Accounting for epistasis improves genomic prediction of phenotypes with univariate and bivariate models across environments

Elaheh Vojgani (vojgani@gwdg.de)
University of Goettingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen https://orcid.org/0000-0003-4375-3531

Torsten Pook
University of Goettingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen

Johannes W.R. Martini
International Maize and Wheat Improvement Center (CIMMYT)

Armin C. Hölker
Plant Breeding, TUM School of Life Sciences Weihennstephan, Technical university of Munich, Freising

Manfred Mayer
Plant Breeding, TM School of Life Sciences Weihenstephan, Technical University of Munich, Freising

Chris-Carolin Schön
Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising

Henner Simianer
University of Goettingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen

Research Article

Keywords: Genomic prediction, GBLUP, Multi-trait models, Epistasis, Interaction

Posted Date: May 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-372414/v1

Accounting for epistasis improves genomic prediction of phenotypes with univariate and bivariate models across environments

Elaheh Vojgani[1*], Torsten Pook[1], Johannes W.R. Martini[2], Armin C. Hölker[3], Manfred Mayer[3], Chris-Carolin Schön[3], Henner Simianer[1]

[1] University of Goettingen, Center for Integrated Breeding Research, Animal Breeding and Genetics Group, Goettingen, Germany; [2] International Maize and Wheat Improvement Center (CIMMYT), Texcoco, State of Mexico, Mexico; [3] Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany

Key Message

The accuracy of genomic prediction of phenotypes can be increased by including the top ranked pairwise SNP interactions into the prediction model.

Abstract

We compared the predictive ability of various prediction models for a maize dataset derived from 910 doubled haploid lines from two European landraces (Kemater Landmais Gelb and Petkuser Ferdinand Rot), which were tested at six locations in Germany and Spain. The compared models were Genomic Best Linear Unbiased Prediction (GBLUP) as an additive model, Epistatic Random Regression BLUP (ERRBLUP) accounting for all pairwise SNP interactions, and selective Epistatic Random Regression BLUP (sERRBLUP) accounting for a selected subset of pairwise SNP interactions. These models have been compared in both univariate and bivariate statistical settings for predictions within and across environments. Our results indicate that modeling all pairwise SNP interactions into the univariate/bivariate model (ERRBLUP) is not superior in predictive ability to the respective additive model (GBLUP). However, incorporating only a selected subset of interactions with the highest effect variances in univariate/bivariate sERRBLUP can increase predictive ability significantly compared to the univariate/bivariate GBLUP. Overall, bivariate models consistently outperform univariate models in predictive ability. Across all studied traits, locations, and landraces, the increase in prediction accuracy from univariate GBLUP to univariate sERRBLUP ranged from 5.9 to 112.4 percent, with an average increase of 47 percent. For bivariate models, the change ranged from -0.3 to +27.9 percent comparing the bivariate sERRBLUP to the bivariate GBLUP, with an average increase of 11 percent. This considerable increase in predictive ability achieved by sERRBLUP may be of interest for "sparse testing" approaches in which only a subset of the lines/hybrids of interest is observed at each location.

Keywords: Genomic prediction, GBLUP, Multi-trait models, Epistasis, Interaction

Declaration

Funding

This work was funded by German Federal Ministry of Education and Research (BMBF) within the scope of the funding initiative “Plant Breeding Research for the Bioeconomy” (MAZE – “Accessing the genomic and functional diversity of maize to improve quantitative traits”; Funding ID: 031B0195)

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethics approval

The authors declare that this study complies with the current laws of the countries in which the experiments were performed.

Consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data and material are available through material transfer agreements upon request.

Code availability

Not applicable

Authors' contributions

EV derived the results, analyzed the data, wrote the initial manuscript; TP proposed epistasis relationship matrices. JWRM proposed epistasis interaction selection; ACH, MM and CCS prepared the material; ACH proposed cross validation strategy in bivariate model; HS proposed the original research question, guided the structure of the research. TP JWRM ACM MM CCS HS read, revised and approved the final manuscript.

Corresponding author

Correspondence to Elaheh Vojgani

Email: vojgani@gwdg.de

ORCID: 0000-0003-4375-3531

Acknowledgements

We are thankful to the technical staff at KWS SAAT SE, Misión Biológica de Galicia, Spanish National Research Council (CSIC), Technical University of Munich, and University of Hohenheim for providing the extensive phenotypic evaluation. We are grateful to the German Federal Ministry of Education and Research (BMBF) for the funding of our project within the scope of the funding initiative “Plant Breeding Research for the Bioeconomy” (MAZE – “Accessing the genomic and functional diversity of maize to improve quantitative traits”; Funding ID: 031B0195).

Introduction

Genomic prediction of phenotypes has been widely explored for crops (Crossa *et al.* 2010), livestock (Daetwyler *et al.* 2013) and clinical research (de los Campos *et al.* 2013). Broad availability and cost effective generation of genomic data had a considerable impact on plant (Bernardo and Yu 2007; de los Campos *et al.* 2009; Crossa *et al.* 2010, 2011; de Los Campos *et al.* 2010; Pérez *et al.* 2010) and animal breeding programs (de los Campos *et al.* 2009; Hayes and Goddard 2010; Daetwyler *et al.* 2013). Genomic prediction relates a set of genome wide markers to the variability in the observed phenotypes and enables the prediction of phenotypes or genetic values of genotyped but unobserved material (Meuwissen *et al.* 2001; Jones 2012; Windhausen *et al.* 2012). This approach has been positively evaluated in most major crop and livestock species (Albrecht *et al.* 2011; Daetwyler *et al.* 2013; Desta and Ortiz 2014) and is becoming a routine tool in commercial and public breeding programs (Stich and Ingheland 2018). In plant breeding, phenotyping is one of the major current bottlenecks and the optimization or minimization of phenotyping costs within breeding programs is needed (Akdemir and Isidro-Sánchez 2019). Therefore, the maximization of genomic prediction accuracy can be directly translated into reduced phenotyping costs (Akdemir and Isidro-Sánchez 2019; Jarquin *et al.* 2020).

Genomic selection and the corresponding prediction of breeding values is based on a covariance matrix describing the (additive) relationship between the considered individuals (Wolc *et al.* 2011; Burgueño *et al.* 2012). This matrix can be constructed from pedigree information, from marker information (VanRaden 2007) or from a combination of pedigree and available genotypic information in a single step approach (Aguilar *et al.* 2010; Legarra *et al.* 2014). It has been broadly demonstrated that marker based relationship matrices enhance the reliability of breeding value estimation on average across traits and compared to pedigree based approaches (Meuwissen *et al.* 2001; VanRaden 2007; Hayes and Goddard 2008; Crossa *et al.* 2010). Since breeding values are additive by definition (Falconer and Mackay 1996), the early development of prediction models exclusively accounted for the additive effects (Filho *et al.* 2016).

Concerning additive models, genomic best linear unbiased prediction (GBLUP, Meuwissen *et al.*, 2001; VanRaden, 2007) is a widely-used linear mixed model (Da *et al.*, 2014; Rönnegård and Shen, 2016; Covarrubias-Pazaran *et al.*, 2018). Although various new approaches such as methods from the Bayesian alphabet (Gianola *et al.* 2009) have been proposed, GBLUP remains the gold standard as new methods typically only perform marginally better, are less robust, require substantially more computing time and are more difficult to implement (Wang *et al.* 2018). Daetwyler *et al.* (2010) showed that BayesB can yield higher accuracy than GBLUP for traits controlled by a small number of quantitative trait nucleotides, emphasizing that the genetic architecture of the trait has an important impact on which method may predict better (Wimmer

et al. 2013; Momen *et al.* 2018). Moreover, the training set size was shown to play a role. For instance, human height prediction using BayesB and BayesC methods in a small reference population (<6,000 individuals) had no advantage over GBLUP. Only when increasing the size of the reference population (>6,000 individuals), these methods outperformed GBLUP (Karaman *et al.* 2016).

Understanding how genetic variation causes phenotypic variation in quantitative traits is still a major challenge of contemporary biology. It has been proven that epistasis as a statistical interaction between two or more loci (Falconer and Mackay 1996) contributes substantially to the genetic variation of quantitative traits (Wright 1931; Carlborg and Haley 2004; Hill *et al.* 2008; Huang *et al.* 2012; Mackay 2014). On the one hand, models which incorporate epistasis have the potential to increase predictive ability (de Los Campos *et al.* 2010; Hu *et al.* 2011; Wang *et al.* 2012; Mackay 2014). On the other hand, accounting for epistasis by modeling interactions explicitly was considered to be computationally challenging (Mackay 2014). In this context, the extended genomic best linear unbiased prediction (EG-BLUP), as an epistasis marker effect model (Jiang and Reif 2015; Martini *et al.* 2016) and reproducing kernel Hilbert space regression (RKHS), as a semi-parametric model (Gianola *et al.* 2006; Gianola and van Kaam 2008; de Los Campos *et al.* 2010) based on Gaussian kernel (Jiang and Reif 2015) were proposed to reduce the computational load by constructing marker-based epistatic relationship matrices (Jiang and Reif 2015; Martini *et al.* 2016). RKHS has shown to be as good as (Jiang and Reif 2015) or better than EG-BLUP (Martini *et al.* 2017). While EG-BLUP is potentially beneficial for genomic prediction, its performance depends on the marker coding (Martini *et al.* 2017, 2019). Moreover, it has been shown that the superiority of epistasis models over the additive GBLUP in terms of predictive ability may vanish when the number of markers increases (Schrauf *et al.* 2020). Also, the Hadamard products of the additive genomic relationship matrices provide only an approximation for the interaction effect model based on interactions between different loci (Martini *et al.* 2020), and more correcting factors are required for interactions of higher degree (Jiang and Reif 2020).

Another downside of epistasis models is that, due to the high number of interactions, a large number of unimportant variables can be introduced into the model (Rönnegård and Shen 2016). This 'noise' might prevent a gain in predictive ability. In this regard, Martini *et al.* (2016) showed that selecting just a subset of the largest epistatic interaction effects has the potential to improve predictive ability. Therefore, reducing the full epistasis model to a model based on a subnetwork of 'most relevant' pairwise SNP interactions may be beneficial for prediction performance (Martini *et al.* 2016).

In addition to the extension from additive effect models to models including epistatic interactions, genomic prediction models can be extended from univariate models to multivariate

models. Univariate models consider each trait separately, while multivariate models treat several traits simultaneously with the objective to exploit the genetic correlation between them to increase predictive ability. Multivariate models which have been first proposed for the prediction of genetic values by Henderson and Quaas (1976) were shown to be potentially beneficial for prediction accuracy when the correlation between traits is strong (He *et al.* 2016; Covarrubias-Pazaran *et al.* 2018; Schulthess *et al.* 2018; Velazco *et al.* 2019). A situation of dealing with multiple environments can also be considered in the framework of a multivariate model by simply considering a trait-in-environment combination as another correlated trait. This is considered as the multi-environment model which is usually employed to assess $\boldsymbol{G} \times \boldsymbol{E}$ interaction (Montesinos-López *et al.* 2016; Hassen *et al.* 2018) and captures the differences in genotypes' performances from one environment to the other as one of the breeders' major challenges in plant breeding (Kang and Gorman 1989). Prediction accuracy could be potentially enhanced through borrowing information across environments by utilizing multi-environment models (Burgueño *et al.* 2012). In addition to multi-environment models, Martini *et al.* (2016) showed that the predictive ability of EG-BLUP as a univariate model can be increased in one environment by variable selection in the other environment under the assumption of a relevant correlation of phenotypes in different environments. This, however, was only demonstrated with a data set of limited size and especially a limited set of markers and, thus, marker interactions.

In the present study, we use a data set of doubled haploid lines derived from two European landraces, to investigate how beneficial the use of subnetworks of interactions in the proposed sERRBLUP framework can be. This was compared in the context of univariate and bivariate models. We assess the optimum proportion of SNP interactions to be kept in the model in the variable selection step. The development of efficient selection strategies which could mitigate costly and time consuming phenotyping of a large number of selection candidates in multiple environments has been a particular focus of research in plant breeding (Jarquin *et al.* 2020). A successful application of our models may reduce the cost of phenotyping by reducing the number of test locations per line.

