Theoretical and Applied Genetics (2024) 137:175
https://doi.org/10.1007/s00122-024-04679-w

ORIGINAL ARTICLE

High-dimensional multi-omics measured in controlled conditions are useful for maize platform and field trait predictions

Baber Ali[1] · Bertrand Huguenin-Bizot[2] · Maxime Laurent[3] · François Chaumont[3] · Laurie C. Maistriaux[3] · Stéphane Nicolas[1] · Hervé Duborjal[4] · Claude Welcker[5] · François Tardieu[5] · Tristan Mary-Huard[1] · Laurence Moreau[1] · Alain Charcosset[1] · Daniel Runcie[6] · Renaud Rincent[1]

Received: 20 March 2024 / Accepted: 15 June 2024

Abstract

Key message **Transcriptomics and proteomics information collected on a platform can predict additive and non-additive effects for platform traits and additive effects for field traits.**

Abstract The effects of climate change in the form of drought, heat stress, and irregular seasonal changes threaten global crop production. The ability of multi-omics data, such as transcripts and proteins, to reflect a plant's response to such climatic factors can be capitalized in prediction models to maximize crop improvement. Implementing multi-omics characterization in field evaluations is challenging due to high costs. It is, however, possible to do it on reference genotypes in controlled conditions. Using omics measured on a platform, we tested different multi-omics-based prediction approaches, using a high dimensional linear mixed model (*MegaLMM*) to predict genotypes for platform traits and agronomic field traits in a panel of 244 maize hybrids. We considered two prediction scenarios: in the first one, new hybrids are predicted (CV-NH), and in the second one, partially observed hybrids are predicted (CV-POH). For both scenarios, all hybrids were characterized for omics on the platform. We observed that omics can predict both additive and non-additive genetic effects for the platform traits, resulting in much higher predictive abilities than *GBLUP*. It highlights their efficiency in capturing regulatory processes in relation to growth conditions. For the field traits, we observed that the additive components of omics only slightly improved predictive abilities for predicting new hybrids (CV-NH, model *MegaGAO*) and for predicting partially observed hybrids (CV-POH, model *GAOxW-BLUP*) in comparison to *GBLUP*. We conclude that measuring the omics in the fields would be of considerable interest in predicting productivity if the costs of omics drop significantly.

Introduction

Climate change affects crop production in several ways, such as significant changes in rainfall and weather patterns, increasing temperature, and drought stress (Lee et al. 2023). Such factors threaten global food security, requiring immediate action to better adapt crops to these challenging conditions. The ever-decreasing cost of high-throughput phenotyping and genotyping technologies offers unprecedented opportunities to accelerate genetic gain by enhancing the efficiency of crop breeding programs. Especially in plant breeding, the interest in multi-environmental genomic predictions is constantly increasing thanks to pedo-climatic data availability (Westhues et al. 2022), which can help to model genotype by environment (GxE) interactions leading to greater predictive abilities (Crossa et al. 2022; Heslot et al. 2014; Jarquín et al. 2014; Rincent et al. 2019). Such environmental data, also known as environmental covariates (ECs),

Communicated by Daniela Bustos-Korts.

✉ Renaud Rincent
renaud.rincent@inrae.fr

1 INRAE, CNRS, AgroParisTech, GQE-Le Moulon, Université Paris-Saclay, 91190 Gif-Sur-Yvette, France

2 Laboratoire Reproduction Et Développement Des Plantes, CNRS, ENS de Lyon-46, Allée d'Italie, 69364 Lyon, France

3 Louvain Institute of Biomolecular Science and Technology, UCLouvain, Louvain-La-Neuve, Belgium

4 Limagrain, Limagrain Fields Seeds, Research Centre, 63720 Chappes, France

5 LEPSE, Univ Montpellier, INRAE, Montpellier, France

6 Department of Plant Sciences, University of California Davis, Davis, CA, USA

Published online: 03 July 2024

can be included in different ways in multi-environment prediction models. For instance, the reaction norm approach by Jarquín et al. (2014) models markers and environmental covariates using covariance structures, where their interactions are modeled through a Kronecker product between the genomic and the environmental covariance matrices. This model relies on a linear statistical relationship between genomic markers and the phenotype.

The advances in multi-omics technologies now make it possible to develop high-dimensional transcriptomic, proteomic, and metabolomic profiles that can efficiently capture the influence of environmental conditions on plants. Most studies that used metabolic data from field plants have proposed several approaches to integrate these in prediction models and have exhibited intermediate to high predictive abilities (de Abreu e Lima et al. 2017; Riedelsheimer et al. 2012; Ward et al. 2015), generally competitive with those obtained when using only genomic markers. Multi-omics data (Hu et al. 2021) are still expensive and cannot be made available for each trial of a breeding program. Breeders generally rely on phenotyping platforms to accurately simulate reproducible stressing conditions to study the multi-omics response of plants. Such data are quite helpful in identifying genes and QTLs related to stress responses (Blein-Nicolas et al. 2020; Jiang et al. 2024; Kilasi et al. 2018; Wen et al. 2019). However, their integration in the genomic prediction models of field traits is still quite challenging due to differences in environmental conditions in the field and platform.

Hu et al. (2020) illustrated that metabolomics and transcriptomics collected from the platform and the field are highly correlated between the same tissue types. This evidence suggests that platform omics data can be helpful for field predictions. In Fu et al. (2012), platform transcriptomic data resulted in promising predictive abilities across genetically distant groups. Other studies have combined different platform omics, such as mRNA and sRNA transcriptomics and metabolomics data measured at an early stage, to improve the predictive abilities of field traits in maize (Schrag et al. 2018). Westhues et al. (2017) collected transcriptomic data from plants in highly controlled conditions and reported improving their prediction models with transcriptomic data and genomic information while predicting maize field traits such as dry matter yield and dry matter content.

As omics abundances are influenced by a plant's response to the growth conditions, we can expect them to be particularly efficient in predicting traits measured on the same plants or plants grown in similar conditions. Omics are intermediary phenotypes, so we hypothesize that they may efficiently capture non-additive genetic effects for predicting growth conditions like those in which the omics were measured. However, in contrast to DNA markers such as SNP, the heritability of omics measurements is lower than one. Moreover, Christensen et al. (2021) stressed that omics can capture genetic and non-genetic effects, which can be an issue when the same individuals are characterized for omics and are also predicted. The model proposed by Christensen et al. (2021), GOBLUP, can trim environmental noise and residuals from the omics expressions. Using this approach to remove the non-genetic part of the omics may be a way to increase the predictive ability of field traits with platform omics at the cost of removing the non-additive genetic part (i.e., the genetic part that is not captured by the additive kinship) of the omics.

Integrating multi-faceted data can also be done by fitting a multivariate linear mixed model (MvLMM). MvLMMs utilize correlations between traits to improve predictions. However, these correlations are not robust if there are more predictors than the number of genotypes ($p >> n$) (Bernardo 2010). Further, the phenotypic correlations between traits are generally a poor representation of genetic correlations. Fitting such MvLMMs is quite computationally inefficient because of the number of iteratively estimated covariance components and their inverses, whose burden increases manifold with the increasing number of traits (Zhou and Stephens 2014). Runcie et al. (2021) proposed the Mega-Scale Linear Mixed Model (*MegaLMM*) as one possible solution that decreases the computational burden of MvLMM by re-parameterizing them as Bayesian sparse factor models. Like Bayesian Sparse Factor Analysis of Genetic Covariance Matrices (*BSFG*) (Runcie and Mukherjee 2013), MegaLMM also considers that a few latent factors can explain covariance among many traits. Following such an approach, MegaLMM has exhibited remarkable predictive abilities and runtime improvements over well-established Bayesian methods like Phenix (Dahl et al. 2016) and *MTG2* (Lee and Van der Werf 2016) in wheat by using hyperspectral reflectance data and in *Arabidopsis thaliana* by using gene expression profiles (Runcie et al. 2021). *MegaLMM* is certainly a model of choice for predicting agronomic traits with multi-omics data.

In the present study, we worked on a panel of 244 maize Dent lines crossed to a Flint tester, characterized for ecophysiological traits, transcriptomics, and proteomics on a high-throughput phenotyping platform in two contrasted water treatments. In addition, grain yield was measured for this same panel of hybrids in 25 trials in different locations, years, and treatment combinations. Our primary objective was to assess the ability of multi-omics data obtained from platform experiments to predict both platform and field traits compared to genomic selection models. We worked both on platform and field traits to evaluate the potential of platform omics to predict non-additive effects for traits measured in similar conditions (platform) or in completely different conditions (fields).

Additionally, we aim to evaluate the effectiveness of platform omics in capturing GxE interactions in the fields. It involved a comparative analysis between models incorporating multi-omics data and conventional genomic prediction models relying solely on genomic information. Our particular focus lies in evaluating the potential advantages of employing the high-dimensional multivariate *MegaLMM* model in conjunction with omics data. Furthermore, we also assessed models combining both multi-omics and ECs.

Material and methods

Plant material and experiments

Maize panel and platform experiments

The platform experiments were conducted on the PhenoArch platform (Cabrera-Bosquet et al. 2016). The platform experiments comprised a hybrid panel of 254 maize dent lines crossed to a common flint parent (UH007) and grown under two contrasting water conditions, i.e., well-watered (WW) and water deficient (WD). The detailed experimental and growing conditions are available in Alvarez-Prado et al. (2018) and Blein-Nicolas et al. (2020). Experiments with two contrasting conditions were carried out in Spring 2012 and 2013 and Winter 2013. Plants grown in WW conditions faced a soil water potential of − 0.05 MPa, while the ones grown under WD conditions faced a range from − 0.3 to − 0.6 MPa. Eight ecophysiological traits related to growth and transpiration were measured during these experiments. The traits included early biomass (*Be*), late biomass (*Bl*), early leaf area (*LAe*), late leaf area (*LAl*), water use efficiency (*WUE*), water use (*WU*), stomatal conductance (*gs*), and transpiration rate (*Trate*). For each of these traits, we have up to three experiments within each treatment (Table S1).

Drought-tolerant yielding plants (DROPS) multi-environment trial

Phenology and production traits, including grain yield, were measured on 254 hybrids of the previously mentioned panel at 25 locations from Eastern to Western Europe and an external site in Graneros, Chile. For each experiment, genotypic means were computed for each hybrid using a mixed model with hybrids and replicates as fixed effects and spatial effects and spatially correlated errors as random effects (Millet et al. 2016). The model was fitted using ASReml-R (Butler et al. 2009; R Core Team 2013), and the best linear unbiased estimators (BLUEs) were used for the subsequent analyzes. Like the platform experiments, DROPS (https://cordis.europa.eu/project/id/244374/reporting/en) trials were also conducted in two a priori contrasting water regimes, i.e., WW and WD. ECs expected to affect maize growth at different phenological stages were measured in these trials, while some of them were computed using the APSIM crop growth model (Hammer et al. 2010; Millet et al. 2019). The ECs include longitude and latitude, soil water potential, maximum temperature, night temperature, radiation intercepted, and vapor pressure deficit during early flowering and grain-filling stages. A detailed overview of the environmental covariates is provided by Millet et al. (2019).

Multi-omics data

Genomic data

All lines were genotyped using the 50 K Infinium HD Illumina array (Ganal et al. 2011), a 600 K Axiom Affymetrix array (Unterseer et al. 2014), and more than 500 K markers obtained through genotype by sequencing method (Negro et al. 2019). As a result, we had access to a dense panel of 978,134 markers spread across the maize genome. In addition, we applied a filter on minor allele frequency (MAF) and removed markers with an MAF lower than 0.02, resulting in 842,165 SNP markers to work with.

