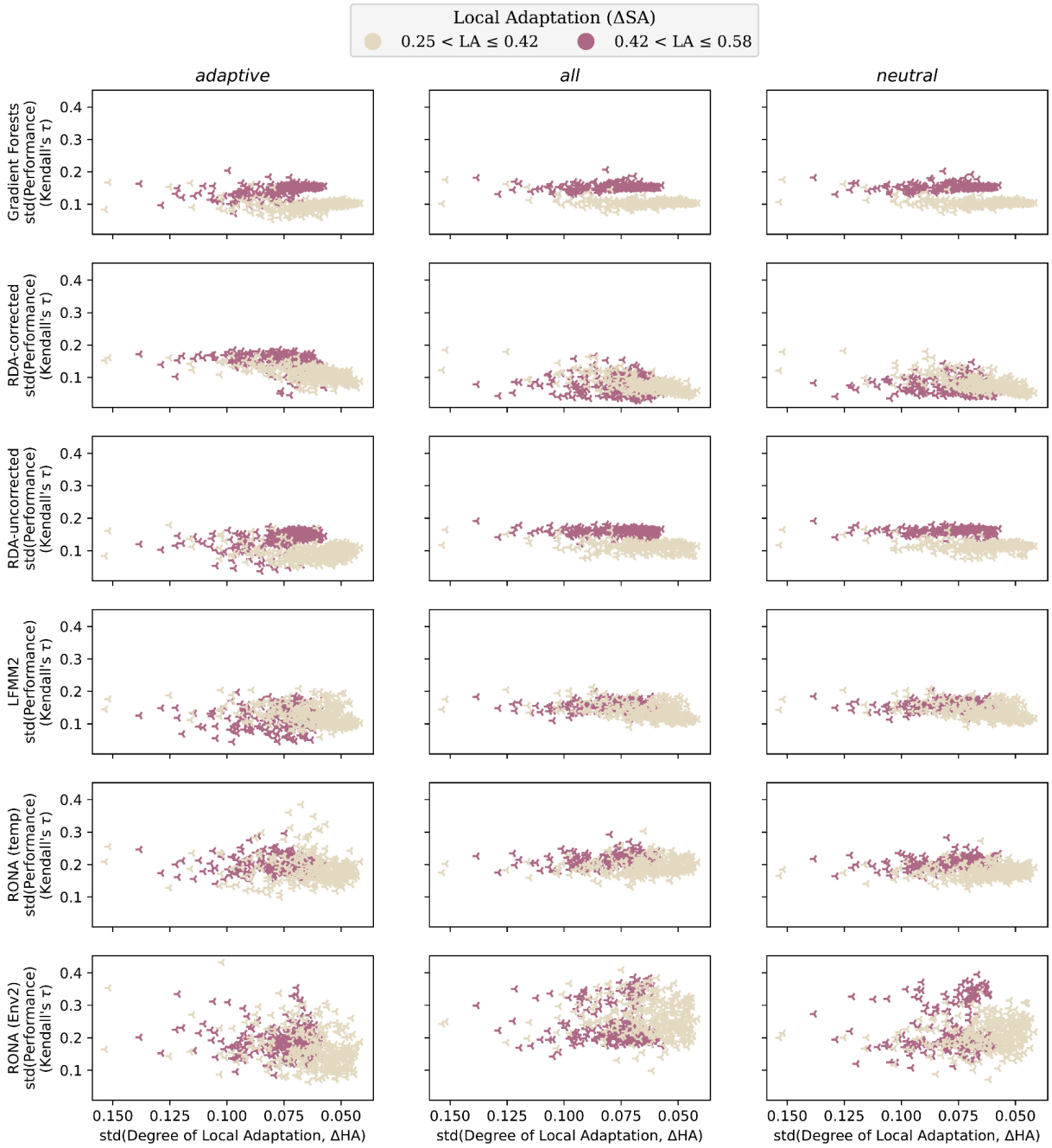


Fig S29 Variability across evaluations of genomic offsets is often unrelated to the variability in the degree of local adaptation across populations. Data included in this figure is from evaluation of 2-trait simulations from *Estuary - Clines* landscapes. For similar figures for *Stepping-Stone - Clines*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

