

When does accounting for gene–environment interactions improve complex trait prediction? A case study with *Drosophila* lifespan

Fabio Morgante ^{1,2,*} Francesco Tiezzi ^{1,3}

¹Institute for Human Genetics, Clemson University, Greenwood, SC 29646, United States

²Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634, United States

³Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Florence 50144, Italy

*Corresponding author: Institute for Human Genetics, Clemson University, 114 Gregor Mendel Circle, Greenwood, SC 29646, USA. Email: fabiom@clemson.edu

Gene–environment interactions (G×E) have been shown to explain a non-negligible proportion of variance for a plethora of complex traits in different species, including livestock, plants, and humans. While several studies have shown that including G×E can improve prediction accuracy in agricultural species, no increase in accuracy has been observed in human studies. In this work, we sought to investigate the scenarios in which accounting for G×E is expected to improve prediction accuracy. Model organisms are useful for studying G×E, since environments can be defined precisely, and genotypes can be replicated across environments, which are ideal conditions to minimize confounding in G×E analyses. Thus, we used data from an experiment in *Drosophila melanogaster*, where researchers measured lifespan in different environments for unrelated inbred lines (i.e. genotypes). We used three different cross-validation (CV) scenarios that mimic different relationships between reference and test populations, and fitted a few statistical models with and without including G×E. The results showed that G×E explained 8% of lifespan variance. Despite that, models accounting for G×E improved prediction accuracy only in CV scenarios where the same genotypes are observed in both the reference and test populations. While these scenarios are common in agriculture, where individuals of the same family or variety appear in both populations, they are not commonly encountered in human studies, where individuals are unrelated. Thus, our work shows in which prediction scenarios we can expect improvements by accounting for G×E, and may provide a potential reason (among others) for results of human studies.

Keywords: G×E; polygenic scores; genomic prediction

Introduction

Estimating the importance of gene–environment interactions (G×E) on complex traits and accounting for it in phenotype prediction is one of the most difficult challenges faced by plant and animal breeders (Bandeira e Sousa et al. 2017). Recently, this topic has also become of interest in human genetics (Kerin and Marchini 2020; Zhou and Lee 2021; Durvasula and Price 2025; Goda et al. 2025; Miao et al. 2025; Tiezzi et al. 2025).

In plant breeding, where G×E has the largest impact and has been studied most extensively, the estimation of such interaction largely relies on multi-environment trials (Xavier et al. 2025). Here, some varieties of a given species are planted in multiple locations, managed under different agronomic techniques and exposed to different pedo-climatological conditions. The estimates of G×E will be non-null when the performance of a given genotype varies across the different locations (Xavier et al. 2025). This implies that the genotype is sensitive to environmental conditions in its expression. G×E can result in different phenotypic variance across genotypes, depending on the environmental conditions in which the genotypes are reared, while the rank of the genotypes remains the same. However, G×E can also affect the rank of the genotypes across environments (Van Eeuwijk et al. 2016).

Non-linearity of the response of the genotypes with respect to the environmental gradient adds to the complexity of G×E (Heslot et al. 2014).

Plant breeders leverage this interaction for developing cultivars that thrive in specific conditions or, perhaps, can perform well in multiple conditions (Elias et al. 2016; Van Eeuwijk et al. 2016). Similarly, animal breeders aim to estimate how the same or similar genotypes can acclimate to different conditions, by using data from related individuals (e.g. paternal half-sibs) that are raised in different environments (Bryant et al. 2005; Hansen 2020). This is done mostly to breed for animals that can be tolerant to stress (e.g. heat stress), but also to develop “the right genotype for that environment”, like ruminants that can adapt well to grazing conditions (Sheahan et al. 2011).

Many statistical approaches for plant and animal breeders have been developed (Hu et al. 2025). These approaches can be summarized as: (i) multiple trait models (MTMs) that consider the trait of interest as a series of correlated traits, each defined by the environment under which the trait is manifested; (ii) random regression models (RRM) that model the change in the performance across environments through the use of covariance functions, with conditions described by one or a few environmental covariates; and (iii) Reproducing Kernel Hilbert Space

regression (RKHS) models that can handle a large number of environmental covariates in interaction with genotypes through the use of similarity matrices (i.e. kernels). All these models have been described in the literature, and the relationships between them can be demonstrated, at least in theory (Jarquín et al. 2014; Bandeira e Sousa et al. 2017; Martini et al. 2020; Hu et al. 2025). It should be noted that MTM is the least parsimonious in terms of parameters to estimate when the number of environments or environmental descriptors becomes large. On the other hand, RKHS is the most parsimonious, as it was developed specifically to handle a large number of environmental covariates (Jarquín et al. 2014). Appropriate modeling of G×E is pivotal in understanding the complexity of this phenomenon, and different strategies have been proposed to evaluate the different methods. These strategies include benchmarking the prediction methods for their ability to rank genotypes regardless of the environmental conditions, or within specific conditions (e.g. within-year-location correlation) (Lopez-Cruz et al. 2023).

Models that incorporate G×E have shown good promise in plant and animal breeding. For example, in plants, Jarquín et al. (2017) showed that G×E improved prediction accuracy when modeling the performance of wheat lines grown across dozens of locations, while Acosta-Pech et al. (2017) successfully modeled general and specific combining ability in interaction with the environment to predict the performance of maize hybrid lines. In addition, Cuevas et al. (2016) showed that G×E implemented in RKHS models produced advantages with non-null correlations between environments. In livestock, both Bohlouli et al. (2019) and Tiezzi et al. (2017) found a clear advantage in modeling G×E using climate data in dairy cattle and RRM or RKHS models, respectively. On the other hand, Bussiman et al. (2025) did not find a clear advantage in the inclusion of G×E in pigs.

