Received: 14 February 2024 | Revised: 19 July 2024 | Accepted: 6 August 2024

DOI: 10.1111/1755-0998.14008

FROM THE COVER

MOLECULAR ECOLOGY RESOURCES WILEY

The accuracy of predicting maladaptation to new environments with genomic data

Brandon M. Lind | **Katie E. Lotterhos**

Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, Nahant, Massachusetts, USA

Correspondence
Brandon M. Lind, Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, 430 Nahant Road, Nahant, MA 01908, USA.
Email: lind.brandon.m@gmail.com

Funding information
National Science Foundation, Grant/Award Number: NSF-2043905

Handling Editor: Luke Browne

Abstract

Rapid environmental change poses unprecedented challenges to species persistence. To understand the extent that continued change could have, 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 the performance of currently available offset methods (gradientForest, the Risk-Of-Non-Adaptedness, redundancy analysis with and without structure correction and LFMM2) using an extensive set of simulated data sets that vary demography, adaptive architecture and the number and spatial patterns of adaptive environments. For each data set, 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 4,849,600 of such evaluations, we find that (1) method performance is largely due to the degree of local adaptation across the metapopulation (*LA*), (2) *adaptive* marker sets provide minimal performance advantages, (3) performance within the species range 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 globally novel climates (i.e. a climate without an analogue within the species range) for metapopulations with greater *LA* than lesser *LA*. We discuss the implications of these results for management, assisted gene flow and assisted migration.

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

1 | INTRODUCTION

The impacts of climate change and habitat loss 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 such impacts, 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 ecological 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 modelling relationships between environmental and genetic variation to predict the 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 evaluations 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). This suggests that realized performance in natural systems may not meet expectations from evaluations in silico. Even so, these empirical results have often shown the expected negative relationship between predicted offset and common garden performance such as measures of juvenile growth (e.g. Fitzpatrick et al., 2021; Lind et al., 2024) and even 52-year mortality at multiple sites (Lind et al., 2024). Furthermore, 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 (Fitzpatrick et al., 2021; Lachmuth, Capblancq, Keller, et al., 2023; Láruson et al., 2022; Lind et al., 2024). Together, these results suggest that genomic offset methods may provide valuable insight for management without needing to identify adaptive loci, but the number of evaluations has been relatively small with few comparisons among methods. Thus, little is known about how robust these methods are across a wide array of realistic empirical scenarios and the comparative performance among available methods. Other signals within marker data, such as the degree to which allele frequencies are clinal across environmental gradients, also require further exploration, particularly for methods that may algorithmically emphasize such patterns over those more relevant to environmental selection, or for those methods that may rely upon such clinal patterns to maintain accuracy in predictions. 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 the performance of a method is affected by aspects of the evolutionary history of sampled populations, whether the type of signals in putatively ideal data sets that may mislead offset inference (e.g. clinal allele frequencies), how important it is to identify the environmental drivers of local adaptation a priori and how consist predictive performance is across the landscape. Finally, because novel climates with no recent analogue 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 novel 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 present ideal circumstances for understanding the opportunities and limits of genomic offsets because there is no error in the measurement of allele frequencies, environmental variables, individual fitness or the drivers of local adaptation. 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 analogue available for model training; Lotterhos et al., 2022).

Here, we used a wide array of previously published, biologically realistic, contrasting and adversarial simulations from Lotterhos (2023a) in an attempt to more precisely define the limits of predictive performance of five implementations of four genomic offset methods (Table 1): gradientForest ($\mathrm{GF}_{\mathrm{offset}}$; sensu Fitzpatrick & Keller, 2015), the Risk Of Non-Adaptedness (RONA, Rellstab et al., 2016), Latent Factor Mixed Models ($\mathrm{LFMM2}_{\mathrm{offset}}$, sensu Gain & François, 2021) and redundancy analysis with and without correction for population structure ($\mathrm{RDA}_{\mathrm{offset}}$, sensu Capblancq &

TABLE 1 Genomic offset methods used for evaluation.

Method	Abbr.	Multivariate?	Structure correction?
gradientForest[1]	GF_{offset}	Yes	No
Redundancy Analysis[2] with population structure correction	RDA-corrected	Yes	Yes, with axes loadings from PCA[a]
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

Note: 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.

[a] Principal component analysis.

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 cite analysis code used to carry out specific analyses in-line with the text. Text S1 and 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 × 10 grid evolving across a heterogeneous landscape (Lotterhos, 2023a). Each data set was simulated under a combination of the following four evolutionary parameters: (i) three landscapes (10 populations × 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 ($n_{traits} \in \{1, 2\}$), presence of pleiotropy (when $n_{traits} = 2$) and variability of selection strength across individual traits (see Figure 1 in Lotterhos, 2023a).

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 (Figure S37). The adaptive trait(s) undergoing selection responded to either a latitudinal temperature gradient (*temp*; $n_{traits} = 1$), or to both *temp* and a longitudinal '*Env2*' gradient ($n_{traits} = 2$). *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 10 linkage groups, and neutral mutations were added to all 20 linkage groups with tree sequencing (for details see Lotterhos, 2023a). Adaptive traits were determined additively by effects of QTNs.