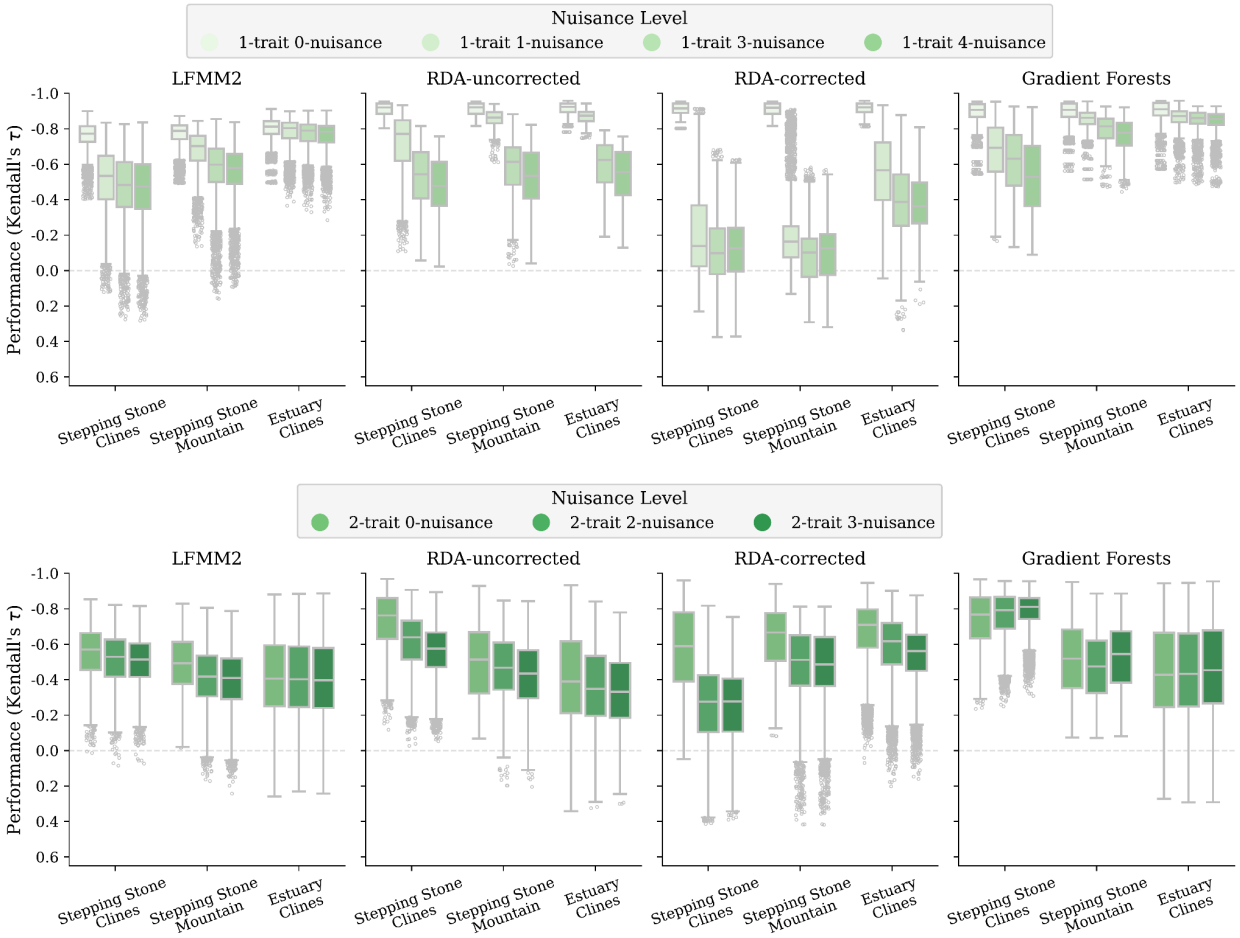


Fig S30 Effect of non-adaptive nuisance environmental variables on offset performance faceted by landscape. Shown are offsets from 1- and 2-trait simulations trained using only adaptive environments (0-nuisance) or with adaptive environments and the addition of N>0 non-adaptive environmental variables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