Proteins

We had access to data on the abundances of 1982 proteins from 254 hybrids grown under two contrasting watering conditions on the PhenoArch platform in the Spring 2012 platform experiment (Blein-Nicolas et al. 2020; Prado et al. 2018). For each hybrid, there were two replicates per condition, resulting in four leaf samples (one per plant) obtained at the pre-flowering stage. Proteins were extracted from the frozen leaves through standard trichloroacetic and acetone solutions-based methods and digested into peptides before mass spectrometry analysis. The proteins were quantified based on either extracted ion currents (XIC) or spectral count (SC). We used normalized and spatially corrected protein abundances obtained from Blein-Nicolas et al. (2020). The broad sense heritabilities of the protein abundance adjusted means over replicates were computed independently for WD and WW treatments (Blein-Nicolas et al. 2020). For the analyzes in this study, we removed proteins with a heritability lower than 0.4 and kept the ones above this threshold. It resulted in the selection of 546 WD proteins and 475 WW proteins for trait predictions. Further details on protein extraction, quantification, spatial correction, and normalization can be found in Blein-Nicolas et al. (2020).

3′ RNA seq gene expression analysis

RNA was extracted from mature leaf tissues on the 1524 plants from 254 different genotypes (254 hybrids × 2 treatments × 2 to 3 repetitions) of the spring 2013 platform experiment. The quantity of RNA was determined by spectrophotometry (Nanodrop®, Ozyme), and the quality was evaluated by capillary electrophoresis (Agilent®) using the RNA Integrity Number (RIN). The concentration of RNA varied between 160 and 1000 ng/ml, and the RIN between 6.5 and 7. The next generation sequencing (NGS) libraries were synthesized using the QuantSeq-3' mRNA-Seq Library Prep Kit FWD (Lexogen, Vienna). The quality and the quantity of each library were assessed using microfluidic electrophoresis on a LabChip GX (Perkin Elmer, Waltham, MA). 1,056 samples were selected among these 1,524 samples based on RNA quality (> 200 ng). For each sequencing batch, 1,056 samples were randomized between and within 11 plates of 96 wells according to four factors: genotype, treatment, genetic group origin, sampling date and hour, and line of sampling in a glasshouse using the "designRandomize" R package (Brien, 2018). Accordingly, treatments, genetic structure, and sampling hour/date were well distributed between and within each plate. Then, 96 libraries of each sample in each plate were equimolarly pooled and sequenced on a Next Seq 550 (Illumina, San Diego, USA) using a High-output flow cell according to the manufacturer's recommendations. The resulting reads were 75 base pairs long. The sequences were first trimmed with Cutadapt v1.8 on single read mode, removing adapter sequences and quality trimming (q 10). The minimum length to keep a read was 50. All samples were individually barcoded during the library process. 80% of the samples had more than 4 million reads, and 92% had more than 3 million reads. Alignment of the sequences was done with the software STAR, using reference genome version 4 of B73 maize. The alignment was improved using an annotation file of the genome, allowing us to consider the splicing junctions. Only reads aligning to a single gene were considered. Moreover, only genes with at least one count per million reads in at least 10% of the samples were considered, which resulted in 20,475 and 20,642 transcripts for WD and WW conditions, respectively. In each condition, the transcripts were normalized with the approach of Trimmed Mean of M-values (Robinson et al. 2009) and transformed with the log2-cpm approach with package edgeR (McCarthy et al. 2012; Robinson et al. 2009).

For a given transcript, the last step was to fit a spatial model to compute adjusted means and estimate heritabilities. For this, we used the following model for each treatment (WW and WD):

$$y_{iklmp} = \mu + h_m + R_p + r_k + c_l + g_i + \varepsilon_{iklmp}, \quad (1)$$

where y_{iklmp} is the observed value for transcript expression of hybrid i sown in row k, column l, repetition p. h_m is the fixed effect of the hour of sampling m, R_p is the fixed effect of repetition p, r_k and c_l are the random effects of row k and column l, respectively, g_i is the fixed genotypic effect of hybrid i, and ε_{iklmp} is the random error effect: $\varepsilon_{iklmp} \sim N\left(0, \boldsymbol{I}\sigma_\varepsilon^2\right)$. Here, $\boldsymbol{I}$ is the identity matrix and σ_ε^2 is the random error variance.

The SPaTS package (Rodriguez-Alvarez et al. 2018) within package statgenSTA (van Rossum et al. 2023) was used for this purpose, allowing for P-spline adjustment of the spatial effects. The number of segments used for adjusting the splines was 20 for rows and columns. The model was first fitted with a fixed genotype effect to compute marginal genetic means and then with a random genotype effect (identity matrix as variance/covariance matrix) to estimate broad sense heritabilities, as in Oakey et al. (2006). After the normalization procedure and merging all the phenotypic and omics data, we obtained a list of 244 genotypes with transcript quantitative expressions. For all further analyzes, we filtered the transcripts based on their broad sense heritabilities to keep only those with heritabilities higher than 0.4. The filter resulted in 7051 WD transcripts and 7043 WW transcripts.

Models

Univariate prediction models

We used the following univariate models for platform traits and field trait predictions.

GBLUP: As a reference, we applied a univariate *GBLUP* model for single trait or environment analysis, defined as follows:

$$y_i = \mu + g_i + \varepsilon_i, \quad (2)$$

where y_i is the phenotypic value of a trait for hybrid i. μ is the overall mean. g_i is the random genetic value of hybrid i. Random genetic effects are considered to follow a normal distribution, i.e., $\boldsymbol{g} \sim N\left(0, \boldsymbol{K_G}\sigma_g^2\right)$. Here, $\boldsymbol{K_G}$ is the genomic kinship matrix computed as $\boldsymbol{K_G} = \boldsymbol{MM'}/2\sum p_s(1-p_s)$ with $\boldsymbol{M}$ as the matrix of centered marker genotypes and p_s as the frequency of reference allele for SNP s (VanRaden 2008). σ_g^2 is the genetic variance. ε_i is the random error effect. Residuals are considered to be independent from $\boldsymbol{g}$ and follow a normal distribution, i.e., $\boldsymbol{\varepsilon} \sim N\left(0, \boldsymbol{I}\sigma_\varepsilon^2\right)$.

OBLUP and G+OBLUP: Kernel-based methods generally offer an easy solution for multi-omics data integration. We used the following formula to calculate omics-based kernels:

$$\boldsymbol{M}_{f_t} = \frac{\boldsymbol{O}_{ft}\boldsymbol{O}'_{ft}}{n_{ft}}, \tag{3}$$

where $\boldsymbol{M}_{f_t}$ is the resulting similarity matrix for omics type f in treatment t. $\boldsymbol{O}_{ft}$ is the centered and scaled matrix of pretreated abundances of omics type f in treatment t, and n_{ft} is the number of features of the concerned type (proteins or transcripts). All kernel matrices, including the genomic kinship matrix, were scaled to have a sample variance of 1 (average of diagonal close to 1) to avoid biased parameter estimations due to different scales (Forni et al. 2011; Kang et al. 2010).

Firstly, we applied a univariate prediction model, hereafter referred to as *OBLUP*, with multiple random effects corresponding to different transcriptomic and proteomic effects as follows:

$$y_i = \mu + t_{r_i} + t_{w_i} + p_{r_i} + p_{w_i} + \varepsilon_i, \tag{4}$$

where $\boldsymbol{y_i}$, μ, and ε_i are the same as defined earlier for Eq. (2). t_{r_i} is the random WD standard-transcriptomic value for hybrid i; $\boldsymbol{t_r} \sim N\left(0, \boldsymbol{M_{t_r}}\sigma^2_{t_r}\right)$. Similarly, t_{w_i} is the random WW standard-transcriptomic value for hybrid i; $\boldsymbol{t_w} \sim N\left(0, \boldsymbol{M_{t_w}}\sigma^2_{t_w}\right)$. p_{r_i} is the random WD standard-proteomic value for hybrid i; $\boldsymbol{p_r} \sim N\left(0, \boldsymbol{M_{p_r}}\sigma^2_{p_r}\right)$. p_{w_i} is the random WW standard-proteomic value for hybrid i; $\boldsymbol{p_w} \sim N\left(0, \boldsymbol{M_{p_w}}\sigma^2_{p_w}\right)$. $\boldsymbol{M_{t_r}}$, $\boldsymbol{M_{t_w}}$, $\boldsymbol{M_{p_r}}$, and $\boldsymbol{M_{p_w}}$ are the standard-omics-based kernel matrices. All random effects are assumed to be independent.

Afterward, we also extended the model to include a random genetic effect (genomic) as defined in Eq. (2). The extended model hereafter will be referred to as *G+OBLUP* and can be represented as follows:

$$y_i = \mu + g_i + t_{r_i} + t_{w_i} + p_{r_i} + p_{w_i} + \varepsilon_i \tag{5}$$

where $\boldsymbol{y_i}$, μ, g_i, and ε_i are the same as defined in Eq. (2). t_{r_i}, t_{w_i}, p_{r_i}, and p_{w_i} are the same as defined in Eq. (4). Like *GBLUP*, we applied both *OBLUP* and *G+OBLUP* on each platform and field trait one by one. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

AOBLUP and G+AOBLUP: We proposed *AOBLUP* and *G+AOBLUP* as extensions of the two previous models, *OBLUP* and *G+OBLUP*. In the first step, we extracted the additive component (BLUPs) of omics features by applying an independent *GBLUP* model to each feature (transcript or protein):

$$O_{m,i} = \mu + u_{m,i} + \varepsilon_{m,i}, \tag{6}$$

where $O_{m,i}$ is the adjusted mean of omics feature m for hybrid i, which can be either a single protein or transcript. $u_{m,i}$ is the random genetic value of genome-wide markers on the omics feature m for hybrid i. The random genetic effects are considered to follow a normal distribution, i.e., $\boldsymbol{u_m} \sim N\left(0, \boldsymbol{K_G}\sigma^2_{u_m}\right)$ with $\boldsymbol{K_G}$ as the genomic kinship matrix and $\sigma^2_{u_m}$ as the genetic variance. $\varepsilon_{m,i}$ is the random error effect independent from $u_{m,i}$ and follows a normal distribution as $\boldsymbol{\varepsilon_m} \sim N\left(0, \boldsymbol{I}\sigma^2_{\varepsilon_m}\right)$ with $\boldsymbol{I}$ as the identity matrix and $\sigma^2_{\varepsilon_m}$ as the residual variance. We also computed Pearson correlations between omics and phenotypic traits and between additive components of omics [predicted by model Eq. (6)] and phenotypic traits to evaluate which omics (standard or additive) are more useful for predicting a trait of interest.