Studies in plants also showed that the benefits of including G×E in prediction models are strongly dependent on the structure of and the relationship between the training and test populations (Crossa et al. 2014; Millet et al. 2019). For example, incorporating G×E showed the greatest advantage when predicting untested combinations of tested genotypes and tested environments (i.e. incomplete field trials). On the other hand, incorporating G×E did not show a marked improvement in prediction accuracy when predicting untested genotypes in tested environments (Burgueño et al. 2012) or tested genotypes in untested environments (Jarquín et al. 2017).

In human genetics, there are additional difficulties to study G×E because researchers cannot design optimal experiments and need to rely on large observational data such as biobanks (Bycroft et al. 2018). These data have a large amount of missing records, rely on noisy self-assessed information about the environment, lack replication of genotypes (i.e. individuals) across environments, and genotypes are non-randomized with respect to environments (Tiezzi et al. 2025). These problems make estimating the magnitude of G×E challenging. Despite that, non-negligible contribution of G×E to the phenotypic variance of several complex traits has been shown (Dahl et al. 2020; Kerin and Marchini 2020; Miao et al. 2025). However, accounting for G×E has generally not resulted in an increase in out-of-sample prediction accuracy (Zhou and Lee 2021; Goda et al. 2025; Tiezzi et al. 2025; Weine et al. 2025). These results disagree with those obtained in agricultural breeding.

To study in which scenarios accounting for G×E is expected to improve prediction accuracy and to investigate the contradictory results between humans and agricultural species, model

organisms can provide valuable insights. As opposed to humans and, to some extent, agricultural species, experiments in model organisms can be carefully controlled. This results in precise environment definitions, accurate phenotypic measurements, and replication of genotypes across environments, which are ideal conditions for G×E analyses and minimize confounding. In this study, we used data from an experiment that measured lifespan for flies from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) inbred lines raised in different environments (Huang et al. 2020). One peculiarity of these lines is that they are largely unrelated (Huang et al. 2014), which differs from populations used in similar studies performed in plants (Burgueño et al. 2012; Jarquín et al. 2017). Using these data and previously devised prediction scenarios, we sought to investigate the conditions in which accounting for G×E in different statistical models improves out-of-sample prediction accuracy.

Materials and methods

Data processing

We used phenotypic data from Huang et al. (2020). Lifespan (in days) was measured for 186 inbred lines of the DGRP for the two sexes at three different temperatures (18°C, 25°C, 28°C). After removing lines that had missing values in at least a sex/temperature combination, we were left with $n = 176$ lines, each measured in $c = 2$ context variables (resulting in $r = 6$ sex/temperature combinations), for a total of $q = 1,056$ records. The genotype data were filtered to remove genetic variants with minor allele frequency (MAF) smaller than 0.05 and missing genotype rate greater than 0.2. These filters retained $p = 1,899,439$ genetic variants.

Statistical models

We analyzed the data using RKHS regression models described in Jarquín et al. (2014):

- G-BLUP. $y_{ij} = \mu + a_i + \varepsilon_{ij}$.
- E-BLUP. $y_{ij} = \mu + e_j + \varepsilon_{ij}$.
- GE-BLUP. $y_{ij} = \mu + a_i + e_j + \varepsilon_{ij}$.
- G×E-BLUP. $y_{ij} = \mu + a_i + e_j + ae_{ij} + \varepsilon_{ij}$.

Here, y_{ij} is the phenotype of line i in environment j , μ is the intercept value, a_i is the random additive genetic value of line i [$\mathbf{a} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{ZGZ}'\sigma_a^2)$, $\mathbf{Z}$ is a $q \times n$ incidence matrix with $q = 1,056$ and $n = 176$, $\mathbf{G}$ is a $n \times n$ genomic relationship matrix (GRM) computed as in VanRaden 2008], e_j is the random environmental value of environment j [$\mathbf{e} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{E}\sigma_e^2)$, $\mathbf{E}$ is a $q \times q$ matrix of similarity based on environmental variables, computed as $\mathbf{E} \propto \mathbf{XX}'$, $\mathbf{X}$ is a $q \times c$ matrix of environmental measurements, with $c = 2$], ae_{ij} is the random gene-environment value of line i in environment j [$\mathbf{ae} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{ZGZ}' \circ \mathbf{E}\sigma_{ae}^2)$], ε_{ij} is the residual value for line i in environment j [$\boldsymbol{\varepsilon} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_n\sigma_\varepsilon^2)$].

These models were fitted in a Restricted Maximum Likelihood (REML) framework, implemented in the sommer R package (Covarrubias-Pazarán 2016).

We also fit a MTM (Calus and Veerkamp 2011):

- mvG-BLUP.
 $\mathbf{Y} = \mathbf{1m}' + \mathbf{A} + \mathbf{R}$

Here, $\mathbf{Y}$ is an $n \times r$ matrix of phenotypic observations in the $r = 6$ environments, $\mathbf{1}$ is an n -vector of ones, $\mathbf{m}$ is an r -vector of intercept values, $\mathbf{A}$ is an $n \times r$ matrix of additive genetic values