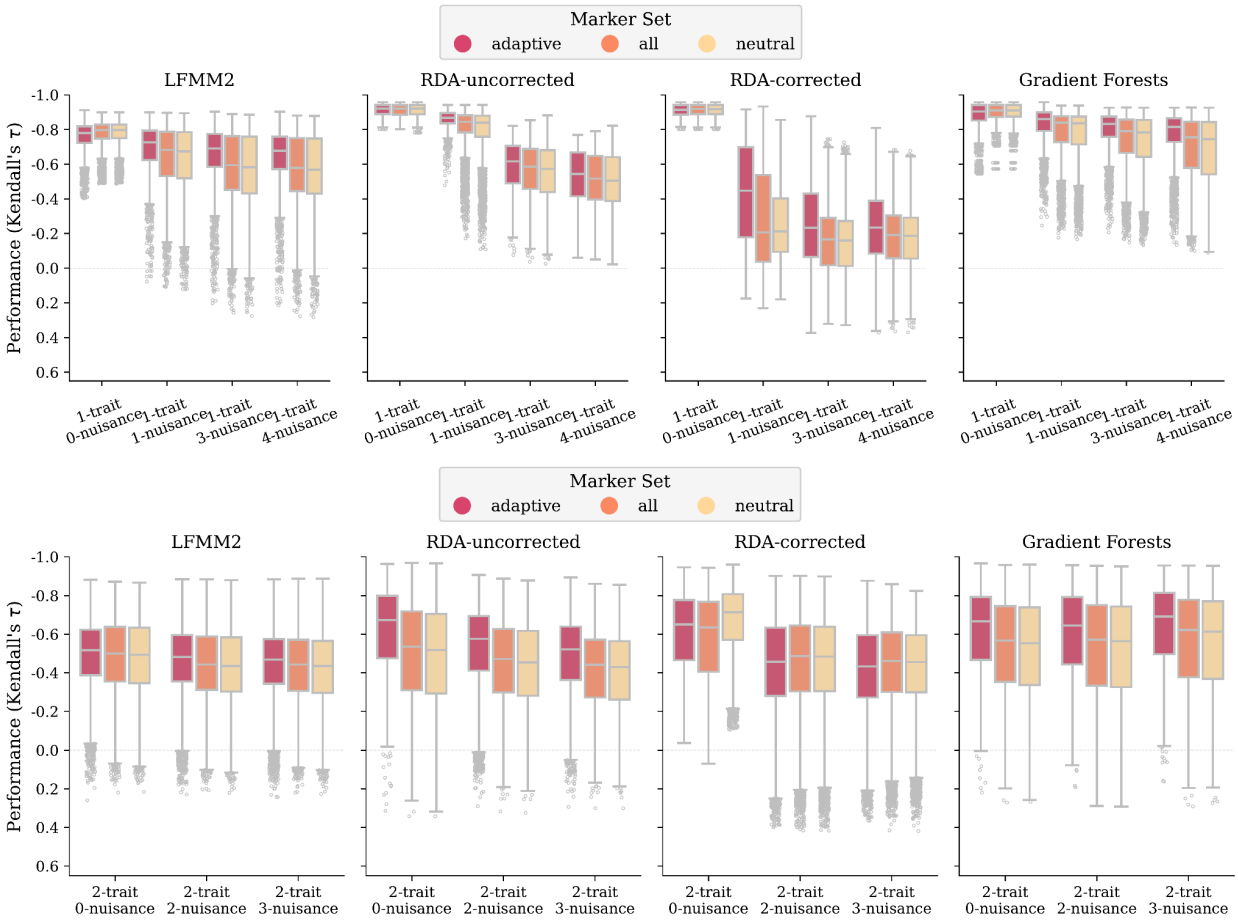
