



# Multi-Trait Genomic Prediction Models Enhance the Predictive Ability of Grain Trace Elements in Rice

Blaise Pascal Muvunyi<sup>1</sup>, Wenli Zou<sup>1</sup>, Junhui Zhan<sup>1</sup>, Sang He<sup>1\*</sup> and Guoyou Ye<sup>1,2\*</sup>

<sup>1</sup>CAAS-IRRI Joint Laboratory for Genomics-Assisted Germplasm Enhancement, Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, <sup>2</sup>Rice Breeding Innovations Platform, International Rice Research Institute, Los Baños, Philippines

## OPEN ACCESS

### Edited by:

Muhammad Sajjad,  
COMSATS University, Pakistan

### Reviewed by:

Andrés J. Cortés,  
Colombian Corporation for Agricultural  
Research (AGROSAVIA), Colombia  
Huihui Li,  
Institute of Crop Sciences (CAAS),  
China

### \*Correspondence:

Sang He  
hesang@caas.cn  
Guoyou Ye  
g.ye@irri.org

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

Received: 25 February 2022

Accepted: 05 May 2022

Published: 22 June 2022

### Citation:

Muvunyi BP, Zou W, Zhan J, He S and  
Ye G (2022) Multi-Trait Genomic  
Prediction Models Enhance the  
Predictive Ability of Grain Trace  
Elements in Rice.  
Front. Genet. 13:883853.  
doi: 10.3389/fgene.2022.883853

Multi-trait (MT) genomic prediction models enable breeders to save phenotyping resources and increase the prediction accuracy of unobserved target traits by exploiting available information from non-target or auxiliary traits. Our study evaluated different MT models using 250 rice accessions from Asian countries genotyped and phenotyped for grain content of zinc (Zn), iron (Fe), copper (Cu), manganese (Mn), and cadmium (Cd). The predictive performance of MT models compared to a traditional single trait (ST) model was assessed by 1) applying different cross-validation strategies (CV1, CV2, and CV3) inferring varied phenotyping patterns and budgets; 2) accounting for local epistatic effects along with the main additive effect in MT models; and 3) using a selective marker panel composed of trait-associated SNPs in MT models. MT models were not statistically significantly ( $p < 0.05$ ) superior to ST model under CV1, where no phenotypic information was available for the accessions in the test set. After including phenotypes from auxiliary traits in both training and test sets (MT-CV2) or simply in the test set (MT-CV3), MT models significantly ( $p < 0.05$ ) outperformed ST model for all the traits. The highest increases in the predictive ability of MT models relative to ST models were 11.1% (Mn), 11.5% (Cd), 33.3% (Fe), 95.2% (Cu) and 126% (Zn). Accounting for the local epistatic effects using a haplotype-based model further improved the predictive ability of MT models by 4.6% (Cu), 3.8% (Zn), and 3.5% (Cd) relative to MT models with only additive effects. The predictive ability of the haplotype-based model was not improved after optimizing the marker panel by only considering the markers associated with the traits. This study first assessed the local epistatic effects and marker optimization strategies in the MT genomic prediction framework and then illustrated the power of the MT model in predicting trace element traits in rice for the effective use of genetic resources to improve the nutritional quality of rice grain.

**Keywords:** rice, trace elements, multi-trait genomic prediction, local epistatic effect, selective marker panel

## INTRODUCTION

Over half of the world's population relies on rice as a staple crop (Bandumula, 2018). Growing and consuming rice has relative merits, as rice is the major dietary source for both toxic and essential trace elements (Yang et al., 2018). For instance, Cd is a potent environmental and human health toxicant (Arao and Ae, 2003; Uruguchi and Fujiwara, 2012; Lien et al., 2021) transported into rice grain *via* the

same channels with other trace elements Zn, Fe, Cu, and Mn (Sasaki et al., 2012; Hao et al., 2018; Han et al., 2021) of essential nutritional and physiological functions to plants, animal and humans species (Miller, 1970; Olivares and Uauy, 1996; White and Broadley, 2009; Aschner and Erikson, 2017; Gao and Xiong, 2018).