In all simulations, phenotypic clines evolved between each trait and the selective environment (Lotterhos, 2023a), where populations became locally adapted to their environment, measured at the metapopulation level as the mean difference of demes in sympatry

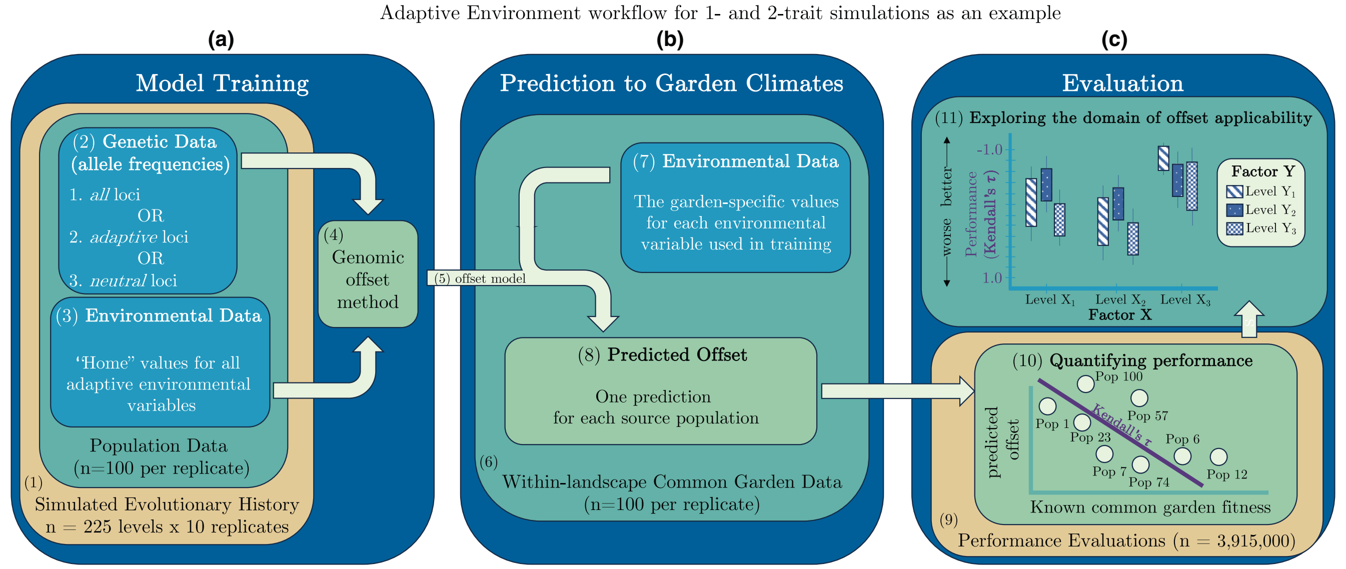

FIGURE 1 Analysis of 1-, 2- and 6-trait simulations included three main phases: (a) model training, (b) model prediction and (c) evaluation of models. The *Adaptive Environment* workflow is shown as an example of the processing of 1- and 2-trait simulation data for genomic offset evaluation. In total,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

minus allopatry ($LA_{\Delta SA}$, Blanquart et al., 2013). $LA_{\Delta SA}$ equates to the average levels of local adaptation at the deme level which can be calculated for each deme by both home-away ($LA_{\Delta HA}$) and local-foreign ($LA_{\Delta LF}$) 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 data sets (Lotterhos, 2023a), 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, 10 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 data sets (225 levels*10 replicates), we also included a non-Wright–Fisher case with range expansion from three refugia and secondary contact (Figure 6 in Lotterhos, 2023a). This simulation evolved variable degrees of admixture across the landscape. Six moderately polygenic environmental traits ($n_{\text{traits}}=6$) 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 (2023a).

2.2 | Evaluation of offset methods

We investigated the performance of five implementations of four genomic offset methods (Table 1): $\text{GF}_{\text{offset}}$, $\text{RDA}_{\text{offset}}$, $\text{LFMM2}_{\text{offset}}$ and RONA. While $\text{GF}_{\text{offset}}$, $\text{RDA}_{\text{offset}}$ and $\text{LFMM2}_{\text{offset}}$ can use multivariate environmental data to train models, RONA is univariate and can only account for a single environmental variable at one time (Table 1). Additionally, while $\text{GF}_{\text{offset}}$ and RONA do not apply correction for population genetic structure, $\text{LFMM2}_{\text{offset}}$ does by default, and structure correction with $\text{RDA}_{\text{offset}}$ is optional. We thus evaluate $\text{RDA}_{\text{offset}}$ 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 Text S1.1–S1.4 and Figures S1–S3.

We varied construction of genomic offset training data sets for each replicate of the 1-, 2-, and 6-trait simulations by varying the marker set used in model training (Figure 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 predicted offset from each model for each population to all 100 within-landscape common gardens from a full factorial in silico reciprocal transplant design (Figure 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, 2023a) and projected population offset (Figure 1c).

TABLE 2 Workflows used to process simulation data for the evaluation of genomic offset methods.

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			3000
Nuisance environment (NE)[b]	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[a]
	6-trait	1 (1)	[$AE_{6\text{-trait}}$ environments + ISO + PSsd + TSsd]			1200[a]
Climate novelty (CN)[c]	1-trait	45 (10)	$AE_{1\text{-trait}}$ environments	Prediction only (using AE training models)	No	64,800[a]
	2-trait	180 (10)	$AE_{2\text{-trait}}$ environments			259,200[a]
	6-trait	1 (1)	$AE_{6\text{-trait}}$ environments			144[a]

Note: Numbers given in column names refer to locations in schematic of Figure 1. The *Adaptive Environment* workflow processes all population data from 1- and 2-trait (example shown in Figure 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 (Figure 1a-5) and evaluates offset in 11 novel environments relative to the range of environments used in training. See Note S3 for details regarding the choice of *Climate Novelty* environmental values and visualizations of climate data in principal component space. See Notes S1 and S2 for descriptions of coding workflows. Total evaluations = 4,849,644. Counts of evaluations were tabulated in SC 02.10.01.

[a] Excludes RONA.

[b] The set of population values for each unique nuisance environment was the same across traits and landscapes.

[c] Includes evaluation of climate centre and 11 *Climate Novelty* scenarios.