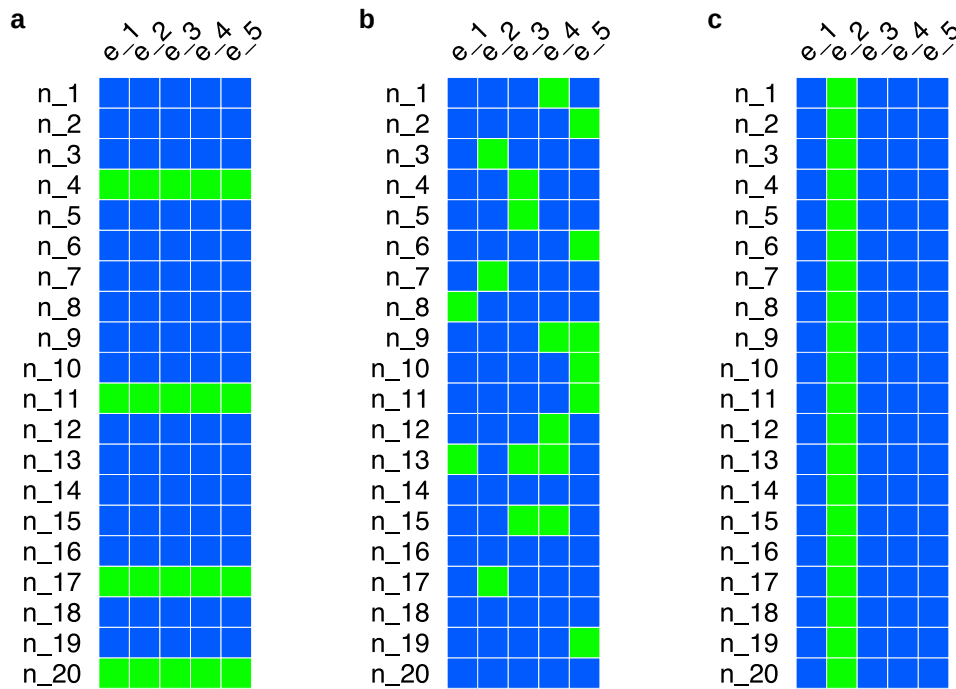


Fig. 1. Cross-validation schemes used to evaluate the prediction accuracy of the different models. a) *Random Lines*. b) *Random Observations*. c) *New Environment*. The blue squares represent observations in the training set and the green squares represent observations in the test set.

$[\mathbf{A} \sim \mathcal{MN}_{n \times r}(\mathbf{0}, \mathbf{G}, \mathbf{\Sigma}_A), \mathbf{\Sigma}_A$ is an $r \times r$ genetic covariance matrix], $\mathbf{R}$ is an $n \times r$ matrix of residual values [$\mathbf{R} \sim \mathcal{MN}_{n \times r}(\mathbf{0}, \mathbf{I}_n, \mathbf{\Sigma}_R)$, $\mathbf{\Sigma}_R$ is an $r \times r$ diagonal residual covariance matrix].

This model was fitted in a Bayesian framework, implemented in the BGLR R package (Pérez and de Los Campos 2014). We ran the sampler for a total of 300,000 iterations, discarding the initial 200,000 iterations as burn-in, followed by thinning every 50 iterations.

We also fit a RRM (Schaeffer 2004):

$$\bullet \text{ RRM. } y_{ij} = \sum_{t=0}^T \beta_t \phi_t(j) + \sum_{i=0}^T a_{it} \phi_t(j) + \varepsilon_{ij}$$

Here, y_{ij} is the phenotype for line i in environment j , $\phi_t(j)$ is the Legendre polynomial of order t for environment j , β_t is the fixed effect of the t th-order Legendre polynomial, a_{it} is the random additive genetic value of the t th-order Legendre polynomial for line i [$\mathbf{a} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G} \otimes \mathbf{\Sigma}_a)$, $\mathbf{\Sigma}_a$ is a $T \times T$ genetic covariance matrix], ε_{ij} is the residual value for line i in environment j [$\varepsilon \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Sigma}_e)$, $\mathbf{\Sigma}_e$ is an $r \times r$ diagonal residual covariance matrix].

We used the mean phenotype across lines within each environment as the environmental value to compute $\phi_t(j)$. After some investigation with polynomials of different orders (results not shown), we chose polynomials of order 1 (i.e. $T = 1$) as they provided better performance than higher order ones. In fact, Supplementary Fig. S1 shows that a linear approximation may suffice for most lines. This model was fitted in a Restricted Maximum Likelihood (REML) framework, implemented in the sommer R package (Covarrubias-Pazarán 2016).

Validation schemes

We implemented three cross-validation (CV) schemes, whereby part of the data (i.e. the training set) was used to train the models and the remaining part (i.e. the test set) was used to evaluate

prediction accuracy. These schemes were used previously in plant studies (Burgueño et al. 2012; Jarquín et al. 2017).

- 1) *Random Lines* (similar to CV1 in Burgueño et al. 2012). We assigned 17% of the lines (for all the combinations of sex and temperature) to the test set, randomly (Fig. 1a). This procedure was repeated 6 times. The peculiarity of this scheme is that the lines in the test set are not represented in the training set. Thus, the training-test transfer of information happens mostly at the environmental level, when genotypes are unrelated.
- 2) *Random Observations* (similar to CV2 in Burgueño et al. 2012). We assigned 17% of the observations (i.e. combinations of line, sex, and temperature) to the test set, randomly (Fig. 1b). This procedure was repeated 6 times. The peculiarity of this scenario is that all the lines, sexes, and temperatures are represented in the training set. Thus, the training-test transfer of information happens at both the genetic and environmental level.
- 3) *New Environment* (similar to CV0 in Jarquín et al. 2017). We assigned all the observations in a specific combination of sex and temperature to the test set (~17% of the data) (Fig. 1c). This procedure was repeated 6 times. The peculiarity of this scenario is that a sex/temperature combination (i.e. an environment) is never seen in the training set. Thus, while technically the training-test transfer of information happens at both the genetic and environmental level, for the latter, it is hampered by not observing the actual environment in which we are trying to predict.