Materials and Methods

Data used for analysis

We used a set of 501 / 409 doubled haploid lines of the European maize landraces Kemater Landmais Gelb / Petkuser Ferdinand Rot genotyped with 501,124 markers using the Affymetrix ® Axiom Maize Genotyping Array (Unterseer *et al.* 2014), out of which 471 and 402 lines were phenotyped for Kemater (KE) and Petkuser (PE), respectively. The performance of the lines has been evaluated by ten separate 10 × 10 lattice designs in four German locations and five separate

10 × 10 lattice designs in two Spanish locations with two replicates. For more details see Hölker *et al.* (2019).

The lines were phenotyped in 2017 for a series of traits in six different environments which were Bernburg (BBG, Germany), Einbeck (EIN, Germany), Oberer Lindenhof (OLI, Germany), Roggenstein (ROG, Germany), Golada (GOL, Spain) and Tomeza (TOM, Spain).

The descriptions of the phenotypic traits, comprising early vigour and mean plant height of three plants of the plot at three growth stages (EV_V3, EV_V4, EV_V6, PH_V4, PH_V6, PH_final), days from sowing until female flowering (FF) and root lodging (RL) are given in the supplementary (Table S1), together with the number of phenotyped lines per location, phenotypic means, standard deviations, and maximum and minimum values. To correct for spatial structure and population effects, Best Linear Unbiased Estimations (BLUEs) were used as input for all considered prediction models. The interested reader is referred to Hölker *et al.* (2019) for details on the correction procedure and the detailed description of the considered traits. E.g., the trait "growth stage V4" indicates the growth stage at which four leaf collars are fully developed (Abendroth *et al.* 2011). In our study, we chose PH_V4 as the main trait for evaluating and illustrating our methods, since it is a relevant metric quantitative trait for early plant development which is suitable for testing our methods. The phenotypic correlations of PH_V4 across all environments are provided in Table 1.

Among the phenotypic traits, root lodging (RL) and female flowering (FF) were not phenotyped in all the environments: RL was only scored in BBG, ROG, OLI and EIN, and FF was phenotyped in all environments except GOL.

Quality control, coding and imputing

As we would not expect any heterozygous calls in DH material, all heterozygous calls were set to missing. Genotype calls were coded according to the allele counts of the B73 AGPv4 reference sequence (Jiao *et al.* 2017) (0 = homozygous for the reference allele, 2 = homozygous for the alternative allele). Imputation of missing values was performed separately for each landrace, using BEAGLE version 4.0 with parameters buildwindow=50, nsamples=50 (Browning and Browning 2007; Pook *et al.* 2020). For the remaining heterozygous calls, the DS (dosage) information of the BEAGLE output was used and genotyped with DS <1 were set to 0 and DS >= 1 to 2.

Linkage disequilibrium pruning

Linkage disequilibrium based SNP pruning with PLINK v1.07 was used to generate a subset of SNPs which are in approximate linkage equilibrium with each other. The parameters: indep 50 5

2 were used, in which 50 is the window size in SNPs, 5 is the number of SNPs to shift the window at each step and 2 is the variance inflation factor $VIF = 1/(1-r^2)$, where r^2 is the squared correlation between single SNPs and linear combinations of all SNPs in the window. All variants in the 50 SNP window which had a VIF > 2 were removed. Then, the window was shifted 5 SNPs forward and the procedure was repeated (Purcell et al. 2007; Chang et al. 2015).

In our study, LD pruning was done separately for each landrace, resulting in data panels containing 25'437 SNPs for KE and 30'212 SNPs for PE.

Univariate statistical models for phenotype prediction

We used three different statistical models to predict phenotypes, which are all based on a linear mixed model (Henderson 1975). We assume that we have in total n lines which are genotyped, and phenotypes are available for a subset of n_1 lines. These n_1 lines are used to train the model and missing phenotypes for the remaining $n_2 = n - n_1$ lines are predicted by using the genotypes of these lines. The basic univariate model is

$$\boldsymbol{y} = \mathbf{1}\mu + (\boldsymbol{I} \quad \boldsymbol{O})\boldsymbol{g} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y}$ is an $n_1 \times 1$ vector of phenotypes, $\mathbf{1}$ is an $n_1 \times 1$ vector with all entries equal to 1 , μ is a scalar fixed effect, $\boldsymbol{I}$ is an identity matrix of dimension $n_1 \times n_1$ and $\boldsymbol{O}$ is a matrix of dimension $n_1 \times n_2$ of zeros. The design matrix $(\boldsymbol{I} \quad \boldsymbol{O})$ is the $n_1 \times (n_1 + n_2)$ matrix resulting from the concatenation of $\boldsymbol{I}$ and $\boldsymbol{O}$. Moreover, $\boldsymbol{g} \sim N(0, \boldsymbol{\Gamma}\sigma_g^2)$ is an $n \times 1$ vector of random genetic effects, and $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{I}\sigma_\epsilon^2)$ is a random error vector, where $\boldsymbol{\Gamma}$ and $\boldsymbol{I}$ are the respective dispersion matrices and σ_g^2 and σ_ε^2 are the corresponding variance components.

With this model, the population mean and the genetic effects $\boldsymbol{g}$ for all lines, including those without phenotypes, are estimated using

$$\begin{bmatrix} \hat{\mu} \\ \widehat{\boldsymbol{g}}_{\mathbf{1}} \\ \widehat{\boldsymbol{g}}_{\mathbf{2}} \end{bmatrix} = \begin{bmatrix} n_1 & \mathbf{1}' & \mathbf{0} \\ \mathbf{1} & \boldsymbol{I} + \boldsymbol{\lambda\Gamma}^{\mathbf{11}} & \boldsymbol{\lambda\Gamma}^{\mathbf{12}} \\ \mathbf{0} & \boldsymbol{\lambda\Gamma}^{\mathbf{21}} & \boldsymbol{\lambda\Gamma}^{\mathbf{22}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'\boldsymbol{y} \\ \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}, \quad \text{(eq. 1)}$$

where $\lambda = \sigma_\varepsilon^2 / \sigma_g^2$, $\boldsymbol{\Gamma}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{\mathbf{11}} & \boldsymbol{\Gamma}^{\mathbf{12}} \\ \boldsymbol{\Gamma}^{\mathbf{21}} & \boldsymbol{\Gamma}^{\mathbf{22}} \end{bmatrix}$ and $\boldsymbol{g} = \begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \end{bmatrix}$ and the indices pertain to the subset of individuals with (index 1) or without (index 2) phenotypes, respectively.

With these estimates, the phenotypes for the set of unphenotyped individuals can be predicted as $\widehat{\boldsymbol{y}}_{\mathbf{2}} = \mathbf{1}_{\mathbf{2}}\hat{\mu} + \widehat{\boldsymbol{g}}_{\mathbf{2}}$, where $\widehat{\boldsymbol{y}}_{\mathbf{2}}$ is the $n_2 \times 1$ vector of predicted phenotypes and $\mathbf{1}_{\mathbf{2}}$ is an $n_2 \times 1$ vector of ones.

For $n = n_1$ and $n_2 = 0$ the solution of eq. 1 provides estimates of genetic effects when all lines are phenotyped and genotyped.

Bivariate statistical models for phenotype prediction

Besides univariate models, we also used bivariate models, where the two variables represent the same trait measured in two different environments.

The basic bivariate model is

$$\boldsymbol{y} = \boldsymbol{X\mu} + \boldsymbol{Zg} + \boldsymbol{e}$$

or, in more detail,

$$\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{1}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{1_2}\end{bmatrix}\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{I}_1 & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{g_1}\\ \boldsymbol{g_2}\end{bmatrix} + \begin{bmatrix}\boldsymbol{e_1}\\ \boldsymbol{e_2}\end{bmatrix}, \quad \text{(eq. 2)}$$

where, $\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}$ is the phenotype vector of length $m = m_1 + m_2$ for environment 1 (m_1) and 2 (m_2), $\boldsymbol{1_1}$ and $\boldsymbol{1_2}$ are respectively $m_1 \times 1$ and $m_2 \times 1$ vectors with all entries equal to 1, $\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$ is the vector of population means for environment 1 and 2, $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ are identity matrices of dimension $m_1 \times m_1$ and $m_2 \times m_2$, respectively assigning genomic values to phenotypes. Moreover, $\begin{bmatrix}\boldsymbol{g_1}\\ \boldsymbol{g_2}\end{bmatrix}$ is the vector of random genomic values which is assumed to have a multivariate normal distribution with mean zero and variance $\boldsymbol{G} = \boldsymbol{H} \otimes \boldsymbol{\Gamma}$, where $\boldsymbol{H} = \begin{bmatrix}\sigma^2_{g_1} & \sigma_{g_{12}}\\ \sigma_{g_{12}} & \sigma^2_{g_2}\end{bmatrix}$, $\boldsymbol{\Gamma}$ is the dispersion matrix of genetic effects and $\otimes$ is the Kronecker product. $\begin{bmatrix}\boldsymbol{e_1}\\ \boldsymbol{e_2}\end{bmatrix}$ is the vector of random errors which is assumed to have a multivariate normal distribution with mean zero and variance $\boldsymbol{R} = \boldsymbol{R_0} \otimes \boldsymbol{I}$, where $\boldsymbol{R_0} = \begin{bmatrix}\sigma^2_{e_1} & \sigma_{e_{12}}\\ \sigma_{e_{12}} & \sigma^2_{e_2}\end{bmatrix}$. $\sigma^2_{g_i}$ and $\sigma^2_{e_i}$ represent the genetic and residual variance of environment $i = 1,2$, and $\sigma_{g_{12}}$ and $\sigma_{e_{12}}$ are the genetic and residual covariance between the environment 1 and 2 (Guo *et al.* 2014). In this model, the phenotypes have to be ordered in the same way in both environments. In case the number of observations in environment 1 and environment 2 is not identical (i.e. in general terms $m_1 \neq m_2$) or different lines are considered in the model, the incidence matrices have to be adapted accordingly.

With this model, the vector of environment specific population means and the vector of genetic effects for all lines are estimated using the standard mixed model equations

$$\begin{bmatrix}\widehat{\boldsymbol{\mu}}\\ \widehat{\boldsymbol{g}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{X'R^{-1}X} & \boldsymbol{X'R^{-1}Z}\\ \boldsymbol{Z'R^{-1}X} & \boldsymbol{Z'R^{-1}Z + G^{-1}}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{X'R^{-1}y}\\ \boldsymbol{Z'R^{-1}y}\end{bmatrix},$$

In analogy to the procedure described in the univariate setting, we consider a setting in which the last l phenotypes for environment 2 are masked and predicted from all observations in environment 1 and the first $k = m_2 - l$ non-masked observations in environment 2.

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_k \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_k \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{I}_1 & 0 & 0 \\ 0 & \boldsymbol{I}_k & 0 \\ 0 & 0 & \boldsymbol{I}_l \end{bmatrix} \begin{bmatrix} \boldsymbol{g_1} \\ \boldsymbol{g_{2k}} \\ \boldsymbol{g_{2l}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_k} \\ \mathbf{0} \end{bmatrix}$$

From the solutions obtained with this model, the phenotypes for the set of unphenotyped individuals in environment 2 can be predicted as $\hat{\boldsymbol{y}}_{\boldsymbol{l}} = \mathbf{1}_{\boldsymbol{l}}\hat{\mu}_2 + \hat{\boldsymbol{g}}_{\boldsymbol{2l}}$, where $\hat{\boldsymbol{y}}_{\boldsymbol{l}}$ is the $l \times 1$ vector of predicted phenotypes and $\mathbf{1}_{\boldsymbol{l}}$ is an $l \times 1$ vector of ones.

The three models compared in this study only differ in the choice of the dispersion matrix $\boldsymbol{\Gamma}$ of the genetic effects.

Model 1: Genomic Best Linear Unbiased Prediction (GBLUP)

In this additive model, we use as $\boldsymbol{\Gamma}$ the genomic relationship matrix which is calculated according to VanRaden (2008) as

$$\boldsymbol{\Gamma}_{\boldsymbol{VR}} = \frac{(\boldsymbol{M} - \boldsymbol{P})(\boldsymbol{M} - \boldsymbol{P})'}{2 \cdot \sum_{i=1}^{m} \left(p_i(1 - p_i)\right)},$$

where $\boldsymbol{M}$ is the $n \times m$ marker matrix which gives m marker values for n lines under the assumption of having n genotyped lines in total. $\boldsymbol{P}$ is a matrix of equal dimension as $\boldsymbol{M}$ with $2 \cdot p_i$ in the i^{th}column, and p_i is the allele frequency of the minor allele of SNP i.