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, Figure 1c-11). We refer to the preceding processing of data as the *Adaptive Environment* workflow (Figure 1, Table 2). Across all workflows, all adaptive environments were included in training and were never excluded.

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 (Figure 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 centre) and evaluate performance in each garden. We refer to this workflow as the *Climate Novelty* workflow. See Text S3 and Figure S5 for details regarding the choice of environmental values for novelty scenarios.

2.3 | Study questions

2.3.1 | 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, 2023a), degree of local adaptation (ΔSA) and common garden ID. Specifically,

$$Y_{ij} = L_i + D_i + GL_i + P_i + p_{cQTN,t,i} + p_{cNeut,t,i} + p_{cQTN,Env2,i} + p_{cNeut,Env2,i} + LA_{\Delta SA,i} + G_j \tag{1}$$

where Y_{ij} is the within-landscape performance (Kendall's τ) 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 Figure 1 of Lotterhos (2023a).

2.3.2 | Q2 – How is offset performance affected by the proportion of clinal alleles in the data?

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 (Equation 2) to test the hypothesis that clinal alleles differentially impact model performance, independent from the other factors from Equation 1:

$$Y_{ij} = p_{cQTN,t,i} + p_{cNeut,t,i} + p_{cQTN,Env2,i} + p_{cNeut,Env2,i} \tag{2}$$

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

2.3.3 | Q3 – Is method performance driven by causal loci or by genome-wide patterns of isolation by environment?

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 ($\overline{N_{adaptive}} = 188$, $\overline{N_{all}} = 33{,}169$, $\overline{N_{neutral}} = 16{,}520$). 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 measured IBE as the rank correlation (Spearman's ρ) between population pairwise F_{ST} (Weir & Cockerham, 1984) and Euclidean climate distance of adaptive environmental variables.

2.3.4 | Q4 – What is the variation of model performance across the landscape?

Within a landscape, offset methods may not have high predictive performance at every site or every environment. Understanding variability 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.

2.3.5 | Q5 – How does the addition of non-adaptive nuisance environments in training affect performance?

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 used nuisance environmental variables from Lotterhos (2023a) 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 (Figure S4). These three nuisance environments differ from previous implementations of such variables (Láruson 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.

2.3.6 | Q6 – How well do offset models extrapolate to novel environments outside the range of environments used in training?

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 (Text S3). We represent these scenarios as standard deviations from the centre 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, 2023a).

3 | RESULTS

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

The ANOVA model (Equation 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; Figure S6). Within the simulations, $LA_{\Delta SA}$ was impacted by pleiotropy, the relative strength of selection and landscape (Figure S7; see also Figure S2a,b in Lotterhos, 2023a), 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 $LA_{\Delta SA}$ (Figure 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*