We computed prediction accuracy as R^2 from the regression of the true phenotypes on the predicted phenotypes, averaged over the 6 replicates. For the *Random Lines* and *Random Observations* scenarios, we also computed R^2 within each environment and then computed the average across environments (Lopez-Cruz et al. 2023).

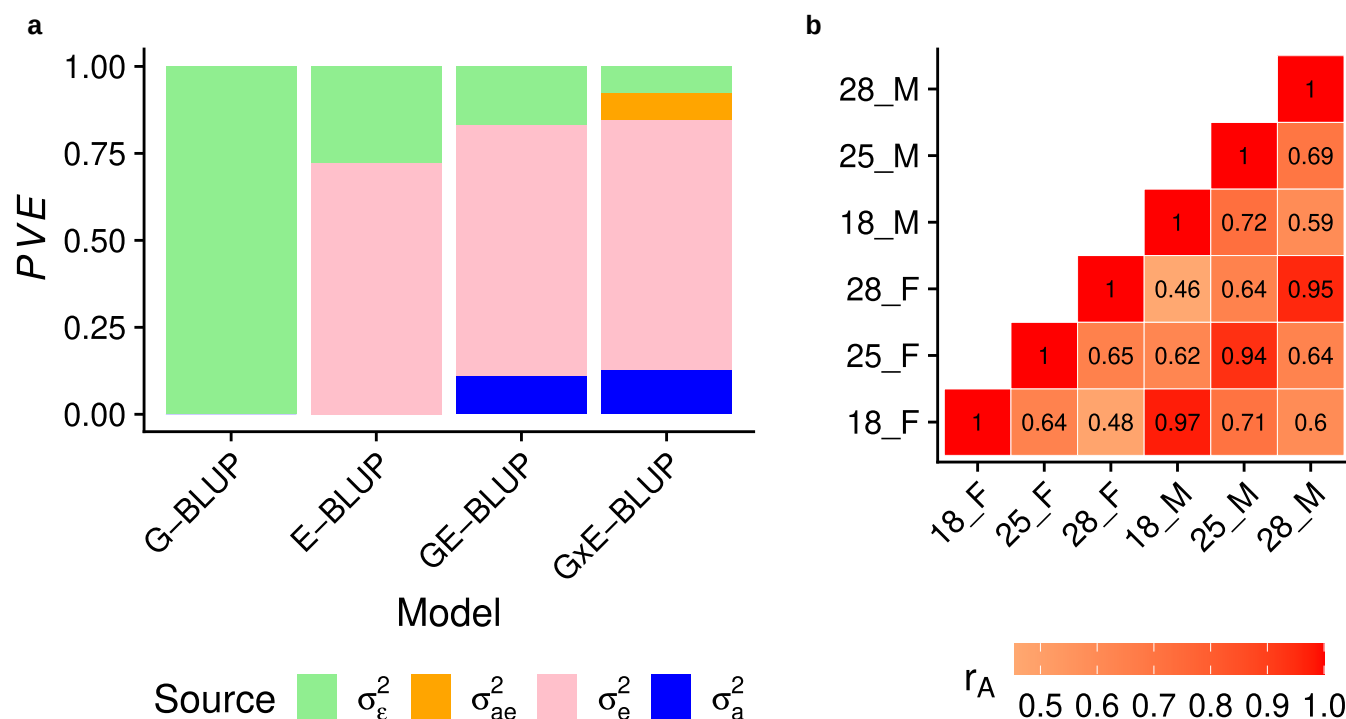


Fig. 2. Variance partition and cross-environment genetic correlations of lifespan. a) Variance partition into genetic (σ_a^2), environmental (σ_e^2), gene-environment interaction (σ_{ae}^2), and residual (σ_ϵ^2) components using RKHS models. The y-axis shows proportion of variance explained (PVE). b) Cross-environment genetic correlations (r_A) estimated using MTM.

Results and discussion

We first partitioned the phenotypic variance into sources of variation attributed to genetics, environment, and gene-environment interactions (Fig. 2a).

The results (Fig. 2a) show that the environment explained the largest amount of lifespan variance ($\sim 72\%$) across all models that include that component. The genetic variance was the second largest contributor in GE-BLUP and GxE-BLUP, explaining $\sim 12\%$ of the variance. However, genetic effects explained no variance in G-BLUP. We attributed this result to the fact that when the environment explains most of the variance and its effect is not modeled, the model struggles to find the (minor) genetic signal in the large amount of unexplained variance. But once the effect of the environment is included in GE-BLUP and GxE-BLUP, it is easier for genetic effects to explain some of the remaining variance (i.e. not accounted for by the environment). GxE explained $\sim 8\%$ of the phenotypic variance. Importantly, when including GxE in the model, it explained variance that would otherwise be included in the residual, as implied by the similar proportion of variance explained (PVE) by genetic effects and environmental effects in GE-BLUP and GxE-BLUP. We also sought to confirm the presence of GxE in a complementary analysis, where we estimated cross-environment genetic correlations (r_A), treating lifespan in each of the six environments (i.e. sex/temperature combinations) as a different trait in mvG-BLUP (Calus and Veerkamp 2011). The results (Fig. 2b) show that the genetic correlations were different from unity for every pair of environments, especially across temperatures. The observation that genetic effects are different across environments agrees with the presence of GxE (Falconer and Mackay 1996).

We then assessed the accuracy of the different models at predicting yet-to-be-observed phenotypes using three CV schemes, illustrated in Fig. 1. The prediction results are shown in Fig. 3.