Recent advancements in genomic research and the increasing number of germplasm resources in gene banks offer a great opportunity to develop safe and nutritious rice varieties cost-effectively. The trait's heritability indicates the potential that a given trait can be genetically improved. Previously, broad sense heritability of grain Zn, Fe, Cu, Mn, and Cd was found to be low (0–0.3), moderate (0.4–0.6) to high (0.6 or higher) (Norton et al., 2010; Pinson et al., 2015; Naik et al., 2020), indicating the practical possibility to improve these traits via selective breeding methods. Furthermore, many molecular genetic studies have identified numerous quantitative trait loci (QTLs) responsible for trace element uptake, transport, and accumulation into different rice tissues through genome-wide association studies (GWAS) or QTL mapping (Lu et al., 2008; Garcia-Oliveira et al., 2009; Ueno et al., 2010; Du et al., 2013; Huang et al., 2016; Meng et al., 2017; Swamy et al., 2018; Yang et al., 2018; Descalsota-Empleo et al., 2019; Frouin et al., 2019; Liu J. et al., 2021). As a genomics-enabled breeding approach, marker-assisted selection (MAS) is useful to improve trace element traits when genes/QTLs with large additive genetic effects exist (Wu et al., 2020). However, prominent non-additive gene action has also been reported for trace element traits, making MAS-based strategies ineffective (Sharma V. et al., 2021). In addition, MAS-based breeding methods are practically ineffective at simultaneously exploiting information from multiple genes (Spindel et al., 2016) or traits (Van Der Straeten et al., 2020).

In contrast, genomic selection (GS) approaches make use of total genome-wide markers with either large additive effects or minor effects to derive the genomic estimated genetic values of genotypes (Meuwissen et al., 2001), which overcomes the constraints of MAS-based methods (de los Campos et al., 2009). Also, GS models can be modified to a multi-trait (MT) form to exploit available information from multiple traits simultaneously. The MT models used in GS heavily rely on genetic correlation between traits (Henderson and Quaas, 1976). This correlation possibly results from the pleiotropic effect (multiple traits controlled by the same QTL) or linkage disequilibrium (LD) between genes (Falconer, 1996). Exploiting multi-trait information in GS has been awarded with an increase in prediction accuracy ranging from 24% to 105% relative to single trait (ST) models (Rutkoski et al., 2016; Sun et al., 2017; Arojju et al., 2020). Besides gains in prediction accuracy, integrating MT models with appropriate cross-validation (CV) schemes compensated for the negative effect of small population size without affecting the prediction accuracy, enabling breeders to minimize phenotyping budgets (Lado et al., 2018; Arojju et al., 2020). The benefits of MT models under various CV schemes are yet to be studied in diverse rice collections. Nevertheless, MT models have shown their potential in

predicting complex traits in rice, such as grain arsenic content (Ahmadi et al., 2021), grain yield (Wang et al., 2017), and root index architecture (Sharma S. et al., 2021).

Most of the MT genomic prediction studies discussed above only modeled the additive genetic effects. Non-additive effects are also essential components of the genetic effect and can benefit the predictive ability of MT models if accommodated (dos Santos et al., 2016; Lyra et al., 2017). However, non-additive effects such as dominance or global epistatic effects may not be conserved during breeding due to chromosomal recombination events (Falconer, 1996; Breseghello and Sorrells, 2006; He et al., 2017). In contrast, the local epistasis that spans short segments of chromosomes can be preserved over generations (Akdemir and Jannink, 2015), as adjacent loci normally hold a strong LD (Ardlie et al., 2002). Earlier GS studies with ST models illustrated that accounting for local epistatic effects along with the main additive model increased the prediction accuracy of agronomic traits in wheat accessions (Akdemir and Jannink, 2015; Akdemir et al., 2017; He et al., 2017; Jiang et al., 2018; He et al., 2019). However, the benefits of modeling local epistasis effects in MT models remain unknown in crop or animal species.

