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

Abstract

We compared the predictive ability of various prediction models for a maize dataset derived from 910 doubled haploid lines from European landraces (Kemater Landmais Gelb and Petkuser Ferdinand Rot), which were tested in 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 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. Over 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. The average increase across traits and locations was 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

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 (Haley and Visscher, 1998; Hayes and Goddard, 2008) 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).

One of the first broad applications of genomic prediction was to select young sires with high breeding values earlier in their life span, without the need for information of the performance of their progeny in dairy cattle breeding (Schaeffer, 2006; VanRaden, 2007). 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). The computational steps involved in GBLUP are much faster than Bayesian methods and it has been difficult to find a method which consistently outperforms GBLUP when predicting complex traits (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 was proposed to reduce the computational load by constructing marker-based epistatic relationship matrices (Jiang and Reif, 2015; Martini *et al.*, 2016). While EG-BLUP is potentially beneficial for genomic prediction, its performance depends on the marker coding (Martini *et al.*, 2017, 2019), and the Hadamard products of the additive genomic relationship matrices provide only an approximation for the interaction effect model based on interactions between individual loci (Martini *et al.*, 2020). 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).

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. Therefore, 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 predictive ability of 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 two doubled haploid populations 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 underlying phenotypic data was generated in multi-location trials, and we assessed whether the different univariate and bivariate models have a potential to provide benefits in across location predictions. 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 costs 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 and Petkuser, respectively.

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 weather data which were obtained from 15th of April to 30th of September 2017 are provided by Hölker *et al.* (2019).

The phenotypic trait description and the mean, standard deviation, maximum and minimum values of each trait across all six locations in 2017 and both landraces are provided in Table 1. Moreover, these values are shown in the supplementary for each environment and for each landrace separately (Table S1). The vegetative growth stage is the corn growth stage based on the leaf collar method representing the number of visible leaf collars. To illustrate this, V4 indicates the growth stage at which four leaf collars are fully developed (Abendroth *et al.*, 2011). In our study, we consider PH_V4 as the main trait for evaluating our methods, since it is a metric quantitative trait for early plant development which is suitable to use for testing our methods. The phenotypic correlations of PH_V4 across all environments are provided in Table 2.

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. FF was phenotyped in all environments except GOL.

160 **Table 1** Phenotypic traits description and the mean, minimum, maximum and standard deviation for all landraces.

Trait	Definition	Mean	Minimum	Maximum	Standard deviation
EV_V3	Early vigour at V3 stage scored on scale from 1 (very poor early vigour) to 9 (very high early vigour)	5.22	0.79	9.01	1.32
EV _V4	Early vigour at V4 stage scored on scale from 1 (very poor early vigour) to 9 (very high early vigour)	5.12	0.7	8.49	1.3
EV _V6	Early vigour at V6 stage scored on scale from 1 (very poor early vigour) to 9 (very high early vigour)	5.36	0.4	8.86	1.25
PH_V4	Mean plant height of three plants of the plot at V4 stage in cm	35.34	6.67	95.35	14.65
PH_V6	Mean plant height of three plants of the plot at V6 stage in cm	65.57	6.73	130.37	20.04
PH_final	Final plant height after flowering in cm	131.73	29.33	241.96	27.05
FF	Days after sowing until female flowering (days until 50% of the plot showed silks)	79.71	59.12	102.22	6.46
RL	Root lodging score from 1 to 9 (1 belonged to no lodging and 9 belonged to severe lodging)	2.83	0.27	9.38	2.29

161

Table 2 Phenotypic correlation across all environments for the trait PH_V4.

Location	EIN	OLI	ROG	GOL	TOM
BBG	0.817	0.660	0.674	0.689	0.575
EIN	-	0.707	0.767	0.747	0.662
OLI	-	-	0.711	0.649	0.503
ROG	-	-	-	0.703	0.575
GOL	-	-	-	-	0.693

Quality control, coding and imputing

In total, 948 DH lines from the landraces Kemater (KE) and Petkuser (PE) were genotyped using the 600k Affymetrix® Axiom® Maize Array (Unterseer et al. 2014). After stringent quality filtering, as described in Hölker *et al.* (2019), the dataset was reduced to 501,124 SNPs and 910 DH lines from KE and PE. Remaining 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). As the dataset only included doubled haploid lines and heterozygous calls were not expected, the DS (dosage) information of the BEAGLE output was used to recode remaining heterozygous calls. The genotype was then coded as 0 if DS < 1 and as 2 if DS >= 1.

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 was the window size in SNPs, 5 was the number of SNPs to shift the window at each step and 2 was the variance inflation factor $VIF = 1/(1 - r^2)$, where r^2 was 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 the same 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

$$\begin{bmatrix} \hat{\mu} \\ \hat{\boldsymbol{g}}_1 \\ \hat{\boldsymbol{g}}_2 \end{bmatrix} = \begin{bmatrix} n_1 & \mathbf{1}' & \mathbf{0} \\ \mathbf{1} & \boldsymbol{I} + \boldsymbol{\lambda}\boldsymbol{\Gamma}^{11} & \boldsymbol{\lambda}\boldsymbol{\Gamma}^{12} \\ \mathbf{0} & \boldsymbol{\lambda}\boldsymbol{\Gamma}^{21} & \boldsymbol{\lambda}\boldsymbol{\Gamma}^{22} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'\boldsymbol{y} \\ \boldsymbol{y} \\ \mathbf{0} \end{bmatrix}, \qquad \text{(eq. 1)}$$

where $\lambda = \sigma_\epsilon^2 / \sigma_g^2$, $\boldsymbol{\Gamma}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}^{11} & \boldsymbol{\Gamma}^{12} \\ \boldsymbol{\Gamma}^{21} & \boldsymbol{\Gamma}^{22} \end{bmatrix}$ 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 solutions, the phenotypes for the set of unphenotyped individuals can be predicted as $\hat{\boldsymbol{y}}_2 = \mathbf{1}_2\hat{\mu} + \hat{\boldsymbol{g}}_2$, where $\hat{\boldsymbol{y}}_2$ is the $n_2 \times 1$ vector of predicted phenotypes and $\mathbf{1}_2$ 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,

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

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), $\mathbf{1_1}$ and $\mathbf{1_2}$ 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 the case that the number of observations in environment 1 and environment 2 not being identical (i.e. in general terms $m_1 \neq m_2$) or not having the same lines in both environments 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)

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

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

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

288 $$\boldsymbol{\Gamma_{sERR}} = \frac{(\boldsymbol{M_{\pi}^{*}} - \boldsymbol{P_{\pi}^{*}})(\boldsymbol{M_{\pi}^{*}} - \boldsymbol{P_{\pi}^{*}})'}{\sum_{i=1}^{\pi m^{*}} \left(p_{\pi i}^{*}(1 - p_{\pi i}^{*})\right)},$$

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

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

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

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

297 In a 5-fold cross validation, the original sample is randomly partitioned into five subsamples of
298 equal size. Out of the five subsamples, each subsample is subsequently considered as the test set
299 for validating the model, and the remaining four subsamples are considered as training data. The
300 training set is used to predict the test (validation) set. By this, all observations are used for both
301 training and validation and each observation is only used once for validation (Utz *et al.*, 2000).
302 We repeated the cross-validation procedure 5 times, using random partitions of the original
303 sample. The results of the 25 repetitions were then averaged (Erbe *et al.*, 2010). We used the
304 Pearson correlation between the predicted genetic value and the observed phenotype in the test
305 set as measure for predictive ability. In our study, predictive ability was assessed for PE and KE
306 for all phenotypic traits separately. In addition, the trait's prediction accuracy was calculated by
307 dividing the obtained predictive ability by the square-root of the respective trait's heritability
308 (Dekkers, 2007). The numbers of KE and PE lines which are available for all combinations of
309 environments are summarized in Table 3. For some traits these numbers can be smaller or even
310 zero for of some environment combinations.

311

312

313

Table 3 Number of Kemater (blue numbers above diagonal) and Petkuser (red numbers below diagonal) lines phenotyped in each pair of environments for trait PH_V4.

Location	BBG	EIN	OLI	ROG	GOL	TOM
BBG	393/461	461	441	461	200	200
EIN	393	393/462	441	461	201	201
OLI	390	390	390/441	441	182	181
ROG	390	390	389	390/461	200	200
GOL	195	195	195	195	204/211	209
TOM	195	195	195	195	204	204/210

We evaluated our univariate and bivariate models as follows:

Assessment of univariate GBLUP and ERRBLUP predictive abilities within environments

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

Assessment of univariate sERRBLUP predictive ability across environments

The basic strategy for sERRBLUP across environments is illustrated in Fig. 1: first, 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). Next, an epistatic relationship matrix for all lines is constructed from the top ranked subset of 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'.

Assessment of bivariate GBLUP, ERRBLUP predictive abilities

The basic strategy for bivariate GBLUP and ERRBLUP is illustrated in Fig. 2: The model is trained jointly on the complete dataset of environment 1 (A) and the training set of environment 2 (B). The test set of environment 2 is predicted, using as dispersion matrix for the genetic effects either $\boldsymbol{\Gamma_{VR}}$ or $\boldsymbol{\Gamma_{ERR}}$.

Assessment of bivariate sERRBLUP predictive ability

The basic strategy of bivariate sERRBLUP is also illustrated in Fig. 1: first, all pairwise SNP interaction effects variances are estimated from univariate ERRBLUP model based on all data in

environment 1, and effects are ordered by effect variance (A). Then, an epistatic relationship matrix $\boldsymbol{\Gamma}_{\boldsymbol{sERR}}$ for all lines is constructed from the top ranked subset of interaction effects (B) This relationship matrix $\boldsymbol{\Gamma}_{\boldsymbol{sERR}}$ is then used to predict the phenotypes of the test set (green) of environment 2 jointly from the training set of environment 2(red) (D) and the complete data set of environment 1 (blue) (E).

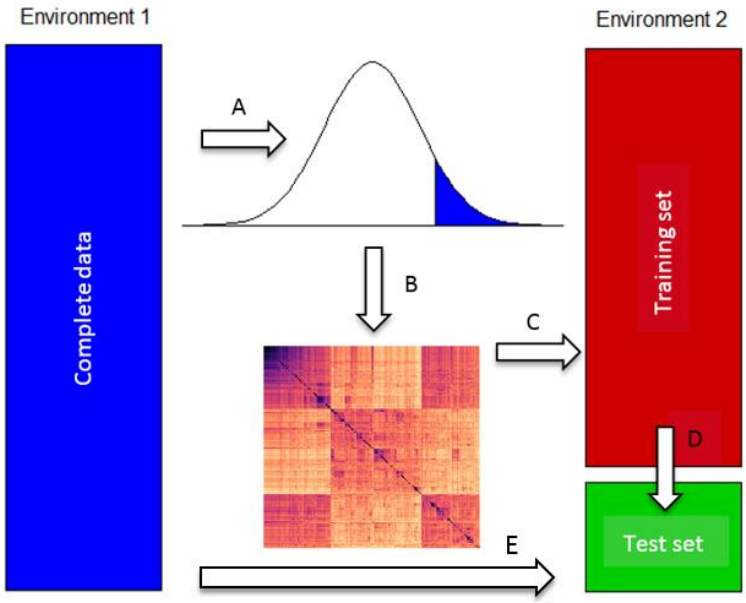


Fig. 1 Basic scheme of uni- and bivariate sERRBLUP across environments. All pairwise SNP interaction effects and their variances are estimated from all data in environment 1, and effects are ordered either by absolute effect size or effect variance (A). Then, an epistatic relationship matrix for all 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.