In the second step, we estimated omics-based kernel matrices following the same formula as Eq. (3) but with the genomic BLUPs of omics features (i.e., the additive part of omics features: $\hat{\boldsymbol{u}}_{\boldsymbol{m}}$) obtained in step one. In the third step, we fitted the *AOBLUP* and *G+AOBLUP* models separately by using additive-omics-based covariance matrices, i.e., $\boldsymbol{M}_{\tilde{f}_t} = \hat{\boldsymbol{O}}_{ft}\hat{\boldsymbol{O}}_{ft}\boldsymbol{\prime}/n_{ft}$, instead of standard-omics-based kernel matrices. $\boldsymbol{M}_{\tilde{f}_t}$ is the resulting relationship kernel of additive omic type $\tilde{f}$ (proteins or transcripts) from treatment t (WD or WW). From now tilde (~) will be used to represent effects associated or modeled with additive components of omics. $\hat{\boldsymbol{O}}_{ft}$ is the matrix of genomic BLUPs (referred to as additive) of omics obtained in the first step. The *AOBLUP* model can be represented as follows:

$$y_i = \mu + \tilde{t}_{r_i} + \tilde{t}_{w_i} + \tilde{p}_{r_i} + \tilde{p}_{w_i} + \varepsilon_i, \tag{7}$$

where y_i, μ, and ε_i are the same as defined earlier for Eq. (2). $\tilde{t}_{r_i}$ is the random WD additive-transcriptomic value; $\tilde{\boldsymbol{t}}_{\boldsymbol{r}} \sim N\left(0, \boldsymbol{M}_{\tilde{\boldsymbol{t}}_{\boldsymbol{r}}}\sigma^2_{\tilde{t}_r}\right)$. Similarly, $\tilde{t}_{w_i}$ is the random WW additive-transcriptomic value; $\tilde{\boldsymbol{t}}_{\boldsymbol{w}} \sim N\left(0, \boldsymbol{M}_{\tilde{\boldsymbol{t}}_{\boldsymbol{w}}}\sigma^2_{\tilde{t}_w}\right)$. $\tilde{p}_{r_i}$ is the random WD additive-proteomic value; $\tilde{\boldsymbol{p}}_{\boldsymbol{r}} \sim N\left(0, \boldsymbol{M}_{\tilde{\boldsymbol{p}}_{\boldsymbol{r}}}\sigma^2_{\tilde{p}_r}\right)$. $\tilde{p}_{w_i}$ is the random WW additive-proteomic value; $\tilde{\boldsymbol{p}}_{\boldsymbol{w}} \sim N\left(0, \boldsymbol{M}_{\tilde{\boldsymbol{p}}_{\boldsymbol{w}}}\sigma^2_{\tilde{p}_w}\right)$. Here, $\boldsymbol{M}_{\tilde{\boldsymbol{t}}_{\boldsymbol{r}}}$, $\boldsymbol{M}_{\tilde{\boldsymbol{t}}_{\boldsymbol{w}}}$, $\boldsymbol{M}_{\tilde{\boldsymbol{p}}_{\boldsymbol{r}}}$, and $\boldsymbol{M}_{\tilde{\boldsymbol{p}}_{\boldsymbol{w}}}$ are additive-omics-based kernel matrices.

While the *G+AOBLUP* can be represented as follows:

$$y_i = \mu + g_i + \tilde{t}_{r_i} + \tilde{t}_{w_i} + \tilde{p}_{r_i} + \tilde{p}_{w_i} + \varepsilon_i, \tag{8}$$

where y_i, μ, g_i, and ε_i are the same as defined in Eq. (2). $\tilde{t}_{r_i}$, $\tilde{t}_{w_i}$, $\tilde{p}_{r_i}$, and $\tilde{p}_{w_i}$ are the same as defined for the Eq. (7). Like *GBLUP*, we applied these models to all platform and field traits one by one. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

Multi-environment prediction models with interactions for field traits prediction

Multi-environment prediction models enable information sharing between different environments, which can help in improving predictive abilities. These models were only used for field trait predictions.

GxE-BLUP, GOxE-BLUP, and GAOxE-BLUP: The model, hereafter referred to as *GxE-BLUP*, integrates a GxE effect as follows:

$$y_{ij} = \mu_j + g_{ij} + \varepsilon_{ij}, \tag{9}$$

with y_{ij} being the adjusted mean of hybrid i in environment j. μ_j is the mean effect of environment j. g_{ij} is the random genetic effect of hybrid i in environment j. ε_{ij} is the residual of hybrid i in environment j. The random effect g_{ij} follows a normal distribution, i.e., $g_{ij} \sim MVR(0, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma} = \boldsymbol{K_G} \otimes \boldsymbol{\Sigma_E}$. Here, $\boldsymbol{K_G}$ is the genomic kinship matrix, $\otimes$ represents the Kronecker product, and $\boldsymbol{\Sigma_E}$ is the between-environment genetic covariance matrix with compound symmetry parametrization:

$$\boldsymbol{\Sigma_E} = \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_2^2 & \sigma_1^2 \end{bmatrix}, \tag{10}$$

with σ_1^2 as the main genetic effect variance plus the GxE variance and σ_2^2 is the covariance between the environments. The compound symmetry model considers the same within environment variance and between environments covariance.

We introduce *GOxE-BLUP* and *GAOxE-BLUP* as extensions of *GxE-BLUP* to integrate the omics data. In these two models, we included both the main omic effect and omic x environment interaction effects of omics. The first extension can be written as:

$$y_{ij} = \mu_j + g_{ij} + t_{w_{ij}} + t_{r_{ij}} + p_{w_{ij}} + p_{r_{ij}} + \varepsilon_{ij}, \tag{11}$$

similar to g_{ij}, $t_{w_{ij}}$, $t_{r_{ij}}$, $p_{w_{ij}}$, and $p_{r_{ij}}$ are the standard-omics-based genetic effects of hybrid i in environment j. Like g_{ij}, these effects also follow a multivariate normal distribution as; $t_{r_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{t_r}}\right)$, with $\boldsymbol{\Sigma_{t_r}} = \boldsymbol{M_{t_r}} \otimes \boldsymbol{\Sigma_{E_{tr}}}$, $t_{w_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{t_w}}\right)$, with $\boldsymbol{\Sigma_{t_w}} = \boldsymbol{M_{t_w}} \otimes \boldsymbol{\Sigma_{E_{tw}}}$, $p_{r_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{p_r}}\right)$, with $\boldsymbol{\Sigma_{p_r}} = \boldsymbol{M_{p_r}} \otimes \boldsymbol{\Sigma_{E_{pr}}}$, $p_{w_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{p_w}}\right)$, with $\boldsymbol{\Sigma_{p_w}} = \boldsymbol{M_{p_w}} \otimes \boldsymbol{\Sigma_{E_{pw}}}$. Here, $\boldsymbol{M_{t_r}}$, $\boldsymbol{M_{t_w}}$, $\boldsymbol{M_{p_r}}$, $\boldsymbol{M_{p_w}}$ are the standard-omics-based kinship matrices. $\boldsymbol{\Sigma_{E_{tr}}}$, $\boldsymbol{\Sigma_{E_{tw}}}$, $\boldsymbol{\Sigma_{E_{pr}}}$, and $\boldsymbol{\Sigma_{E_{pw}}}$ are the respective between environment genetic covariance matrices following compound symmetry parameterizations like $\boldsymbol{\Sigma_E}$.

The second extension can be written as:

$$y_{ij} = \mu_j + g_{ij} + \tilde{t}_{r_{ij}} + \tilde{t}_{w_{ij}} + \tilde{p}_{r_{ij}} + \tilde{p}_{w_{ij}} + \varepsilon_{ij}, \tag{12}$$

here, $\tilde{t}_{r_{ij}}$, $\tilde{t}_{w_{ij}}$, $\tilde{p}_{r_{ij}}$, and $\tilde{p}_{w_{ij}}$ are the additive-omics-based genetic effects of hybrid i in environment j. Like before in Eq. (11), these effects also follow a multivariate normal distribution as; $\tilde{t}_{r_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{\tilde{t}_r}}\right)$, with $\boldsymbol{\Sigma_{\tilde{t}_r}} = \boldsymbol{M_{\tilde{t}_r}} \otimes \boldsymbol{\Sigma_{E_{\tilde{t}r}}}$, $\tilde{t}_{w_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{\tilde{t}_w}}\right)$, with $\boldsymbol{\Sigma_{\tilde{t}_w}} = \boldsymbol{M_{\tilde{t}_w}} \otimes \boldsymbol{\Sigma_{E_{\tilde{t}w}}}$, $\tilde{p}_{r_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{\tilde{p}_r}}\right)$, with $\boldsymbol{\Sigma_{\tilde{p}_r}} = \boldsymbol{M_{\tilde{p}_r}} \otimes \boldsymbol{\Sigma_{E_{\tilde{p}r}}}$, $\tilde{p}_{w_{ij}} \sim MVR\left(0, \boldsymbol{\Sigma_{\tilde{p}_w}}\right)$, with $\boldsymbol{\Sigma_{\tilde{p}_w}} = \boldsymbol{M_{\tilde{p}_w}} \otimes \boldsymbol{\Sigma_{E_{\tilde{p}w}}}$. Here, $\boldsymbol{M_{\tilde{t}_r}}$, $\boldsymbol{M_{\tilde{t}_w}}$, $\boldsymbol{M_{\tilde{p}_r}}$, $\boldsymbol{M_{\tilde{p}_w}}$ are the additive-omics-based kernel matrices. $\boldsymbol{\Sigma_{E_{\tilde{t}r}}}$, $\boldsymbol{\Sigma_{E_{\tilde{t}w}}}$, $\boldsymbol{\Sigma_{E_{\tilde{p}r}}}$, and $\boldsymbol{\Sigma_{E_{\tilde{p}w}}}$ are the respective between environment genetic covariance matrices following compound symmetry parameterizations similar to $\boldsymbol{\Sigma_E}$. We applied these models using the R package Sommer (Covarrubias-Pazaran 2016).

GxW-BLUP, GOxW-BLUP, and GAOxW-BLUP: ECs can also be used to describe environmental conditions that can impact and interact with the genotypes (Jarquín et al. 2014). We fitted the following model to include the effect of standardized ECs on traits (Jarquín et al. 2014):

$$y_{ij} = \mu + g_i + E_j + gw_{ij} + \varepsilon_{ij}, \tag{13}$$

where μ is the overall mean, g_i is the effect of hybrid genotype i and E_j is the main effect of environment j. The term gw_{ij} is the genotype x ECs interaction term, which can be defined as $gw \sim N\left(0, \boldsymbol{K_G} \otimes \boldsymbol{W}\sigma^2_{gw}\right)$, where $\boldsymbol{W}$ is the inter-environmental covariance matrix defined as $\boldsymbol{W} = 1 - \frac{ED}{\max(ED)}$. ED here represents the Euclidean distance between different environments calculated as $ED_{jj'} = \sqrt{\sum_{q=1}^{n}\left(EC_{jq} - EC_{j'q}\right)^2}$, EC_{jq} is the qth EC in the environment j. We followed the same approach as Rincent et al. (2019) for estimating $\boldsymbol{W}$.

Similar to *GOxE-BLUP* and *GAOxE-BLUP*, we extended the *GxW-BLUP* model into *GOxW-BLUP* [Eq. (14)] and *GAOxW-BLUP* [Eq. (15)] to account for standard and additive-omics data, respectively.

$$\begin{aligned} y_{ij} = \mu + g_i + E_j + gw_{ij} + t_{r_i} + t_r w_{ij} + t_{w_i} \\ + t_w w_{ij} + p_{r_i} + p_r w_{ij} + p_{w_i} + p_w w_{ij} + \varepsilon_{ij}, \end{aligned} \tag{14}$$

and

$$\begin{aligned} y_{ij} = \mu + g_i + E_j + gw_{ij} + \tilde{t}_{r_j} + \tilde{t}_r w_{ij} + \tilde{t}_{w_j} \\ + \tilde{t}_w w_{ij} + \tilde{p}_{r_i} + \tilde{p}_r w_{ij} + \tilde{p}_{w_i} + \tilde{p}_w w_{ij} + \varepsilon_{ij}, \end{aligned} \tag{15}$$

where $t_r w_{ij}$, $t_w w_{ij}$, $p_r w_{ij}$, and $p_{wr} w_{ij}$ are the interactions between different standard-omics-based values and ECs, which can be defined as $t_r w \sim N\left(0, \left[\boldsymbol{M_{t_r}}\right] \otimes \boldsymbol{W}\sigma^2_{t_r w}\right)$, $t_w w \sim N\left(0, \left[\boldsymbol{M_{t_w}}\right] \otimes \boldsymbol{W}\sigma^2_{t_w w}\right)$, $p_r w \sim N\left(0, \left[\boldsymbol{M_{p_r}}\right] \otimes \boldsymbol{W}\sigma^2_{p_r w}\right)$

and $p_w w \sim N\left(0, \left[\boldsymbol{M}_{\boldsymbol{p}_w}\right] \otimes \mathbf{W}\sigma^2_{p_w w}\right)$, respectively. While, $\tilde{t}_w w_{ij}$, $\tilde{t}_r w_{ij}$, $\tilde{p}_w w_{ij}$, and $\tilde{p}_r w_{ij}$ are the interactions between different additive-omics-based values and ECs, which can be defined as $\tilde{t}_r w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{t}}_r}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{t}_r w}\right)$, $\tilde{t}_w w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{t}}_w}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{t}_w w}\right)$, $\tilde{p}_r w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{p}}_r}\right] \otimes \boldsymbol{W}\sigma^2_{\tilde{p}_r w}\right)$ and $\tilde{p}_w w \sim N\left(0, \left[\boldsymbol{M}_{\tilde{\boldsymbol{p}}_w}\right] \otimes \mathbf{W}\sigma^2_{\tilde{p}_w w}\right)$, respectively. We fitted these models using the R package BGLR (Pérez and de los Campos 2014).