Genomic prediction models can be extended to incorporate markers associated with causal QTLs, such as trait-associated SNPs (TA-SNPs), bridging the gap between biology and mechanistic GS models using uninformative genome-wide markers. Also, genomic prediction with markers derived from functional QTL is less reliant on LD patterns shared by training and target populations, possibly allowing robust prediction, especially across unrelated populations where LD decays more rapidly (Snelling et al., 2013). Simulation and empirical studies have shown that accounting for known QTLs improves the performance of genomic prediction models compared to models using uninformative genome-wide markers (Bernardo, 2014; Owens et al., 2014). Alemu et al. (2021); Zhou et al. (2021) reported a two- to four-fold gain in prediction accuracy using GS + *de novo* GWAS (Spindel et al., 2016), in which the most significant TA-SNPs from a GWAS conducted on the training population are fitted as fixed effects in the model along with the polygenic background. Other groups (Bhandari et al., 2019; Ahmadi et al., 2021) also reported gains in prediction accuracy ranging from 16% to 32% by exploiting GWAS-derived TA-SNPs using trait-specific genomic relationship matrices (Zhang et al., 2014) in which markers with stronger association signals are assigned higher weights than markers with weaker associations. However, the application of the above methods has not always been beneficial (Veerkamp et al., 2016; Rice and Lipka, 2019) and has been shown to depend on the genetic architecture of the traits of interest, trait heritability, the number of underlying causal mutations, and their effect sizes (Huang and Mackay, 2016). In addition, the use of TA-SNPs in genomic prediction has been scarcely investigated in models accounting for the non-additive effects. The potential of GWAS-derived TA-SNPs on the predictive ability of MT models accounting for the local epistatic effects in diverse rice populations is yet to be demonstrated and worth inspecting.

There is a great scope for applying MT models to evaluate trace elements in large germplasm collections such as those archived in gene banks. Earlier studies using ST models showed that GS is a robust and cost-efficient tool to predict the genetic merit of individuals in large germplasm collections for various agronomic traits, such as grain yield in rice (Tanaka et al., 2021), biomass yield in sorghum (Yu et al., 2016), oil, protein, and yield in soybean (Jarquin et al., 2016), total root length in maize (Pace et al., 2015), and days to head and days to maturity in wheat (Crossa et al., 2017). However, the application of MT or even ST models to predict the concentrations of trace elements in food crops is still limited to a few studies involving arsenic (Frouin et al., 2019; Ahmadi et al., 2021), Mn (Leplat et al., 2016), and Zn (Guo et al., 2020) in rice, barley, and maize grain, respectively. Therefore, the overall goal of the present study is to compare the robustness of ST and MT models in predicting concentrations of four essential trace elements, Fe, Zn, Cu, and Mn, and one toxic metal, Cd, in rice grain. Different CV schemes, implying varied phenotyping patterns and costs, were examined in our study to seek the most efficient phenotyping strategy when multiple traits are planned to be measured. In addition, we investigated whether incorporating local epistatic effects and using a selective marker panel of TA-SNPs derived from GWAS into MT models could further enhance the predictive ability of MT models.

## MATERIALS AND METHODS

### Rice Materials

Our study used 250 rice accessions, including *indica* and *japonica* ecotype accessions from Asian countries (Supplementary Table S1). Accessions from China are mainly landrace *indica* varieties mostly cultivated on Cd-polluted soils in Guangdong province, China (Long et al., 2014).

### Plant Cultivation and Quantification of Trace Elements in Rice Grain

The procedures followed for growing the 250 accessions and determining concentrations of trace elements in rice grain were as previously described by Liu S. et al. (2021). Briefly, seeds from the 250 accessions were first cultivated in pots filled with soil collected from the experimental station of the Agricultural Genomics Institute at Shenzhen, China. Next, germinated seeds were selected and cultivated in seedling trays for 4 weeks. Healthy seedlings were then transferred into pots containing soil amended with an initial concentration of Cd of 0.5 mg kg<sup>-1</sup>. Finally, all the seedlings were planted in an augmented randomized complete block design with two replicates of 25 accessions from 20 July 2019 to 2 October 2019. Our study was limited to a single environment. Multi-environment data would be essential for understanding the environmental correlations and their stability and genotype effects by environment interactions (GxE). To

determine grain concentrations of Zn, Fe, Cu, Mn, and Cd, grain samples were first peeled and dried at 65°C for 3 days. The dried samples were then crushed, wet-digested in concentrated nitric acid (HNO<sub>3</sub>) at 120°C for 30 min, and further digested with perchloric acid (HClO<sub>4</sub>) at 180°C until the samples became transparent. The samples were then diluted with ultrapure water. Finally, the grain concentration of each trace element was determined using the Inductively Coupled Plasma Mass Spectrometry (ICP-MS) machine (Houk et al., 1980).