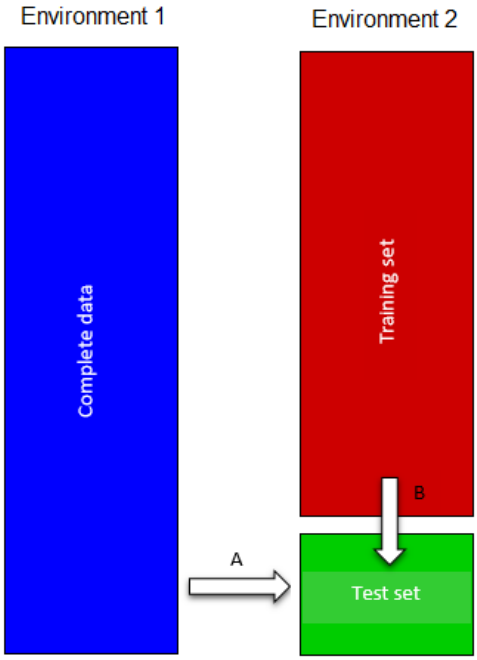


Fig. 2 Basic scheme of bivariate GBLUP and ERRBLUP illustrating that the model is trained in both the complete data set of environment 1 (A) and the training set of environment 2 (B) to predict the test set in environment 2.

Assessment of univariate sERRBLUP and bivariate GBLUP, ERRBLUP and sERRBLUP predictive ability across five environments jointly

In addition to assessing the predictive ability of univariate sERRBLUP when borrowing information from a single alternative environment, we considered five environments jointly for borrowing information to predict the sixth environment. Based on this approach, we combined five environments' phenotypes by considering the mean of all five centered environments' mean phenotypic values. The combined phenotypes for predicting BBG from EIN, OLI, ROG, GOL and TOM were jointly calculated as

$$\boldsymbol{Y}_{\boldsymbol{1\ldots5}} = \frac{\sum_{i=1}^{5}(\boldsymbol{Y_i} - mean(\boldsymbol{Y_i}))}{5},$$

where $\boldsymbol{Y}_{\boldsymbol{1\ldots5}}$ is the averaged phenotypes of EIN ($\boldsymbol{Y_1}$), OLI ($\boldsymbol{Y_2}$), ROG ($\boldsymbol{Y_3}$), GOL ($\boldsymbol{Y_4}$), TOM ($\boldsymbol{Y_5}$).

Borrowing information for variable selection across five environments jointly for univariate sERRBLUP follows the same strategy as borrowing information from a single environment for univariate sERRBLUP. Therefore, the newly generated phenotypes based on five environments (i.e. $\boldsymbol{Y}_{\boldsymbol{1\ldots5}}$) were used to estimate the interaction effects and their variances, then the

corresponding relationship matrices were constructed to be used for a prediction within the sixth (i.e. BBG ($\boldsymbol{Y_6}$)) environment. This approach henceforth is termed 'univariate sERRBLUP across multiple environments jointly'.

This approach was also applied to bivariate GBLUP, ERRBLUP and sERRBLUP by using the combined phenotypes of the other five environments as the additional environment instead of a single environment's phenotypes which are then termed 'bivariate GBLUP across multiple environments jointly', 'bivariate ERRBLUP across multiple environments jointly' and 'bivariate sERRBLUP across multiple environments jointly'.

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, we needed to specify a strategy to deal with such cases in an automated manner, given the huge number of analyses we ran. 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. 3). 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. 4), 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 converged in 100 iterations are detailed in the supplementary (Table S2 – S9).

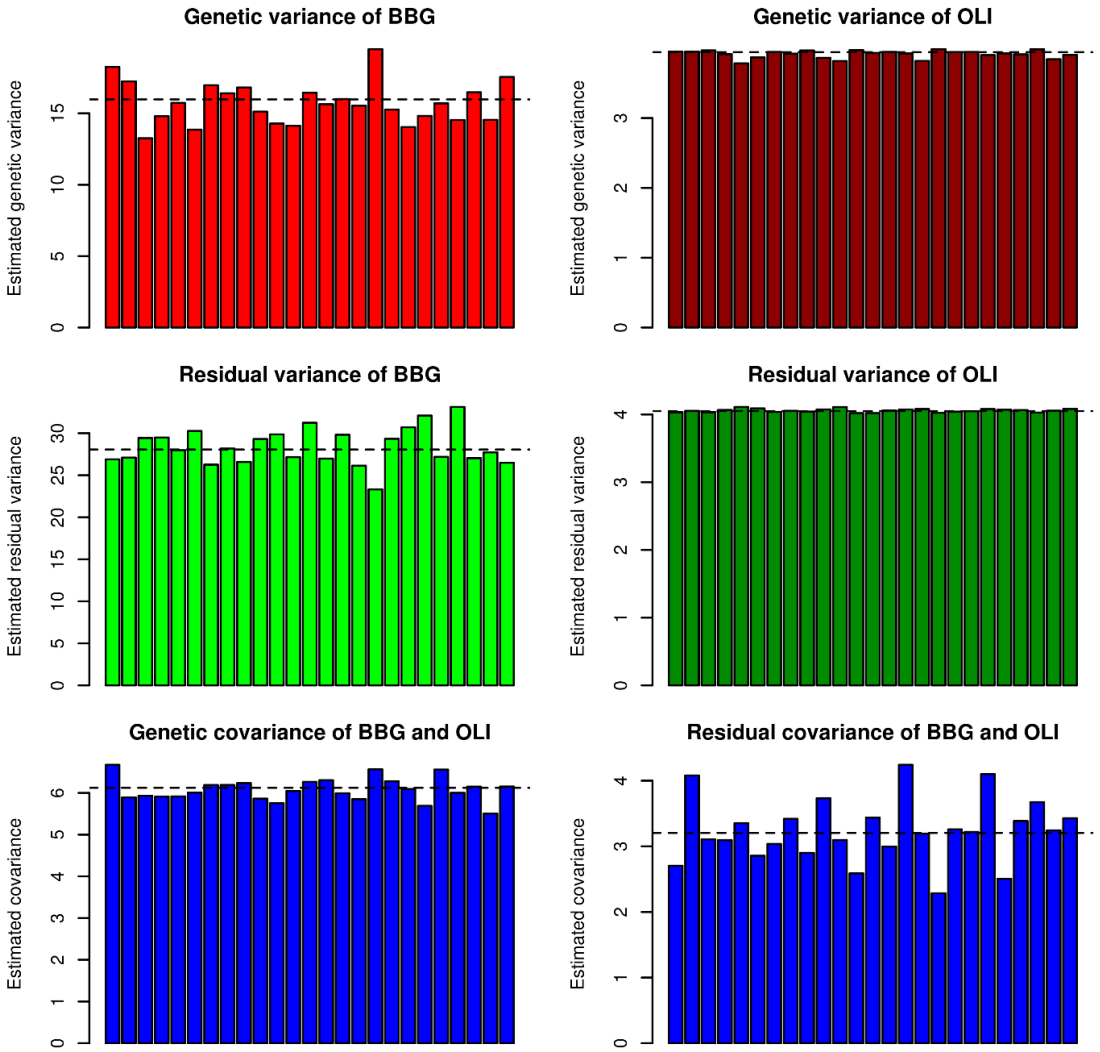


Fig. 3 Comparison of pre estimated genetic and residual variances and covariances of converged bivariate sERRBLUP model (top 5%) based on the full dataset (dashed horizontal lines) and estimated genetic and residual variances and covariances of converged bivariate sERRBLUP (top 5%) based on training set in each run of 5-fold cross validation with 5 replicates (colored bars) for predicting BBG when the additional environment is OLI in Kemater for trait PH-V4.

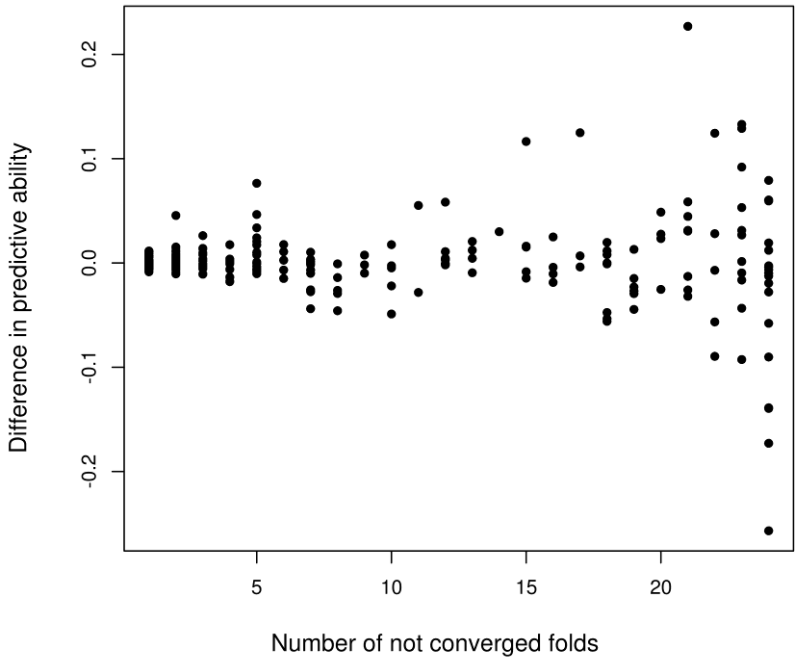
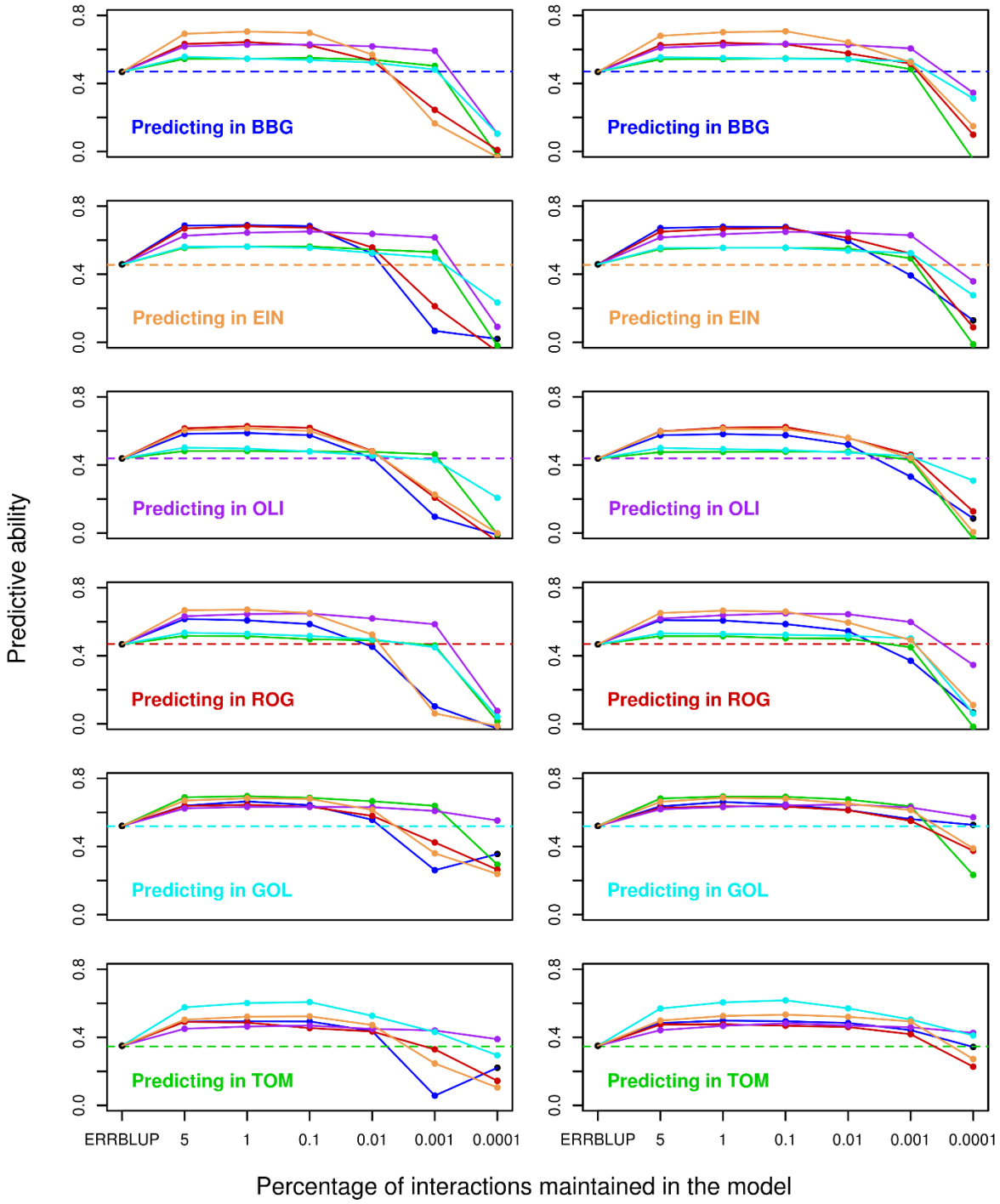