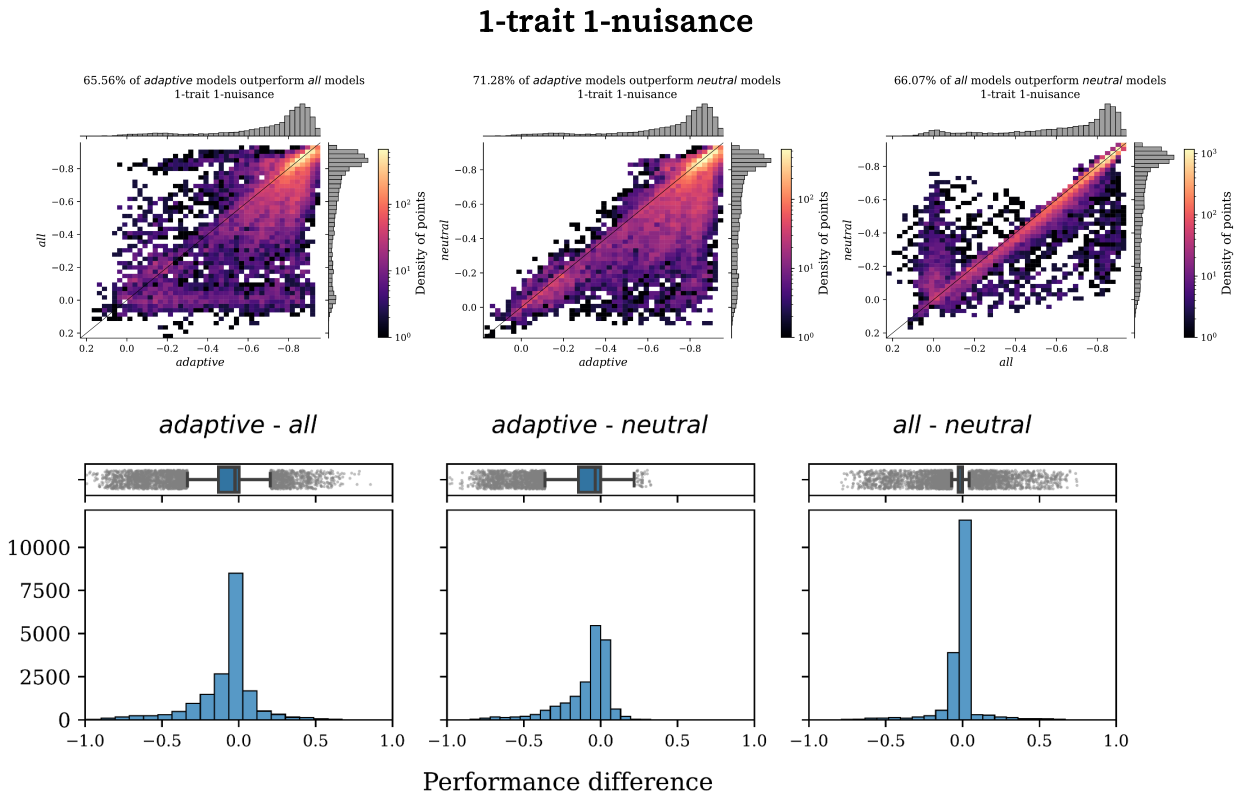