### Genotyping

The 250 accessions were genotyped following the re-sequencing and variants-calling procedures of the rice 3K project as reported by Wang et al. (2018). The following steps were implemented for all the genotypes to merge the variants-calling: First, raw reads were aligned to the R498 reference genome (Du et al., 2017) using the program bwamem alignment software (Li, 2013). Next, the PCR duplicates were identified with Picard software, version 2.9.0 (<http://broadinstitute.github.io/picard/>), and discarded. Following that, the GATK HaplotypeCaller engine (McKenna et al., 2010), with the option “-ERC GVCF,” was used to call genotypes at each site. The resulting genomic variants called format (gVCF) for each genotype were combined using the GATK Genotype GVCFs engine. Next, the GATK hard filter pipeline was used to individually call SNP and INDEL variants from the population variant file. All the variants within 5 bp of an INDEL were discarded. A variant was confirmed if at least one genotype supported it with a QUAL parameter greater than 30. After that, VCF tools indicated (Danecek et al., 2011) sites for which genotypes were not called in at least 20% of the used genotypes. The above procedures yielded 30,089,814 bi-allelic SNPs for the 250 genotypes. SNP quality control steps were implemented using PLINK software (Chang et al., 2015) with standards that remove SNPs with 1) minor allele frequency lower than 0.5, 2) call rate less than 0.9, and 3) pairwise LD ( $r^2$ ) greater than 0.01. Finally, 36,171 SNPs were available for the 250 accessions.

### Estimation of Genomic Heritability and Traits Genomic Correlation

The mixed linear model was used to estimate genomic heritability as follows:

$$y = \mathbf{1}_n\mu + \mathbf{g} + \mathbf{e}$$

Where  $y$  is the vector of concentration of trace element under consideration,  $\mathbf{1}_n$  is the vector of ones,  $n$  is the number of genotyped cultivars,  $\mu$  is the intercept,  $\mathbf{g}$  is the vector of the genetic effects of accessions,  $\mathbf{e}$  is the residual vector.  $\mathbf{g}$  and  $\mathbf{e}$  are assumed as random effects, respectively, following  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_g^2)$  and  $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$  where  $\mathbf{G}$  is genomic relationship matrix estimated following (Yang et al., 2010) and  $\mathbf{I}$  is identity matrix. The genomic heritability was estimated as

$$h_g^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}$$

Where  $\sigma_g^2$  is the additive genetic variance component and  $\sigma_e^2$  is the residual variance component. GCTA software (Yang et al., 2011) was used to compute the genomic relationship matrix and genomic heritability.

The genomic correlation between traits was estimated using the formula:  $cor = \frac{cov_{i,j}}{\sigma_i \sigma_j}$  where  $cov_{i,j}$  is the genetic covariance between  $i^{th}$  trait and  $j^{th}$  trait,  $\sigma_i$  and  $\sigma_j$  are the square root of the genetic variance of  $i^{th}$  trait and  $j^{th}$  trait. The genetic covariances and variances were estimated using the R package MTM (De los Campos and Grüneberg, 2016).

### Genetic Diversity

A hierarchical cluster analysis based on the Euclidean's distance matrix computed with the SNP genomic profiles was performed to inspect the genetic diversity among the 250 genotypes. In addition, a heat plot based on the cluster analysis was drawn to visualize the genetic dissimilarities.