In the *Random Lines* scenario (Fig. 3a), G-BLUP performed poorly. This result is expected, since the lines assigned to the test were not present in the training set, DGRP lines are largely unrelated, and G-BLUP was shown to explain no variance in the whole data-set. On the other hand, E-BLUP had high prediction accuracy ($R^2 \approx 0.72$). Here, sex and temperature were shown to explain $\sim 72\%$ of the variance and all the sex/temperature combinations were observed in the training set, allowing to obtain precise estimates of their effects. The prediction accuracy provided by GE-BLUP and GxE-BLUP was similar to (although slightly higher than) that of E-BLUP, showing that the inclusion of genetic effects and gene-environment interaction effects did not result in a substantial improvement of the predictions in this scenario. mvG-BLUP provided similar accuracy to E-BLUP, GE-BLUP, and GxE-BLUP. This result stems from the fact that mvG-BLUP fits lifespan in each environment as a different trait and thus estimates environment-specific intercepts, genetic and residual variances as well as genetic and residual correlations across environments. The difference in phenotype due to the different environmental conditions is accounted for by the intercepts in mvG-BLUP. The genomic heritability ranged $\sim 72 - 80\%$ across environments, showing that genetic effects could explain a large proportion of within-environment phenotypic variance. mvG-BLUP also leveraged the medium-high cross-environment genetic correlations to achieve good prediction accuracy. On the other hand, RRM had a substantially lower accuracy compared to the other environment-aware models. These results are in partial disagreement with other studies that showed RRM to be competitive with the other models (Sun et al. 2017). An explanation for this observation is that our data might be suboptimal for the application of RRM, since the environments appear to be better defined as discrete rather than stratified on a continuous scale. RRM might also be better suited for the analysis of a larger number of

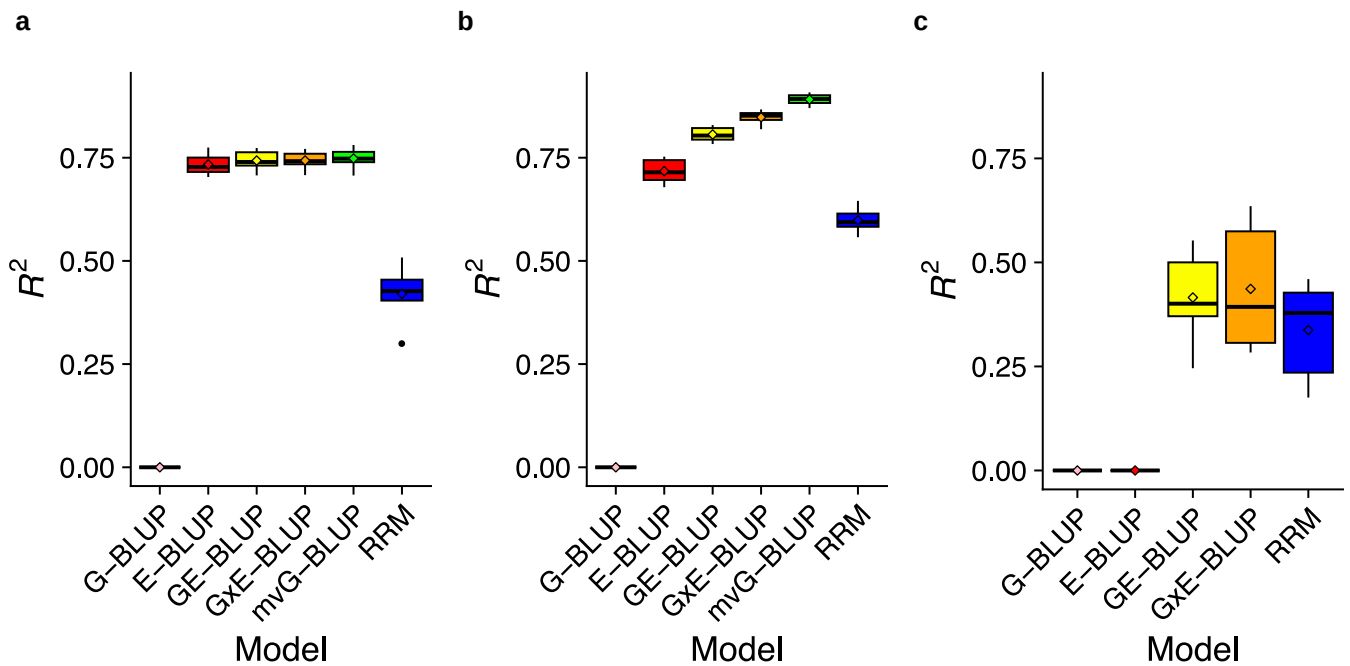


Fig. 3. Prediction accuracy in the different cross-validation schemes. a) *Random Lines*. b) *Random Observations*. c) *New Environment*.

environments than in our study, which allows for a better estimation of trajectories (Jarquín et al. 2017). In fact, RRM were developed for modeling longitudinal data such as test day milk production in dairy cattle (Schaeffer 2004).

In the *Random Observations* scenario (Fig. 3b), prediction accuracies were generally higher than in the *Random Lines* scenario. This is expected since all the lines and environments were present in the training set—only specific line/environment combinations were not observed in the training set. Thus, *Random Observations* is a considerably less challenging scenario than *Random Lines* for prediction. Similar to the *Random Lines* scenario, G-BLUP achieved null accuracy, whereas E-BLUP performed well ($R^2 \approx 0.72$). However, GE-BLUP improved accuracy substantially over E-BLUP, providing $R^2 \approx 0.81$ in this scenario. This shows that once the large proportion of variance explained by the environment is accounted for, genetic effects can explain additional variance and improve prediction accuracy in this scenario. Accounting for G×E in G×E-BLUP increased accuracy further ($R^2 \approx 0.85$), showing that this full model can achieve high accuracy at predicting specific genotype/environment combinations by borrowing information across lines and across environments. The best performing model was mvG-BLUP, which achieved $R^2 \approx 0.89$. This is a remarkable result since R^2 from this model closely approached the PVE of the full model (i.e. G×E-BLUP) in the whole dataset (~ 0.92). Thus, a model that explicitly accounts for the specificity of genetic effects across environments, while also leveraging their similarities, seems to be the best choice. However, this model has some limitations: (1) it requires discrete, well defined environments, with a low to moderate number of classes; (2) it conflates the contributions of genetics and G×E to the phenotypic value into a single term; (3) it does not allow for the prediction in unobserved environments, unless estimates of the genetic and residual (co)variances between observed and unobserved environments, and the phenotypic mean in the unobserved environments are available. Again, RRM had a much lower accuracy compared to the other environment-aware models in this scenario.