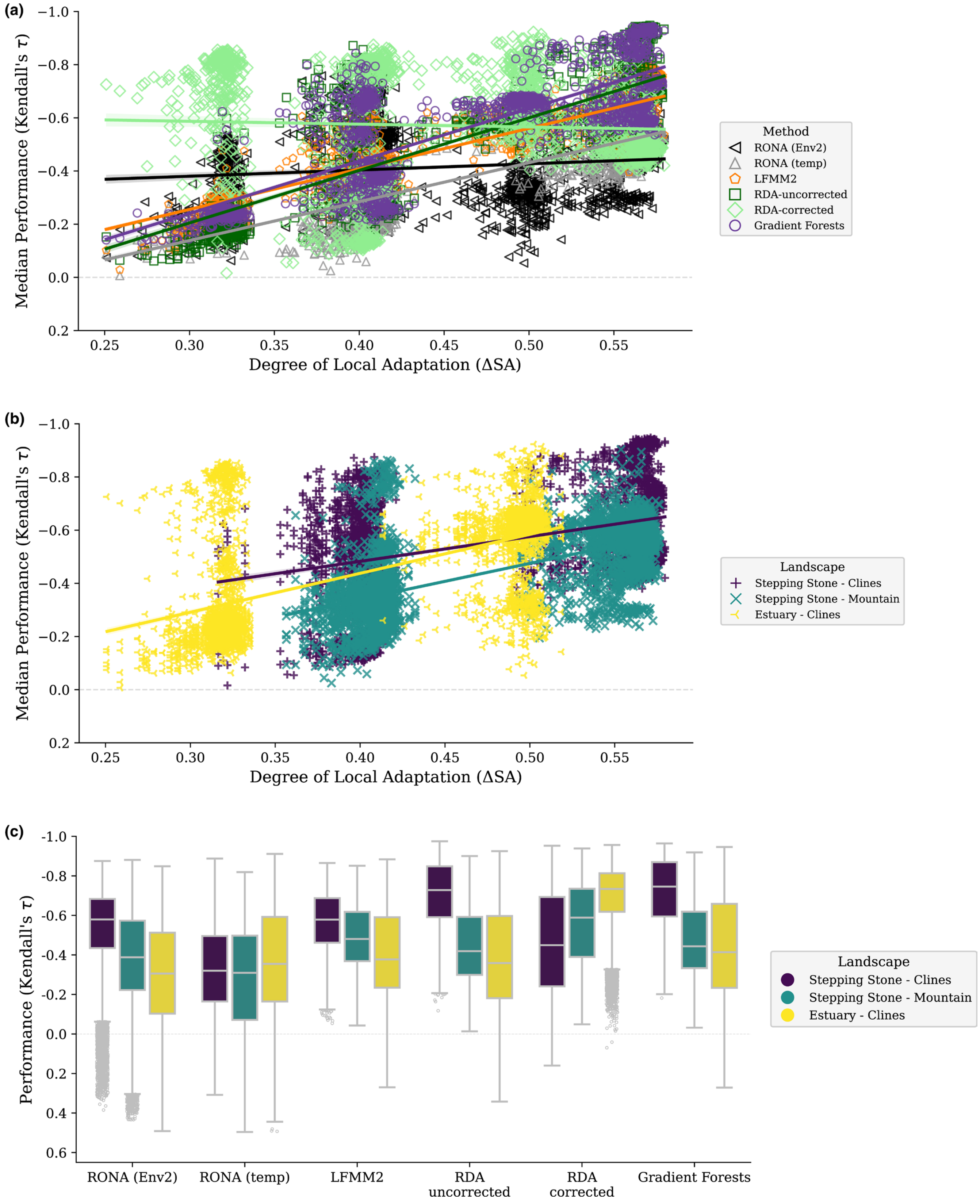

FIGURE 2 Predictive performance of genomic offset models (y-axes) is driven by the degree of local adaptation (a) and the spatial patterns of adaptive environments across the landscape (b, c). For each model, a median value from performance scores from 100 common gardens is shown for (a and b); c shows 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 were processed through the *Adaptive Environment* workflow but only include 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.

– *Mountain* landscapes (Figure 2b) despite similar levels of $LA_{\Delta SA}$ (Figure 2a). Offset methods also generally performed better in the two *Stepping Stone* landscapes than the *Estuary – Clines* landscape (Figure 2b). However, there were some interactions between method and landscape (Figure 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 (Figure S8). Additionally, while demography primarily influenced population differentiation across the landscape with little impact on $LA_{\Delta SA}$ within simulations (Table S2 in Lotterhos, 2023a), migration breaks between populations and latitudinal clines in population size generally decreased offset performance for $\text{LFMM2}_{\text{offset}}$, $\text{GF}_{\text{offset}}$ and RDA-uncorrected (Figure S9).

3.2 | Q2 – How is offset performance affected by the proportion of clinal alleles in the data?

The sum of squares from Equation 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 (Equation 2) indicated that p_{cNeut} accounted for 4.14–9.65 times the variation than did p_{cQTN} for $\text{GF}_{\text{offset}}$, $\text{LFMM2}_{\text{offset}}$ and RDA-corrected. For $\text{GF}_{\text{offset}}$ and RDA-uncorrected, $p_{cNeut,Env2}$ accounted for >16% of the sum of squares (Table S2, Figure S10).

Overall, relationships between performance and p_{cNeut} (second column, Figure S11) were stronger than between performance and p_{cQTN} (first column, Figure S11). However, sometimes performance increased with p_{cNeut} and sometimes it decreased, depending on the method (Figure 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 (Figure S12) drove relationships with p_{cNeut} (Figure S13) which in turn drove relationships with performance (Figures S11 and S14).

3.3 | Q3 – Is method performance driven by causal loci or by genome-wide patterns of isolation-by-environment?

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 (Figure 3a–c). For RDA-corrected, the *neutral* markers performed slightly better than either *adaptive* or *all* markers in 2-trait evaluations (Figure 3e). *Adaptive* markers from

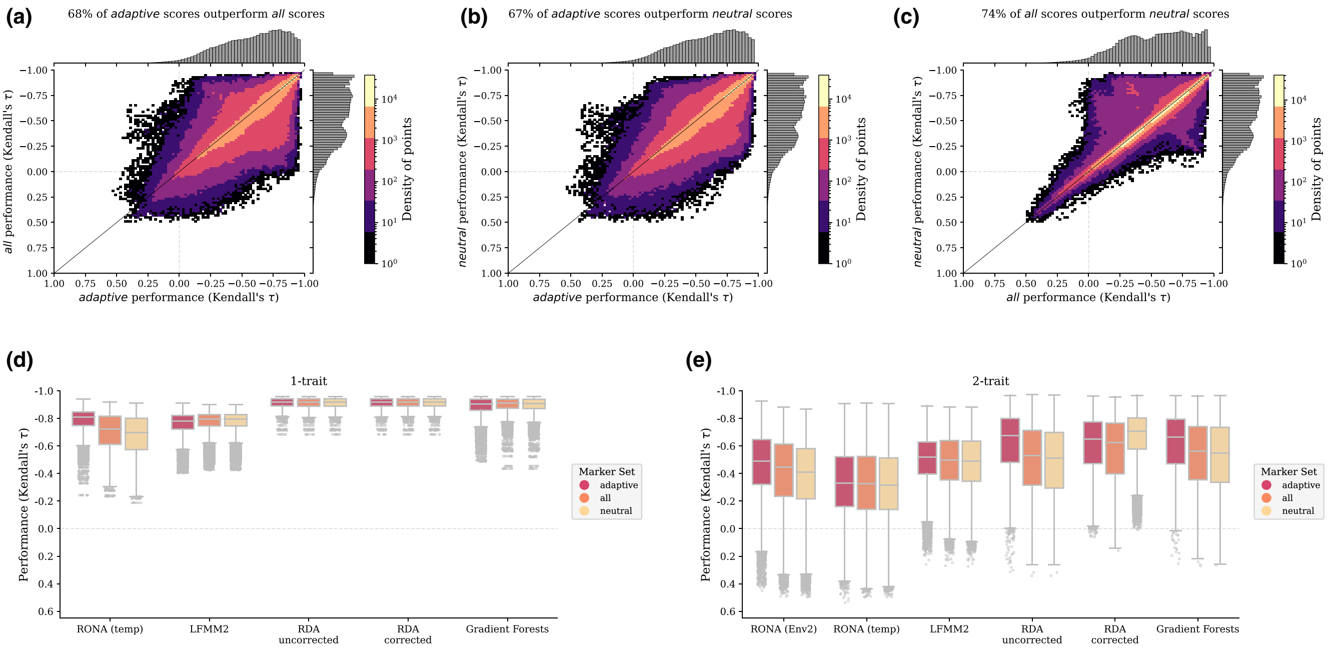

FIGURE 3 Comparison of marker choice across genomic offset methods for 1- and 2-trait simulations. (a–c) Scatterplots of pairwise comparisons of performance between marker sets (histograms in each margin) from both 1- and 2-trait models where density of points is indicated by colour in legend (note colour scale is different for each figure to accentuate patterns in data). (d, e) 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.