Fig. 4 The difference between the mean predictive ability of only the converged folds and the mean predictive ability of all folds in 5-fold cross validation with 5 replicates vs. the number of the folds (1 to 24) which did not converge across all traits in all combinations for both Kemater and Petkuser.

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. 5 and 6 for KE and PE, respectively. 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. 5 and 6 show 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 were kept in the univariate sERRBLUP model regardless of the interaction selection strategy. 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. 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 over all environments which is more pronounced in KE then PE. Moreover, the predictive abilities of sERRBLUP when the selection of pairwise SNP interaction was based on estimated absolute effect sizes (left hand side of the panel) were compared to the predictive abilities of sERRBLUP when

SNP interactions were selected based on estimated effects variances (right hand side of the panel). This comparison reveals that the interaction selection based on the effect variances is more robust than interaction selection based on the absolute effect sizes, especially when the top 0.001 percent of interactions are maintained in the model for KE. However, this comparison in Fig. 6 does not display a considerable difference for the interaction selection strategy for PE. Due to the higher robustness of the approach based on interaction variances as selection criterion for KE, we used this strategy for the series of all other traits in both landraces, for which results are shown in the supplementary (Fig. S1a – S7a).



438

439 **Fig. 5** Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate
440 ERRBLUP within environment (black filled circle) and univariate sERRBLUP across environments (solid
441 colored lines) when SNP interaction selections are based on estimated effects sizes (left side) and
442 estimated effects variances (right side) for trait PH-V4 in Kemater. In each panel, the solid lines' color
443 indicates the environment in which the relationship matrices were determined by variable selection.

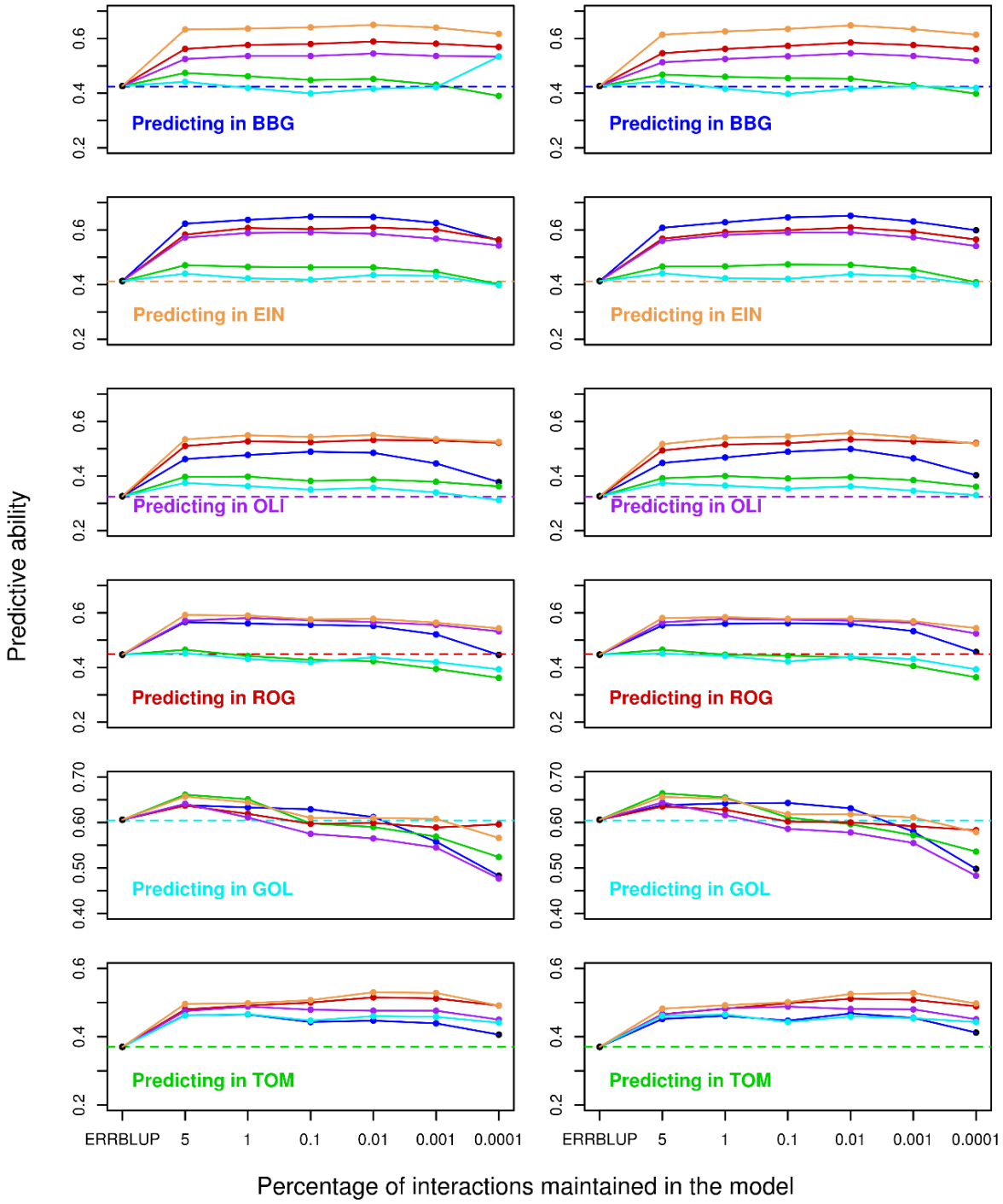
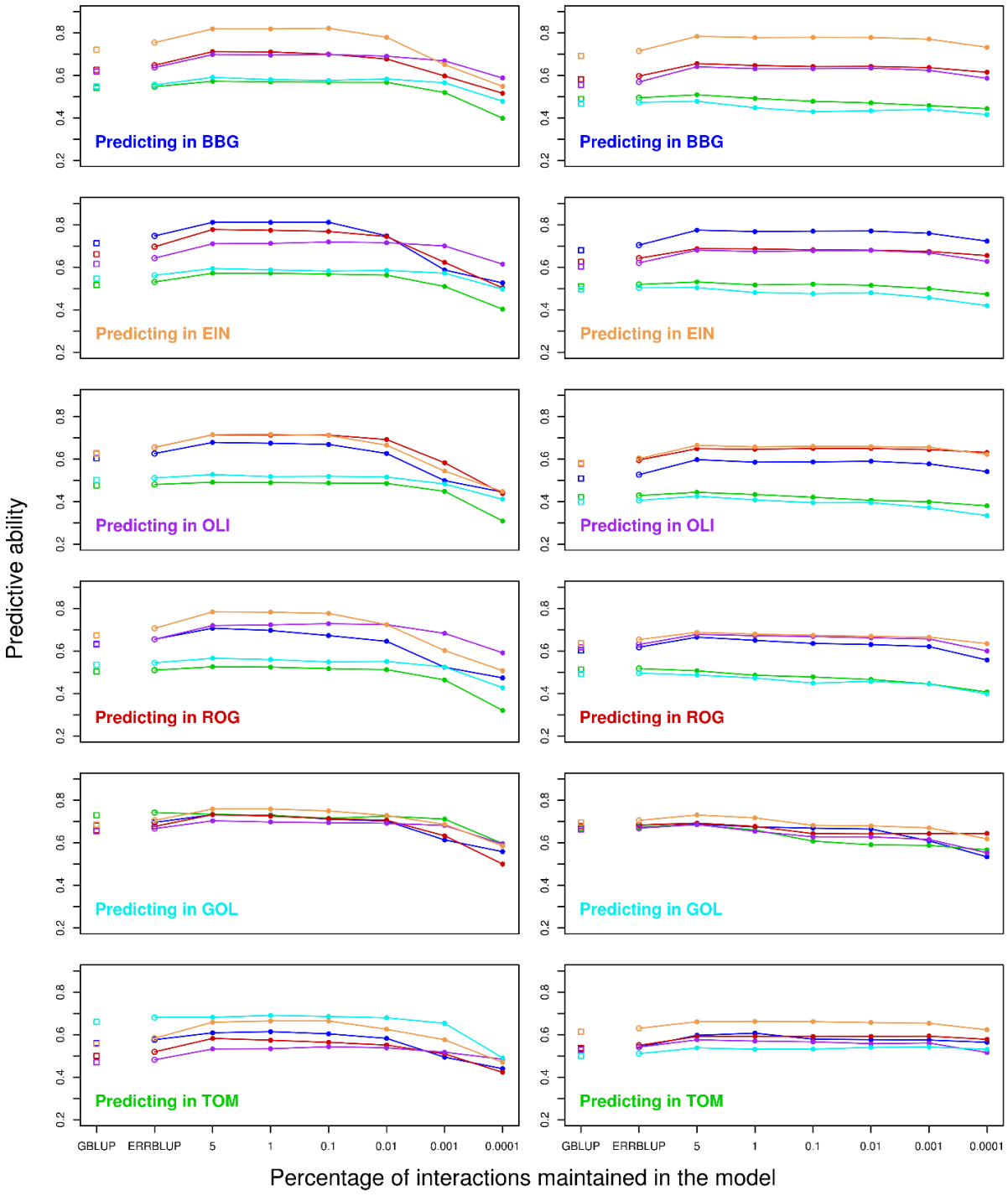


Fig. 6 Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (black filled circle) and univariate sERRBLUP across environments (solid colored lines) when SNP interaction selections are based on estimated effects sizes (left side) and estimated effects variances (right side) for trait PH-V4 in Petkuser. In each panel, the solid lines' color indicates the environment in which the relationship matrices were determined by variable selection.