For the these two CV schemes, we also computed within-environment R^2 . This metric is particularly relevant to plant studies, where environments are well defined. The results (Supplementary Fig. S2) show that prediction accuracy was much lower with this metric compared to its pooled-environments counterpart, especially in the *Random Lines* scenario. This result can be explained by the fact that within-environment R^2 is not influenced by the difference in phenotypic means between environments, and here the environment was the largest contributor to the phenotypic variance. The general patterns observed with pooled-environments R^2 (Fig. 3) were also observed with within-environment R^2 . In the *Random Lines* scenario, accounting for G×E did not improve prediction accuracy. In the *Random Observations* scenario, accounting for G×E improved prediction accuracy when using mvG-BLUP. However, G×E-BLUP performed as well as GE-BLUP, which is in contrast with the results of the pooled-environments R^2 .

The *New Environment* scenario (Fig. 3c) was the most challenging of the CV schemes, resulting in considerably lower prediction accuracies compared to the other scenarios. In fact, the environment in which we are trying to predict is not observed in the training set, and the environmental conditions explain the majority of the phenotypic variance. This is reflected in E-BLUP yielding null accuracy in this scenario. However, GE-BLUP achieved moderate accuracy ($R^2 \approx 0.42$), presumably due to a better disentanglement of genetic and environmental effects when estimating parameters in the training set. G×E-BLUP provided slightly improved accuracy ($R^2 \approx 0.44$) compared to GE-BLUP, confirming the utility of accounting for G×E when predicting in unobserved environments. In this scenario, mvG-BLUP could not be fitted because of reason (3) discussed in the previous paragraph. While RRM performed worse than GE-BLUP and G×E-BLUP on average, its accuracy ($R^2 \approx 0.34$) was much closer to the accuracies of those models in this scenario. In the *New Environment* scenario, there were pronounced differences in prediction accuracy depending on which environment the predictions were made (Table 1). Overall, at least one of the models accounting for G×E (i.e. G×E-BLUP and RRM)

Table 1. Prediction accuracy (R^2) in each environment (Sex_Temperature) in the *New Environment* cross-validation scenario.

Env	G-BLUP	E-BLUP	GE-BLUP	GxE-BLUP	RRM
F_18°C	0.00	0.00	0.55	0.63	0.37
M_18°C	0.00	0.00	0.52	0.63	0.44
F_25°C	0.00	0.00	0.37	0.37	0.17
M_25°C	0.00	0.00	0.43	0.42	0.19
F_28°C	0.00	0.00	0.25	0.29	0.46
M_28°C	0.00	0.00	0.37	0.28	0.39

performed as well as or better than GE-BLUP in every environment. As expected given the estimates of the cross-environment genetic correlations, differences in prediction accuracy were larger among temperatures than between sexes (within temperature). At 18°C, GxE-BLUP greatly outperformed GE-BLUP. At 25°C, GxE-BLUP and GE-BLUP yielded similar accuracy. RRM performed substantially worse than the other methods at both 18°C and 25°C. On the other hand, at 28°C, RRM achieved higher accuracy than GE-BLUP and GxE-BLUP.

Overall, our results showed that accounting for GxE can be helpful for improving predictions. However, the presence and magnitude of the improvement depended on the CV scenario, as shown in previous studies (Burgueño et al. 2012; Jarquín et al. 2017). If the goal is to predict phenotypes for unknown individuals in known environments (our *Random Lines* scenario), accounting for GxE may not improve prediction accuracy. An explanation for this result is the lack of information sharing at the genetic level between the training set and the test set, when the individuals in the test set are largely unrelated to the individuals in the training set. In fact, Jarquín et al. (2017) found that a model with GxE outperformed the models without it in a similar CV scenario in wheat, which may be due to higher genetic similarity between training and test sets.

At the other end of the spectrum, if the goal is to predict phenotypes for a set of known individuals in a new environment (our *New Environment* scenario), accounting for GxE may improve prediction accuracy. A plausible explanation for this observation is that GxE-aware models can predict the adaptability of each genotype to several conditions. This scenario is particularly relevant for agricultural breeding, where breeders need to know how available breeds/cultivars would fare in a new environment (Reynolds et al. 2018; Tiezzi and Maltecca 2022), and how accurately models can predict such performance. However, the models' predictive ability to new environments can depend on the environment itself in at least two ways. First, phenotypes could be easier to predict in certain environments, if these show stronger genetic correlations with the known environments (Burgueño et al. 2012) and the models can borrow information across environments (Crossa et al. 2014). This is also in agreement with Lopez-Cruz et al. (2023), who showed that the within-year-location accuracy had some spread, indicating that the predictive ability depended on the environmental conditions. Second, different GxE-aware methods may outperform others, depending on which environment we aim to predict in (as seen in Table 1).