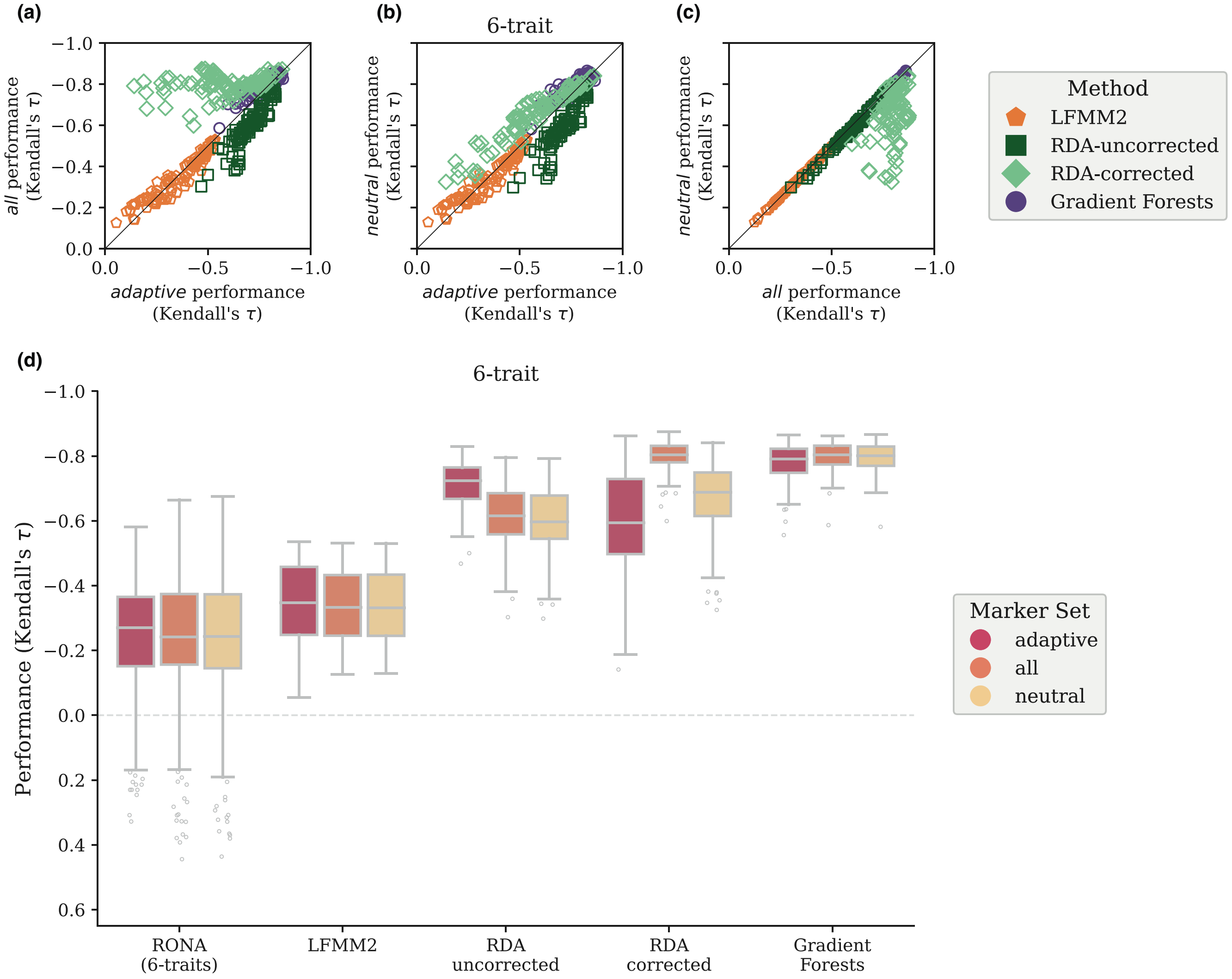

FIGURE 4 Comparison of marker choice across genomic offset methods for the 6-trait simulation. (a–c) Scatterplots of pairwise comparisons of performance between marker sets (RONA is not shown, except in SN 02.05.10). (d) Boxplots from the same data in a–c ($\text{RONA}_{6\text{-traits}}$ is the combined 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.

6-trait evaluations provided varied performance advantages across methods (Figure 4; Figure S37).

The *adaptive* marker sets had relatively elevated levels of *IBE* compared to sets of *neutral* or *all* markers in 1- and 2-trait simulations, but levels of *IBE* were nonetheless quite similar between marker sets (Figure 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 (Figure S16).

Intriguingly, levels of *IBE* found within a landscape (Figure S17a) did not correspond to the degree of $LA_{\Delta SA}$ that developed (Figure 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 (Figure S18). As such, *IBE* from *all* markers explained very little variation in performance when added as a factor to the ANOVA model from Equation 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 were generally unrelated to the differences in *IBE* between the two marker sets used to train the models (Figure S19).

3.4 | Q4 – What is the variation of model performance across the landscape?

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

(Figures S20–S23). Variation in performance was also observed for 6-trait models (Figure 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 (Figures S24–S26). Ultimately, we found no strong indicator for predicting when a model will be highly variable. Indeed, while performance generally increased with $LA_{\Delta SA}$ (Figure 2), variability in performance was not strongly related to the variability in deme-level *LA* on the landscape (Figures S27–S29). Despite $LA_{\Delta SA}$ driving performance more generally (from Q1), this indicates that variation in model performance across the landscape was not strongly driven by metapopulation levels of, nor deme-level variation in, *LA*.