Model 2: Epistatic Random Regression BLUP (ERRBLUP)

This model accounts for all possible SNP interactions in the prediction model. With m markers and fully inbred lines, we have two possible genotypes at a single locus, i.e. 0 or 2 when coded as the counts of the minor allele. For each pair of loci, we have four different possible genotype combinations: {00, 02, 20, 22}. The total number of pairs of loci is $\frac{m \times (m+1)}{2}$ if we allow for interaction of a locus with itself. Since for each of these pairs we have four possible genotype combinations, the total number of combinations to be considered as dummy variables is

$$m^* = 4 \times \frac{m \times (m+1)}{2} = 2m \times (m+1).$$

We define a marker combination matrix $\boldsymbol{M}^*$ of dimension $n \times m^*$ whose element i, j is 1 if genotype combination j is present in individual i and is 0 otherwise. We further define for column i of this matrix the average value p_i^*, giving the frequency of the respective genotype

combination in the population, and a matrix $\boldsymbol{P}^*$ being of equal dimension as $\boldsymbol{M}^*$ with p_i^* in the i^{th}column.

Then, the relationship matrix based on all SNP interactions was calculated according to VanRaden (2008) as

$$\boldsymbol{\Gamma_{ERR}} = \frac{(\boldsymbol{M}^* - \boldsymbol{P}^*)(\boldsymbol{M}^* - \boldsymbol{P}^*)'}{\sum_{i=1}^{m^*}\left(p_i^*(1-p_i^*)\right)}$$

and this matrix was used in ERRBLUP as dispersion matrix for the genetic effects, which now are based on epistatic interaction effects. It should be noted that including the interaction of each locus with itself replaces the additive effect, so that it is not necessary to use a model that separately accounts for additive and epistatic effects. This model had been introduced earlier as "categorical epistasis model" (Martini *et al.* 2017).

Model 3: selective Epistatic Random Regression BLUP (sERRBLUP)

sERRBLUP is based on the same approach as ERRBLUP, but here the $\boldsymbol{\Gamma}$ -matrix is constructed from a selected subset of genotype interactions. We decided to use those interactions with the highest estimated marker effects variances. Selection based on highest absolute effects (as used by Martini *et al.* (2016) in the framework of the EGBLUP epistasis model) was also considered, but lead to similar to slightly worse results. For this, it was necessary to backsolve interaction effects $\hat{\boldsymbol{t}}$ and effects variances $\widehat{\boldsymbol{\sigma}}^2$ from the ERRBLUP model using (Mrode 2014)

$$\hat{\boldsymbol{t}} = \frac{\widehat{\sigma_g^*}^2}{\sum_{i=1}^{m^*}(p_i^*(1-p_i^*))}(\boldsymbol{M}^* - \boldsymbol{P}^*)'\left(\widehat{\sigma_g^*}^2\boldsymbol{\Gamma_{ERR}} + \widehat{\sigma_\epsilon^*}^2\boldsymbol{I}\right)^{-1}(\boldsymbol{y} - \boldsymbol{1}\widehat{\mu}),$$

$$\widehat{\boldsymbol{\sigma}}^2 = (\hat{\boldsymbol{t}} \circ \hat{\boldsymbol{t}})2\boldsymbol{P}^*(1-\boldsymbol{P}^*),$$

with ∘ denoting the Hadamard product.

After estimating SNP interaction effects in $\hat{\boldsymbol{t}}$ and effects variances in $\widehat{\boldsymbol{\sigma}}^2$, we selected those interactions whose absolute estimated effects or effect variances were in the top $\pi = 0.05, 0.01, 0.001, 0.0001, 0.00001$ or 0.000001 proportion of all interactions, respectively. These proportions were chosen since it was observed in preliminary analyses that they cover the most relevant range. For each of these subsets, we generated reduced matrices $\boldsymbol{M}_{\boldsymbol{\pi}}^*$ and $\boldsymbol{P}_{\boldsymbol{\pi}}^*$ of dimension $n \times \pi m^*$, containing only those columns of $\boldsymbol{M}^*$ and $\boldsymbol{P}^*$ pertaining to the selected subset of genotype interactions, and then set up the dispersion matrix in analogy to VanRaden (2008) as

317 $$\boldsymbol{\Gamma_{sERR}} = \frac{(\boldsymbol{M_{\pi}^{*}} - \boldsymbol{P_{\pi}^{*}})(\boldsymbol{M_{\pi}^{*}} - \boldsymbol{P_{\pi}^{*}})'}{\sum_{i=1}^{\pi m^{*}} \left(p_{\pi i}^{*}(1 - p_{\pi i}^{*})\right)},$$

318 where $p_{\pi i}^{*}$ are the mean frequencies of the selected genotype combinations.

319 Note here that even for the univariate model, information from another environment is used for
320 the prediction, namely for variable selection and the definition of $\boldsymbol{\Gamma_{sERR}}$. However, having used
321 the information from another environment to define the subset of interactions and to derive the
322 relationship matrix $\boldsymbol{\Gamma_{sERR}}$, the actual prediction is within the considered environment from the
323 training to the test set.

324 We used the miraculix package (Schlather 2020) to efficiently calculate $\boldsymbol{\Gamma_{ERR}}$, $\hat{\boldsymbol{t}}$ and $\boldsymbol{\Gamma_{sERR}}$.

325 **Assessment of predictive ability via 5-fold random cross validation with 5 replicates**

326 In a 5-fold cross validation, the original sample is randomly partitioned into five subsamples of
327 equal size. Out of the five subsamples, each subsample is subsequently considered as the test set
328 for validating the model, and the remaining four subsamples are considered as training data. The
329 training set is used to predict the test set. By this, all observations are used for both training and
330 testing and each observation is only used once for testing (Utz *et al.* 2000). We repeated the
331 cross-validation procedure 5 times, using random partitions of the original sample. The results of
332 the 25 repetitions were then averaged (Erbe *et al.* 2010). We used the Pearson correlation
333 between the predicted genetic value and the observed phenotype in the test set as the measure
334 for predictive ability. In our study, predictive ability was assessed for PE and KE for all phenotypic
335 traits separately. In addition, the trait's prediction accuracy was calculated by dividing the
336 obtained predictive ability by the square-root of the respective trait's heritability (Dekkers 2007).
337 The numbers of KE and PE lines which are available for all combinations of environments are
338 summarized in Table 2. For some traits these numbers can be smaller or even zero for some
339 environment combinations. We evaluated our univariate and bivariate models as follows:

340 **Assessment of GBLUP, ERRBLUP and sERRBLUP predictive abilities**

341 The univariate GBLUP and ERRBLUP within environments were evaluated by training the model
342 in the same environment as the test set was sampled from.

343 The basic strategy for univariate and bivariate sERRBLUP across environments is illustrated in Fig.
344 1: first, all pairwise SNP interaction effects and their variances are estimated from all data in
345 environment 1 and effects are ordered either by absolute effect size or effect variance (A). Next,
346 an epistatic relationship matrix for all lines is constructed from the top ranked subset of
347 interaction effects (B). Then, this matrix is used in environment 2 (C) to predict phenotypes of

the test set (green) from the respective training set (red) (D). This approach henceforth is termed 'sERRBLUP across environments'. In the case of bivariate sERRBLUP both the full data panel from environment 1 and the training set from environment 2 are used in a bivariate prediction model.

The basic strategy for bivariate GBLUP and ERRBLUP can also be illustrated in Fig. 1 when the model is trained jointly on the complete dataset of environment 1 (E) and the training set of environment 2 (D). The test set of environment 2 is then predicted, using as dispersion matrix for the genetic effects either $\mathbf{\Gamma}_{VR}$ or $\mathbf{\Gamma}_{ERR}$.

Use of multiple environments jointly

In addition to considering each environment separately, we used the average of all environments, except the current target environment, as an additional environment. This was considered for univariate sEERBLUP and all bivariate models.

Estimation of variance and covariance components

Since we aimed at estimating variance components in each replicate of the cross-validation from the training data, but variance component estimation with ASREML has a certain risk of non-convergence in particular in models with a high number of parameters such as the models proposed here. Therefore, we needed to specify a strategy to deal with such cases in an automated manner. In univariate analyses, variance components were estimated using EMMREML (Akdemir and Godfrey 2015) in each run of a 5-fold cross validation based on the training set. In bivariate analyses, the variance components were estimated using ASReml-R (Butler *et al.* 2018). In the bivariate ERRBLUP and sERRBLUP models, the genetic and residual variance and covariance were estimated first from the full data set in a bivariate ASReml-R model for each combination of environments in each trait. If the estimation of variance components didn't converge after 100 iterations, then the computation was stopped and the genetic and residual variance and covariance estimates at the last iteration (100) were extracted. These estimates were defined as the initial starting values of the bivariate ASReml-R model in each run of a 5-fold cross validation, followed by a re-estimation of the variance and covariance components based on the training set in the cross validation. If the estimation of variance components did not converge at 50 iterations in each fold, the pre-estimated variance and covariance components based on the full dataset, which was defined as the initial start values of the model, were used as fixed values, so that the breeding values were estimated based on these pre-estimated parameters. It was verified from converged estimates that variance and covariance components estimated from the training set deviated only little from the variances and covariances from the full set (see Fig. S1). Also, the mean result obtained from just the converged replicates and the mean results of all replicates including the ones where variance and

covariance components were fixed were rather similar (Fig. S2), only when the majority (>20) of replicates failed to converge, substantial random fluctuation was observed. Thus, we argue that this strategy appears justifiable, but still the number of cases where estimates did not converge in 5-fold cross validation with 5 replicates and the combinations whose pre estimation of variance components also did not converge in 100 iterations are detailed in the supplementary (Table S2 – S9).

Results

Predictive abilities of univariate sERRBLUP across environments compared to univariate ERRBLUP and univariate GBLUP within environments for the trait PH_V4 are shown in Fig. 2 for KE and PE. Univariate GBLUP within the environment is used as a reference and is compared to results obtained with univariate ERRBLUP within environments and univariate sERRBLUP when the top 5, 1, 0.1, 0.01, 0.001 and 0.0001 percent of pairwise SNP interactions are maintained in the model. Fig. 2 shows that the predictive abilities of univariate GBLUP and univariate ERRBLUP within the environment are almost identical (the highest deviation observed was 0.004). A considerable increase in predictive ability was observed when the top 1 or 0.1 percent of SNP interactions, selected based on their effect variances, were kept in the univariate sERRBLUP model. A more stringent selection, i.e. by considering only the top 0.01, 0.001 and 0.0001 percent of SNP interactions in the model, often led to a reduction in predictive ability, such that for the most stringent selection of 0.001 and 0.0001 percent, the predictive ability was sometimes even below the univariate GBLUP reference. This pattern is observed across all environments and is more pronounced in KE than PE. Results for the other traits are given in the Supplementary (Fig. S3a – S9a). In this study, estimated effect variances were identified as the best selection criteria in sERRBLUP, since sERRBLUP predictive abilities were observed to be more robust when the selection of pairwise SNP interaction was based on the effect variances compared to absolute effect sizes, especially when the top 0.001 and 0.0001 percent of interactions are maintained in the model (Fig. S10 and S11). In addition, the maximum predictive ability obtained from univariate sERRBLUP are almost identical when selecting SNP interactions based on absolute effect sizes or effect variances for both KE and PE (Fig. S12).

In the context of univariate models, we also investigated the predictive ability of univariate sERRBLUP when the variable selection was based on the training set from the same environment as the test set. This was exemplarily done within Bernburg for the trait PH_V4 (Fig. S13), illustrating that the predictive ability obtained from univariate sERRBLUP is marginally higher than univariate GBLUP only when the top 0.01 percent of interactions are kept in the model. When the selection of effects is too strict, with only 0.001 percent of interactions used, the

predictive ability of univariate sERRBLUP within Bernburg is smaller than the one obtained with GBLUP, especially if the selection is based on effect sizes.