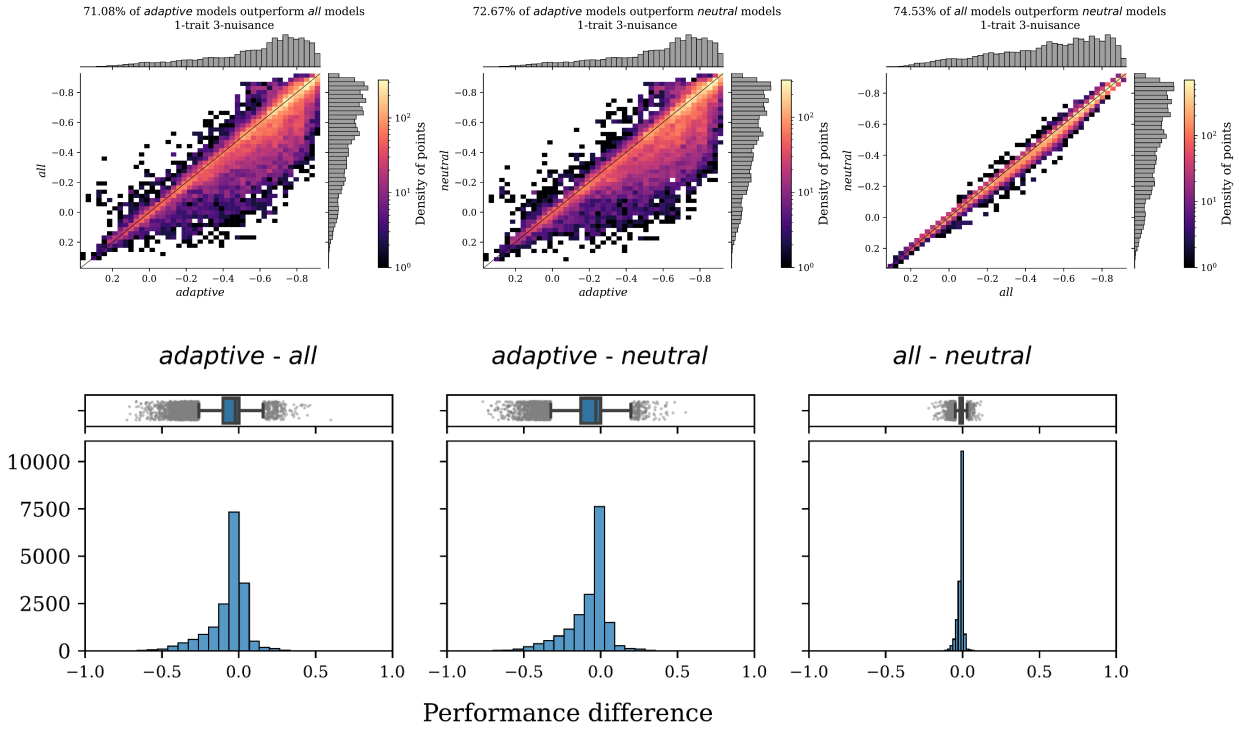
Fig S31 Effect of non-adaptive nuisance environmental variables on offset performance faceted by marker set. Shown are offsets from 1- (A) and 2-trait (B) simulations trained using only adaptive environments (0-nuisance) or with adaptive environments and the addition of N>0 non-adaptive environmental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