### Genomic Prediction Approaches

The genomic prediction models used were the ST and MT models. The ST model only captured additive genetic effects, while the MT models accommodated both additive and local epistatic effects. ST model was the commonly used genomic best linear unbiased prediction (GBLUP) model and same as the mixed model estimating genomic heritability:  $y = \mathbf{1}_n \mu + \mathbf{g} + \mathbf{e}$ , where  $y$ ,  $\mathbf{1}_n$ ,  $\mu$  and  $\mathbf{e}$  were exactly as afore denoted,  $\mathbf{g}$  is the vector of additive genetic effect in genotype-based model or additive plus local epistatic effects in the haplotype-based model. In the genotype-based model, we assumed  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{G}\sigma_a^2)$ , where  $\mathbf{G}$  is an  $n \times n$ -dimensional additive genomic relationship matrix, and  $\sigma_a^2$  is the additive genetic variance component. In the haplotype-based model, we assumed  $\mathbf{g} \sim N(\mathbf{0}, \mathbf{H}\sigma_h^2)$ , where  $\mathbf{H}$  is an  $n \times n$ -dimensional haplotypic relationship matrix derived from the haplotypic profile matrix with values 0, 1, and 2 indicating the number of copies for a specific haplotypic allele (Jiang et al., 2018; He et al., 2019). To obtain the haplotypic profile matrix, the genotypic data of SNPs were phased using software SHAPEIT (Delaneau et al., 2012) with default augment settings. The phased genotypic data was recoded to haplotypic profiles using a fixed-length haplotype of 2, 3, 4, or 5 SNPs.  $\mathbf{G}$  and  $\mathbf{H}$  were established using software GCTA (Yang et al., 2010; Yang et al., 2011) based on the genotypic and haplotypic data, respectively.

For MT models, we used two approaches considering no correlation between traits; the Bayesian multi-output regressor stacking (BMORS) proposed by Montesinos-López et al. (2019); the MT-GBLUP; and two methods accommodating correlation between traits; the factorial analytic (FA) model and the unstructured (UN) model. For genotype-based approaches, the MT-GBLUP model was formulated as  $\mathbf{y} = \mu + \mathbf{u} + \mathbf{e}$ , where

$\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_m)'$ ,  $\mu = (\mu'_1, \mu'_2, \dots, \mu'_m)'$ ,  $\mathbf{u} = (\mathbf{g}'_1, \mathbf{g}'_2, \dots, \mathbf{g}'_m)'$ ,  $\mathbf{e} = (\mathbf{e}'_1, \mathbf{e}'_2, \dots, \mathbf{e}'_m)'$ ,  $m$  is the number of traits included in the model. We assumed  $\mathbf{u} \sim N((\mathbf{0}', \mathbf{0}', \dots, \mathbf{0}')', \Psi_u \otimes \mathbf{G})$ ,  $\mathbf{e} \sim N((\mathbf{0}', \mathbf{0}', \dots, \mathbf{0}')', \Psi_e \otimes \mathbf{I})$

where  $\Psi_u = \begin{pmatrix} \sigma_{g_1}^2 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{g_j}^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \sigma_{g_m}^2 \end{pmatrix}$ ,

$\Psi_e = \begin{pmatrix} \sigma_{\epsilon_1}^2 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{\epsilon_j}^2 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \sigma_{\epsilon_m}^2 \end{pmatrix}$ ,  $\sigma_{g_j}^2$  and  $\sigma_{\epsilon_j}^2$  are respectively the

genetic and residual variance of  $j^{th}$  trait, and  $\otimes$  denotes Kronecker product of matrices.

BMORS was a two-stage process. The first stage is the same as the MT-GBLUP, but instead of directly using the GBLUP predicted values as the final output, BMORS implemented a second stage that integrated the GBLUP predicted values of each trait in the first stage and fitted a ridge regression model. In this way, the prediction of a single trait could be corrected by the predictions of other traits in the first stage using the second stage model (Spyromitros-Xioufis et al., 2012; Spyromitros-Xioufis et al., 2016; Montesinos-López et al., 2019). The FA model was also based on the formula of the MT-GBLUP model but assuming a covariance structure between traits, that is