Finally, if the goal is to predict unobserved phenotypes for a set of known individuals in known environments (our *Random Observations* scenario), that is the scenario where we can expect the largest increase in accuracy when accounting for GxE. This is in agreement with previous studies (Burgueño et al. 2012; Jarquín et al. 2017). In this scenario, a multivariate model treating phenotypes in different environments as different traits seemed

to provide the highest accuracy. The *Random Observations* scenario is particularly relevant to precision medicine in humans, where there is interest in predicting medically relevant phenotypes (e.g. blood pressure) after a change in the environment (e.g. switching from a high fat diet to a low fat diet). However, this scenario cannot be evaluated with human data, as individuals are present in only one level of the environmental variable at any given time (e.g. a person either smokes or does not, but not both). This peculiarity might also contribute to the reasons why accounting for GxE—despite explaining non-negligible variance—has generally not resulted in improved prediction accuracy for human traits (Zhou and Lee 2021; Goda et al. 2025; Tiezzi et al. 2025), as well as some livestock species (Bussiman et al. 2025). In fact, assigning some individuals to the test set and trying to predict their phenotype in a known environment (as done in human studies) is equivalent to our *Random Lines*. In this scenario, we have shown that including GxE does not improve predictions, when training individuals and test individuals are largely unrelated as it is common in human studies.

However, it should be noted that human studies, which are observational in nature, have additional complications that could contribute to the lack of prediction accuracy improvement from accounting for GxE. First, environments are not clearly defined, which may result in environmental heterogeneity. Second, phenotypic and, especially, environmental measures are noisy. Thus, considering that interaction effects are usually smaller than main effects, there may be limited statistical power to estimate GxE effects precisely (which is necessary for accurate predictions) with current sample sizes of human studies. Third, non-random allocation of individuals (i.e. genotypes) to environments might result in reverse causality or genotype–environment correlation, which confound estimates of GxE (Tiezzi et al. 2025).

Our study has some important limitations. First, the DGRP has a small sample size, limiting the accuracy with which genetic effects can be estimated. Second, all the models used in this study do not perform variable selection, implying that all genetic variants, all environmental variables, and all the interactions between them have an effect on lifespan. Using methods that perform variable selection has the potential to increase prediction accuracy further. Third, our study focused on only one trait, as this was the only one available for a well designed and controlled experiment including unrelated genotypes, which allowed us to avoid confounding effects. Thus, our results will need to be confirmed on additional traits from equally well designed, but larger experiments.

Data availability

The genotype data are available from Huang et al. (2014). The phenotype data are available from Huang et al. (2020). Code used for the analyses is available at https://github.com/morgantelab/dgrp_lifespan_gxe.

Supplemental material available at G3 online.

Funding

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM146868 to FM. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

The authors declare that no conflict of interest exists.