(Fig. S31)

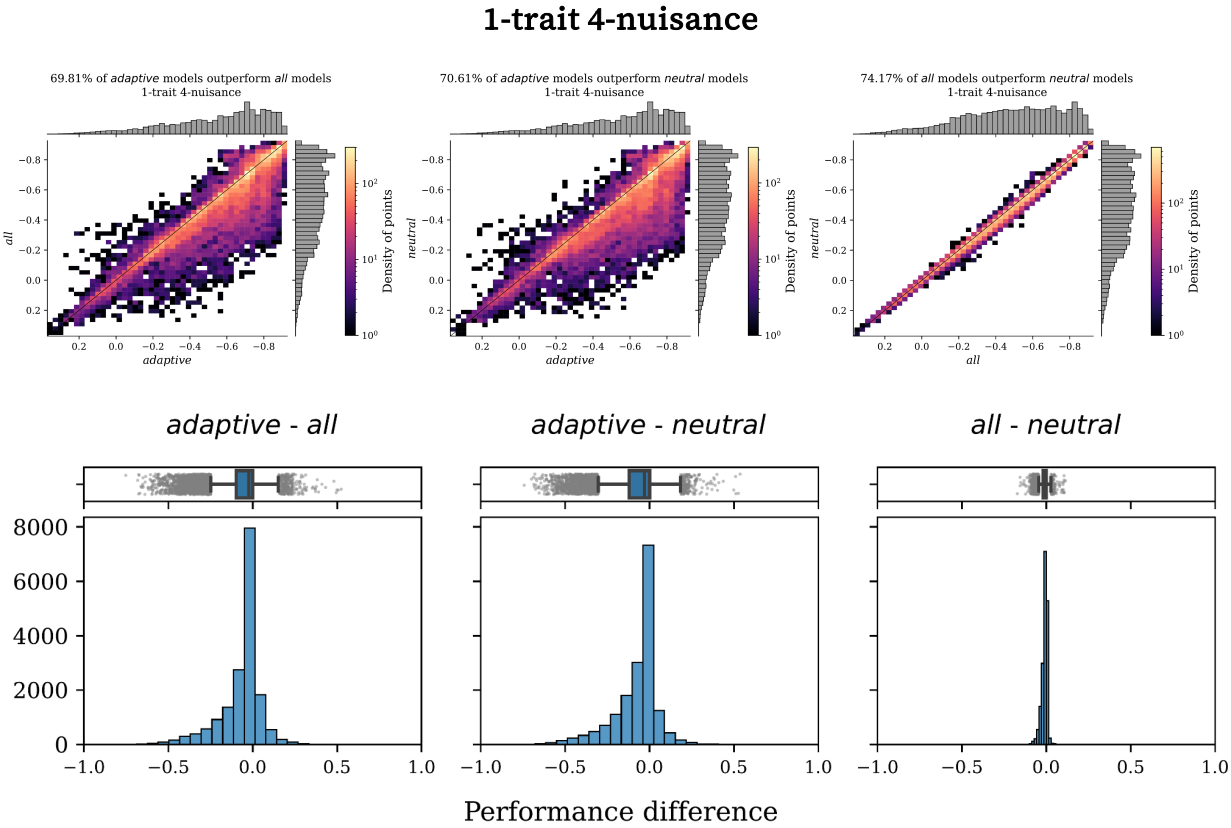

1-trait 1-nuisance
65.56% of adaptive models outperform all models
1-trait 1-nuisance
71.28% of adaptive models outperform neutral models
1-trait 1-nuisance
66.07% of all models outperform neutral models
1-trait 1-nuisance
adaptive
all
neutral
Density of points
adaptive - all
adaptive - neutral
all - neutral
Performance difference

(Fig. S31 continued)