The predictive abilities of bivariate GBLUP, ERRBLUP and sERRBLUP when SNP interactions were selected based on estimated effect variances are compared for trait PH_V4 in KE and PE in Fig. 3. Fig. 3 shows that the bivariate ERRBLUP increases the predictive ability slightly compared to bivariate GBLUP with the maximum absolute increase of 0.03 in KE and 0.02 in PE across all environments' combinations. A considerable increase in predictive ability is obtained in bivariate sERRBLUP mostly when the top 5 or 1 percent of interactions are maintained in the model. However, the bivariate sERRBLUP predictive abilities decrease dramatically for too stringent selection of pairwise SNP interactions such as 0.01, 0.001 or 0.0001 percent. Moreover, the reduction in predictive ability with too stringent factor selection is more severe for KE than for PE. This pattern is observed for the majority of environments for both landraces and the results for other traits are shown in the supplementary (Fig. S3b – S9b)

The relative increase in prediction accuracy of the best univariate sERRBLUP across environments compared to univariate GBLUP within environments for all traits and all locations is shown in form of a heat map in Fig. 4 for both landraces. The maximum relative increase in prediction accuracy among all traits and all environments in KE is 85.6 percent (PH_V6 in OLI) and in PE it is 112.4 percent (EV_V3 in EIN). Those highest increases in accuracy were found in traits and environment combinations where the univariate GBLUP prediction accuracy was particularly low. An increase is observed in each studied trait by location combination, with the smallest increase in both landraces for PH_final in BBG (20.1 percent in KE) or in GOL (5.9 percent in PE). In general, both plots in Fig. 4 demonstrate that for the majority of traits and environments, there is more than a 30 percent increase in prediction accuracy from univariate GBLUP within environments to the best univariate sERRBLUP across environments. The average increase across all combinations in KE is 47.1 percent and in PE is 46.7 percent. Note that this increase is somewhat inflated as a single GBLUP accuracy is compared against the best prediction from a set of various models (environment / selection proportions). However, even when using a set environment and a fixed proportion of interactions, there are still substantial gain. Exemplary, EIN with a proportion of 0.1 still lead to an increase of 43.1 percent in KE and 36.9 percent in PE (Fig. S14). The choice of EIN was made as it had the highest number of phenotyped lines (Table S1), while 0.1 in general led to stable models. Results using any other location or reasonable choice of the share of included interactions were very similar. The absolute increase in prediction accuracy is also shown as a heat map in Supplementary Fig. S15a, which indicates the average absolute increase of 0.204 in KE and 0.181 in PE.

Fig. 5 also shows the relative increase in prediction accuracy from the best bivariate GBLUP to the best bivariate sERRBLUP for all traits and all locations. The maximum increase in prediction accuracy among all traits and all environments is 21.1 percent (EV_V6 in ROG) in KE and 27.9 percent (EV_V3 in BBG) in PE. There is an increase across all studied traits in all environments except for the trait PH_final in PE which shows a relative decrease of 0.3 percent. The minimum increase in prediction accuracy in KE was also observed for PH_final (1.7 percent). In general, Fig. 5 shows that the relative increase in prediction accuracy from the best bivariate GBLUP to the best bivariate sERRBLUP is more than 7 percent for the majority of trait by location combinations in both landraces with an average increase of 10.9 percent in KE and 10.5 in PE across all combinations. The absolute increase in prediction accuracy of bivariate models is also shown as a heat map in supplementary (Fig. S15b) indicating an average absolute increase of 0.1 across all traits, environment combinations, and landraces.

In addition to assessing the predictive ability of univariate sERRBLUP based on a single environment, Fig. 6 displays the comparison between the predictive ability obtained from univariate GBLUP and univariate ERRBLUP within environments, and univariate sERRBLUP across multiple environments jointly for trait PH_V4 in KE and PE. It is demonstrated that univariate sERRBLUP has a higher predictive ability than univariate GBLUP when interactions are selected based on all the other five environments jointly. The preliminary analysis also reveals the robustness of the selection strategy based on the effects variance compared to selection strategy based on the absolute effects sizes in univariate sERRBLUP across multiple environments jointly for KE (Fig. S16), while for PE it does not show a significant difference for the interaction selection strategy (Fig. S17). Fig. 6 demonstrates that the predictive ability of univariate sERRBLUP across multiple environments jointly is as good as or better than using a single environment with few exceptions when the selection of effects is not too strict. With less than 0.1 percent of interactions used, predictive abilities deteriorate (especially so in KE) and selection from combined environments turns out to be worse than selection from single environments.

Fig. 7 illustrates the comparison between the predictive ability of bivariate GBLUP, ERRBLUP and sERRBLUP across multiple environments jointly and the maximum predictive ability of bivariate GBLUP and ERRBLUP and all the predictive abilities of sERRBLUP when a single environment is considered as an additional environment for the trait PH_V4 in both KE and PE. The results indicate that bivariate sERRBLUP across multiple environments jointly increases the predictive ability compared to bivariate GBLUP and ERRBLUP across multiple environments jointly. In most cases, bivariate GBLUP, ERRBLUP and sERRBLUP across multiple environments jointly performs as good as or better than when using a single environment.

Discussion

The accuracy of genomic prediction when incorporating epistasis interactions in the model compared to prediction models with only main effects has been widely discussed over the last years. In particular, it was found that accounting for epistasis can increase predictive ability (Carlborg and Haley 2004; Hu *et al.* 2011; Huang *et al.* 2012; Wang *et al.* 2012; Mackay 2014; Jiang and Reif 2015; Ober *et al.* 2015; Rönnegård and Shen 2016).

A major concern in utilizing epistasis models has been the high computational load (Mackay 2014) which has been reduced for the full model including all interactions by utilizing marker based epistasis relationship matrices derived from Hadamard products of additive genomic relationship matrices (Jiang and Reif 2015; Ober *et al.* 2015; Martini *et al.* 2016). The key advantage of this approach is that the number of random effects in the model is reduced from the number of SNP interactions to the number of genotypes. While the approaches of Jiang and Reif (2015) and Martini *et al.* (2016), only capture the interactions whose products differ from zero (i.e. {22} genotype combinations for 0, 2 coded markers), our approach captures all possible genotype combinations ({00}, {02}, {20}, and {22}). Further, these epistasis relationship matrices and interaction effects were computed by bit-wise computations via the R-package miraculix (Schlather 2020), which carries out matrix multiplications about 15 times faster than regular matrix multiplications on genotype data in EpiGP R-package (Vojgani *et al.* 2021). In the analyzed datasets containing up to 30'212 SNPs (and thus 456'397'578 interactions), the computing time required to set up the sERRBLUP relationship matrix was about 810 minutes out of which around 330 minutes were required to estimate all pairwise SNP interaction effects and 480 minutes were required to set up the sERRBLUP relationship matrix for selected proportion of interactions by utilizing the R-package miraculix with 15 cores on a server cluster with Intel E5-2650 (2X12 core 2.2GHz) processors. Computing times for sERRBLUP scale approximately quadratic in the number of markers considered. The released EpiGP R-package (Vojgani *et al.* 2021), which is available at https://github.com/evojgani/EpiGP, has been utilized for ERRBLUP and sERRBLUP genomic prediction of phenotypes.

Our proposed epistasis model eventually can generate a considerable prohibitive computational load if the number of SNPs grows to hundreds of thousands (Vojgani, *et al.*, 2019). The computing time for sERRBLUP exhibits quadratic growth with increasing number of SNPs. A potential strategy to overcome these limitations is to achieve a feature reduction by SNP pruning, as was implemented in our maize dataset (Purcell *et al.* 2007; Chang *et al.* 2015). Another option to obtain an even stronger variable reduction than pruning might be the use of haplotype blocks (Pook *et al.* 2019). Although sERRBLUP model can be computationally challenging by increasing the number of SNPs, its predictive ability is constantly higher than the models such as RKHS, which reduces the computational time considerably (Table S10).

In this study, we showed that the predictive ability obtained by use of GBLUP and a full epistasis model with all pairwise SNP interactions included (ERRBLUP) was almost identical. In contrast, it was shown that the use of sERRBLUP increases predictive ability when only the most relevant SNP interactions are taken into account, regardless of the choice of the training environment, which is likely a result of enriching for true causal variant combinations among the list of all variant combinations used to construct the genetic covariance matrix. In our study, the maximum predictive ability with sERRBLUP was obtained by incorporating the top 5, 1 or 0.1 percent of pairwise SNP interactions, while a too strict selection of SNP interactions such as the top 0.01, 0.001 and 0.0001 percent often reduced the predictive ability. A similar loss in predictive ability with a too strict selection of interactions to be included in the model was also observed by Ober et al. (2015). The difference in interaction selection can be explained by the fact that the absolute number of interaction effects in the model is more important than the percentage of interaction effects. To illustrate this, the absolute numbers of interactions maintained in the model for the top 0.001 and 0.0001 percent of interactions in KE are respectively 3'235 and 323, which is less than the number of additive effects in KE (25'437) where the obtained sERRBLUP predictive ability is lower than GBLUP predictive ability. In addition, the possible differences in linkage can also lead to different redundancy patterns of interactions. Here we also saw the only major systematic difference between the two selection criteria: when SNP interactions were selected based on the magnitude of their estimated (absolute) effects, the loss in predictive ability when selecting too few interactions was much more severe than when SNP interactions were selected based on the variance associated with them. This phenomenon has been more prevalent in KE than in PE (Fig. S10-S11), and is valid in both scenarios, using information either from a single environment or from the average of all other environments (Fig. S16-S17). A potential reason for this is that the few interactions that remain in the model are highly linked and thus no proper representation of the overall population structure is possible anymore. This effect was even more pronounced when selecting based on effect sizes. Thus, we recommend the use of effect variances as a selection criterion in sERRBLUP applications since this should be conceptually more robust.

The bivariate models exhibited a considerably higher predictive ability than univariate models. In consequence, the bivariate GBLUP performed slightly better than the best univariate sERRBLUP in most cases (Fig. S18). Across all studied traits, the increase in prediction accuracy from GBLUP to sERRBLUP displays a similar pattern in both univariate and bivariate models. It has to be noted that this increase in predictive ability is exclusively caused by the modelling of epistasis in a bivariate statistical setting, while it is caused by both modelling of epistasis and borrowing information across environments through variable selection in the univariate statistical setting.

In general, it is expected that the predictive ability for phenotypes should be higher with higher heritability. In this study, the correlation between the heritability of all traits, which have been calculated on an entry-mean basis within each landraces (Hallauer *et al.* 2010) over all environments, was 0.296 with univariate GBLUP within environments and 0.543 with maximum univariate sERRBLUP across environments (Fig. S19a). Corresponding correlations were higher in the bivariate statistical setting of the respective models, with an increase in the respective correlation from maximum bivariate GBLUP (0.537) to maximum bivariate sERRBLUP (0.647) (Fig. S19b).

When comparing sERRBLUP to a traditional GxE model (Kang and Gorman 1989), the modelling approach is quite different. In sERRBLUP, the selection of marker-by-marker interactions is done based on a second environment. However, for the final estimation of the actual effect size, the data from the same environment is used. Thus, effect sizes can substantially differ between environments. In contrast to this, traditional GxE model will assign effects to specific marker-by-environment combinations. As included interactions between different environments in sERRBLUP are different, it is not possible to put concrete GxE effects on specific markers or marker-by-environment interactions, which would be the essence of traditional GxE models. As prediction performances are increasing quite substantially by the use of sERRBLUP, this still can be seen as an indication that effect regions are similar between environment (although effect sizes might differ).

Our results indicate that a higher number of phenotyped lines (in particular overlapping between environments) and including information from a more similar second environments were beneficial for prediction. E.g., when the two Spanish locations GOL or TOM were used as the second environment to predict a German environment, prediction accuracies were lower as these environments have substantially different climate and for some traits lower overlap between phenotyped lines. On the other hand, the best prediction results for GOL were obtained when using TOM as second environment and vice versa.

In both univariate and bivariate models, it was shown that the obtained predictive ability across multiple environments jointly was mostly equivalent or higher than the maximum predictive ability obtained based on a single environment. Thus, using an average across all other environments should be a robust alternative which in most cases will yield a result that is as good as or even better than the best single environment.

Overall, our results demonstrate that bivariate models can outperform univariate models and epistatic interactions can substantially increase the predictive ability. In the context of univariate models, it was shown that selecting a suitable subset of interactions based on other environments where phenotypic data of the full set of lines are available can substantially

increase the predictive ability. As the ideal share π of interactions to be included in sERRBLUP is not known in practice, one could consider to run a testing scheme with an additional validation set for the identification of a suitable π. As results were quite robust as long as a reasonable fraction (between 5 and 0.1 percent) of interactions were included in the model and this introduces further computational load, this should however usually not be required.