MegaLMM: With MegaLMM, all traits (platform and yield) and the omics are considered jointly in the **Y** matrix, each column of **Y** being a trait or an omic feature. *MegaLMM* utilizes latent factors to model joint effects of all predictors, including fixed, random, and residuals (Runcie et al. 2021). *MegaLMM* decomposes trait variation into two components, the first component is dependent on the *K* latent factors, which model correlations between the *t* traits, and the second component is trait-specific variation uncorrelated between the traits. In other words, *MegaLMM* decomposes the variation of *t* traits into correlated and uncorrelated components. Further details about *MegaLMM* can be found in Runcie et al. (2021). In our case the traits used as input of MegaLMM were grain yield in the different environments, or platform traits, and potentially omics data, depending on the prediction scenario.

The final model implemented in this study can be written as follows:

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\Lambda} + \mathbf{E}, \tag{16}$$

With

$$\forall \boldsymbol{v}, \boldsymbol{F}_{\boldsymbol{v}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{g} + \boldsymbol{\varepsilon}, \tag{17}$$

where **Y** is the *n x t* matrix for *n* hybrids and *t* traits (can be field traits, platform traits, and omics). **F** is a *n x K* matrix of *n* hybrids and *K* latent factors (*K* is defined by the user). **Λ** is the *K x t* loading matrix whose elements, i.e. λ_{kj}, determines the relation between the trait *k* and the corresponding factor *j*. **E** is a matrix of independent residuals for each trait. The *K* latent factors present in **F**, i.e., F_v for latent factor *v*, are further individually decomposed into fixed effect $\boldsymbol{\beta}$ and random effect $\boldsymbol{g}$, with $\boldsymbol{X}$ and $\boldsymbol{Z}$, respectively, as their incidence matrices. $\boldsymbol{g}$ is modeled as follows: $\boldsymbol{g} \sim N\left(0, \mathbf{K}_{\mathbf{G}}\sigma^2_g\right)$. $\boldsymbol{\varepsilon}$ is the residual of each independent factor modeled as $\boldsymbol{\varepsilon} \sim N\left(0, \mathbf{I}\sigma^2_{\varepsilon}\right)$.

We implemented three versions of *MegaLMM*. The first version contained only platform traits or field grain yield in the **Y** matrix. From now on, we will refer to this model as MegaLMM. The second version contained traits, i.e., platform or field grain yield and standard-omics expressions and abundances. From now on, we will refer to this version as MegaGO. The third version contained traits, i.e., platform or field grain yield and additive-omics expressions and abundances. From now on, we will refer to this version as MegaGAO. We specified 10 latent factors (*K*) for *MegaLMM* models and 150 latent factors (*K*) for models including omics, i.e., *MegaGO* and *MegaGAO*. *MegaLMM*-based models can estimate genetic correlations between all the traits (phenotypic traits and platform omics). These genetic correlations allow *MegaLMM* to use more information than univariate models, that consider each trait independently. The *MegaLMM* model uses Bayesian priors to regularize different effects. We used Automatic Relevance Determination (ARD) (Mbuvha et al. 2020) for the factor loadings matrix. In high-dimensional multi-omics datasets with many features, ARD priors can help identify and emphasize important features, improving predictive ability and model interpretability. They introduce sparsity to the model by setting many predictor coefficients close to zero, aiding in feature selection and preventing overfitting. ARD priors are particularly useful when dealing with collinearity and limited sample sizes, as they enhance data efficiency and model stability. We fitted the *MegaLMM* models using the MegaLMM R-package (Runcie et al. 2021).

Prediction scenarios

We compared two prediction scenarios that plant breeders classically face to evaluate the efficiency of the different models and omics data. The first one corresponds to a classical situation in which breeders predict the performance of hybrids unobserved for the target trait (CV-NH). The second one predicts the performance of partially observed hybrids (CV-POH), similar to sparse testing. For both scenarios, all hybrids (training and test) had omics characterization available (Fig. 1). Moreover, we used the same five-fold and five-repeat divisions of hybrids for all platform traits and grain yield trials for both CV-NH and CV-POH. Comparing the predictive abilities of the different models for platform traits and field traits allows us to evaluate the efficiency of omics data to predict non-additive effects when predicting traits measured in similar conditions (platform traits in a different experiment) or in completely different conditions (field) than those in which the omics were measured (the same platform experiment).

CV-NH for platform traits

In CV-NH, we performed fivefold with 5-repeats cross-validations for all the platform traits (Fig. 1). Meaning that for each iteration of the cross-validation, 20% of the hybrids (test set) were masked for all the platform traits. This scheme was applied with *MegaLMM*, *MegaGO*, and *MegaGAO*. The

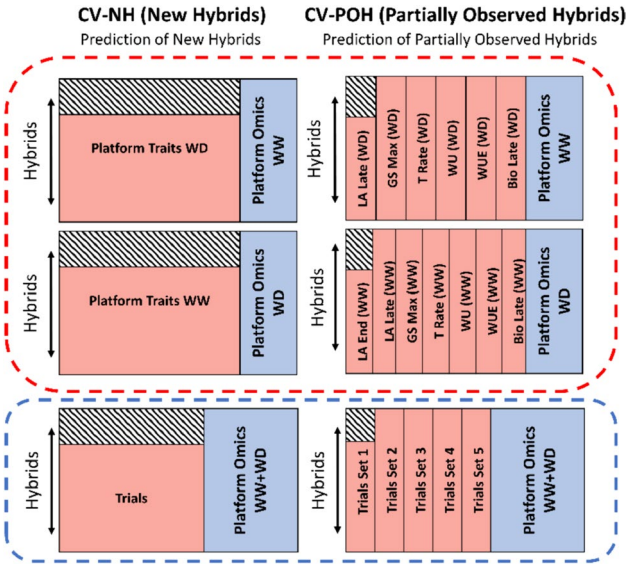

Fig. 1 Description of the fivefold and 5-repeat cross-validation schemes. CV-NH is the scenario where we predict hybrids having no phenotypic observations but only platform omics data. CV-POH is the scenario where we predict hybrids with partially observed phenotypes and complete platform omics data. The red (dashed) square represents the cross-validation settings of platform traits, while the blue (dashed) square represents the cross-validation settings of field grain yield. The pink color represents phenotypes available for training the models (platform traits or field grain yield), and the blue represents platform omics used as predictors. The hatched rectangle represents the hybrids/traits that are to be predicted (colour figure online)

omics information of both the training and test hybrids were used in models capitalizing on omics.

CV-POH for platform traits

In the CV-POH scheme, we first categorized platform traits into trait sets. Each set contained different combinations of years and seasons for the same trait and the same water treatment. We formed six distinct trait sets for the WD platform traits, corresponding to traits *LAe*, *gs*, *Trate*, *WU*, *WUE*, and *Bl*, with three year-season combinations (three experiments) each. We created seven distinct trait sets for the WW platform traits, where the first two sets comprised one and two-year + season combinations of *LAe* and *LAl*, respectively. The remaining 5 comprised three-year + season combinations of 5 platform traits: gs, Trate, *WU*, *WUE*, and *Bl* (Fig. 1). We performed fivefold and 5-repeat cross-validation for each trait set and pooled all predictive abilities. It means that for each iteration of the cross-validation, 20% of the hybrids (test set) were masked for the trait set (one at a time) that had to be predicted. We applied this scheme to *MegaLMM*, *MegaGO*, and *MegaGAO*, with the last two also using omics information as predictors.

CV-NH for field traits (grain yield)

In CV-NH, we applied fivefold cross-validation with 5 repeats, in which we masked the phenotypes of the test hybrids across all the field trials (Fig. 1). We applied this validation scheme with the multienvironment and multivariate models defined above, i.e., *GxE-BLUP*, *GxW-BLUP*, *MegaLMM*, *GOxE-BLUP*, *GOxW-BLUP*, *MegaGO*, *GAOxE-BLUP*, *GAOxW-BLUP*, and *MegaGAO*. We also included WW and WD omics information for both training and test hybrids for the models capitalizing on omics.

CV-POH for field traits (grain yield)

In CV-POH, we first randomly assigned the 25 yield trials in five sets of five trials each, referred to as Trial Sets 1 to 5. We performed five-fold and five-repeat cross-validation for each trial set (Fig. 1). For each cross-validation iteration, 20% of the hybrids were masked for the trial set that had to be predicted. We applied this validation scheme with the same multivariate models mentioned in the previous section. We also included omics information for both training and test hybrids for the models capitalizing on omics.

As univariate models (*GBLUP*, *OBLUP*, *G* + *OBLUP*, *AOBLUP*, and *G* + *AOBLUP*) consider a single platform trait or a grain yield trial at a time, they resulted in the same predictive abilities in CV-NH and in CV-POH. In addition, models that included a GxE component or ECs were only applied to grain yield field trials.

An important point to note is that, unlike grain yield, we independently predicted the two platform trait treatments (WW and WD). We predicted WW platform traits with omics originating from WD treatment and predicted WD platform traits with omics originating from WW treatment (Fig. 1). We implemented this precaution to ensure that the predictive capabilities based on omics solely depended on genetic effects rather than any potential non-genetic effects that might arise from obtaining omics data and traits from the same plants.

Predictive ability was defined as the Pearson correlation between the predictions and the adjusted means of the test hybrids. As this study relies on model comparison, we used gain compared to the reference model (*GBLUP*) as a measure of model performance. We defined gain as the predictive ability of each model minus the ability of univariate *GBLUP*, which only uses information from within the single predicted platform trait or grain yield trial, as was the case for all univariate models.

Table 1 Predictive abilities of different genomic and omics models for predicting WD & WW platform traits and field traits (grain yield) in CV-NH and CV-POH cross-validation schemes

Model type	Model	Platform traits				Field traits (GY)	
		WD		WW		CV-NH	CV-POH
		CV-NH	CV-POH	CV-NH	CV-POH		
Genomic	GBLUP	0.26 ± 0.11		0.24 ± 0.10		0.57 ± 0.09	
	GxE-BLUP	–	–	–	–	0.58 ± 0.09	0.72 ± 0.06
	GxW-BLUP	–	–	–	–	0.58 ± 0.05	0.72 ± 0.08
	MegaLMM	0.25 ± 0.11	0.88 ± 0.03	0.25 ± 0.10	0.89 ± 0.03	0.59 ± 0.09	0.77 ± 0.05
Standard Multi-omics	OBLUP	0.44 ± 0.09		0.42 ± 0.09		0.55 ± 0.10	
	G + OBLUP	0.43 ± 0.09		0.42 ± 0.09		0.54 ± 0.10	
	GOxE-BLUP	–	–	–	–	0.57 ± 0.10	0.72 ± 0.06
	GOxW-BLUP	–	–	–	–	0.59 ± 0.10	0.72 ± 0.06
	MegaGO	0.15 ± 0.14	0.58 ± 0.12	0.13 ± 0.15	0.59 ± 0.09	0.58 ± 0.09	0.61 ± 0.10
Additive Multi-omics	AOBLUP	0.40 ± 0.11		0.40 ± 0.11		0.58 ± 0.10	
	GAOBLUP	0.38 ± 0.12		0.36 ± 0.12		0.58 ± 0.10	
	GAOxE-BLUP	–	–	–	–	0.58 ± 0.09	0.72 ± 0.06
	GAOxW-BLUP	–	–	–	–	0.59 ± 0.09	0.74 ± 0.06
	MegaGAO	0.30 ± 0.13	0.35 ± 0.12	0.30 ± 0.15	0.39 ± 0.12	0.60 ± 0.09	0.60 ± 0.10

Model type represents whether the model has genomic, standard multi-omics, or additive multi-omics information. Univariate models presented were tested only trait by trait for platform and environment by environment for grain yield under five-fold and five-repeat CV, resulting in the same predictive abilities for CV-NH and CV-POH. Standard deviations are computed individually for each platform trait and grain yield, followed by a mean over platform traits or field trials

Results

Univariate and multivariate models for platform trait prediction without transcript and protein information

CV-NH predictions

We predicted platform traits with different univariate and multivariate models under CV-NH (Table 1; Fig. 2). Firstly, in the case of WD traits, the reference univariate *GBLUP* model resulted in a mean predictive ability of 0.26 over all the traits. The multivariate *MegaLMM* model resulted in a predictive ability of 0.25 with negligible difference compared to the reference *GBLUP* model. The results of WW platform traits are similar to the WD platform traits. For instance, the reference *GBLUP* model resulted in a mean predictive ability of 0.24, while the multivariate *MegaLMM* model resulted in 0.25.