In the context of univariate models, we also investigated the predictive ability of univariate sERRBLUP when variable selection was based on the training set of the same environment as the test set was sampled from. This was exemplarily done within Bernburg for the trait PH-V4 (see Fig. S8), 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 selection of effects is too strict, with only 0.001 percent of interactions used, predictive ability of univariate sERRBLUP within Bernburg falls short of the one obtained with GBLUP, especially if selection is based on effect sizes. Consequently, we did not include this model variant in our model comparison due to its poor performance which was obtained at a cost of high computing time, required for setting up a different univariate sERRBLUP relationship matrix for each training set.

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. 7. Fig. 7 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. S1b – S7b).



472

473 **Fig. 7** Predictive ability for bivariate GBLUP (open squares), bivariate ERRBLUP (open circles) and bivariate
474 sERRBLUP (filled circles and solid lines) when SNP interaction selections are based on estimated effects
475 variances in Kemater (left side) and Petkuser (right side) for trait PH-V4. In each panel, the solid lines' color
476 indicates the additional environment used to predict the target environment.

Fig. 8 provides a comparison of predictive ability obtained from univariate GBLUP within environments and the maximum predictive ability obtained from univariate sERRBLUP across environments based on both strategies of interaction selection for trait PH_V4. Both panels in Fig. 8 demonstrate that for both landraces in each environment univariate sERRBLUP across environments outperforms univariate GBLUP regardless of the selection strategy. It also demonstrates that the maximum predictive ability obtained when selecting interactions based on absolute effect sizes and their variances are almost identical.

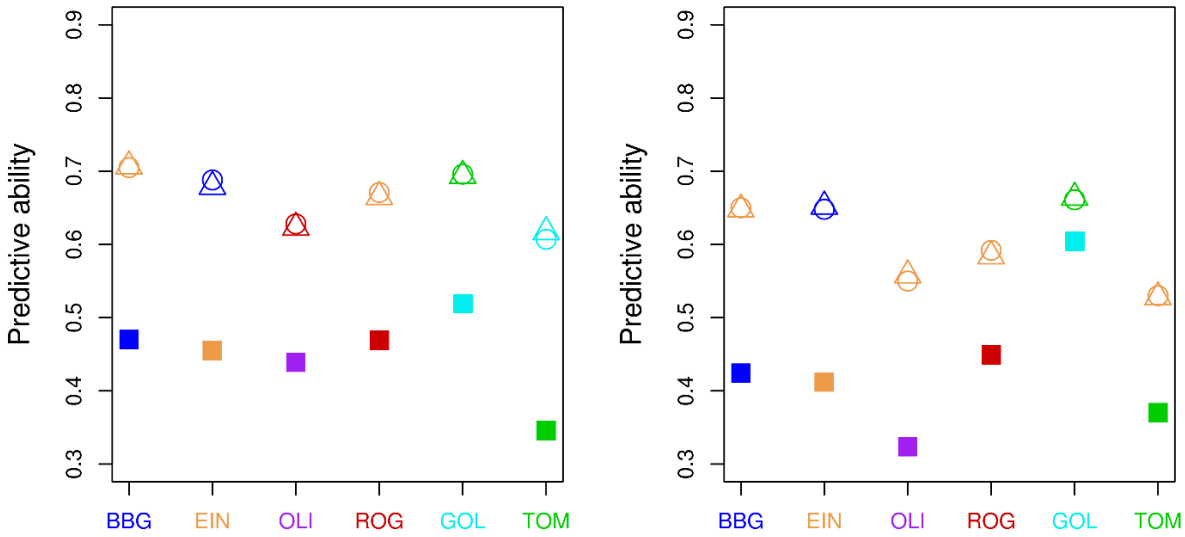


Fig. 8 Comparison of predictive ability of univariate GBLUP within environments (filled squares) and the maximum predictive ability of univariate sERRBLUP across environments when the SNP interaction selections are based on estimated effects sizes (circles) and estimated effects variances (triangles) for trait PH-V4 in Kemater (left side plot) and in Petkuser (right side plot). The colors dark blue, green, red, purple, light blue and orange represent the environments BBG, EIN, OLI, ROG, GOL and TOM, respectively. The circles' and triangles' colors indicate the environment which had the maximum predictive ability for this respective target environment

The maximum predictive ability obtained from bivariate GBLUP is also compared with the maximum predictive ability obtained from bivariate sERRBLUP when pairwise SNP interaction selections were based on estimated effects variances in Fig. 9 for the trait PH_V4. Both panels of Fig. 9 illustrate that the predictive ability in each environment is enhanced from bivariate GBLUP to bivariate sERRBLUP in both landraces.

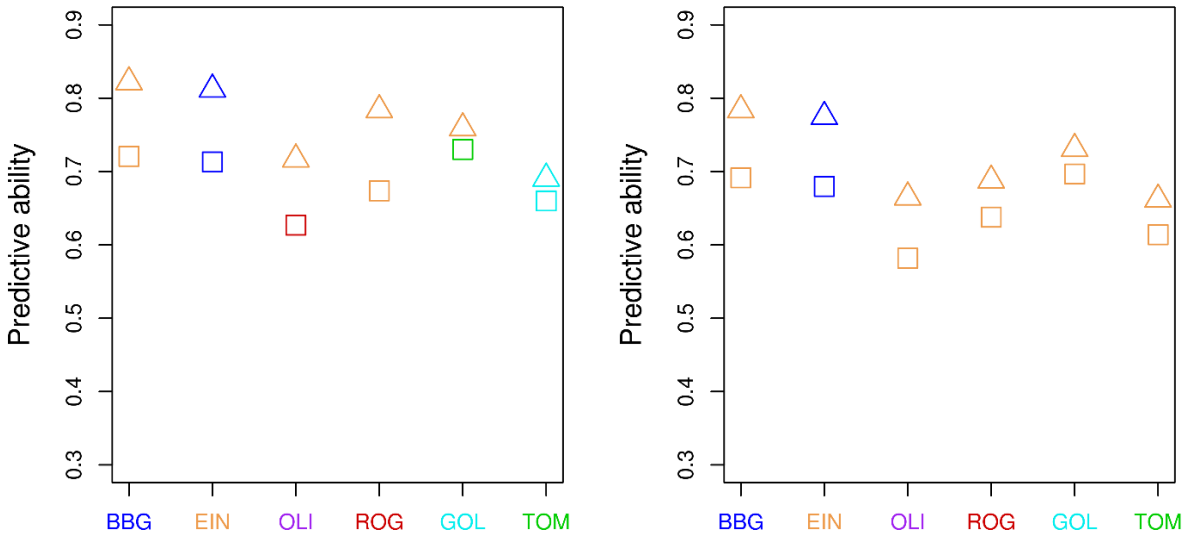


Fig. 9 Comparison of maximum predictive ability of bivariate GBLUP (squares) to the maximum predictive ability of bivariate sERRBLUP (triangles) when the SNP interactions are selected based on estimated effects variances for trait PH-V4 in Kemater (left side plot) and in Petkuser (right side plot). The colors dark blue, green, red, purple, light blue and orange represent the environments BBG, EIN, OLI, ROG, GOL and TOM, respectively. The colors of the squares and triangles indicate the environment leading to the maximal predictive ability for this respective target environment.

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. 10 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 environments 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. 10 demonstrate that for the majority of traits and environments, there is more than 30 percent increase in prediction accuracy from univariate GBLUP within environments to the best univariate sERRBLUP across environments. Average increase across all combinations in KE is 47.1 percent and in PE is 46.7 percent. Fig. 10 also shows the average increase in prediction accuracy for each environment across series of all traits and for each trait across series of all environments. The maximum average increase in prediction accuracy among all traits was found for EV_V4 in KE (59.7 percent) and EV_V3 in PE (75.0 percent). Regarding environments, the maximum average increase in prediction accuracy obtained for univariate sERRBLUP was in TOM for KE (63.3 percent) and in OLI for PE (71.9 percent). Additionally, the

absolute increase in prediction accuracy is shown as a heat map in supplementary (Fig. S9a) which indicates the average absolute increase of 0.204 in KE and 0.181 in PE.