3.5 | Q5 – How does the addition of non-adaptive nuisance environments in training affect performance?

Training offset models with the addition of non-adaptive nuisance environmental variables generally reduced offset method performance (Figure 5). This decline was most dramatic for offset trained on 1-trait simulations (Figure 5a) compared to the decline observed for 2-trait (Figure 5b) and 6-trait (Figure 5c) simulations. The only instances for which median performance did not decrease monotonically with nuisance level were for 2-trait simulations evaluated with $\mathrm{GF}_{\mathrm{offset}}$ (Figure S30).

Overall, landscape had the most influence over performance differences due to non-adaptive nuisance environments (Figure 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 data sets where the advantages were more substantial compared to 2-trait data sets (Figures S31 and 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).

3.6 | Q6 – How well do offset models extrapolate to novel environments outside the range of environments used in training?

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, Figure 6). Importantly, declines in performance for data sets 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; Text S4, Figure S33). Despite little change in the median performance for data sets with low levels of LA, most performance scores from these data sets did not exceed Kendall's $\tau = -0.5$, and therefore had little predictive value in novelty scenarios. Performance of individual methods followed the overall trend presented in Figure 6 (Figure S35).

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

4 | DISCUSSION

In the last decade, genomic offset methods have been identified as a complement to other ecological forecasting models because they incorporate intraspecific variation (Capblancq et al., 2020; Fitzpatrick & Keller, 2015; Rellstab et al., 2021). Our evaluations show that offset methods may be impacted by both the evolutionary history of sampled populations as well as the decisions made during model training. Our analyses emphasize that performance of offset methods is highest when applied to locally adapted populations with known drivers of environmental selection, and when restricting offset projections to climates similar to those used in training. These results suggest that there may be a number of scenarios for which genomic offset methods may provide valuable insight for management, but also highlight that they will not yield accurate results in every situation. 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 $LA_{\Delta SA} \sim 0.5$ 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 meta-analyses 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 & Fischer, 2008), we may therefore expect some genomic offset methods to do reasonably well when predictions are made for environmental values most similar to those used in training, and when local adaptation in the metapopulation is high (e.g. when $LA_{\Delta SA} > 0.5$, the

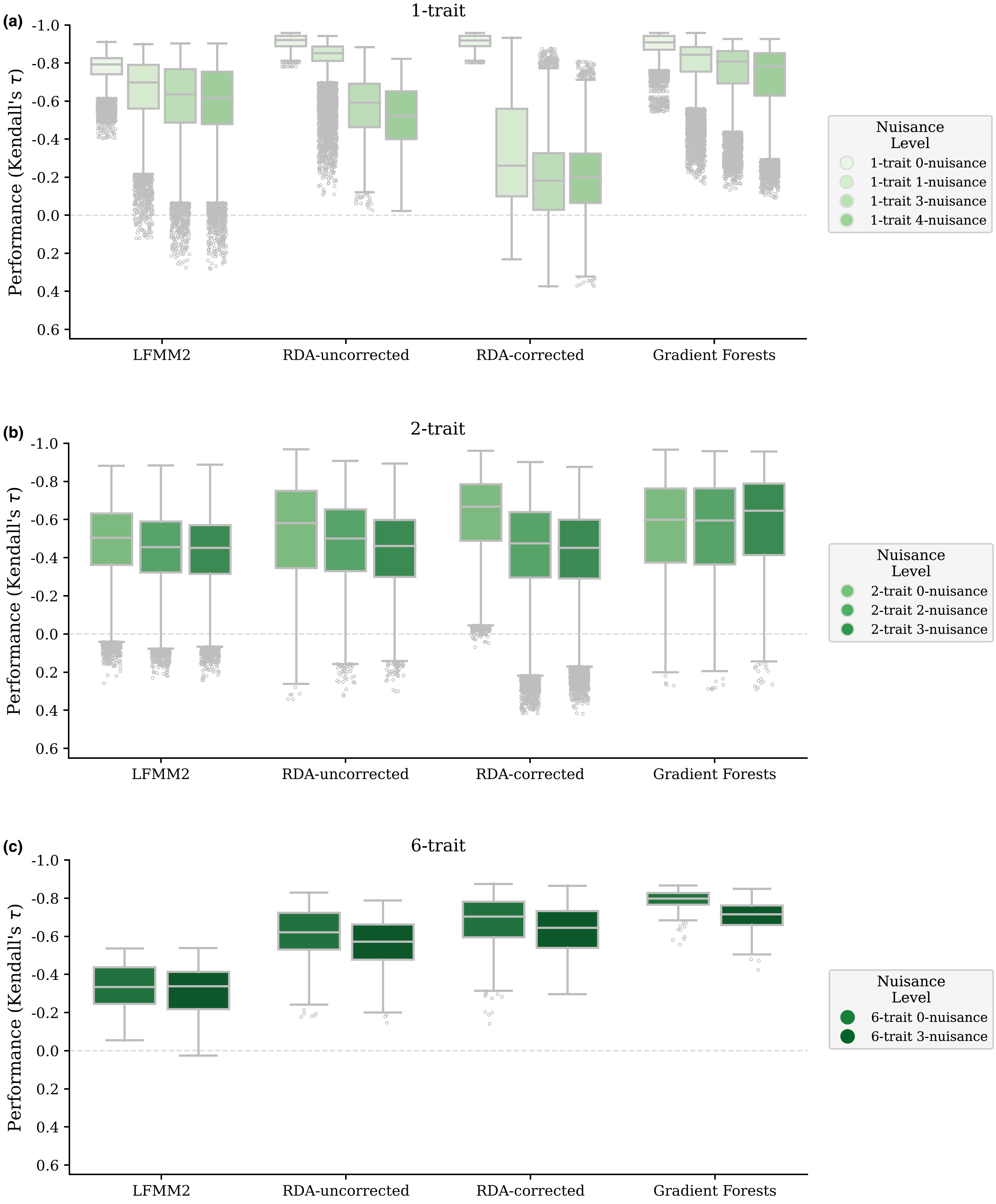

FIGURE 5 Effect of non-adaptive nuisance environmental variables on offset performance. Shown are evaluations of offsets from 1- and 2-trait models trained using only adaptive environments (*0-nuisance*) or with adaptive environments and the addition of $N > 0$ non-adaptive environmental variables (*N-nuisance*). 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.