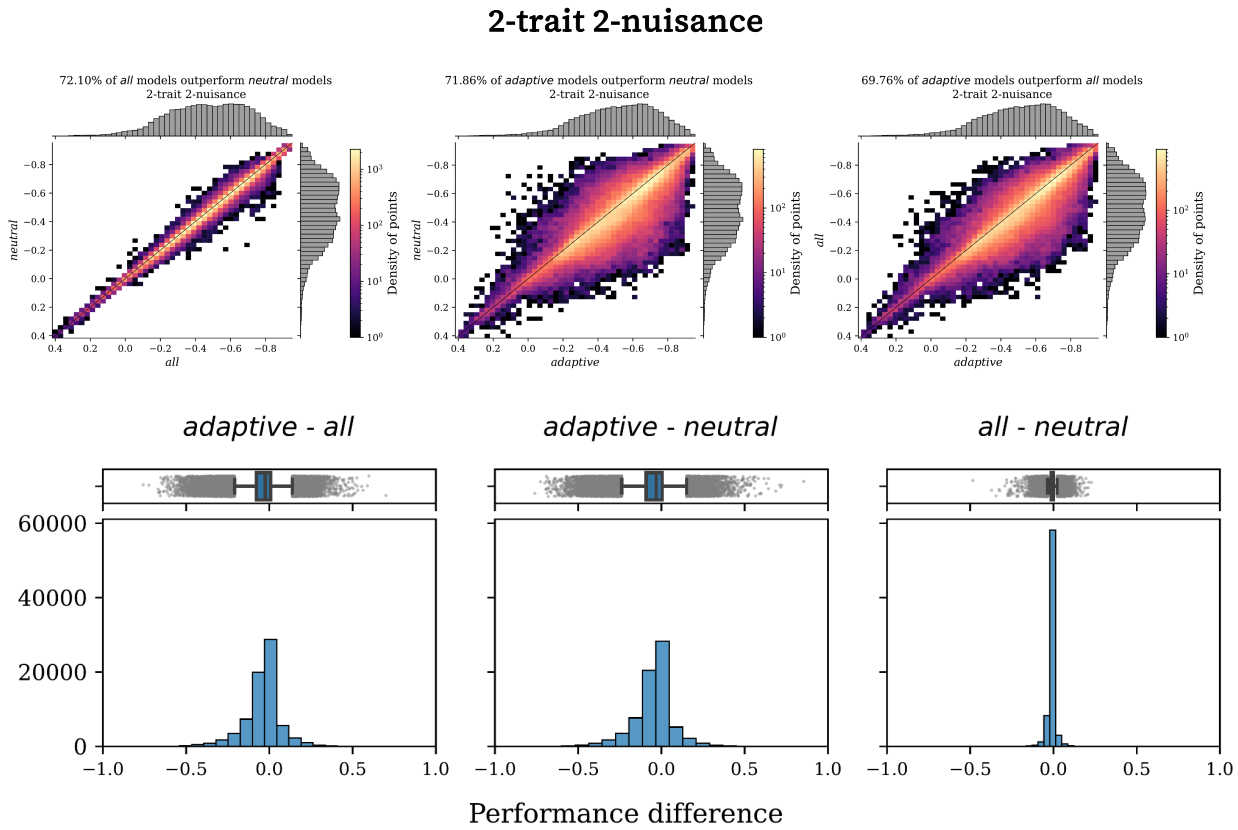
1-trait 3-nuisance



(Fig. S31 continued)



(Fig. S31 continued)



(Fig. S31 continued)