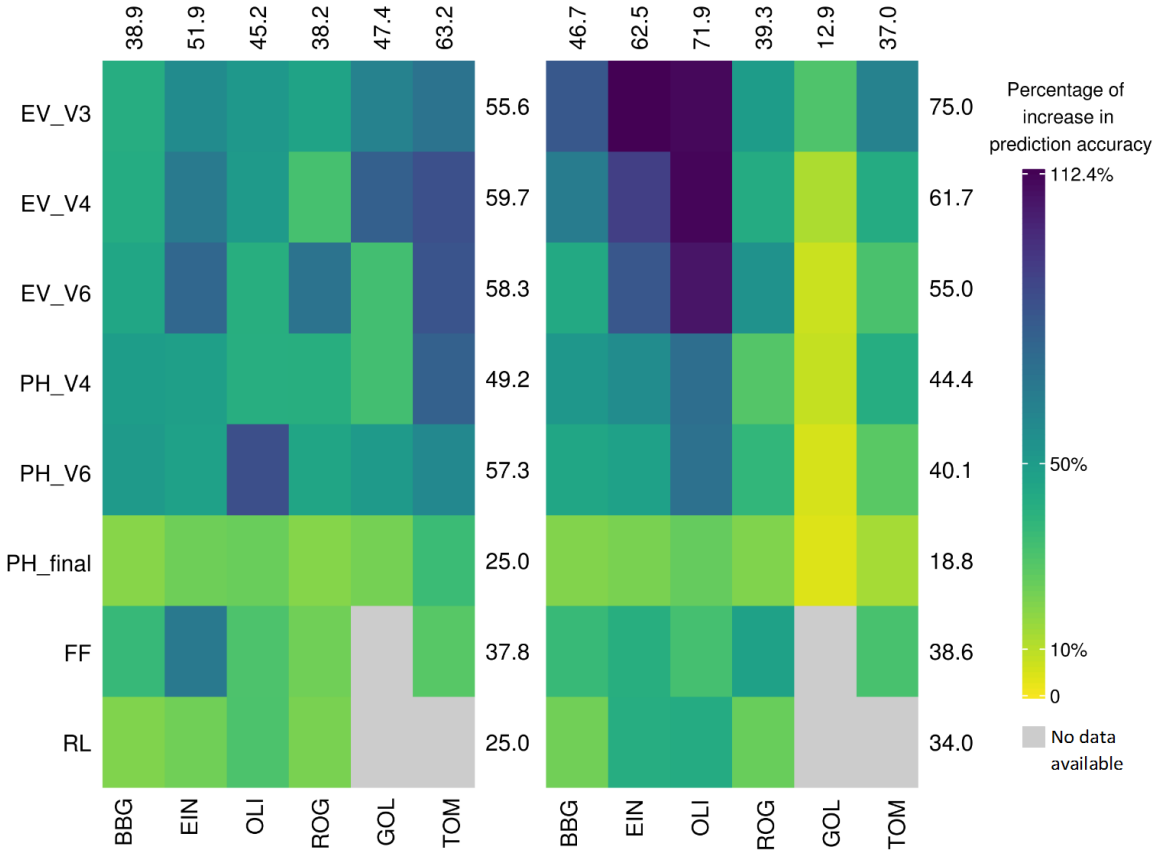
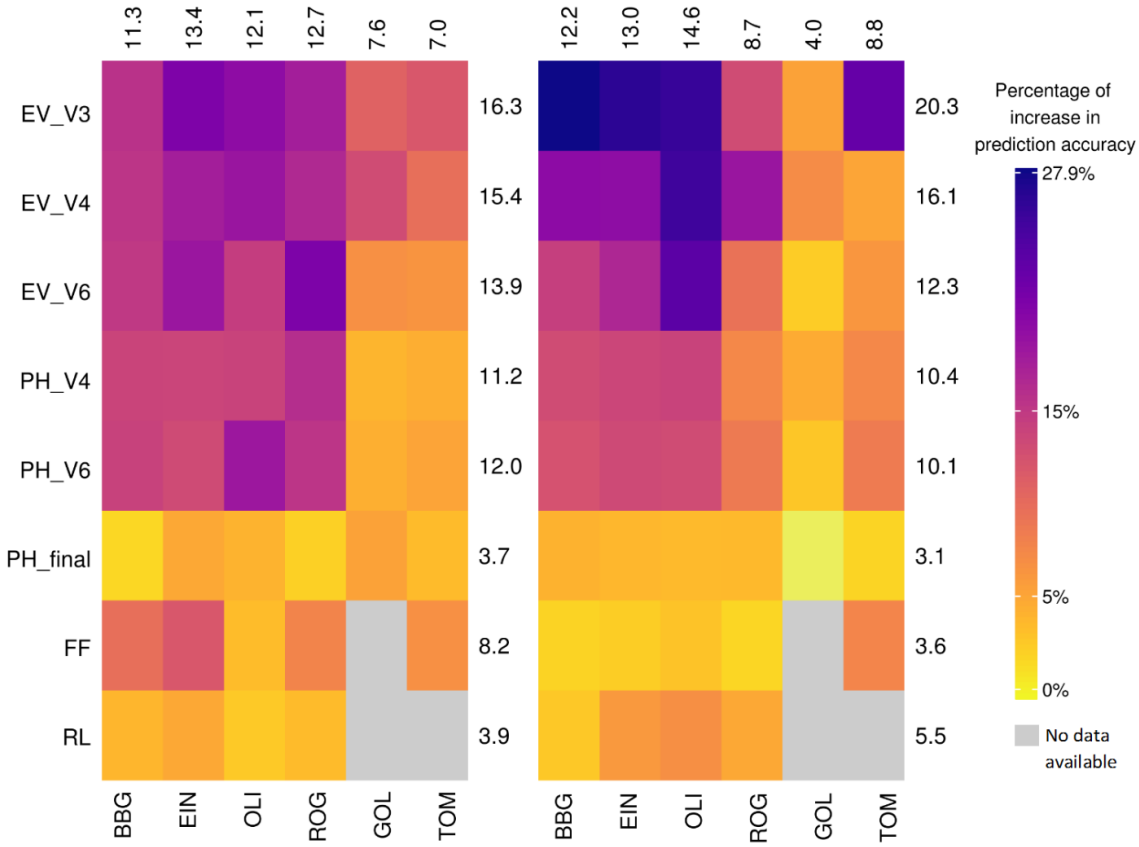


Fig. 10 Percentage of increase in prediction accuracy from univariate GBLUP within environments to the maximum prediction accuracy of univariate sERRBLUP across environments when the SNP interaction selections are based on estimated effects variances in Kemater (left side plot) and in Petkuser (right side plot). The average percentage of increase in prediction accuracy for each trait and environments are displayed in rows and columns, respectively.

Fig. 11 also shows the relative increase in prediction accuracy from the best bivariate GBLUP to the best bivariate sERRBLUP for all traits and all locations in a form of heat map. 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 0.257 percent. The minimum increase in prediction accuracy in KE was also observed for PH_final (1.7 percent). In general, Fig. 11 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

538 10.5 in PE across all combinations. The maximum average increase in prediction accuracy among
539 all traits was observed for EV_V3 in both landraces (16.3 percent in KE and 20.3 percent in PE).
540 Regarding environments, the maximum average increase across all environments was found in
541 EIN for KE (13.4 percent) and in OLI for PE (14.6 percent). In addition, the absolute increase in
542 prediction accuracy of bivariate models is shown as a heat map in supplementary (Fig. S9b)
543 indicating an average absolute increase of 0.1 across all traits, environments combinations, and
544 landraces.



545

546 **Fig. 11** Percentage of increase in prediction accuracy from the maximum bivariate GBLUP to the maximum
547 prediction accuracy of bivariate sERRBLUP when the SNP interaction selections are based on estimated
548 effects variances in Kemater (left side plot) and in Petkuser (right side plot). The average percentage of
549 increase in prediction accuracy for each trait and environments are displayed in rows and columns,
550 respectively.

551 In general, we assume that predictive ability for phenotypes should be higher with higher
552 heritability. This is confirmed by the results in both Fig. 12 and 13, depicting the correlation
553 between the heritability of all traits and the predictive ability obtained from univariate and
554 bivariate models, respectively. The traits' heritabilities were calculated on an entry-mean basis

within each KE and PE landraces (Hallauer *et al.*, 2010) over all six environments in the year 2017. Fig. 12 shows the correlation between the heritability of all traits and the univariate GBLUP within environments and maximum predictive ability of univariate sERRBLUP across environments. It is shown that this correlation increased from 0.296 with the univariate GBLUP model to 0.543 with the univariate sERRBLUP model. Likewise, Fig. 13 shows the correlation between the heritability of all traits and the maximum predictive ability of bivariate GBLUP and maximum predictive ability of bivariate sERRBLUP. It is shown that this correlation in bivariate GBLUP is 0.537 which is lower than respective correlation in univariate sERRBLUP and it is increased to 0.647 in bivariate sERRBLUP.

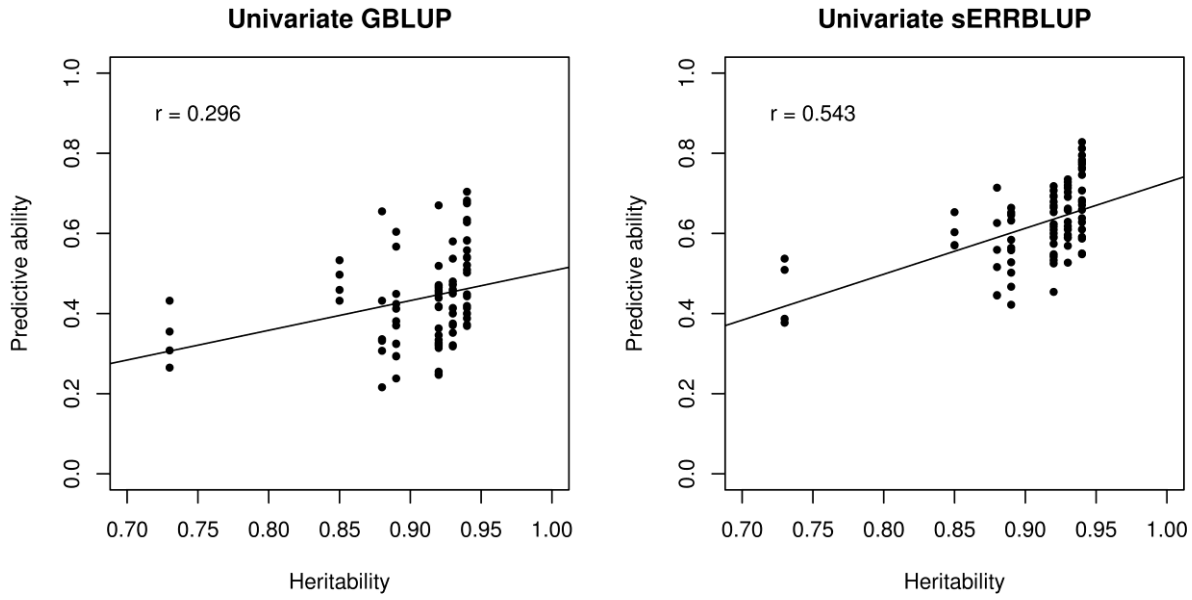


Fig. 12 The correlation between all eight traits' heritabilities and predictive abilities of univariate GBLUP within all six environments (left side plot) and maximum predictive abilities of univariate sERRBLUP across environments (right side plot) for all traits in both landraces. The black lines indicate the overall linear regression lines.

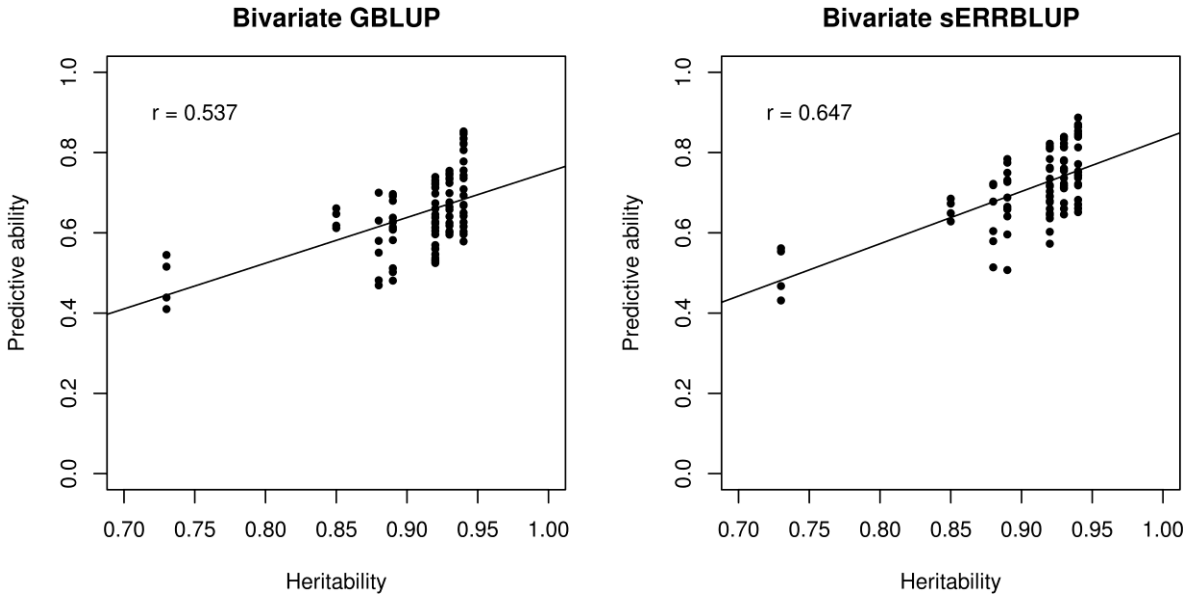


Fig. 13 The correlation between all eight traits' heritabilities and the maximum predictive abilities of bivariate GBLUP (left side plot) and maximum predictive abilities of bivariate sERRBLUP (right side plot) for all six environments in both landraces. The black lines indicate the overall linear regression lines.