The presented approach can substantially improve the phenotype prediction accuracy in another environment by 'borrowing' information on effect regions from another variable. In our case, this other variable were phenotypes of the same trait grown in different environments. However, one could also imagine using data from another growing season or even from a highly correlated second trait. This can be useful in sparse testing designs, e.g. where not all lines are grown in all environments. The suggested approach can be used to 'impute' missing phenotypes with a much increased accuracy compared to conventional approaches.

Figures Captions

Fig. 1 Basic scheme of uni- and bivariate sERRBLUP across environments. All pairwise SNP interaction effects and their variances are estimated from all data in environment 1, and effects are ordered either by absolute effect size or effect variance (A). Then, an epistatic relationship matrix for all lines is constructed from the top ranked subset of interaction effects (B) which in the univariate model is used in environment 2 (C) to predict phenotypes of the test set (green) from the respective training set (red, D). In the bivariate model, this information is combined with the complete data from environment 1 (blue, E) to predict the test set.

Fig. 2 Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (black filled circle) and univariate sERRBLUP across environments (solid colored lines) when SNP interaction selections are based on estimated effects variances in KE (left side) and PE (right side) for trait PH_V4. In each panel, the solid lines' color indicates the environment in which the relationship matrices were determined by variable selection.

Fig. 3 Predictive ability for bivariate GBLUP (open squares), bivariate ERRBLUP (open circles) and bivariate sERRBLUP (filled circles and solid lines) when SNP interaction selections are based on estimated effects variances in KE (left side) and PE (right side) for trait PH_V4. In each panel, the solid lines' color indicates the additional environment used to predict the target environment.

Fig. 4 Percentage of increase in prediction accuracy from univariate GBLUP within environments to the maximum prediction accuracy of univariate sERRBLUP across environments when the SNP interaction selections are based on estimated effects variances in KE (left side) and in PE (right side). The average percentage of increase in prediction accuracy for each trait and environments are displayed in rows and columns, respectively.

Fig. 5 Percentage of increase in prediction accuracy from the maximum bivariate GBLUP to the maximum prediction accuracy of bivariate sERRBLUP when the SNP interaction selections are based on estimated effects variances in KE (left side) and in PE (right side). The average percentage of increase in prediction accuracy for each trait and environments are displayed in rows and columns, respectively.

Fig. 6 Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (gray open circle), univariate sERRBLUP using a single environment for selecting the SNP interactions (gray open circles) and univariate sERRBLUP using all 5 environments jointly (filled black circles and solid line) for the SNP interaction selection based on estimated effects variances for trait PH_V4 in KE (left side) and PE (right side).

Fig. 7 Predictive ability for bivariate GBLUP (black dashed horizontal line), bivariate ERRBLUP and bivariate sERRBLUP (filled black circles) for the SNP interaction selection based on estimated effects variances using all 5 environments jointly for trait PH-V4 in KE (left side) and PE (right side). In each panel, gray horizontal line and first gray open circles refer to maximum bivariate GBLUP and maximum bivariate ERRBLUP, and the gray open circles at the top 5, 1, 0.1, 0.01, 0.001, 0.0001 quantiles refer to bivariate sERRBLUP using a single environment as an additional environment.

Tables Captions

Table 1 Phenotypic correlation across all environments for the trait PH_V4 in KE (blue numbers above diagonal) and PE (red numbers below diagonal) which are highly significant (p_values < 0.001).

Table 2 Number of KE (blue numbers above diagonal) and PE (red numbers below diagonal) phenotyped lines in each pair of environments for trait PH_V4.

References

Abendroth LJ, Elmore RW, Boyer MJ, and Marlay SK (2011) Corn Growth and Development. *PMR 1009. Iowa State University of Science and Technology, Cooperative Extension Service, Ames, Iowa.*

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, and Lawlor TJ (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93(2): 743–752. https://doi.org/10.3168/jds.2009-2730

Akdemir D and Godfrey OU (2015) EMMREML: Fitting Mixed Models with Known Covariance Structures. Available at: https://cran.r-project.org/package=EMMREML

Akdemir D and Isidro-Sánchez J (2019) Design of training populations for selective phenotyping in genomic prediction. *Scientific Reports* 9(1446). https://doi.org/https://doi.org/10.1038/s41598-018-38081-6

Albrecht T, Wimmer V, Auinger H-J, Erbe M, Knaak C, Ouzunova M, Simianer H, and Schön C-C (2011) Genome-based prediction of testcross values in maize. *Theoretical and Applied Genetics* 123(339): 339–350. https://doi.org/10.1007/s00122-011-1587-7

Bernardo R and Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47: 1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Browning SR and Browning BL (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81(5): 1084–1097. https://doi.org/10.1086/521987

Burgueño J, Campos G de los, Weigel K, and Crossa J (2012) Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. *Crop Science* 52(2): 707–719. https://doi.org/10.2135/cropsci2011.06.0299

Butler DG, Cullis BR, Gilmour AR, Gogel BJ, and Thompson R (2018) ASReml-R Reference Manual Version 4. VSN International Ltd., Hemel Hempstead

Carlborg Ö and Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* 5: 618–625. https://doi.org/10.1038/nrg1407

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, and Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4(7). https://doi.org/10.1186/s13742-015-0047-8

Covarrubias-Pazaran G, Schlautman B, Diaz-Garcia L, Grygleski E, Polashock J, Johnson-Cicalese J, Vorsa N, Iorizzo M, and Zalapa J (2018) Multivariate GBLUP Improves Accuracy of Genomic Selection for Yield and Fruit Weight in Biparental Populations of Vaccinium macrocarpon Ait.

Frontiers in Plant Science 9(1310). https://doi.org/10.3389/fpls.2018.01310

Crossa J, Campos G de los, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, and Braun H-J (2010) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186(2): 713–724. https://doi.org/10.1534/genetics.110.118521

Crossa J, Pérez P, Campos G de los, Mahuku G, Dreisigacker S, and Magorokosho C (2011) Genomic Selection and Prediction in Plant Breeding. *Crop Improvement* 25(3): 239–261

Da Y, Wang C, Wang S, and Hu G (2014) Mixed Model Methods for Genomic Prediction and Variance Component Estimation of Additive and Dominance Effects Using SNP Markers. *PLOS ONE* 9(1). https://doi.org/10.1371/journal.pone.0087666

Daetwyler HD, Pong-Wong R, Villanueva B, and Woolliams JA (2010) The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185(3): 1021–1031. https://doi.org/10.1534/genetics.110.116855

Daetwyler HD, Calus MPL, Pong-Wong R, Campos G de los, and Hickey JM (2013) Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and Benchmarking. *Genetics* 193: 347–365. https://doi.org/10.1534/genetics.112.147983

Dekkers JCM (2007) Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* 124: 331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

Desta ZA and Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* 19(9): 592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Erbe M, Pimentel E, Sharifi AR, and Simianer H (2010) Assessment of cross-validation strategies for genomic prediction in cattle. *9th World Congress of Genetics Applied to Livestock Production* 129–132

Falconer DS and Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman. Essex Engl.

Filho J de A, Guimarães J, Silva F e, Resende M de, Muñoz P, Kirst M, and Jr MR (2016) The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity (Edinb)* 117(1): 33–41. https://doi.org/10.1038/hdy.2016.23

Gianola D, de los Campos G, Hill WG, Manfredi E, and Fernando R (2009) Additive Genetic Variability and the Bayesian Alphabet. *Genetics* 183(1): 347–363. https://doi.org/10.1534/genetics.109.103952

Gianola D, Fernando RL, and Stella A (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173(3): 1761–1776. https://doi.org/10.1534/genetics.105.049510

Gianola D and van Kaam JBCHM (2008) Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* 178(4): 2289–2303. https://doi.org/10.1534/genetics.107.084285

Guo G, Zhao F, Wang Y, Zhang Y, Du L, and Su G (2014) Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genetics* 15(1): 30. https://doi.org/10.1186/1471-2156-15-30

Hallauer AR, Carena MJ, and Miranda Filho JB (2010) *Quantitative genetics in maize breeding*. *Springer*. Berlin

Hassen M Ben, Bartholomé J, Valè G, Cao T-V, and Ahmadi N (2018) Genomic Prediction Accounting for Genotype by Environment Interaction Offers an Effective Framework for Breeding Simultaneously for Adaptation to an Abiotic Stress and Performance Under Normal Cropping Conditions in Rice. *G3 (Bethesda, Md.)* 8(7): 2319–2332. https://doi.org/10.1534/g3.118.200098

Hayes B and Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome*. NRC Research Press 53(11): 876–883. https://doi.org/https://doi.org/10.1139/G10-076

Hayes BJ and Goddard ME (2008) Technical note: Prediction of breeding values using marker-derived relationship matrices. *Journal of Animal Science* 86(9): 2089–2092. https://doi.org/10.2527/jas.2007-0733

He D, Kuhn D, and Parida L (2016) Novel applications of multitask learning and multiple output regression to multiple genetic trait prediction. *Bioinformatics* 32(12): i37–i43. https://doi.org/10.1093/bioinformatics/btw249

Henderson CR and Quaas RL (1976) Multiple Trait Evaluation Using Relatives' Records. *Journal of Animal Science* 43(6): 1188–1197. https://doi.org/10.2527/jas1976.4361188x

Hill WG, Goddard ME, and Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4(2). https://doi.org/10.1371/journal.pgen.1000008

Hölker AC, Mayer M, Presterl T, Bolduan T, Bauer E, Ordas B, Brauner PC, Ouzunova M, Melchinger AE, and Schön C-C (2019) European maize landraces made accessible for plant breeding and genome-based studies. *Theoretical and Applied Genetics* 132(12): 3333–3345. https://doi.org/10.1007/s00122-019-03428-8

Hu Z, Li Y, Song X, Han Y, Cai X, Xu S, and Li W (2011) Genomic value prediction for quantitative traits under the epistatic model. *BMC Genet* 12(15). https://doi.org/https://doi.org/10.1186/1471-2156-12-15

Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, Warner CB, Blankenburg K, Han Y, Javaid M, Jayaseelan J, Jhangiani

SN, Muzny D, Ongeri F, Perales L, Wu Y-Q, Zhang Y, Zou X, Stone EA, Gibbs RA, and Mackay TFC (2012) Epistasis dominates the genetic architecture of Drosophila quantitative traits. *Proceedings of the National Academy of Sciences of the United States of America* 109(39): 15553–15559. https://doi.org/10.1073/pnas.1213423109

Jarquin D, Howard R, Crossa J, Beyene Y, Gowda M, Martini JWR, Covarrubias Pazaran G, Burgueño J, Pacheco A, Grondona M, Wimmer V, and Prasanna BM (2020) Genomic Prediction Enhanced Sparse Testing for Multi-environment Trials. *G3: Genes|Genomes|Genetics* 10(8): 2725–2739. https://doi.org/10.1534/g3.120.401349

Jiang Y and Reif JC (2015) Modeling Epistasis in Genomic Selection. *Genetics* 201(2): 759–768. https://doi.org/10.1534/genetics.115.177907

Jiang Y and Reif JC (2020) Efficient Algorithms for Calculating Epistatic Genomic Relationship Matrices. *Genetics* 216(3): 651–669. https://doi.org/10.1534/genetics.120.303459

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, and Ware D (2017) Improved maize reference genome with single-molecule technologies. *Nature* 546: 524–527. https://doi.org/10.1038/nature22971

Jones B (2012) Predicting phenotypes. *Nature Reviews Genetics* 13(7): 450. https://doi.org/10.1038/nrg3267

Kang MS and Gorman DP (1989) Genotype × Environment Interaction in Maize. *Agronomy Journal* 81(4): 662–664. https://doi.org/10.2134/agronj1989.00021962008100040020x

Karaman E, Cheng H, Firat MZ, Garrick DJ, and Fernando RL (2016) An Upper Bound for Accuracy of Prediction Using GBLUP. *PLOS ONE* 11(8): e0161054. https://doi.org/10.1371/journal.pone.0161054

Legarra A, Christensen OF, Aguilar I, and Misztal I (2014) Single Step, a general approach for genomic selection. *Livestock Science* 166: 54–65. https://doi.org/10.1016/j.livsci.2014.04.029

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, and Cotes JM (2009) Predicting Quantitative Traits With Regression Models for Dense Molecular Markers and Pedigree. *Genetics* 182(1): 375–385. https://doi.org/10.1534/genetics.109.101501

de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, and Sorensen D (2013) Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics* 9(7). https://doi.org/10.1371/journal.pgen.1003608

de Los Campos G, Gianila D, Rosa G, Weigel K, and Crossa J (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* 92(4): 295–308. https://doi.org/10.1017/S0016672310000285