CV-POH predictions

In the case of CV-POH (Table 1; Fig. 2) with WD traits, the *MegaLMM* model resulted in a predictive ability of 0.88, 0.62 units higher than the reference *GBLUP*. Similar results were also observed with WW platform traits in CV-POH with *MegaLMM*, resulting in a mean predictive ability of 0.89. In the CV-POH scenario, *MegaLMM* was the best model for predicting platform traits.

Univariate and multivariate models for platform trait prediction with transcript and protein information

CV-NH predictions

We used WW omics as predictors for WD platform traits (Table 1; Fig. 2). The *OBLUP* model applied to WD traits resulted in a mean predictive ability of 0.44, which is 0.18 units higher than the benchmark *GBLUP* model. Adding genomic information in addition to transcripts and proteins, i.e., in the *G + OBLUP* model, we observed a similar mean predictive ability of 0.43. We observed a slight decrease in predictive abilities when we shifted to additive-omics. The model solely with transcripts and proteins as predictors, i.e., *AOBLUP*, resulted in a mean predictive ability of 0.40. At the same time, the model that also used genomic information (*G + AOBLUP*) resulted in a mean predictive ability of 0.38.

The predictive abilities of models predicting WW platform traits were similar to those of WD platform traits. We used WD omics as predictors for WW platform traits (Table 1; Fig. 2). The OBLUP model resulted in a mean predictive ability of 0.42. Adding genomic information

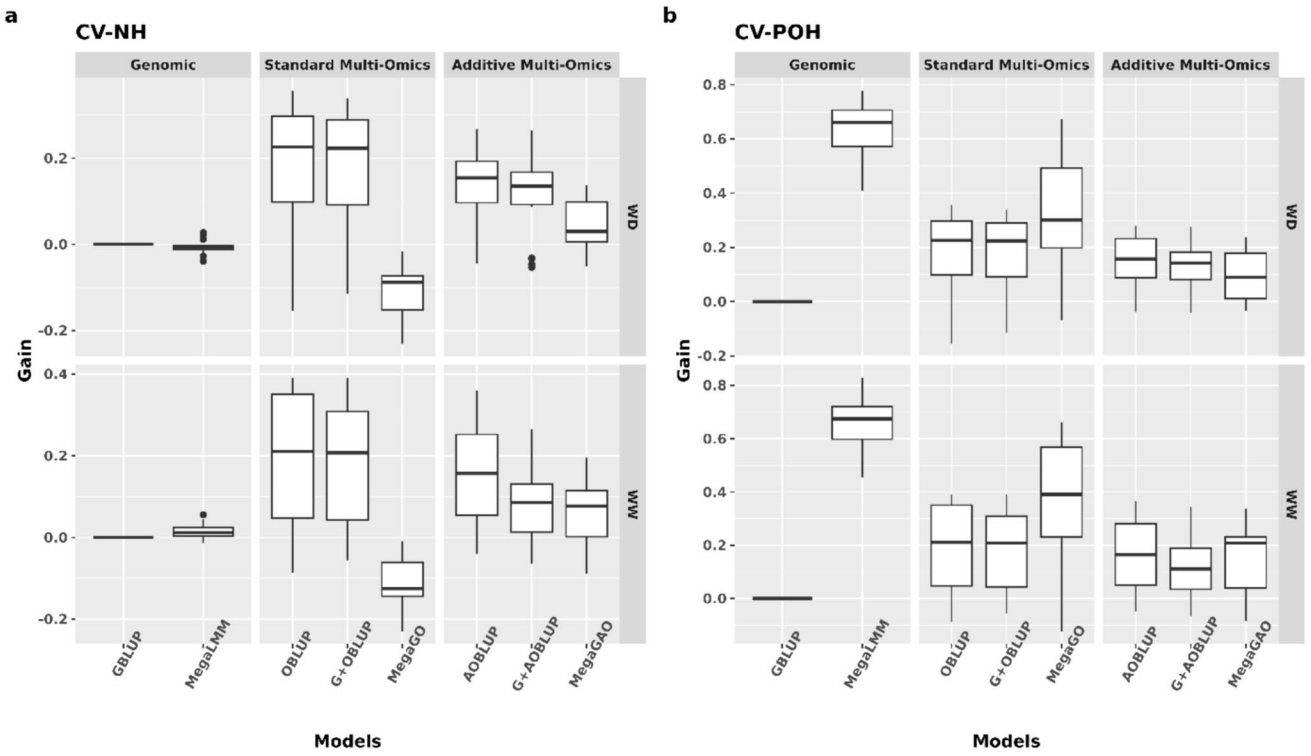

Fig. 2 Boxplots of predictive gains of different models compared to GBLUP in different cross-validation schemes for platform traits. **a** CV-NH Scheme for predicting new hybrids, **b** CV-POH Scheme for predicting partially observed hybrids. Here, WD represents WD platform traits predicted with WW omics, and WW represents WW traits predicted with WD omics. "Genomic" represents the scenarios without transcripts and proteins, "Standard Multi-omics" represents the case where standard transcripts and proteins are used, and "Additive Multi-omics" represents the case where the additive component of transcripts and proteins was used

($G+OBLUP$) did not change the predictive ability. As before, we observed a decrease in mean predictive abilities with additive-omics. For example, the *AOBLUP* model resulted in a predictive ability of 0.40, while $G+AOBLUP$ resulted in an even lower predictive ability of 0.36.

Now, we will compare different multivariate prediction models with omics information for predicting platform traits (Table 1; Fig. 2). In the case of WD traits, *MegaGO*, the model with standard-omics, resulted in a mean predictive ability of 0.15, lower than any other prediction model for the same traits. Shifting to additive-omics with *MegaGAO* resulted in a mean predictive ability of 0.30, higher than *MegaGO* but quite lower than all other univariate models for the WD platform traits.

The performances of multivariate *MegaGO* and *MegaGAO* for predicting WW platform traits (Table 1; Fig. 2) were similar to their respective cases with WD platform traits. For example, both resulted in mean predictive abilities of 0.13 and 0.30, respectively, again falling behind univariate models for the same traits.

CV-POH predictions

Under CV-POH, we tested two models capitalizing on omics information, *MegaGO* and *MegaGAO* (Table 1; Fig. 2), with the former utilizing standard-omics while the latter utilizing additive-omics. The *MegaGO* model applied to WD platform traits resulted in a mean predictive ability of 0.58, indicating a gain of 0.32 units over the reference *GBLUP* model but a loss of 0.30 units compared to *MegaLMM*, the model without omics. Shifting to additive-omics made the case worse. *MegaGAO* resulted in a mean predictive ability of 0.35, comparable to the reference *GBLUP* and lower than *MegaLMM* and *MegaGO*. We also applied *MegaGO* and *MegaGAO* models

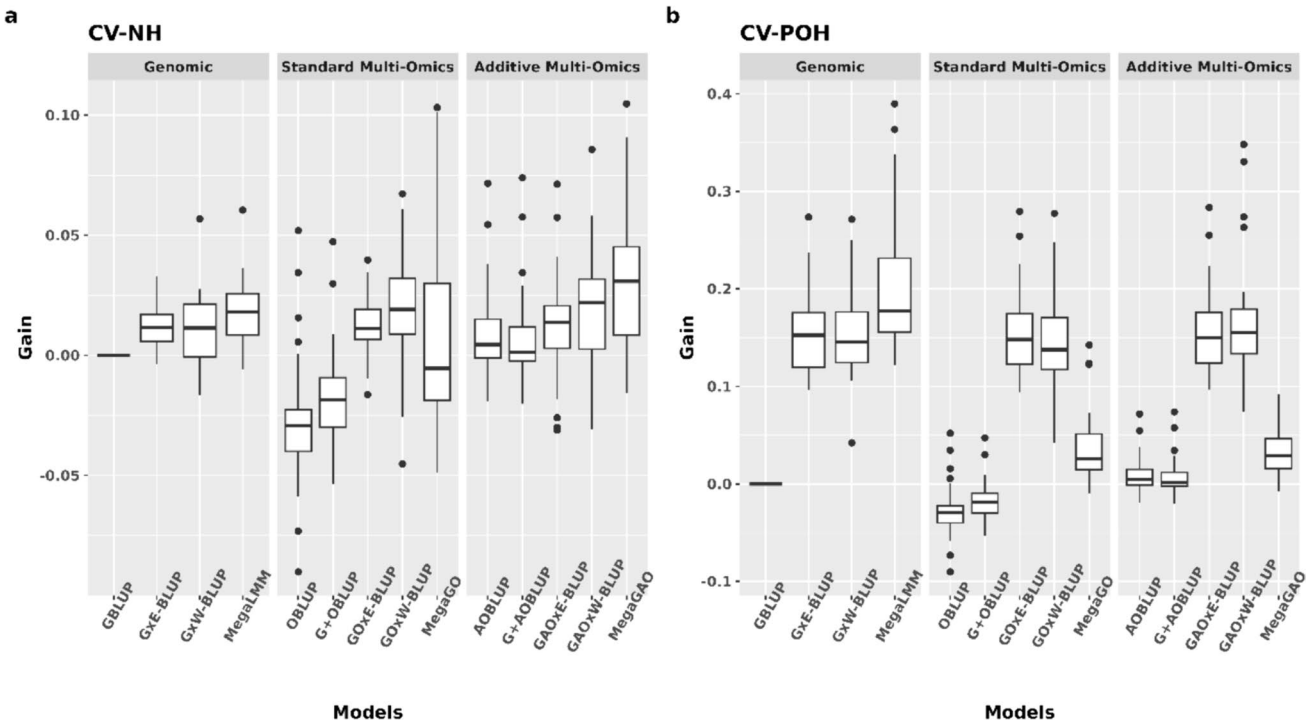

Fig. 3 Boxplots of predictive gains of different models compared to GBLUP in different cross-validation schemes for field grain yield. **a** CV-NH Scheme for predicting new hybrids, **b** CV-POH Scheme for predicting partially observed hybrids, "Genomic" represents the scenarios without transcripts and proteins, "Standard Multi-omics" represents the case where standard transcripts and proteins are used, and "Additive Multi-omics" represents the case where the additive components of transcripts and proteins were used

to WW platform predictions (Table 1; Fig. 2). Both models performed similarly as in the predictions of WD platform traits, resulting in mean predictive abilities of 0.59 and 0.39, respectively.