In addition to assessing the predictive ability of univariate sERRBLUP based on a single environment, Fig. 14 and 15 display 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, respectively. In each Figure, the left hand side of the panel represent predictive abilities of univariate sERRBLUP across multiple environments jointly when pairwise SNP interaction selections were based on estimated absolute effect sizes, and the right hand side of the panel represent predictive abilities of univariate sERRBLUP across multiple environments jointly when pairwise SNP interaction selections were based on estimated effects variances. 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. Fig. 14 also reveals the robustness of the selection strategy based on the effects variance compared to selection strategy based on the absolute effects sizes in KE, while Fig. 15 does not show a significant difference for the interaction selection strategy for PE. Fig. 14 and 15 demonstrate 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 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. The comparison between the

maximum predictive ability of univariate sERRBLUP across a single environment and multiple environments jointly are shown in the supplementary (Fig. S10).

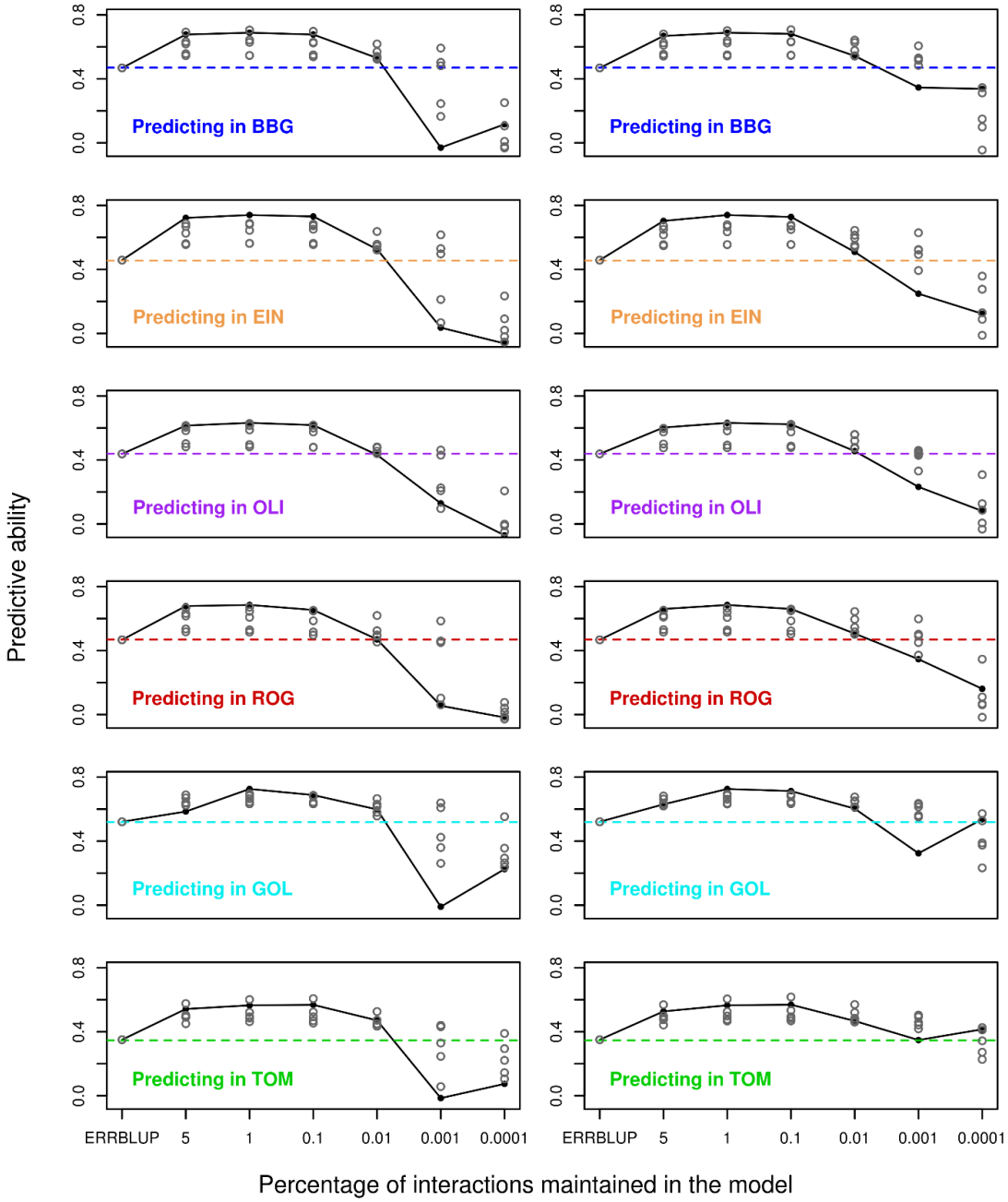
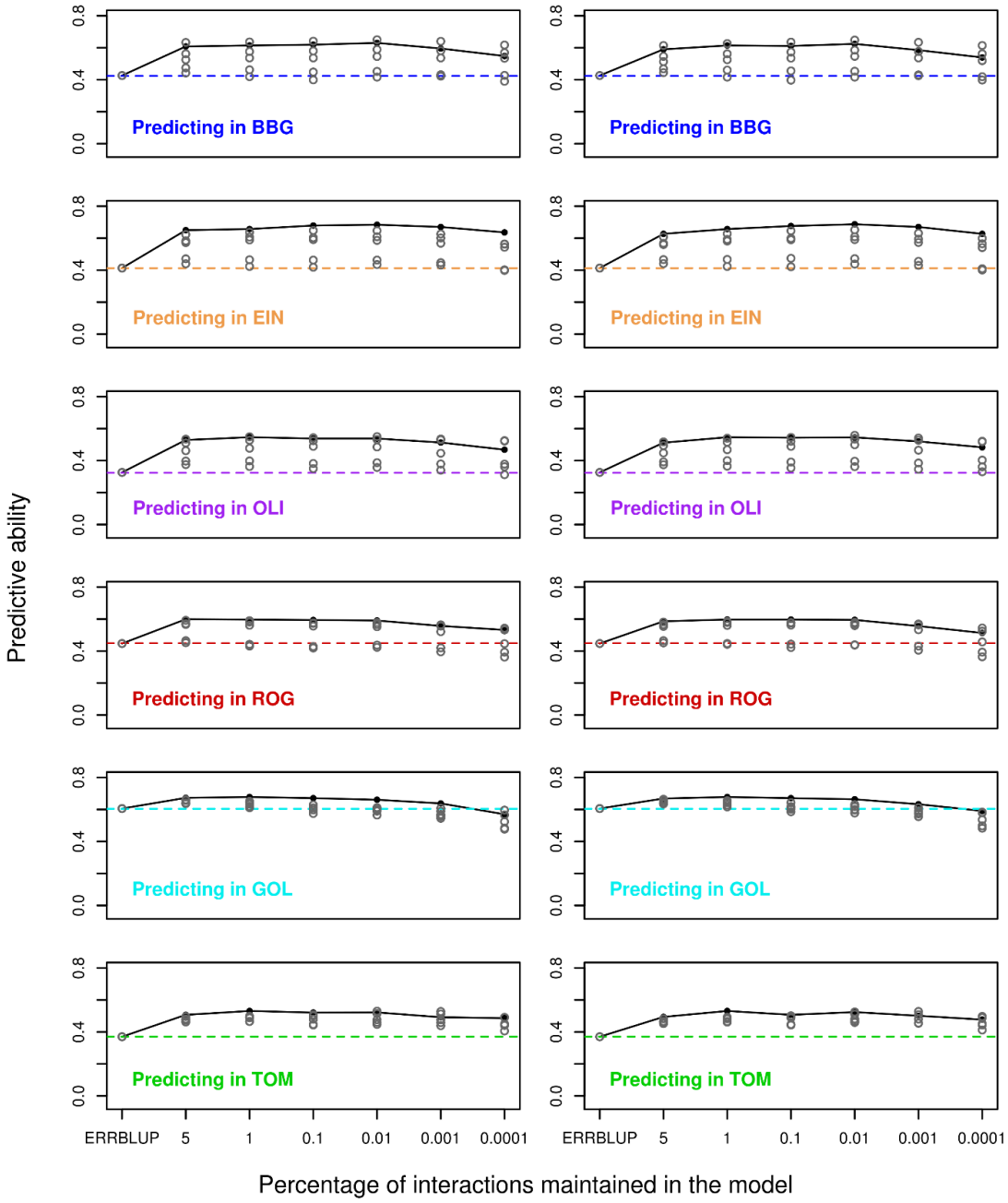


Fig. 14 Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate ERRBLUP within environment (gray open circle), univariate sERRBLUP using a single environment for selecting the SNP interactions (gray open circles) and univariate sERRBLUP using all 5 environments jointly (filled black circles and solid line) for the SNP interaction selection based on estimated effects sizes (left side) and estimated effects variances (right side) for trait PH-V4 in Kemater.



599

600 **Fig. 15** Predictive ability for univariate GBLUP within environment (dashed horizontal line), univariate
601 ERRBLUP within environment (gray open circle), univariate sERRBLUP using a single environment for
602 selecting the SNP interactions (gray open circles) and univariate sERRBLUP using all 5 environments jointly
603 (filled black circles and solid line) for the SNP interaction selection based on estimated effects sizes (left
604 side) and estimated effects variances (right side) for trait PH-V4 in Petkuser.

605 Fig. 16 illustrates the comparison between the predictive ability of bivariate GBLUP, ERRBLUP
606 and sERRBLUP across multiple environments jointly and the maximum predictive ability of

bivariate GBLUP and ERRBLUP and all the predicative 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. Moreover, in most cases bivariate GBLUP, ERRBLUP and sERRBLUP across multiple environments jointly are as good as or better than bivariate GBLUP, ERRBLUP and sERRBLUP (when the selection of effects is not too strict especially in KE) joint with a single environment, respectively. With less than 0.1 percent of pairwise SNP interactions used in the model, the predictive ability of the bivariate sERRBLUP across multiple environments jointly decreases considerably in KE.