784 Mackay TFC (2014) Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-
785 Gene Interactions. *Nat Rev Genet.* 15(1): 22–33. https://doi.org/10.1038/nrg3627

786 Martini JWR, Wimmer V, Erbe M, and Simianer H (2016) Epistasis and covariance: how gene
787 interaction translates into genomic relationship. *Theoretical and Applied Genetics* 129(5): 963–
788 976. https://doi.org/10.1007/s00122-016-2675-5

789 Martini JWR, Gao N, Cardoso DF, Wimmer V, Erbe M, Cantet RJC, and Henner S (2017)
790 Genomic prediction with epistasis models: on the marker-coding-dependent performance of
791 the extended GBLUP and properties of the categorical epistasis model (CE). *BMC Bioinformatics*
792 18(3). https://doi.org/10.1186/s12859-016-1439-1

793 Martini JWR, Rosales F, Ha N-T, Heise J, Wimmer V, and Kneib T (2019) Lost in Translation: On
794 the Problem of Data Coding in Penalized Whole Genome Regression with Interactions. *G3:*
795 *Genes*|*Genomes*|*Genetics* 9(4): 1117–1129. https://doi.org/10.1534/g3.118.200961

796 Martini JWR, Toledo FH, and Crossa J (2020) On the approximation of interaction effect models
797 by Hadamard powers of the additive genomic relationship. *Theoretical population biology* 132:
798 16–23. https://doi.org/10.1016/j.tpb.2020.01.004

799 Meuwissen THE, Hayes BJ, and Goddard ME (2001) Prediction of total genetic value using
800 genome-wide dense marker maps. *Genetics* 157(4): 1819–1829

801 Momen M, Mehrgardi AA, Sheikhi A, Kranis A, Tusell L, Morota G, Rosa GJM, and Gianola D
802 (2018) Predictive ability of genome-assisted statistical models under various forms of gene
803 action. *Scientific Reports* 8(12309). https://doi.org/10.1038/s41598-018-30089-2

804 Montesinos-López OA, Montesinos-López A, Crossa J, Toledo FH, Pérez-Hernández O, Eskridge
805 KM, and Rutkoski J (2016) A Genomic Bayesian Multi-trait and Multi-environment Model. *G3:*
806 *Genes*|*Genomes*|*Genetics* 6(9): 2725–2744. https://doi.org/10.1534/g3.116.032359

807 Mrode RA (2014) *Linear models for the prediction of animal breeding values*. Third edit.
808 Wallingford, OXon, UK: CAB International. https://doi.org/10.1079/9780851990002.0000

809 Ober U, Huang W, Magwire M, Schlather M, Simianer H, and Mackay TFC (2015) Accounting for
810 Genetic Architecture Improves Sequence Based Genomic Prediction for a Drosophila Fitness
811 Trait. *PLOS ONE*. Public Library of Science 10(5): e0126880.
812 https://doi.org/10.1371/journal.pone.0126880

813 Pérez P, de Los Campos G, Crossa J, and Gianola D (2010) Genomic-Enabled Prediction Based
814 on Molecular Markers and Pedigree Using the Bayesian Linear Regression Package in R. *The*
815 *Plant Genome* 3(2): 106–116. https://doi.org/10.3835/plantgenome2010.04.0005

816 Pook T, Schlather M, de los Campos G, Mayer M, Schoen CC, and Simianer H (2019)
817 HaploBlocker: Creation of subgroup specific haplotype blocks and libraries. *GENETICS* 212(4):
818 1045–1061. https://doi.org/10.1534/genetics.119.302283

Pook T, Mayer M, Geibel J, Weigend S, Cavero D, Schoen CC, and Simianer H (2020) Improving Imputation Quality in BEAGLE for Crop and Livestock Data. *G3: Genes|Genomes|Genetics* 10(1): 177–188. https://doi.org/10.1534/g3.119.400798

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW de, Daly MJ, and Sham PC (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* 81(3): 559–575. https://doi.org/10.1086/519795

Rönnegård L and Shen X (2016) Genomic prediction and estimation of marker interaction effects. *bioRxiv* 38935. https://doi.org/10.1101/038935

Schlather M (2020) Efficient Calculation of the Genomic Relationship Matrix. *bioRxiv*. https://doi.org/10.1101/2020.01.12.903146

Schrauf MF, Martini JWR, Simianer H, and de Los Campos G (2020) Phantom Epistasis in Genomic Selection: On the predictive ability of epistatic models. *G3: Genes|Genomes|Genetics* 10(9): 3137–3145. https://doi.org/10.1534/g3.120.401300

Schulthess AW, Zhao Y, Longin CFH, and Reif JC (2018) Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (Triticum aestivum L.). *Theoretical and Applied Genetics* 131(3): 685–701. https://doi.org/10.1007/s00122-017-3029-7

Stich B and Ingheland D Van (2018) Prospects and Potential Uses of Genomic Prediction of Key Performance Traits in Tetraploid Potato. *Frontiers in Plant Science* 9(159). https://doi.org/10.3389/fpls.2018.00159

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, Bertani C, Davassi A, Mayer KF, and Schön C-C (2014) A powerful tool for genome analysis in maize: 584 development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* 15(823). https://doi.org/10.1186/1471-2164-15-823

Utz HF, Melchinger AE, and Schön CC (2000) Bias and Sampling Error of the Estimated Proportion of Genotypic Variance Explained by Quantitative Trait Loci Determined From Experimental Data in Maize Using Cross Validation and Validation With Independent Samples. *Genetics* 154(4): 1839–1849. Available at: http://www.genetics.org/content/154/4/1839.abstract

VanRaden P (2007) Efficient estimation of breeding values from dense genomic data. *Journal of Dairy Science* 90: 374 –375

VanRaden P (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91(11): 4414–4423. https://doi.org/10.3168/jds.2007-0980

Velazco JG, Jordan DR, Mace ES, Hunt CH, Malosetti M, and van Eeuwijk FA (2019) Genomic

Prediction of Grain Yield and Drought-Adaptation Capacity in Sorghum Is Enhanced by Multi-Trait Analysis. *Frontiers in Plant Science* 997. Available at: https://www.frontiersin.org/article/10.3389/fpls.2019.00997

Vojgani E, Pook T, and Simianer H (2021) Phenotype Prediction under Epistasis. in KC, W. (ed.) *Epistasis: Methods and Protocols*. Springer. https://doi.org/10.1007/978-1-0716-0947-7_8

Wang D, El-Basyoni IS, Baenziger PS, Crossa J, Eskridge KM, and Dweikat I (2012) Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. *Heredity* 109(5): 313–319. https://doi.org/10.1038/hdy.2012.44

Wang J, Zhou Z, Zhang Zhe, Li H, Liu D, Zhang Q, Bradbury PJ, Buckler ES, and Zhang Zhiwu (2018) Expanding the BLUP alphabet for genomic prediction adaptable to the genetic architectures of complex traits. *Heredity* 121(6): 648–662. https://doi.org/10.1038/s41437-018-0075-0

Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, and Schön C-C (2013) Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. *Genetics* 195(2): 573–587. https://doi.org/10.1534/genetics.113.150078

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, and Melchinger AE (2012) Effectiveness of Genomic Prediction of Maize Hybrid Performance in Different Breeding Populations and Environments. *G3 (Bethesda)* 2(11): 1427–1436. https://doi.org/10.1534/g3.112.003699

Wolc A, Stricker C, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Lamont SJ, and Dekkers JCM (2011) Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution* 43(5). https://doi.org/10.1186/1297-9686-43-5

Wright S (1931) Evolution in Mendelian populations. *Genetics* 16(2): 97–159

Figures

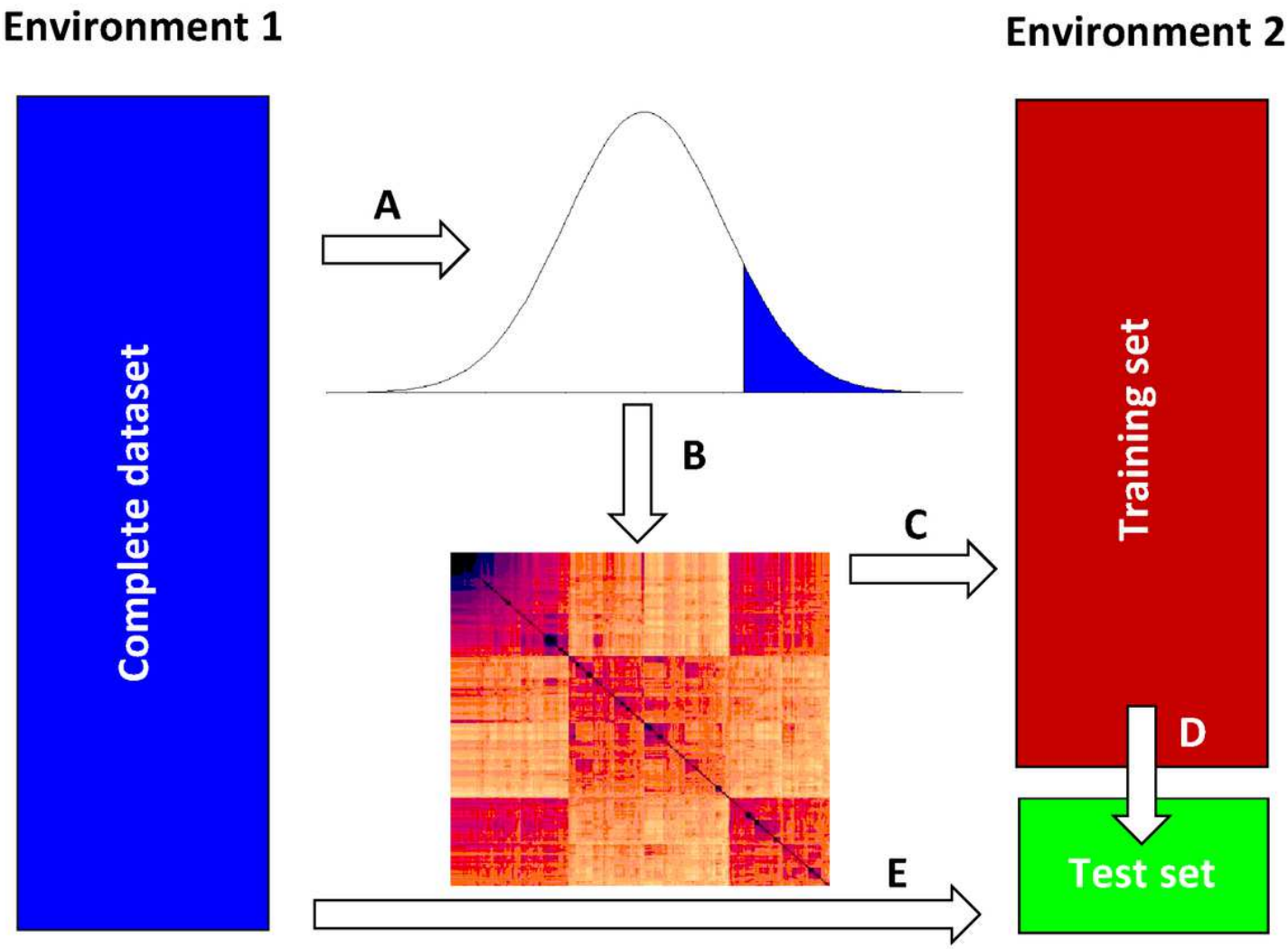

Figure 1

Basic scheme of uni- and bivariate sERRBLUP across environments. All pairwise SNP interaction effects and their variances are estimated from all data in environment 1, and effects are ordered either by absolute effect size or effect variance (A). Then, an epistatic relationship matrix for all lines is constructed from the top ranked subset of interaction effects (B) which in the univariate model is used in environment 2 (C) to predict phenotypes of the test set (green) from the respective training set (red, D). In the bivariate model, this information is combined with the complete data from environment 1 (blue, E) to predict the test set.