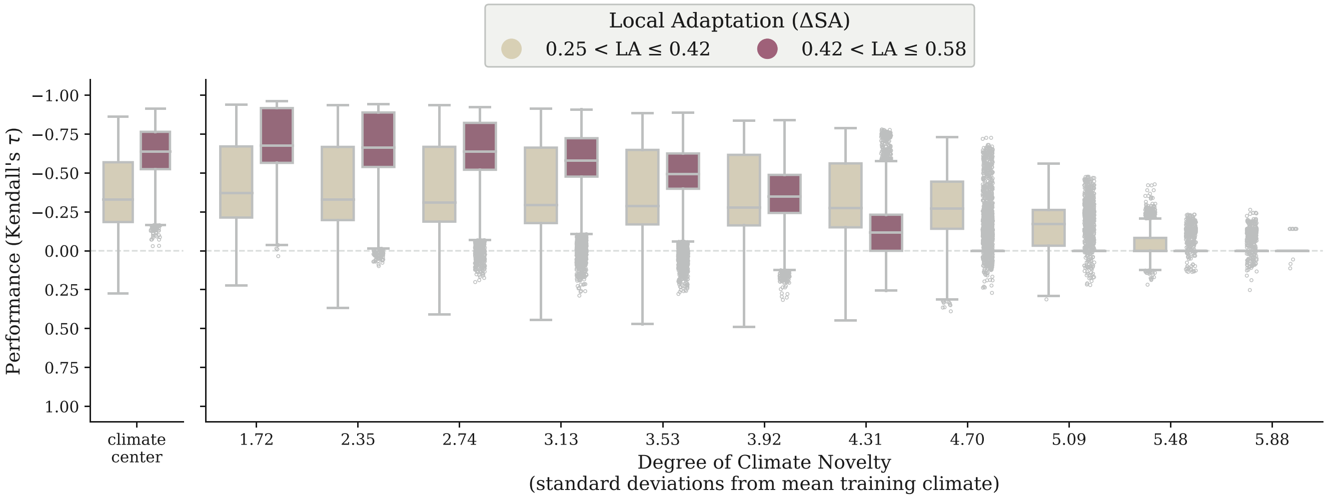

FIGURE 6 Performance decays with climate novelty relative to training data. Shown is model performance (y-axes) across methods at climate centre and across common gardens each representing increasing degrees of climate novelty relative to training data (right-most x-axis) where all 100 populations have been transplanted. The standard deviation values 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 Text S3. Data included in this figure are from models trained using 1- and 2-trait simulations from the *Climate Novelty* workflow, and excludes both RONA$_{\text{temp}}$ and RONA$_{\text{Env2}}$. Code used to create this figure can be found in SC 02.04.05.

best performing methods, RDA-uncorrected and GF$_{\text{offset}}$, had an average correlation between offset and fitness of ~−0.6). However, even moderately high levels of local adaptation had substantial decline in performance (e.g. when $LA_{\Delta SA} \approx 0.3$, the best performing methods, RDA-uncorrected and GF$_{\text{offset}}$, had an average correlation between GO and common garden fitness of ~−0.2). Future studies should seek to demonstrate evidence for strong local adaptation in the system prior to applying genomic offset methods, but ultimately the level of predictable performance that is acceptable will depend on the stakeholders and management goals in the system. In the absence of evidence for local adaptation for the target system, additional data that can be used to verify offset predictions (such as common garden[s], demographic data or remote sensing data) should be used to verify offset predictions (Lind et al., 2024).

4.2 | The importance of the signals within genomic marker sets

Initial implementations of genomic offset models focused on putatively adaptive markers where the signal of local adaptation may be strongest (Fitzpatrick & Keller, 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) and our results are consistent with these studies. We found that the *adaptive* marker sets provided a slight 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 similar to all markers is that genome-wide data capture 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. 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.

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, we measured IBE as the correlation between pairwise environmental distance and pairwise F_{ST}, and did not incorporate geographic distance. In doing so, our measure of IBE distilled genetic distance down to a single value from a large number of loci and gave less weight to loci with rare alleles. In future studies, creating a fourth marker set based on loci with the highest IBE signals to compare with other marker sets offers another opportunity to understand the impact of IBE on performance. Empirical data sets 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 have varied impacts on performance, sometimes being positively related to performance and sometimes negatively depending on the context. In addition to the effect of clinal allele patterns, our results also suggest that nonmonotonic relationships between environmental gradients and allele frequencies will be an important element to consider during model building, as the methods that strictly modelled linear relationships (RONA, LFMM2) had on average poorer performance than other methods we evaluated ($\mathrm{GF}_{\mathrm{offset}}$, $\mathrm{RDA}_{\mathrm{offset}}$). In contrast, we found very similar performance across *adaptive*, *neutral* and *all* loci. Although, while the similarity in performance among marker sets evaluated here may indicate that it is often unnecessary to subset loci for model training to those that are putatively adaptive, testing multiple marker sets during model exploration will be important to understand model sensitivity (Lind et al., 2024). These and other signals within data that could improve or mislead offset models, including the choice of how and when to correct for population structure, also warrant further investigation.