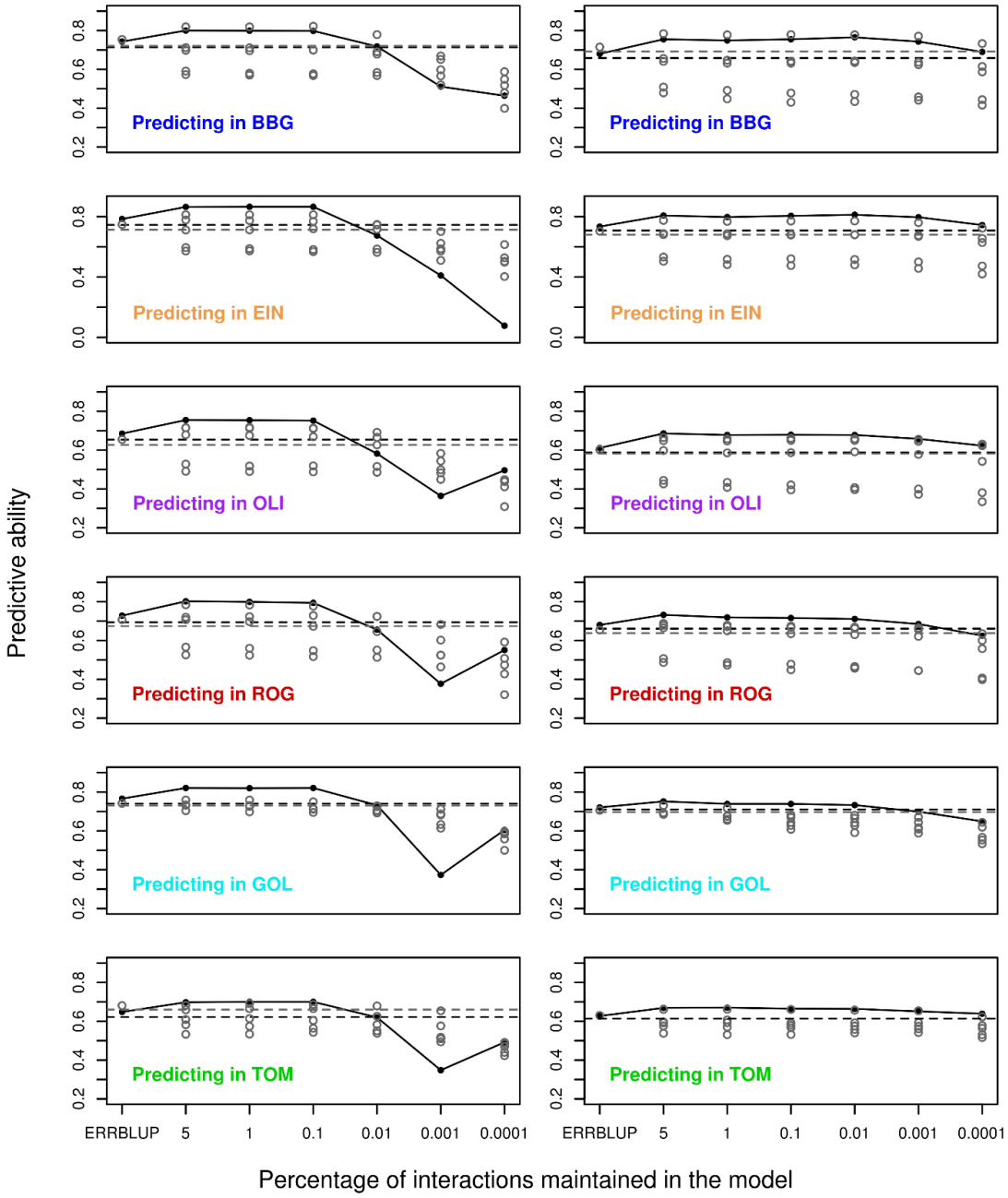


Fig. 16 Predictive ability for bivariate GBLUP (black dashed horizontal line), bivariate ERRBLUP and bivariate sERRBLUP (filled black circles) for the SNP interaction selection based on estimated effects variances using all 5 environments jointly for trait PH-V4 in Kemater (left side) and Petkuser (right 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.

Discussion

The accuracy of the genomic prediction when incorporating epistasis interactions in the model compared to prediction models with only main effects have 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).

The 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 (Jiang and Reif, 2015; Martini *et al.*, 2016). In this work, epistasis relationship matrices were constructed by using the package miraculix (Schlather, 2020) which carries out matrix multiplications about 15 times faster than regular matrix multiplications on genotype data in R. In the analyzed datasets containing up to 30'212 SNPs, 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 EpiGP R-package has been released for genomic prediction of phenotypes based on ERRBLUP and sERRBLUP (Vojgani *et al.*, 2019a).

The predictive ability of EG-BLUP depends on marker coding and the symmetric coding seems to perform best. In contrast, the ERRBLUP (called "categorical epistasis (CE)" by Martini *et al.* (2017)) does not rely on marker coding and performs as good as EG-BLUP with symmetrically coded markers (Martini *et al.*, 2017). Although the marker matrix generated with ERRBLUP and sERRBLUP is 4 times as large than the EG-BLUP interaction marker matrix, the computing time is comparable to EG-BLUP. Still, our proposed epistasis models 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. Potential strategies to overcome these limitations are 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 might be the use of haplotype blocks (Pook *et al.*, 2019).

In this study, we showed that the predictive ability obtained from univariate GBLUP and from a univariate full epistasis model with all pairwise SNP interactions included (ERRBLUP) was almost identical. In contrast, it was shown that the univariate sERRBLUP across environments increases predictive ability when the most relevant SNP interactions are taken into account. Based on the

preliminary analysis in our datasets, univariate sERRBLUP across environments did not provide a considerable increase in predictive ability until 95 percent of pairwise SNP interactions were removed which means around 30 million interactions still remained in the model. It was also demonstrated in a wheat dataset (Crossa *et al.*, 2010; Pérez and de los Campos, 2014) that the highest gain in predictive ability was obtained when over 80 percent were removed (Martini *et al.*, 2016). Ober *et al.* (2015) concluded that the improvement of predictive ability when selecting only the top interaction - as we did in sERRBLUP - is likely a result of enriching for true causal variants among the list of variants used to construct the genetic covariance matrix. In our study, the maximum predictive ability with univariate sERRBLUP was obtained by incorporating the top 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 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 may be explained by the absolute number of interaction effects in the model being more important than the percentage as well as potential differences in linkage leading 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 (see Fig. 5). This phenomenon has been more prevalent in KE than in PE (Fig. 5 vs. Fig. 6), and is valid in both scenarios, using information either from a single environment or from the average of all other environments (Fig. 14 and Fig. 15). Conceptually, the SNP interaction effect variance appears to be the better choice, since extreme estimates of SNP interaction effects can result in cases where some interaction classes are just represented by very few lines. This is balanced by taking into account the frequency in the SNP interaction variance. Due to the conceptual advantage and the more robust performance we recommend the latter to be used as selection criterion in sERRBLUP applications.

The bivariate models exhibited a considerably higher predictive ability than univariate models, in consequence the maximum bivariate GBLUP performed slightly better than the maximum univariate sERRBLUP in most cases (Fig. S11), and the predictive ability increased further in the bivariate sERRBLUP when the top 5 or 1 percent of pairwise SNP interactions are selected based on the effects variances in most cases. Over all studied traits, the increase in prediction accuracy from GBLUP to sERRBLUP displays a similar pattern in both univariate and bivariate models. It should be noted, though, that the increase in prediction accuracy is limited by the prediction accuracy obtained with GBLUP in relation to the trait heritability, which is illustrated with the following example: for the trait final plant height (PH_final) the GBLUP prediction accuracy is

0.73, while the maximum possible value is the square root of the trait heritability, which is 0.97 in this case. Thus, the maximum possible absolute increase would be 0.24 which is 32.9 per cent of the prediction accuracy with GBLUP. So, the higher the GBLUP prediction accuracy is, the less room for improvement remains. However, the increase in predictive ability from bivariate GBLUP to bivariate sERRBLUP is only caused by the modelling of epistasis.

Cross validation in multi-trait genomic prediction models utilizing the secondary trait's full dataset for prediction of the test set in the focal trait was shown to bias prediction accuracy (Runcie and Cheng, 2019), the main explanation being the existence of non-genetic covariance between the two traits observed in the same individuals, which is not properly accounted for in the prediction model. This scenario, however, differs from the cross-validation scenario studied in our case, since here the two 'traits' are actually the same biological trait observed in the same genotype, but in two completely separated environments. This setting should not give rise to non-genetic correlations between the two 'traits', e.g. caused by identical weather conditions affecting both traits simultaneously. Runcie and Cheng (2019) observed no systematic bias when the individuals do not share the same source of non-genetic variation, and thus, we assume that this source of bias is not relevant in our study.

In our scenario, the prediction accuracy can be reduced, though, if the second environment has a smaller number of phenotyped lines than the target environment indicating less overlap of genotypes with the target environment. In tendency this was confirmed in our study when GOL or TOM were modeled as the second environment for predicting the unobserved lines in other environments (e.g. BBG), since these two environments have around half the number of lines compared to other environments.

It has been shown that prediction accuracy is positively correlated to trait heritability. For instance, it was reported that prediction accuracies for resistance to yellow and stem rust in wheat was related to their heritability, respectively (Zhao *et al.*, 2013; Momen *et al.*, 2018). Similarly it was reported in maize that grain yield with low heritability has a prediction accuracy of 0.58, while grain moisture with high heritability has a prediction accuracy of 0.90 (Technow *et al.*, 2014; Momen *et al.*, 2018). Across the multitude of traits, environments, lines, and prediction methods, we could also find a substantial positive correlation between the heritabilities and the predictive abilities (see Fig. 12 – 13), which was more pronounced for sERRBLUP compared to GBLUP. The smallest correlation between heritability and the predictive ability was obtained with the univariate GBLUP model, while the maximum correlation of these two quantities was obtained with the bivariate sERRBLUP model.

Predictive ability in each environment can be increased by borrowing information especially from environments which have high phenotypic and genetic correlations to the target environment, which should be related to geographic and climatic similarities of the environments. In our study, the maximum predictive ability in each environment was obtained by sERRBLUP selecting interaction terms in environments with similar geographic features. This was observed for the majority of environments across series of phenotypic traits. The PCA based on environmental features (Fig. S12) also confirms that the environments which were closely located shared more common geographical and climatic features resulting in higher accuracy for prediction across the respective environments. To illustrate this, BBG and EIN which are closely located in Germany are close to each other in the PCA as well. This was also observed for ROG and OLI in Germany. GOL and TOM which are also closely located in Spain are close to each other in the PCA, too. For instance, our results indicate that in most cases the highest gain in phenotype prediction of e.g. BBG was obtained for sERRBLUP when borrowing information from EIN and vice versa. This was also observed in the majority of cases in bivariate models such that if the additional environment (i.e. EIN) which was added to the model shared more common geographical and climatic features with the target environment (i.e. BBG), the obtained predictive ability was higher compared to adding other environments as an additional environment (i.e. GOL) joint with target environment (i.e. BBG).