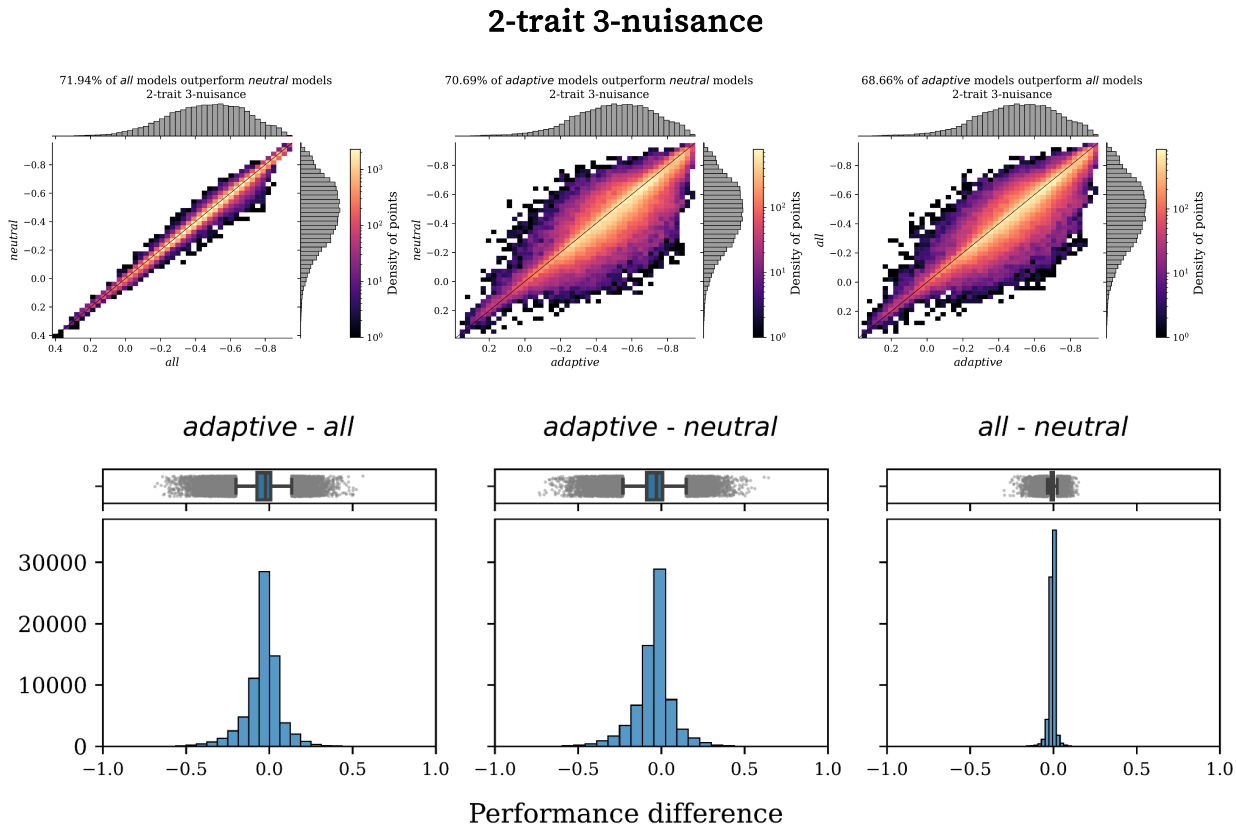


Fig S32 Pairwise comparison of performance differences between marker sets for *Nuisance Environment* scenarios. The first row for each nuisance level (*N-trait N-nuisance*) are scatterplots of pairwise comparisons of performance between marker sets (histograms in each margin) from both 1- and 2-trait 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.

Fig S33 is in Supplemental Note S4

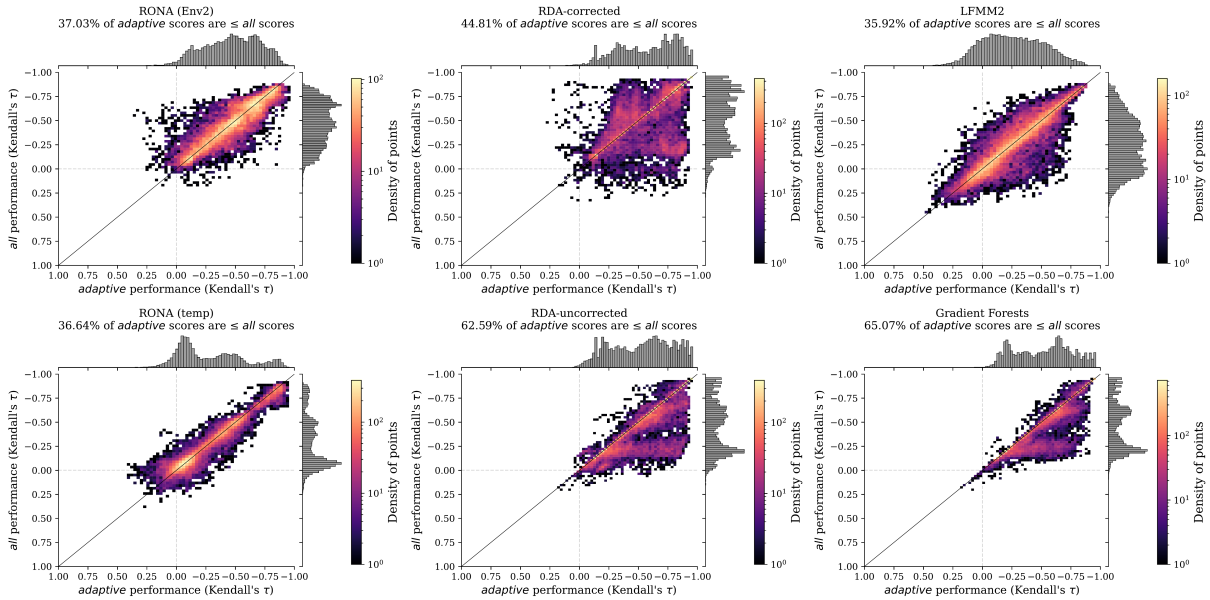


Fig S34 Pairwise comparison of performance differences between marker sets for *Climate Novelty* scenarios. Shown are scatterplots of pairwise comparisons of performance between marker sets (histograms in each margin) from both 1- and 2-trait models where density of points is indicated by color in legend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

Supplemental References

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.13722

Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906

Lotterhos, K. E. (2023). The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. *Proceedings of the National Academy of Sciences*, *120*(12). https://doi.org/10.1073/pnas.2220313120