Univariate and multivariate models for field trait predictions without transcript and protein information

CV-NH predictions

At first, we applied both univariate and multivariate models to predict grain yield without adding multi-omics information (Table 1; Fig. 3). The reference *GBLUP* model resulted in a mean predictive ability of 0.57 over all the folds and repeats of all 25 field trials. The two interactions-based models, i.e., *GxE-BLUP* and *GxW-BLUP*, resulted in similar mean predictive abilities of 0.59 and 0.58, respectively. Similarly, the multivariate *MegaLMM* model also resulted in a mean predictive ability of 0.59. Even though the average predictive abilities of *MegaLMM* appear to be closer to the reference model, there are several trials where important gains were observed, for instance, 0.06 points in the case of *Karlsruhe.2012.OPT* in contrast to the minimum gain equal to 0 in *Campagnola.2012.WD* (Fig. S1).

CV-POH predictions

We tested *GxE-BLUP*, *GxW-BLUP*, and *MegaLMM* in a less challenging CV-POH situation for predicting grain yield (Table 1; Fig. 3). In this cross-validation scenario, *GxE-BLUP* and *GxW-BLUP* models resulted in a similar predictive ability of 0.72, which is 0.15 units higher than *GBLUP*. *MegaLMM* performed better than the reference *GBLUP* and the two interaction-based models, with a mean predictive ability of 0.77. We also observed considerable gains with *MegaLMM* for many environments. For instance, *Craiova.2012.WD* improved by 0.39 units in comparison to GBLUP (Fig. S2).

Univariate and multivariate models for field trait predictions with transcript and protein information

CV-NH predictions

Firstly, we tested the predictive ability of the omics data to predict field traits one by one using univariate models (Table 1; Fig. 3). We observed that adding standard-omics information in the univariate models decreased the predictive abilities compared to *GBLUP* (Table 1; Fig. 4). For example, *OBLUP* and *G + OBLUP* resulted in mean

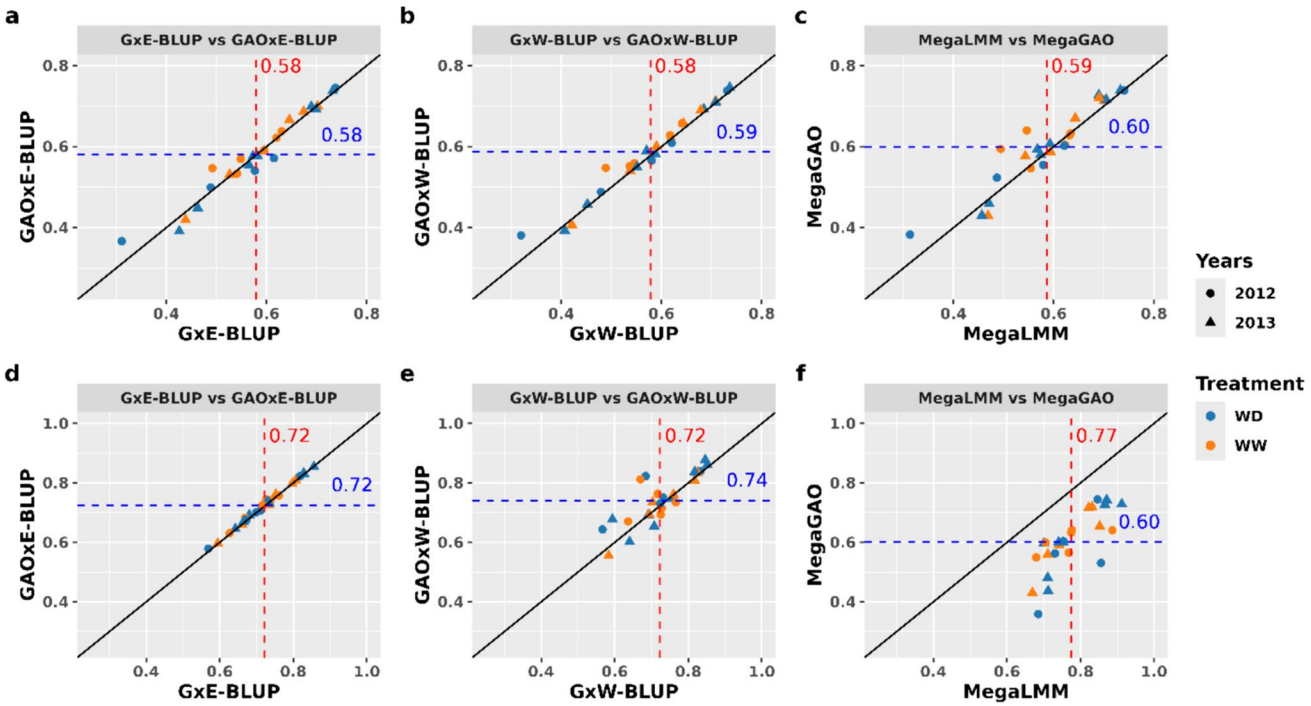

Fig. 4 Comparison of predictions of different additive-omics-based models with their versions with only genomics in CV-NH **a–c** and CV-POH **d–f** for grain yield. The X-axis represents models with genomics, and the Y-axis represents models involving additive-omics. The red dashed line represents the mean predictive ability of the model without omics (x-axis), while the blue dashed line represents the mean predictive ability of the model involving additive-omics (y-axis). The black line represents the diagonal. Colors and the shapes of points refer to the treatment and years of field trials

predictive abilities of 0.55 and 0.54, respectively, lower than the *GBLUP* benchmark model (0.57). Shifting to additive-omics with *AOBLUP* and *G* + *AOBLUP* models, we observed the same predictive ability of 0.58 with both models. Even though their predictive ability is somewhat higher than *OBLUP* and *G* + *OBLUP*, it is similar to that of the reference *GBLUP*.

The two interaction-based models with standard-omics, *GOxE-BLUP* and *GOxW-BLUP*, resulted in predictive abilities of 0.57 and 0.59, respectively, similar to their predecessors, *GxE-BLUP* and *GxW-BLUP*. The two interaction-based models with additive-omics, *GAOxE-BLUP* and *GAOxW-BLUP*, resulted in comparable predictive abilities of 0.58 and 0.59, respectively. Even though the performance is similar, *GAOxW-BLUP* resulted in important predictive gains in some trials, as high as 0.09 (*Craiova.2012.WD*), with a maximum loss of –0.03 (*Campagnola.2013.OPT*), as observed in Fig. 4.

The predictive ability of multivariate *MegaGO* (0.58) was also quite similar to the reference *GBLUP* and *MegaLMM* models. However, using additive-omics information, *MegaGAO* showed some improvement over *GBLUP* and *MegaLMM* with a mean predictive ability of 0.60. Even though we could not observe overall important improvements by using additive-omics data in *MegaGAO*, some environments, such as *Craoiva.2012.OPT* and *Campagnola.2012.OPT*, indicated gains as high as 0.10 compared to *MegaLMM*, while the maximum loss was –0.04 in *Campagnola.2013.OPT* (Fig. 4). Overall, in the challenging situation of CV-NH, all the models, including the additive-omics information, exhibited gains relative to their predecessors with standard-omics information, although these gains were small.

CV-POH predictions

The interaction-based models involving standard-omics, *GOxE-BLUP* and *GOxW-BLUP*, resulted in the same predictive abilities (0.72) as *GxE-BLUP* and *GxW-BLUP*, suggesting that omics information was not beneficial in the CV-POH scenario. The additive-omics-based models, *GAOxE-BLUP* and *GAOxW-BLUP*, resulted in mean predictive abilities of 0.72 and 0.74, respectively. Even though the gain with *GAOxW-BLUP* compared to *GxW-BLUP* is small, there are several trials where we observed gains as high as 0.09 (*Craiova.2012.WD*) and a maximum loss of 0.03 (*Campagnola.2013.OPT*) (Fig. 4). Multivariate *MegaGO* with standard-omics information resulted in a mean predictive ability of 0.61, indicating a sharp decrease compared to 0.77 of *MegaLMM* (Table 1; Fig. 3). It suggests that adding

standard-omics in the model was worsening the situation. This situation is also true with the multivariate *MegaGAO* model containing additive-omics information.

Discussion

The interest in utilizing omics data in trait predictions is increasing due to their ability to capture interactions and regulation processes in response to the environment. However, due to the high costs of multi-omics characterization, it is currently impossible for a breeder to collect such data for all field plots each year. It is possible to measure omics in controlled conditions on phenotyping platforms. Even though omics data originating from the platform help detect stress QTLs (Blein-Nicolas et al. 2020), their capability to predict platform and field traits needs further evaluation. Therefore, we evaluated different univariate and multivariate models with or without omics data. Our main objectives were to assess the efficiency of those omics and of a high dimensional multivariate model, *MegaLMM* (Runcie et al. 2021), to predict traits measured in similar conditions (platform), or to predict productivity traits measured in completely different conditions (the fields), than those in which the omics were measured (a platform experiment). For this, we have considered standard univariate models such as *GBLUP*, *OBLUP*, *AOBLUP*, *G + OBLUP*, and *G + AOBLUP*, predicting each platform trait or each grain yield trial one by one. We also fitted *GxE-BLUP* and *GxW-BLUP* models that were developed to capture GxE in multi-environment settings. We extended these models to include standard-omics, *GOxE-BLUP*, and *GOxW-BLUP*, or additive-omics, *GAOxE-BLUP* and *GAOxW-BLUP*. We have compared the efficiency of the multi-trait models in two different prediction schemes, CV-NH and CV-POH. CV-NH focuses on predicting new hybrids that are not phenotyped for platform or field traits but have multi-omics characterization. CV-POH predicts hybrids partially observed for some platform traits or field trials. It is generally easier to predict based on CV-POH than on CV-NH due to the fact that test hybrids are partially observed for platform traits or field trials in CV-POH.

We first tested and compared models without using the omics information to predict platform traits. In the case of CV-NH, *MegaLMM* performs similarly to *GBLUP* for both WW and WD platform traits. It is consistent with studies showing that multi-trait models are not much better than univariate models if none of the traits is phenotyped for the test set. In the case of CV-POH, *MegaLMM* performed extremely well (0.77) compared to *GBLUP* (0.57, Table 1). In this case, the *MegaLMM* model benefits from the genetic correlations available for the test hybrids because of unmasked platform traits. *MegaLMM* trait loadings matrix (*K x t*) captures the correlated variation in the phenotypic data (Runcie et al. 2021). Genetically correlated traits tend to load onto the same latent factors (*K*) (Fig. S3). *MegaLMM* can effectively use these genetic correlations between traits, especially when information about the test set is available for unmasked trait sets as in the case of CV-POH. We observed high genetic correlations between platform traits as modeled by *MegaLMM* (Fig. S3). This scenario, similar to trait-assisted prediction, has been shown to result in highly accurate predictions when predicted and observed traits are correlated (Calus and Veerkamp 2011; Fernandes et al. 2018; Henderson and Quaas 1976; Jia and Jannink 2012; Pszczola et al. 2013; Robert et al. 2020).

Platform Omics are very efficient in predicting additive and non-additive genetic effects for platform traits

We also tested different standard-omics and additive-omics-based prediction models on platform traits. *OBLUP* and *G + OBLUP*, the two univariate prediction models with standard-omics information, were much better than the reference GBLUP model (Table 1; Fig. 2), with a gain in predictive ability as high as 0.36 (*Spring_2013.WD.WU*) and 0.34 (*Spring_2012.WD.WU*), respectively. Meaning that omics are very efficient in making predictions in environmental conditions similar to those in which they were measured. This is of particular interest for prediction approaches based on physiological traits measured in controlled conditions (Bouidghaghen et al. 2023). Note that we ensured that the improved predictive abilities of models *OBLUP* and *G + OBLUP* were only due to genetics (and not because of some shared environmental factors) by predicting traits measured on different plants than those used for omics measurements. More precisely, we used omics measured in the WD treatment to predict plants grown under WW treatment and the reverse. It means that the only non-genetic effect that could be involved is the quality of the seed lots, as the same seed lots were used for both treatments.