Rather than using a single environment to select SNP interactions for univariate sERRBLUP prediction, we also used an approach in which SNP interaction selection was based on all five environments jointly to evaluate the model in the sixth environment. This was also done for bivariate GBLUP, ERRBLUP and sERRBLUP such that the additional environment was the combination of all the other five environments instead of a single environment. 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. Using an average across all other environments appears to be a robust alternative which in most cases will yield a result that is as good or even better than the best single environment, at the same time avoiding the risk of compromising prediction quality by choosing the 'wrong' training 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. Additionally, selecting a suitable subset of interactions based on all the other environments jointly performs as well as selecting a suitable subset of interactions

763 based on the single environment with highest phenotypic correlation with the target
764 environment.

765 The presented approach can be useful in cases, where different lines are grown in multiple
766 environments, which can be either simultaneously, i.e. in the same season, or in subsequent
767 seasons. We have shown, that 'borrowing' information, however just in the selection of the most
768 relevant interactions, can substantially improve the phenotype prediction accuracy in another
769 environment. This can be useful in sparse testing designs, e.g. where not all lines are grown in all
770 environments. The suggested approach can be used to 'impute' missing phenotypes with a much
771 increased accuracy compared to conventional approaches.

772

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 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 manuscript.

Corresponding author

799 Correspondence to Elaheh Vojgani

800 Email: vojgani@gwdg.de

801 ORCID: 0000-0003-4375-3531

802 **Acknowledgements**

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

809

References

Abendroth, L. J., Elmore, R. W., Boyer, M. J. and Marlay, S. K. (2011) Corn Growth and Development. *PMR 1009. Iowa State University of Science and Technology, Cooperative Extension Service, Ames, Iowa.*

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S. and Lawlor, T. J. (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*. Elsevier, 93(2): 743–752. https://doi.org/10.3168/jds.2009-2730

Akdemir, D. and Godfrey, O. U. (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*. John Wiley & Sons, Ltd, 47(3): 1082–1090. https://doi.org/10.2135/cropsci2006.11.0690

Browning, S. R. and Browning, B. L. (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, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J. and Thompson, R. (2018) ASReml-R Reference Manual Version 4. VSN International Ltd., Hemel Hempstead

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

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. and Lee, J. J. (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-

843 Cicalese, J., Vorsa, N., Iorizzo, M. and Zalapa, J. (2018) Multivariate GBLUP Improves Accuracy of
844 Genomic Selection for Yield and Fruit Weight in Biparental Populations of Vaccinium
845 macrocarpon Ait. *Frontiers in Plant Science*, 9(1310).
846 https://doi.org/https://doi.org/10.3389/fpls.2018.01310

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

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

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

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

859 Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., Campos, G. de los and Hickey, J. M. (2013)
860 Genomic Prediction in Animals and Plants: Simulation of Data, Validation, Reporting, and
861 Benchmarking. *Genetics*, 193: 347–365. https://doi.org/10.1534/genetics.112.147983

862 Dekkers, J. C. M. (2007) Prediction of response to marker-assisted and genomic selection using
863 selection index theory. *Journal of Animal Breeding and Genetics*. John Wiley & Sons, Ltd, 124(6):
864 331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x

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

868 Erbe, M., Pimentel, E., Sharifi, A. R. and Simianer, H. (2010) Assessment of cross-validation
869 strategies for genomic prediction in cattle. *Proceedings of the World Congress on Genetics*
870 *Applied to Livestock Production*, Methods an: 553

871 Falconer, D. S. and Mackay, T. F. C. (1996) *Introduction to Quantitative Genetics*. Longman.
872 Essex Engl.

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

876 Guo, G., Zhao, F., Wang, Y., Zhang, Y., Du, L. and Su, G. (2014) Comparison of single-trait and

877 multiple-trait genomic prediction models. *BMC Genetics*, 15(1): 30.
878 https://doi.org/10.1186/1471-2156-15-30

879 Haley, C. S. and Visscher, P. M. (1998) Strategies to Utilize Marker-Quantitative Trait Loci
880 Associations. *Journal of Dairy Science*, 81: 85–97.
881 https://doi.org/https://doi.org/10.3168/jds.S0022-0302(98)70157-2

882 Hallauer, A. R., Carena, M. J. and Miranda Filho, J. B. (2010) *Quantitative genetics in maize*
883 *breeding*. *Springer*. Berlin

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

886 Hayes, B. J. and Goddard, M. E. (2008) Technical note: Prediction of breeding values using
887 marker-derived relationship matrices. *Journal of Animal Science*, 86(9): 2089–2092.
888 https://doi.org/10.2527/jas.2007-0733

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

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

894 Hill, W. G., Goddard, M. E. and Visscher, P. M. (2008) Data and theory point to mainly additive
895 genetic variance for complex traits. *PLoS Genetics*, 4(2).
896 https://doi.org/10.1371/journal.pgen.1000008

897 Hölker, A. C., Mayer, M., Presterl, T., Bolduan, T., Bauer, E., Ordas, B., Brauner, P. C., Ouzunova,
898 M., Melchinger, A. E. and Schön, C.-C. (2019) European maize landraces made accessible for
899 plant breeding and genome-based studies. *Theoretical and Applied Genetics*, 132(12): 3333–
900 3345. https://doi.org/10.1007/s00122-019-03428-8

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

904 Huang, W., Richards, S., Carbone, M. A., Zhu, D., Anholt, R. R. H., Ayroles, J. F., Duncan, L.,
905 Jordan, K. W., Lawrence, F., Magwire, M. M., Warner, C. B., Blankenburg, K., Han, Y., Javaid, M.,
906 Jayaseelan, J., Jhangiani, S. N., Muzny, D., Ongeri, F., Perales, L., Wu, Y.-Q., Zhang, Y., Zou, X.,
907 Stone, E. A., Gibbs, R. A. and Mackay, T. F. C. (2012) Epistasis dominates the genetic
908 architecture of Drosophila quantitative traits. *Proceedings of the National Academy of Sciences*
909 *of the United States of America*. National Academy of Sciences, 109(39): 15553–15559.
910 https://doi.org/10.1073/pnas.1213423109

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

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

917 Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., Campbell, M. S., Stein, J. C., Wei, X.,
918 Chin, C.-S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K. L., Wolfgruber, T.
919 K., May, M. R., Springer, N. M., Antoniou, E., McCombie, W. R., Presting, G. G., McMullen, M.,
920 Ross-Ibarra, J., Dawe, R. K., Hastie, A., Rank, D. R. and Ware, D. (2017) Improved maize
921 reference genome with single-molecule technologies. *Nature*. The Author(s), 546: 524–527.
922 https://doi.org/https://doi.org/10.1038/nature22971

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

925 Karaman, E., Cheng, H., Firat, M. Z., Garrick, D. J. and Fernando, R. L. (2016) An Upper Bound for
926 Accuracy of Prediction Using GBLUP. *PLOS ONE*, 11(8): e0161054.
927 https://doi.org/https://doi.org/10.1371/journal.pone.0161054

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

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

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

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

942 Mackay, T. F. C. (2014) Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-
943 Gene Interactions. *Nat Rev Genet.*, 15(1): 22–33. https://doi.org/10.1038/nrg3627

944 Martini, J. W. R., Wimmer, V., Erbe, M. and Simianer, H. (2016) Epistasis and covariance: how
945 gene interaction translates into genomic relationship. *Theoretical and Applied Genetics*, 129(5):

963–976. https://doi.org/10.1007/s00122-016-2675-5

Martini, J. W. R., Gao, N., Cardoso, D. F., Wimmer, V., Erbe, M., Cantet, R. J. C. and Henner, S. (2017) Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). *BMC Bioinformatics*, 18(3). https://doi.org/https://doi.org/10.1186/s12859-016-1439-1

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

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

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4): 1819–1829

Momen, M., Mehrgardi, A. A., Sheikhi, A., Kranis, A., Tusell, L., Morota, G., Rosa, G. J. M. and Gianola, D. (2018) Predictive ability of genome-assisted statistical models under various forms of gene action. *Scientific Reports*, 8(12309)

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

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

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

Pérez, P. and de los Campos, G. (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics*. 2014/07/09. Genetics Society of America, 198(2): 483–495. https://doi.org/10.1534/genetics.114.164442

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

Pook, T., Mayer, M., Geibel, J., Weigend, S., Cavero, D., Schoen, C. C. and Simianer, H. (2020) Improving Imputation Quality in BEAGLE for Crop and Livestock Data. *G3:*

Genes|*Genomes*|*Genetics*, 10(1): 177 LP – 188. https://doi.org/10.1534/g3.119.400798

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., Bakker, P. I. W. de, Daly, M. J. and Sham, P. C. (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/https://doi.org/10.1101/038935

Runcie, D. and Cheng, H. (2019) Pitfalls and Remedies for Cross Validation with Multi-trait Genomic Prediction Methods. *G3: Genes*|*Genomes*|*Genetics*, 9(11): 3727 LP – 3741. https://doi.org/10.1534/g3.119.400598

Schaeffer, L. R. (2006) Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding & Genetics*, 123(4): 218–223

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

Schrauf, M. F., Martini, J. W. R., 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

Schulthess, A. W., Zhao, Y., Longin, C. F. H. and Reif, J. C. (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

Technow, F., Schrag, T. A., Schipprack, W., Bauer, E., Simianer, H. and Melchinger, A. E. (2014) Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. *GENETICS*, 197(4): 1343–1355

Unterseer, S., Author, E. B., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T. M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K. F. 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, H. F., Melchinger, A. E. and Schön, C. C. (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, J. G., Jordan, D. R., Mace, E. S., Hunt, C. H., Malosetti, M. and van Eeuwijk, F. A. (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. (2019a) EpiGP: Epistatic relationship matrix based genomic prediction of phenotypes. Available at: https://github.com/evojgani/EpiGP

Vojgani, E., Pook, T. and Simianer, H. (2019b) Phenotype Prediction under Epistasis. in *Methods in Molecular Biology*. Humana Press

Wang, D., El-Basyoni, I. S., Baenziger, P. S., Crossa, J., Eskridge, K. M. 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, P. J., Buckler, E. S. 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 LP – 587. https://doi.org/10.1534/genetics.113.150078

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., Raman, B., Cairns, J. E., Tarekegne, A., Semagn, K., Beyene, Y., Grudloyma, P., Technow, F., Riedelsheimer, C. and Melchinger, A. E. (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, J. E., O'Sullivan, N. P., Preisinger, R., Habier, D., Fernando, R., Garrick, D. J., Lamont, S. J. and Dekkers, J. C. M. (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-

1049 43-5

1050 Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, 16(2): 97–159

1051 Zhao, Y., Zeng, J., Fernando, R. and Reif, J. C. (2013) Genomic Prediction of Hybrid Wheat
1052 Performance. *Crop Breeding & Genetics*, 53(3): 802–810

1053