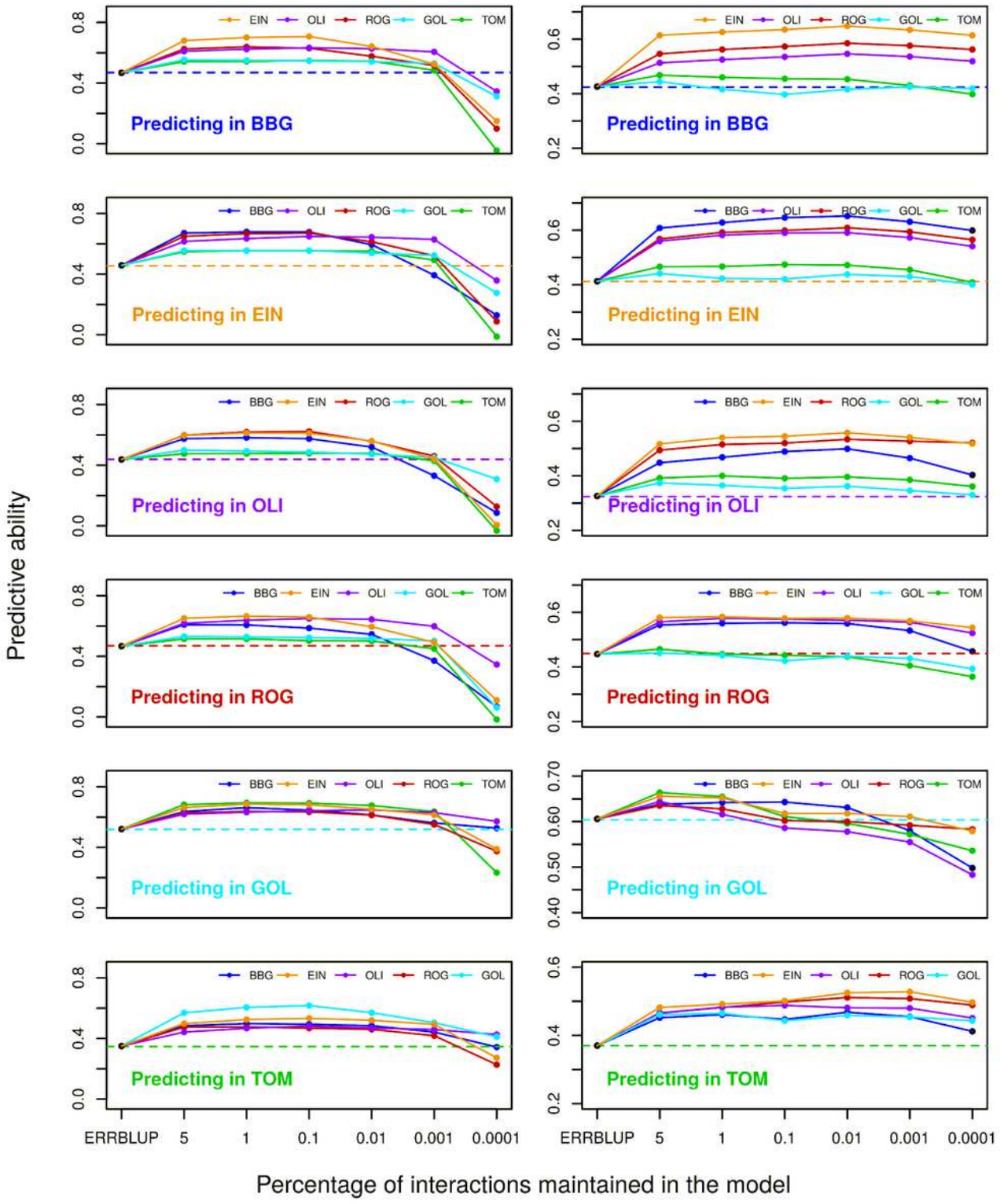

Figure 2

Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (black filled circle) and univariate sERRBLUP across environments (solid colored lines) when SNP interaction selections are based on estimated effects variances in KE (left side) and PE (right side) for trait PH_V4. In each panel, the solid lines' color indicates the environment in which the relationship matrices were determined by variable selection.

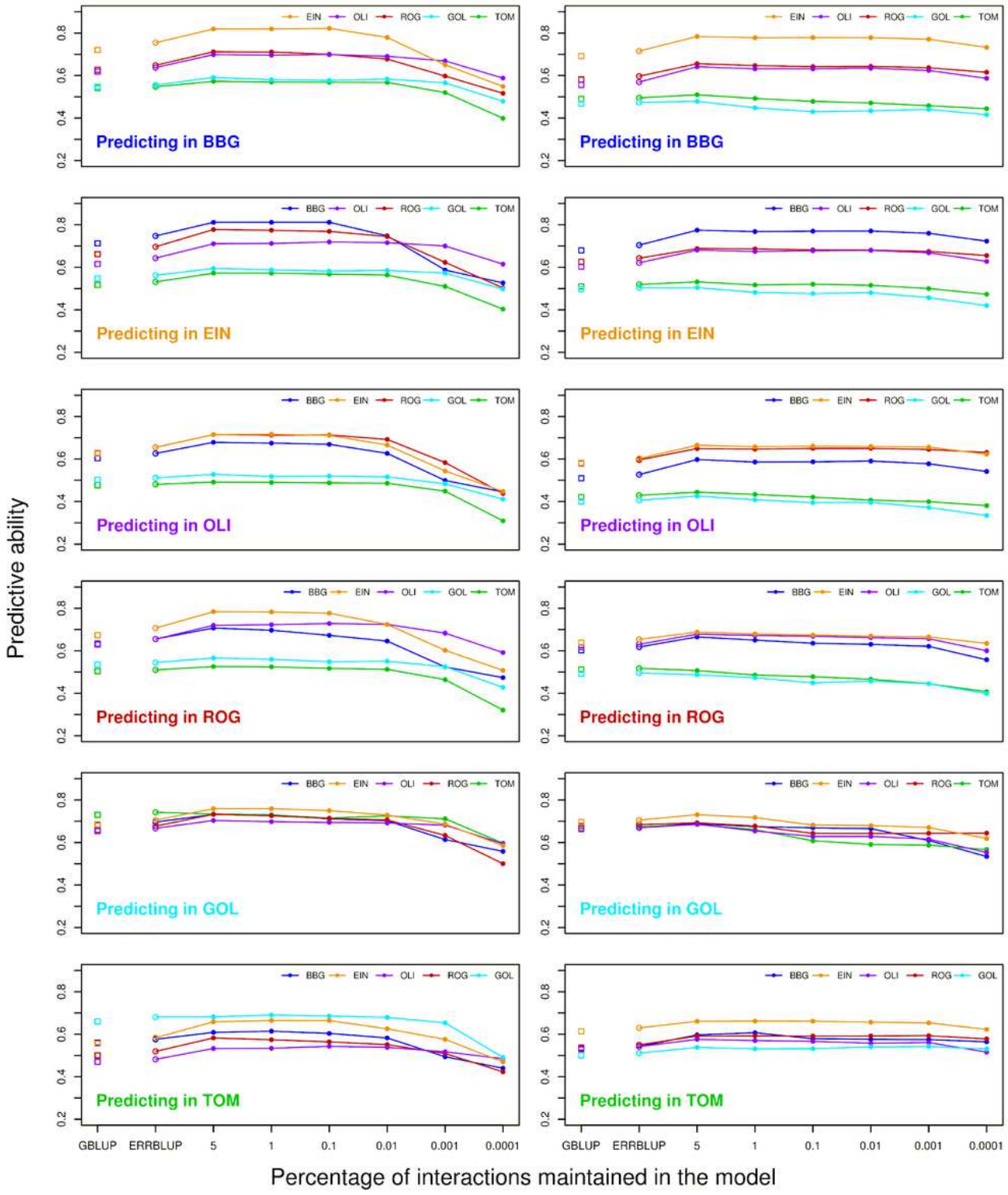

Figure 3

Predictive ability for bivariate GBLUP (open squares), bivariate ERRBLUP (open circles) and bivariate sERRBLUP (filled circles and solid lines) when SNP interaction selections are based on estimated effects variances in KE (left side) and PE (right side) for trait PH_V4. In each panel, the solid lines' color indicates the additional environment used to predict the target environment.

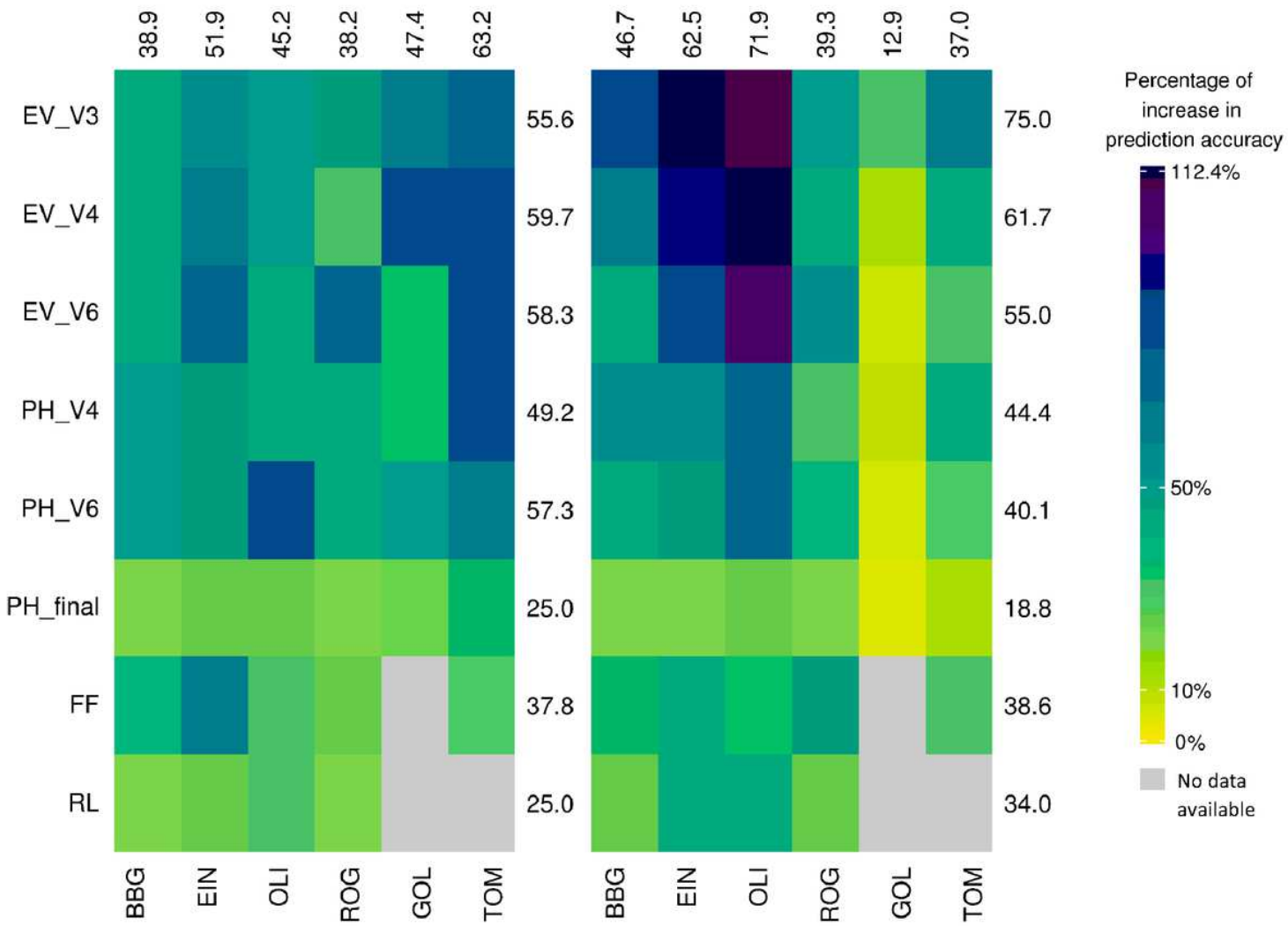

Figure 4

Percentage of increase in prediction accuracy from univariate GBLUP within environments to the maximum prediction accuracy of univariate sERRBLUP across environments when the SNP interaction selections are based on estimated effects variances in KE (left side) and in PE (right side). The average percentage of increase in prediction accuracy for each trait and environments are displayed in rows and columns, respectively.