One key result is that upon fitting the models with additive-omics, we observed an important decrease in predictive abilities with *AOBLUP* and *G + AOBLUP* compared to *OBLUP* and *G + OBLUP*. It means that the non-additive genetic part of the omics was very useful in predicting platform traits. Another important result to note is that omics measured in similar conditions as the predicted trait (but different water treatments) can capture some non-additive genetic variation like epistasis and responses to growth conditions. We can indeed hypothesize that responses to local conditions are driven by regulatory processes, with epistasis consequently. These findings are also in line with the suggestions of Li et al. (2019) regarding the importance of omics expressions in capturing non-additive genetic variation. The potential ability of *OBLUP* to capture non-additive genetics

works when conditions are similar because the epistatic variation is likely to be more correlated among the platform traits than between platform and field traits.

GxE-BLUP and *GxW-BLUP* models are useful for predicting grain yield

The benchmark *GBLUP* model (Table 1; Fig. 3) resulted in a mean predictive ability of 0.57 for grain yield (25 trials). In both CV-NH and CV-POH validation schemes, multi-environment models *GxE-BLUP* and *GxW-BLUP* models performed better than *GBLUP*. As expected, the gains with models in CV-NH are not as prominent as in CV-POH. These results are comparable to Burgueño et al. (2012), where authors observed a significant improvement in CV-POH but not in CV-NH. Even though our *GxE-BLUP* model performs extremely well, it has one disadvantage due to its underlying assumption of the same variance and covariance among different trials. Furthermore, we expected the model with envirotyping data, i.e., *GxW-BLUP*, to perform better than *GxE-BLUP*, but this was not the case. The potential advantage of using *GxW-BLUP* is to give more weight to trials with similar environmental conditions. Literature indicates that including environmental covariates is not always advantageous. For instance, Monteverde et al. (2019); Nguyen et al. (2023) reported that adding environmental covariates resulted in worse predictions than *GBLUP* in rice while predicting new environments, which could also be attributed to higher inter-environmental differences. In the studies predicting new genotypes, Jarquín et al. (2014); Malosetti et al. (2016), and Rincent et al. (2019) observed little to no improvement with a GxW model.

MegaLMM is an efficient model for predicting field traits in multi-environment scenarios

We also implemented a two-level hierarchical model, *MegaLMM* (Runcie et al. 2021). *MegaLMM* was the best model for both CV-NH and CV-POH prediction schemes (before multi-omics inclusion). The gain was marginal for CV-NH, but *MegaLMM* seemed to be very efficient at dissecting genetic correlations (Fig. S4) between trials and benefit from those, particularly in the CV-POH scenario, with a gain of 0.21 units in comparison to *GBLUP* and 0.05 compared to *GxE-BLUP* and *GxW-BLUP*. In previous studies, *MegaLMM* was shown to result in a high predictive gain compared to *GBLUP* in sparse testing scenarios (Runcie et al. 2021). However, it has not been tested in the more challenging situation of CV-NH before. These results confirm that multi-trait models are very helpful when some information on the predicted set is available, for example, in the case of trait-assisted predictions or sparse trial networks.

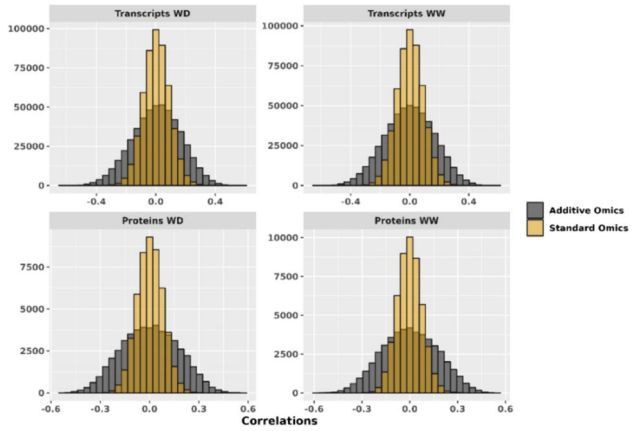

Fig. 5 The distribution of the correlations between omics and grain yield from all 25 trials. The gray color refers to the additive component of omics, and the yellow color refers to the standard-omics. WD represents omics measured in water-deficit and WW represents omics measured in well-watered conditions. The top row corresponds to transcripts and the bottom row corresponds to proteins

The inclusion of platform-based omics can only slightly improve the predictive ability of field traits

We also fitted different prediction models with standard-omics or with the additive part of the omics. We hypothesized that extracting the additive genetic component of omics might be more beneficial for field predictions because it is a way of removing part of the error (Christensen et al. 2021) and because non-additive genetic effects are probably different on the platform than in the field, and so not useful for field trait predictions. It is clearly visible in Fig. 5 that the additive-omics expressions have a higher Pearson correlation with grain yield than the standard-omics, which was not the case for platform traits (Fig. S5). In CV-NH (Fig. 3), none of the models sufficiently benefited from adding the standard-omics. However, with the addition of additive-omics instead of standard-omics, most of the models exhibited small to intermediate gains. For instance, the *AO-BLUP* model appeared to be 0.03 units better than *OBLUP* but similar to *GBLUP*, meaning that the additive part of omics alone can predict similarly to genomic markers. It is likely that additive-omics-based kernel matrices get closer to the genomic kinship matrix, which is evident from the higher correlations between additive-omics kernels and genomic kinship matrix (0.68, 0.61, 0.62, and 0.60 for $\boldsymbol{M}_{\tilde{t}_r}$, $\boldsymbol{M}_{\tilde{t}_w}$, $\boldsymbol{M}_{\tilde{p}_r}$, and $\boldsymbol{M}_{\tilde{p}_w}$, respectively) compared to standard-omics kernels and genomic kinship matrix (0.59, 0.56, 0.58, and 0.57 for $\boldsymbol{M}_{t_r}$, $\boldsymbol{M}_{t_w}$, $\boldsymbol{M}_{p_r}$, and $\boldsymbol{M}_{p_w}$, respectively). The *AO-BLUP* model does not include genomic markers in the regression step on field traits but includes them in the first step, where we remove non-additive genetic and residual noise from

the omics. Even though gains with omics information were low on average, predictions of some environments showed important improvements, whereas just a few environments exhibited small losses (Fig. 4).

As the omics capture a response to the environmental conditions, we also hypothesized that transcriptomics and proteomics could be useful to model GxE interactions, which are important in field trials as they can account for a significant part of phenotypic variance, sometimes even more than the genetic variance (Rogers et al. 2021). In our dataset, the amount of variance explained by GxE is similar to the variance explained by G alone (Millet et al. 2016). We proposed two multi-environment models with omics information, and according to our knowledge, our approach to model interactions between omics and environmental covariates in a multi-environment setting is the first implementation in the literature. In the CV-NH scheme, the *GAOxW-BLUP* model gains slightly compared to the same model without omics (Table 1). The gain compared to *GxW-BLUP* is higher in CV-POH, where *GxW-BLUP* predictive ability stands at 0.72 and *GAOxW-BLUP* at 0.74, with some environments gaining as high as 0.14 units (Fig. 4). If we also compare these results to *GxE-BLUP* and *GAOxE-BLUP*, the improvement with *GAOxW-BLUP* might likely have occurred due to the interactions between additive-omics and ECs.

For CV-NH, the best model was the one based on the additive part of omics and *MegaLMM*, i.e., *MegaGAO*, which was expected since trimming off residual variance and noise from the omics increased correlations with the grain yield (Fig. 5). Upon removing the non-additive part from the omics data, we observed a strong increase in the number of features with an absolute correlation above 0.4. Overall, additive-omics are useful in CV-NH because they offer some phenotypic information about test hybrids, which is better than having no information at all. The performance of *MegaGAO* in CV-POH was lower than expected. *MegaGAO* loses in CV-POH despite unmasked field trials probably because their information gets diluted within the high dimensions of omics information, and the genetic correlations between grain yield trials and omics are lower than among grain yield trials. This case of *MegaGAO* in CV-POH was also true for *MegaGO*.

The scientific community's interest in implementing biological information in genomic prediction models to boost genetic gain is consistently increasing (MacLeod et al. 2016). Our work indicates that omics data can be considerably useful for making predictions in conditions similar to those in which they were measured. In contrast, when predicting field traits with platform omics, the gain was only marginal, and only the additive part of the omics was useful. In comparison to genomic selection, the inclusion of the multi-omics profiles in the breeding process could result in a delay due to the necessity to grow plants to measure omics abundances. However, it is possible to collect high-dimensional omics profiles from the nurseries when double haploid (DH) lines undergo rapid seed multiplication through selfing. Our work lays hypothetical grounds that these high-dimensional omics data could potentially be helpful for field trait predictions. We conclude from our study, that measuring the omics in field conditions is necessary to capture non-additive effects and boost predictive ability. It was for instance proposed in a similar way, that near-infrared data measured in the fields could be useful for predicting non additive effects and in particular GxE interaction (Robert et al. 2022a, 2022b).This may also be possible with molecular omics in the near future thanks to the important decrease of cost that are ongoing.

Conclusion

In this study, we integrated omics information collected from the platform to predict field and platform traits using different univariate and multi-environmental models and a mega-scale multivariate model, i.e., *MegaLMM*. We also introduced a novel multi-environmental model to integrate omics and ECs simultaneously (*GOxW-BLUP*). Platform omics data could dramatically increase predictive ability compared to *GBLUP* for traits measured in similar conditions (platform). Still, the gain was marginal for predicting traits in completely different conditions (fields). The main reason was that platform omics were able to predict non-additive effects only for plants grown in similar conditions. In this case, the non-additive component of omics becomes important due to its ability to capture non-additive genetic variations like epistasis related to the response to environmental conditions. We observed that trimming omics data off their non-additive genetic and residual components for field traits can improve field predictions (especially while predicting new hybrids) because such components cannot explain the field phenotypic variation. We also observed that *MegaLMM* could extract information from additive-omics in predicting field traits, particularly for predicting new hybrids, which were only evaluated for omics information. Finally, in the absence of omics, we confirm the very high predictive abilities of the multi-trait model *MegaLMM* in the sparse testing scenario (CV-POH) due to its ability to capitalize on genetic correlations between traits.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-024-04679-w.

Acknowledgements We thank all the collaborators and funding of the PIA Amaizing (ANR-10-BTBR-01) and the DROPS (FP7-244374) projects for the production of these very rich datasets. We are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale Bioinformatics Facility, https://doi.org/10.15454/1.5572390655343293E12) for providing help and/or computing and/or storage

resources. We thank Melisande Blein-Nicolas, Michel Zivy, and the platform PAPPSO for introducing us to the world of proteomics.

Author contribution BA performed the statistical modeling and analyzes under the supervision of RR, DR, AC, TM-H, and LM. BA wrote the manuscript. RR edited the manuscript. DR supported BA for the MegaLMM analyzes. FC, LCM, SN, FT, CW and HD designed the 3'RNAseq experiment. HD pretreated and mapped the 3' RNAseq data, BH-B and ML treated and analyzed the 3' RNAseq data under the supervision of SN, RR, and FC. All authors approved the submission of the final version of the manuscript.

Funding We thank Ecole Doctorale Agriculture, alimentation, biologie, environnement et santé (ED ABIES) for the funding of B. Ali's PhD. We also thank SPS – Sciences des Plantes de Saclay and ED ABIES for providing mobility funding for B. Ali's research internship in D. Runcie's laboratory at the University of California, Davis. D. Runcie was supported by Agriculture and Food Research Initiative Grants Nos. 2020-67013-30904 and 2018-67015-27957 from the USDA National Institute of Food and Agriculture and by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA), Hatch Project 1010469. The 3'RNAseq experiment was funded by INRAE within the Selgen metaprogram. We thank all the collaborators and funding of the PIA Amaizing (ANR-10-BTBR-01) and the DROPS (FP7-244374) projects for the production of these very rich datasets.