$$\Psi_u = \begin{pmatrix} \sigma_{g_1}^2 & \dots & cov_{g_1, g_j} & \dots & cov_{g_1, g_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ cov_{g_j, g_1} & \dots & \sigma_{g_j}^2 & \dots & cov_{g_j, g_m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ cov_{g_m, g_1} & \dots & cov_{g_m, g_j} & \dots & \sigma_{g_m}^2 \end{pmatrix} = (\Lambda\Lambda' + \mathbf{I}) = FA(k)$$

where  $k$  is the number of latent factors,  $\Lambda$  is a  $j \times k$  dimensional matrix containing trait loadings,  $\mathbf{I}$  is a  $j \times j$  diagonal matrix (Burgueño et al., 2012). Theoretically, the FA model requires at least three traits to be simultaneously included in the model. The UN model (Burgueño et al., 2012; Cuevas et al., 2017) tried to estimate all variances and covariances in  $\Psi_u$ , i.e.,  $\sigma_{g_j}^2$  and  $cov_{g_i, g_j}$ ,  $i, j \in \{1, \dots, m\}$ , which may cause convergence problems when a large number of traits are considered. The haplotype-based approach was only implemented in the MT-UN model by replacing the relationship matrix  $\mathbf{G}$  by  $\mathbf{H}$ .

The ST-GBLUP model was implemented in R (R Core Team, 2016) using the BGLR package (De los Campos and Pérez-Rodríguez, 2015). The MT-GBLUP, FA, and UN approaches were realized using the R package MTM (De los Campos and Grüneberg 2016). BMORS was fitted using the R package BMTME (Montesinos-López et al., 2019). The number of

**TABLE 1** | Investigated single trait (ST) and multi-trait (MT) cross-validation (CV) schemes.

	Training set		Test set	
	Target traits	Auxiliary traits	Target traits	Auxiliary traits
ST-CV1	<b>Zn/Cu/Fe/Cd/Mn</b>	—	<b>Zn/Cu/Fe/Cd/Mn</b>	—
MT-CV1	<b>Zn</b>	Mn, Fe, Cu, Cd	<b>Zn</b>	
	<b>Cu</b>	Mn, Fe, Zn, Cd	<b>Cu</b>	
	<b>Fe</b>	Mn, Cu, Zn, Cd	<b>Fe</b>	
	<b>Cd</b>	Mn, Fe, Cu, Zn	<b>Cd</b>	
	<b>Mn</b>	Fe, Cu, Zn, Cd	<b>Mn</b>	
MT-CV2	<b>Zn</b>	Mn, Fe, Cu, Cd	<b>Zn</b>	Mn, Fe, Cu, Cd
	<b>Cu</b>	Mn, Fe, Zn, Cd	<b>Cu</b>	Mn, Fe, Zn, Cd
	<b>Fe</b>	Mn, Cu, Zn, Cd	<b>Fe</b>	Mn, Cu, Zn, Cd
	<b>Cd</b>	Mn, Fe, Cu, Zn	<b>Cd</b>	Mn, Fe, Cu, Zn
	<b>Mn</b>	Fe, Cu, Zn, Cd	<b>Mn</b>	Fe, Cu, Zn, Cd
MT-CV3	<b>Zn</b>	—	<b>Zn</b>	Mn, Fe, Cu, Cd
	<b>Cu</b>		<b>Cu</b>	Mn, Fe, Zn, Cd
	<b>Fe</b>		<b>Fe</b>	Mn, Cu, Zn, Cd
	<b>Cd</b>		<b>Cd</b>	Mn, Fe, Cu, Zn
	<b>Mn</b>		<b>Mn</b>	Fe, Cu, Zn, Cd

Phenotypes for auxiliary traits are not available. The unobserved target traits to be predicted are highlighted with bold font.

iterations of all models was set to 20,000, and the first 12,000 were discarded as burn-in.