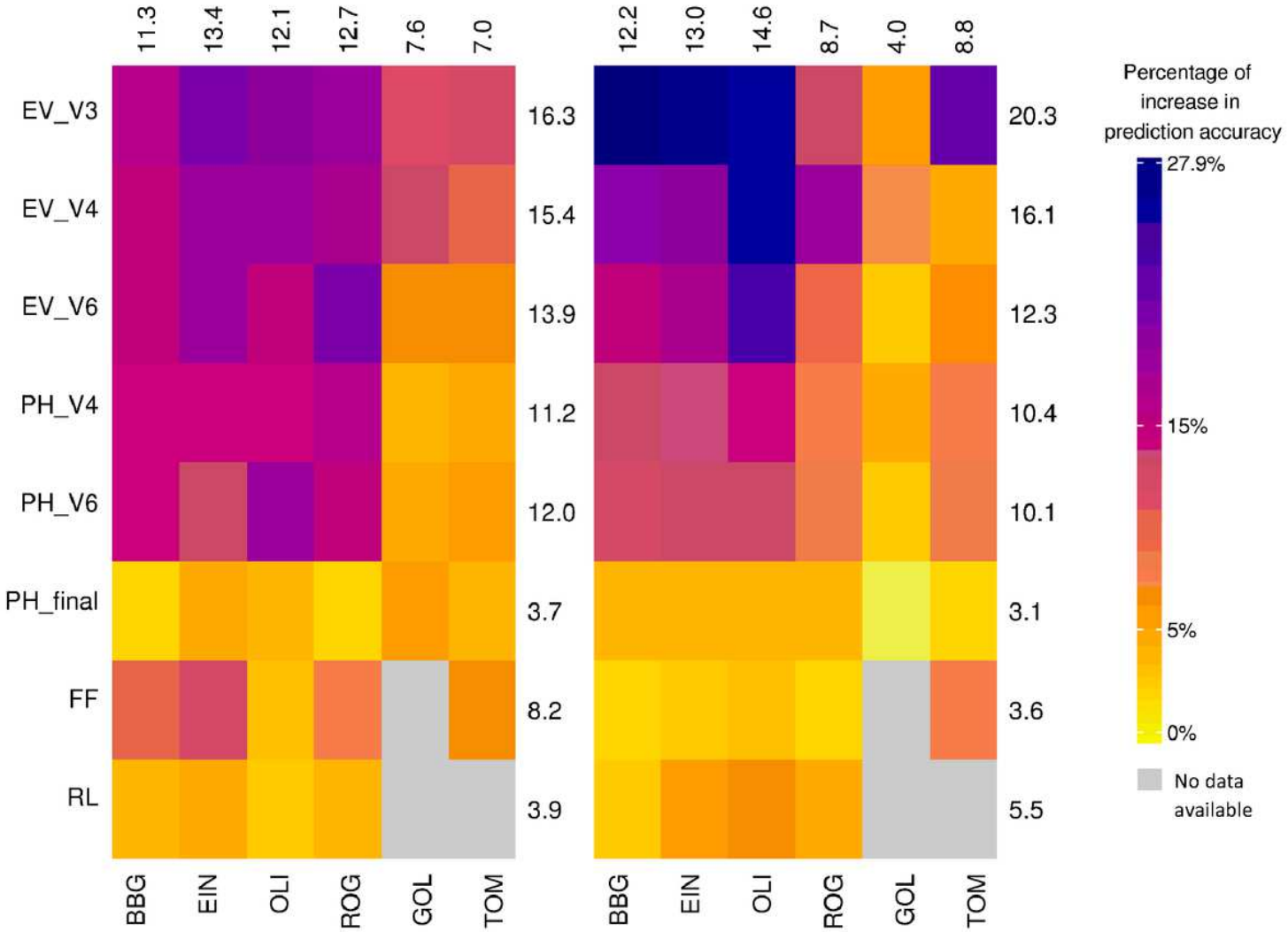

Figure 5

Percentage of increase in prediction accuracy from the maximum bivariate GBLUP to the maximum prediction accuracy of bivariate sERRBLUP when the SNP interaction selections are based on estimated effects variances in KE (left side) and in PE (right side). The average percentage of increase in prediction accuracy for each trait and environments are displayed in rows and columns, respectively.

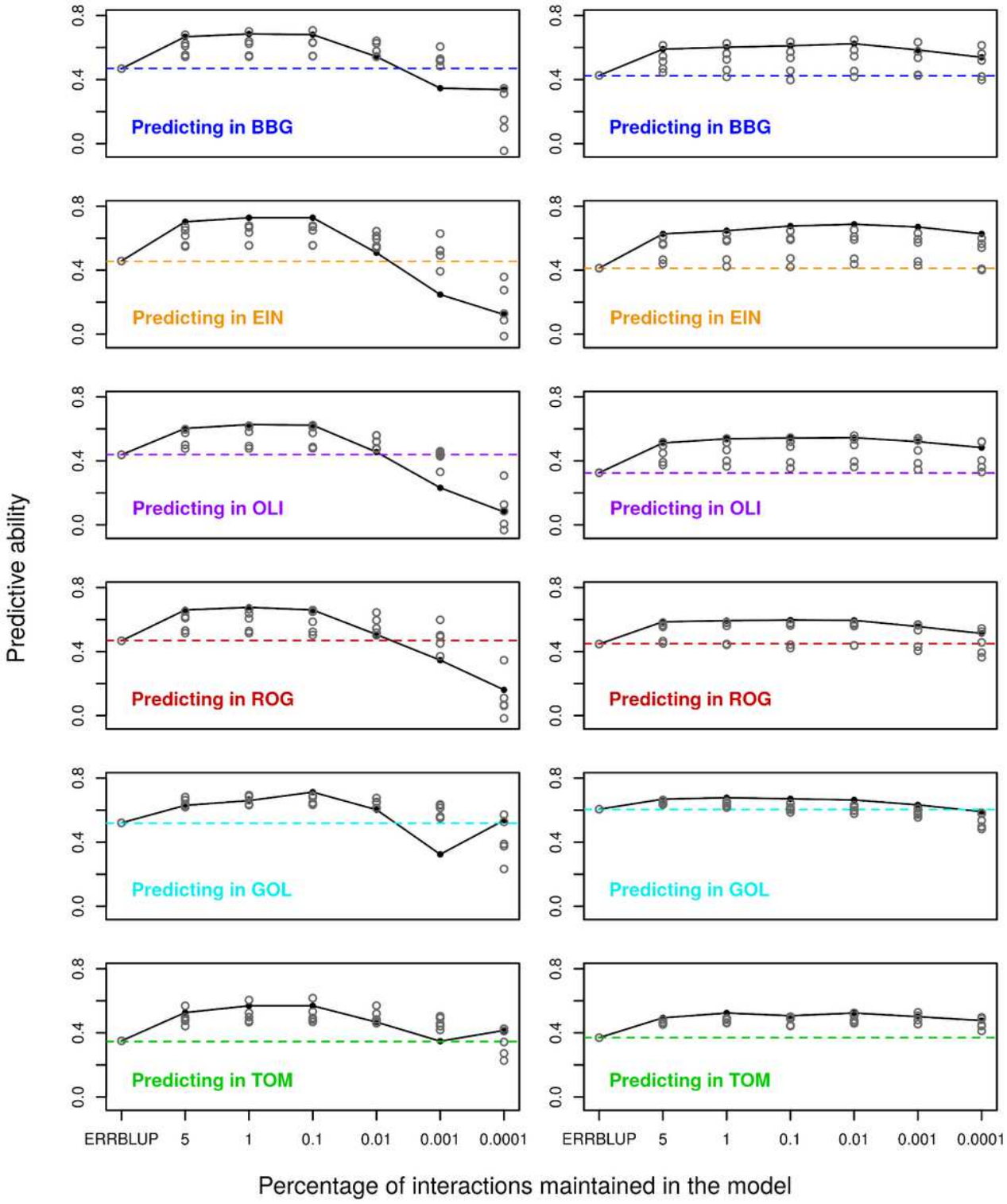

Figure 6

Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (gray open circle), univariate sERRBLUP using a single environment for selecting the SNP interactions (gray open circles) and univariate sERRBLUP using all 5 environments jointly (filled black circles and solid line) for the SNP interaction selection based on estimated effects variances for trait PH_V4 in KE (left side) and PE (right side).

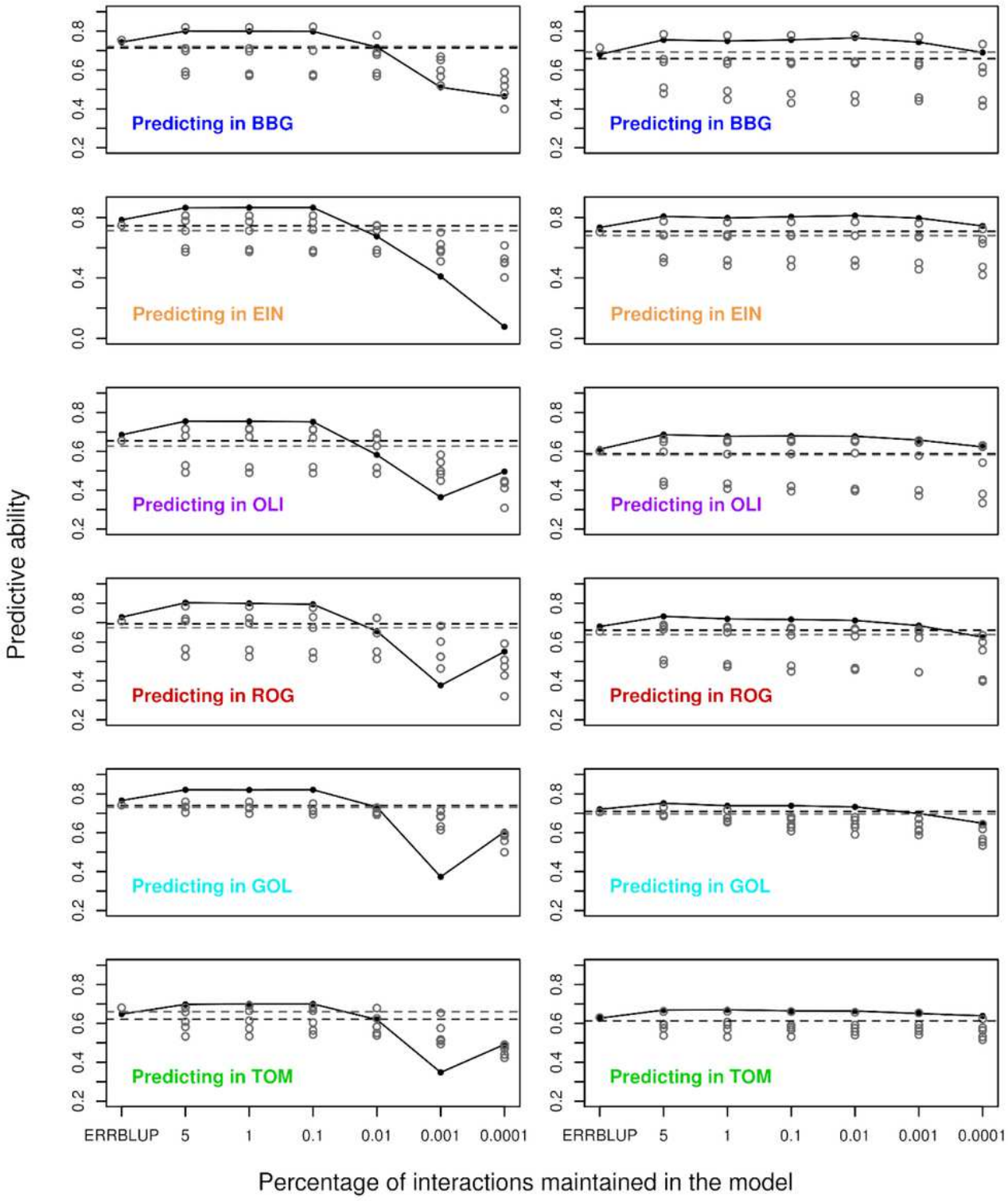

Figure 7

Predictive ability for bivariate GBLUP (black dashed horizontal line), bivariate ERRBLUP and bivariate sERRBLUP (filled black circles) for the SNP interaction selection based on estimated effects variances using all 5 environments jointly for trait PH-V4 in KE (left side) and PE (right side). In each panel, gray horizontal line and first gray open circles refer to maximum bivariate GBLUP and maximum bivariate

ERRBLUP, and the gray open circles at the top 5, 1, 0.1, 0.01, 0.001, 0.0001 quantiles refer to bivariate sERRBLUP using a single environment as an additional environment.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ElectronicSupplementaryMaterial.pdf