Data availability Field grain yield and environmental covariates can be found at: https://doi.org/10.1104/pp.16.00621. Platform phenotypic data are available online using the PHIS information system (Neveu et al. 2019) at http://www.phis.inra.fr/openphis/web/index.php?r=site%2Flogin-as-guest. From this webpage, data are accessible by clicking on Experimental organization > Project > Systems genetics for maize drought tolerance (Amaizing project). For proteomics data, The raw MS output files were submitted to PROTICdb (http://moulon.inra.fr/protic/amaizing) (Ferry-Dumazet et al. 2005; Langella et al. 2007, 2013), the MassIVE database (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) under accession number MSV000085594, and the ProteomeXchange database (http://www.proteomexchange.org) under accession number PXD019804. Detailed information on all peptides and proteins identified in the LC–MS/MS runs, as well as peptide intensities and protein abundances obtained for each sample, are also freely available on PROTICdb at the same URL. The genotyping data can be found at: https://doi.org/10.15454/AEC4BN. Transcriptomics data are available to anyone upon reasonable request. R scripts to run the analysis are available at: https://github.com/rrincent/MegaGO.

Declarations

Conflict of interest The authors declare no conflict of interest or personal relationships that could appear to influence the work reported in this paper.

References

Bernardo R (2010) Breeding for Quantitative Traits in Plants, Vol. 2. Stemma Press, Woodbury, MN, p 390

Blein-Nicolas M, Negro SS, Balliau T, Welcker C, Cabrera-Bosquet L, Nicolas SD, Charcosset A, Zivy M (2020) A systems genetics approach reveals environment-dependent associations between SNPs, protein coexpression, and drought-related traits in maize. Genome Res 30:1593–1604

Bouidghaghen J, Moreau L, Beauchêne K, Chapuis R, Mangel N, Cabrera-Bosquet L, Welcker C, Bogard M, Tardieu F (2023) Robotized indoor phenotyping allows genomic prediction of adaptive traits in the field. Nat Commun 14:6603

Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. Crop Sci 52:707–719

Butler D, Cullis B, Gilmour A, Gogel B, Thompson R (2009) ASReml user guide, release 3.0. VSN International, Hemel Hempstead

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F (2016) High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. New Phytol 212:269–281

Calus MP, Veerkamp RF (2011) Accuracy of multi-trait genomic selection using different methods. Genet Sel Evol 43:1–14

Christensen OF, Börner V, Varona L, Legarra A (2021) Genetic evaluation including intermediate omics features. Genetics 219:iyab130

Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS ONE 11:e0156744

Crossa J, Montesinos-López O, Pérez-Rodríguez P, Costa-Neto G, Fritsche-Neto R, Ortiz R, Martini JW, Lillemo M, Montesinos-Lopez A, Jarquin D (2022) Genome and environment based prediction models and methods of complex traits incorporating genotype× environment interaction. Methods Mol Biol (clifton, NJ) 2467:245–283

Dahl A, Iotchkova V, Baud A, Johansson Å, Gyllensten U, Soranzo N, Mott R, Kranis A, Marchini J (2016) A multiple-phenotype imputation method for genetic studies. Nat Genet 48:466–472

deAbreueLima F, Westhues M, Cuadros-Inostroza Á, Willmitzer L, Melchinger AE, Nikoloski Z (2017) Metabolic robustness in young roots underpins a predictive model of maize hybrid performance in the field. Plant J 90:319–329

Fernandes SB, Dias KO, Ferreira DF, Brown PJ (2018) Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. Theor Appl Genet 131:747–755

Forni S, Aguilar I, Misztal I (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genet Sel Evol 43:1

Fu J, Falke KC, Thiemann A, Schrag TA, Melchinger AE, Scholten S, Frisch M (2012) Partial least squares regression, support vector machine regression, and transcriptome-based distances for prediction of maize hybrid performance with gene expression data. Theor Appl Genet 124:825–833

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE 6:e28334

Hammer GL, van Oosterom E, McLean G, Chapman SC, Broad I, Harland P, Muchow RC (2010) Adapting APSIM to model the physiology and genetics of complex adaptive traits in field crops. J Exp Bot 61:2185–2202

Henderson C, Quaas R (1976) Multiple trait evaluation using relatives' records. J Anim Sci 43:1188–1197

Heslot N, Akdemir D, Sorrells ME, Jannink J-L (2014) Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor Appl Genet 127:463–480

Hu H, Gutierrez-Gonzalez JJ, Liu X, Yeats TH, Garvin DF, Hoekenga OA, Sorrells ME, Gore MA, Jannink J-L (2020) Heritable temporal gene expression patterns correlate with metabolomic seed content in developing hexaploid oat seed. Plant Biotechnol J 18:1211–1222

Hu H, Campbell MT, Yeats TH, Zheng X, Runcie DE, Covarrubias-Pazaran G, Broeckling C, Yao L, Caffe-Treml M, La G (2021) Multi-omics prediction of oat agronomic and seed nutritional traits across environments and in distantly related populations. Theor Appl Genet 134:4043–4054

Jarquín D, Crossa J, Lacaze X, Du Cheyron P, Daucourt J, Lorgeou J, Piraux F, Guerreiro L, Pérez P, Calus M (2014) A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor Appl Genet 127:595–607

Jia Y, Jannink J-L (2012) Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192:1513–1522

Jiang W, Liu Y, Zhang C, Pan L, Wang W, Zhao C, Zhao T, Li Y (2024) Identification of major QTLs for drought tolerance in soybean, together with a novel candidate gene, GmUAA6. J Exp Bot 75:1852–1871

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-Y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354

Kilasi NL, Singh J, Vallejos CE, Ye C, Jagadish SK, Kusolwa P, Rathinasabapathi B (2018) Heat stress tolerance in rice (*Oryza sativa* L.): Identification of quantitative trait loci and candidate genes for seedling growth under heat stress. Front Plant Sci 9:1578

Lee SH, Van der Werf JH (2016) MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32:1420–1422

Lee H, Calvin K, Dasgupta D, Krinner G, Mukherji A, Thorne P, Trisos C, Romero J, Aldunce P, Barret K (2023) IPCC, 2023: climate change 2023: synthesis report, summary for policymakers. In: Lee H, Romero J (eds.) Contribution of working groups I, II and III to the 6th assessment report of the intergovernmental panel on climate change [Core Writing Team]. IPCC, Geneva

Li Z, Gao N, Martini JW, Simianer H (2019) Integrating gene expression data into genomic prediction. Front Genet 10:126

MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME (2016) Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. BMC Genom 17:144

Malosetti M, Bustos-Korts D, Boer MP, van Eeuwijk FA (2016) Predicting responses in multiple environments: issues in relation to genotype× environment interactions. Crop Sci 56:2210–2222

Mbuvha R, Boulkaibet I, Marwala T (2020) An automatic relevance determination prior bayesian neural network for controlled variable selection. arXiv preprint arXiv:2001.01765

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucl Acids Res 40:4288–4297

Millet EJ, Welcker C, Kruijer W, Negro S, Coupel-Ledru A, Nicolas SD, Laborde J, Bauland C, Praud S, Ranc N (2016) Genome-wide analysis of yield in Europe: allelic effects vary with drought and heat scenarios. Plant Physiol 172:749–764

Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F (2019) Genomic prediction of maize yield across European environmental conditions. Nat Genet 51:952–956

Monteverde E, Gutierrez L, Blanco P, Pérez de Vida F, Rosas JE, Bonnecarrère V, Quero G, McCouch S (2019) Integrating molecular markers and environmental covariates to interpret genotype by environment interaction in rice (*Oryza sativa* L.) grown in subtropical areas. G3: Genes. Genom, Genet 9:1519–1531

Negro SS, Millet EJ, Madur D, Bauland C, Combes V, Welcker C, Tardieu F, Charcosset A, Nicolas SD (2019) Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biol 19:1–22

Nguyen VH, Morantte RIZ, Lopena V, Verdeprado H, Murori R, Ndayiragije A, Katiyar SK, Islam MR, Juma RU, Flandez-Galvez H (2023) Multi-environment genomic selection in rice elite breeding lines. Rice 16:7

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. Theor Appl Genet 113:809–819

Pérez P, de los Campos G, (2014) BGLR: a statistical package for whole genome regression and prediction. Genetics 198:483–495

Prado SA, Cabrera-Bosquet L, Grau A, Coupel-Ledru A, Millet EJ, Welcker C, Tardieu F (2018) Phenomics allows identification of genomic regions affecting maize stomatal conductance with conditional effects of water deficit and evaporative demand. Plant, Cell Environ 41:314–326

Pszczola M, Veerkamp R, De Haas Y, Wall E, Strabel T, Calus M (2013) Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. Animal 7:1759–1768

R Core Team R (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet 44:217–220

Rincent R, Malosetti M, Ababaei B, Touzy G, Mini A, Bogard M, Martre P, Le Gouis J, Van Eeuwijk F (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. Theor Appl Genet 132:3399–3411

Robert P, Brault C, Rincent R, Segura V (2022a) Phenomic Selection: A New and Efficient Alternative to Genomic Selection (GS). In: Ahmadi N, Bartholomé J (eds) Genomic Prediction of Complex Traits: Methods and Protocols. Springer US, New York, NY, pp 397–420

Robert P, Goudemand E, Auzanneau J, Oury F-X, Rolland B, Heumez E, Bouchet S, Caillebotte A, Mary-Huard T, Le Gouis J (2022b) Phenomic selection in wheat breeding: prediction of the genotype-by-environment interaction in multi-environment breeding trials. Theor Appl Genet 135:3337–3356

Robert P, Le Gouis J, Consortium B, Rincent R (2020) Combining crop growth modeling with trait-assisted prediction improved the prediction of genotype by environment interactions. Front Plant Sci 11:827

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Rodriguez-Alvarez MX, Boer MP, van Eeuwijk FA, Eilers PH (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat Stat 23:52–71

Rogers AR, Dunne JC, Romay C, Bohn M, Buckler ES, Ciampitti IA, Edwards J, Ertl D, Flint-Garcia S, Gore MA, Graham C, Hirsch CN, Hood E, Hooker DC, Knoll J, Lee EC, Lorenz A, Lynch JP, McKay J, Moose SP, Murray SC, Nelson R, Rocheford T, Schnable JC, Schnable PS, Sekhon R, Singh M, Smith M, Springer N, Thelen K, Thomison P, Thompson A, Tuinstra M, Wallace J, Wisser RJ, Xu W, Gilmour AR, Kaeppler SM, De Leon N, Holland JB (2021) The importance of dominance and genotype-by-environment interactions on grain yield variation in a large-scale public cooperative maize experiment. G3 Genes|Genomes|Genetics 11:jkaa050

Runcie DE, Mukherjee S (2013) Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. Genetics 194:753–767

Runcie DE, Qu J, Cheng H, Crawford L (2021) MegaLMM: mega-scale linear mixed models for genomic predictions with thousands of traits. Genom Biol 22:1–25

Schrag TA, Westhues M, Schipprack W, Seifert F, Thiemann A, Scholten S, Melchinger AE (2018) Beyond genomic prediction: combining different types of omics data can improve prediction of hybrid performance in maize. Genetics 208:1373–1385

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. BMC Genom 15:1–15

van Rossum B-J, van Eeuwijk F, Boer M (2023) statgenSTA: single trial analysis (STA) of Field Trials. R Package version 1.0.12

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Ward J, Rakszegi M, Bedő Z, Shewry PR, Mackay I (2015) Differentially penalized regression to predict agronomic traits from metabolites and markers in wheat. BMC Genet 16:1–7

Wen J, Jiang F, Weng Y, Sun M, Shi X, Zhou Y, Yu L, Wu Z (2019) Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol 19:1–17

Westhues M, Schrag TA, Heuer C, Thaller G, Utz HF, Schipprack W, Thiemann A, Seifert F, Ehret A, Schlereth A (2017) Omics-based hybrid prediction in maize. Theor Appl Genet 130:1927–1939

Westhues CC, Simianer H, Beissinger TM (2022) learnMET: an R package to apply machine learning methods for genomic prediction using multi-environment trial data. G3 12:226

Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11:407–409