Literature cited

- Acosta-Pech R et al. 2017. Genomic models with genotypex environment interaction for predicting hybrid performance: an application in maize hybrids. *Theor Appl Genet.* 130:1431–1440. <https://doi.org/10.1007/s00122-017-2898-0>.
- Bandeira e Sousa M et al. 2017. Genomic-enabled prediction in maize using kernel models with genotypex environment interaction. *G3 (Bethesda)*. 7:1995–2014. <https://doi.org/10.1534/g3.117.042341>.
- Bohlouli M, Alijani S, Naderi S, Yin T, König S. 2019. Prediction accuracies and genetic parameters for test-day traits from genomic and pedigree-based random regression models with or without heat stress interactions. *J Dairy Sci.* 102:488–502. <https://doi.org/10.3168/jds.2018-15329>.
- Bryant J, López-Villalobos N, Holmes C, Pryce J. 2005. Simulation modelling of dairy cattle performance based on knowledge of genotype, environment and genotype by environment interactions: current status. *Agric Syst.* 86:121–143. <https://doi.org/10.1016/j.agsy.2004.09.004>.
- Burgueño J, de los Campos G, Weigel K, Crossa J. 2012. Genomic prediction of breeding values when modeling genotypex environment interaction using pedigree and dense molecular markers. *Crop Sci.* 52:707–719. <https://doi.org/10.2135/cropsci2011.06.0299>.
- Bussiman F et al. 2025. Genotype-by-environment interaction with high-dimensional environmental data: an example in pigs. *Genet Sel Evol.* 57:28. <https://doi.org/10.1186/s12711-025-00974-2>.
- Bycroft C et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 562:203–209. <https://doi.org/10.1038/s41586-018-0579-z>.
- Calus MP, Veerkamp RF. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 43:1–14. <https://doi.org/10.1186/1297-9686-43-1>.
- Covarrubias-Pazarán G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One.* 11:e0156744. <https://doi.org/10.1371/journal.pone.0156744>.
- Crossa J et al. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity (Edinb)*. 112:48–60. <https://doi.org/10.1038/hdy.2013.16>.
- Cuevas J et al. 2016. Genomic prediction of genotypex environment interaction kernel regression models. *Plant Genome.* 9:plantgenome2016–03. <https://doi.org/10.3835/plantgenome2016.03.0024>.
- Dahl A et al. 2020. A robust method uncovers significant context-specific heritability in diverse complex traits. *Am J Hum Genet.* 106:71–91. <https://doi.org/10.1016/j.ajhg.2019.11.015>.
- Durvasula A, Price AL. 2025. Distinct explanations underlie gene-environment interactions in the UK Biobank. *Am J Hum Genet.* 112:644–658. <https://doi.org/10.1016/j.ajhg.2025.01.014>.
- Elias AA, Robbins KR, Doerge R, Tuinstra MR. 2016. Half a century of studying genotypex environment interactions in plant breeding experiments. *Crop Sci.* 56:2090–2105. <https://doi.org/10.2135/cropsci2015.01.0061>.
- Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. Pearson Education.
- Goda K, Arango NK, Tiezzi F, Mackay T, Morgante F. 2025 Jul 3. Gene-environment interactions contribute to blood pressure variation across global populations [preprint]. *medRxiv* 25330727. <https://doi.org/10.1101/2025.07.02.25330727>.
- Hansen PJ. 2020. Prospects for gene introgression or gene editing as a strategy for reduction of the impact of heat stress on production and reproduction in cattle. *Theriogenology.* 154:190–202. <https://doi.org/10.1016/j.theriogenology.2020.05.010>.
- Heslot N, Akdemir D, Sorrells ME, Jannink JL. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet.* 127:463–480. <https://doi.org/10.1007/s00122-013-2231-5>.
- Hu H, Rincet R, Runcie DE. 2025. Megalmm improves genomic predictions in new environments using environmental covariates. *Genetics.* 229:1–41. <https://doi.org/10.1093/genetics/iyae171>.
- Huang W et al. 2014. Natural variation in genome architecture among 205 drosophila melanogaster genetic reference panel lines. *Genome Res.* 24:1193–1208. <https://doi.org/10.1101/gr.171546.113>.
- Huang W et al. 2020. Context-dependent genetic architecture of drosophila life span. *PLoS Biol.* 18:e3000645. <https://doi.org/10.1371/journal.pbio.3000645>.
- Jarquín D et al. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor Appl Genet.* 127:595–607. <https://doi.org/10.1007/s00122-013-2243-1>.
- Jarquín D. et al. 2017. Increasing genomic-enabled prediction accuracy by modeling genotypex environment interactions in Kansas wheat. *Plant Genome.* 10:plantgenome2016–12. <https://doi.org/10.3835/plantgenome2016.12.0130>.
- Kerin M, Marchini J. 2020. Inferring gene-by-environment interactions with a Bayesian Whole-Genome regression model. *Am J Hum Genet.* 107:698–713. <https://doi.org/10.1016/j.ajhg.2020.08.009>.
- Lopez-Cruz M et al. 2023. Leveraging data from the genomes-to-fields initiative to investigate genotype-by-environment interactions in maize in North America. *Nat Commun.* 14:6904. <https://doi.org/10.1038/s41467-023-42687-4>.
- Martini JW, Crossa J, Toledo FH, Cuevas J. 2020. On hadamard and kronecker products in covariance structures for genotypex environment interaction. *Plant Genome.* 13:e20033. <https://doi.org/10.1002/tpg2.v13.3>.
- Miao J et al. 2025. Pigeon: a statistical framework for estimating gene-environment interaction for polygenic traits. *Nat Hum Behav.* 9:1654–1668. <https://doi.org/10.1038/s41562-024-02066-5>.
- Millet EJ et al. 2019. Genomic prediction of maize yield across European environmental conditions. *Nat Genet.* 51:952–956. <https://doi.org/10.1038/s41588-019-0414-y>.
- Pérez P, de Los Campos G. 2014. Genome-wide regression and prediction with the bgrr statistical package. *Genetics.* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>.
- Reynolds M et al. 2018. Role of modelling in international crop research: overview and some case studies. *Agronomy.* 8:291. <https://doi.org/10.3390/agronomy8120291>.
- Schaeffer LR. 2004. Application of random regression models in animal breeding. *Livest Prod Sci.* 86:35–45. [https://doi.org/10.1016/S0301-6226\(03\)00151-9](https://doi.org/10.1016/S0301-6226(03)00151-9).
- Sheahan A, Kolver E, Roche J. 2011. Genetic strain and diet effects on grazing behavior, pasture intake, and milk production. *J Dairy Sci.* 94:3583–3591. <https://doi.org/10.3168/jds.2010-4089>.
- Sun J et al. 2017. Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome.* 10:plantgenome2016–11. <https://doi.org/10.3835/plantgenome2016.11.0111>.
- Tiezi F, de Los Campos G, Gaddis KP, Maltecca C. 2017. Genotype by environment (climate) interaction improves genomic prediction

- for production traits in us holstein cattle. *J Dairy Sci.* 100: 2042–2056. <https://doi.org/10.3168/jds.2016-11543>.
- Tiezzi F, Goda K, Morgante F. 2025. Improvement of polygenic modeling of blood pressure traits using lifestyle information in the UK Biobank. *Genetics.* 230(3):iyaf089. <https://doi.org/10.1093/genetics/iyaf089>.
- Tiezzi F, Maltecca C. 2022. Genotype by environment interactions in livestock farming. In: Meyers RA, editor. *Encyclopedia of Sustainability Science and Technology*. Springer Science+Business Media, LLC, part of Springer Nature. https://doi.org/10.1007/978-1-4939-2493-6_1115-1.
- Van Eeuwijk FA, Bustos-Korts DV, Malosetti M. 2016. What should students in plant breeding know about the statistical aspects of genotypex environment interactions? *Crop Sci.* 56:2119–2140. <https://doi.org/10.2135/cropsci2015.06.0375>.
- VanRaden PM. 2008. Efficient methods to compute genomic predictions. *J Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- Weine E, Smith SP, Knowlton RK, Harpak A. 2025. Trade-offs in modeling context dependency in complex trait genetics. *Elife.* 13: RP99210. <https://doi.org/10.7554/eLife.99210>.
- Xavier A, Runcie D, Habier D. 2025. Megavariable methods capture complex genotype-by-environment interactions. *Genetics.* 229: iyae179. <https://doi.org/10.1093/genetics/iyae179>.
- Zhou X, Lee SH. 2021. An integrative analysis of genomic and exposomic data for complex traits and phenotypic prediction. *Sci Rep.* 11:21495. <https://doi.org/10.1038/s41598-021-00427-y>.

Editor: D.-J. de Koning