4.3 | The importance of adaptive environmental variables

In empirical settings, the environmental 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 $\mathrm{RDA}_{\mathrm{offset}}$) 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 data sets 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 $\mathrm{RDA}_{\mathrm{offset}}$ are likely to perform worse in, or less uniformly across, realistic empirical settings than what our current findings suggest.

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 data sets, depending on which marker set was used. These importance values ultimately affect the model predictions. Including all available environmental variables may therefore negatively impact $\mathrm{GF}_{\mathrm{offset}}$ 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 data sets with the greatest levels of $LA_{\Delta SA}$ were also the data sets 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, it will be important for future studies to clearly show what offset predictions have been extrapolated to environments that have no analogue in the data used to train the model. This is particularly relevant for applications of offset methods that attempt to estimate the in situ risk of climate change where the environment is expected to be increasingly novel. Global climate novelty is often measured with respect to historical variability across a species range (e.g. Lotterhos et al., 2021; Mahony et al., 2017; Williams et al., 2007). Terrestrial systems could experience change in excess of three standard deviations relative to historic values (Williams et al., 2007), with similar indices in marine systems indicating potential for even greater novelty (Lotterhos et al., 2021). We observed performance declines below the analogous $z_E = 3.13$ standard deviations *Climate Novelty* scenario, indicating offset predictions could 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 for an empirical example that highlights model extrapolations).

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 Figure 5 in Capblancq et al., 2020). The simulations used here assumed a linear relationship between environment (whether current or novel) and optimal trait value, without a cost for the production of extreme traits. Thus, by assuming linearity in the simulations, we presented methods with a straightforward relationship to extrapolate to novel climates. In empirical settings, the relationship between the selective environment and optimal trait value could be nonlinear. Developing a clearer understanding of the relationship between the environment and optimal trait value, as well as various nonlinear methods for extrapolation (e.g. see Text S1), are important areas of future work.

4.5 | Genomic offsets in practice

Our evaluations show that genomic offset methods hold promise for predicting maladaptation to environmental change for metapopulations that evolve strong local adaptation, and within environments similar to those in the data used to train the model. However, our analyses also emphasize the limits of these methods in some scenarios and the variability of performance across methods. Although $\text{GF}_{\text{offset}}$ and RDA-uncorrected generally had greater performance than RDA-corrected, LFMM2 or RONA, there was no single method that outperformed the others across all situations. For instance, unlike other methods that performed best in *Stepping Stones – Clines* landscapes, RDA-corrected had highest performance in *Estuary – Clines* landscapes. RDA-corrected also had greater performance than other methods in many of the data sets that developed low levels of $LA_{\Delta SA}$, but this was not the case when local adaptation was high. 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 programmes 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 (e.g. from marker sets and the populations used; 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.

AUTHOR CONTRIBUTIONS

Katie E. Lotterhos received funding. Katie E. Lotterhos and Brandon M. Lind conceptualized the project and methodology. With input, editing, and feedback from Katie E. Lotterhos, Brandon M. Lind wrote code to train and evaluate offset models, created figures, curated coding and records for archiving and wrote the manuscript.

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. We thank four anonymous reviewers for helpful comments that improved the clarity and presentation of our results.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

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. yml files are available to reconstruct the coding environments for the Rv3.5.1 (r35.yml) and python v3.8.5 (mvp_env.yml) environments in the Zenodo repository. The yml file to reconstruct the R v4.0.3 environment (MVP_env_R4.0.3.yml) is archived in Lotterhos (2023c). 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 Text S1 and S2. Data used for analysis have been archived previously (Lotterhos, 2023b). Code has been archived on Zenodo (Lind, 2024), which mirrors the GitHub repository at https://github.com/ModelValidationProgram/MVP-offsets.

ORCID

Brandon M. Lind https://orcid.org/0000-0002-8560-5417
Katie E. Lotterhos https://orcid.org/0000-0001-7529-2771

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*, *12*, 2298–2309. 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., Pielke, R., 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. (2018). 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.14138

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*, *21*, 2749–2765. 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.12376

Gain, C., & François, O. (2021). 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. 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*, *11*, 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), e05877. 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. (2024). *GitHub.com/ModelValidationProgram/MVP-offsets. Revision release (v1.0.1)*. Zenodo. https://doi.org/10.5281/zenodo.11209812

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? *Global Change Biology*, *30*, e17227. https://doi.org/10.1111/gcb.17227

Lotterhos, K. (2023b). *Output model data from paradox of adaptive trait clines with non-clinal patterns in the underlying genes (Model Validation Program project)*. Biological and chemical oceanography data management office (BCO-DMO). (version 1) version date 2023-02-13 [if applicable, indicate subset used]. https://doi.org/10.26008/1912/bco-dmo.889769.1 [access date].

Lotterhos, K. E. (2023a). The paradox of adaptive trait clines with non-clinal patterns in the underlying genes. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(12), e2220313120. https://doi.org/10.1073/pnas.2220313120

Lotterhos, K. E. (2023c). Code from: "The paradox of adaptive trait clines with non-clinal patterns in the underlying genes" (v1.0). Zenodo.. https://doi.org/10.5281/zenodo.7622893

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/10.1146/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), 4–18. 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 of the United States of America*, *104*(14), 5738–5742. https://doi.org/10.1073/pnas.0606292104

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lind, B. M., & Lotterhos, K. E. (2025). The accuracy of predicting maladaptation to new environments with genomic data. *Molecular Ecology Resources*, *25*, e14008. https://doi.org/10.1111/1755-0998.14008