## Cross-Validation Schemes and Evaluation of Genomic Prediction Accuracy

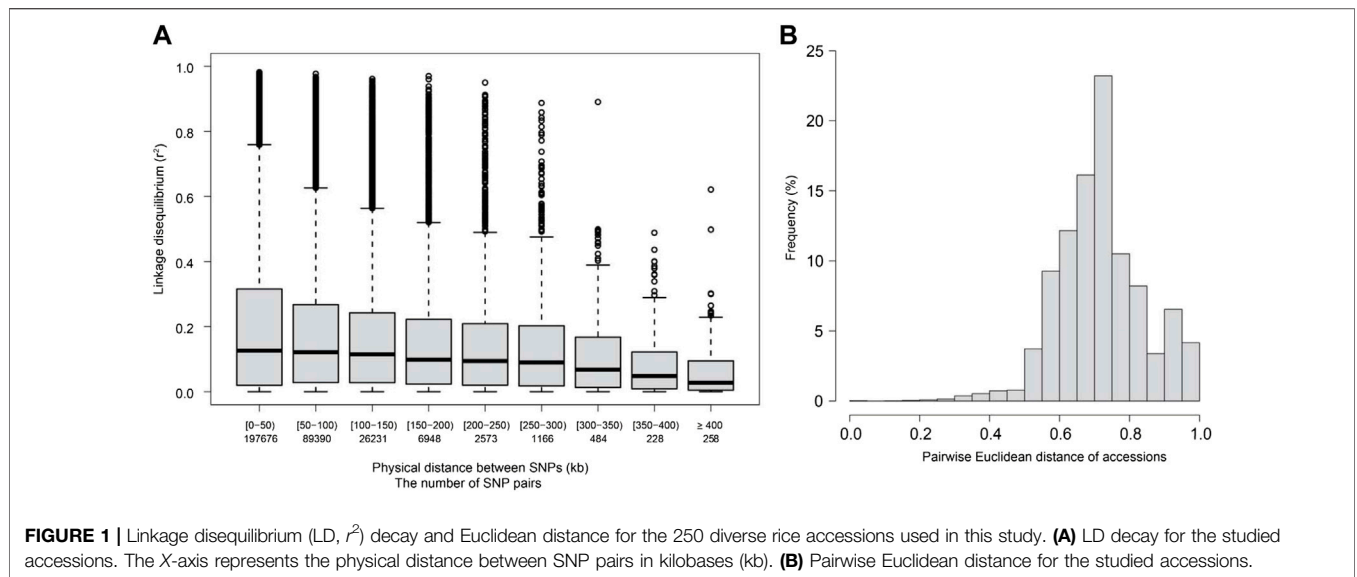
Four different CV schemes, referring to those reported by Lado et al. (2018) and Arojju et al. (2020) were used in our study (Table 1). CV1 was applied to both ST-GBLUP and MT models, referring to a scenario where the target trait was predicted without the support of auxiliary traits (ST-CV1) or with auxiliary traits only available in the training set (MT-CV1). CV2 and CV3 were only assessed for MT models. Under CV2 scheme, genotypes in both training and test sets had phenotypic data for all the auxiliary traits. Under CV3, phenotypes of the auxiliary traits were only available in the test set.

To assess the genomic prediction accuracy across the above CV schemes, the entire population of 250 genotypes was randomly divided into five equal-sized folds. Four folds collectively constituted the training set, and the remaining fold was the test set. Stochastic partitioning of training and test sets was repeated 20 times, yielding one hundred times (5 folds × 20 replicates) calibrations and predictions. The genomic prediction accuracy of the target trait was estimated using the Pearson correlation coefficient between the genomic predicted genetic values and the observed phenotypic values of 250 accessions when incorporated in the five test sets of each repeat of CV. The Student *t*-test was used to test the statistical difference in genomic prediction accuracies among the prediction models.

## Selective SNP Marker Panel

To investigate whether the predictive ability of MT genomic prediction on rice grain trace elements concentration could be

boosted by optimizing the SNP marker panel, we applied a (GWAS) to identify the trait-associated SNPs (TA-SNPs) and establish the selective marker panel. The CV scenarios in which the MT-UN haplotype-based models disregarding the length of haplotypes (two to five SNPs) constantly showed statistically significantly ( $p < 0.05$ , *t*-test) higher prediction accuracies than their genotype-based counterparts and the ST-GBLUP model were used to validate the efficacy of using the TA-SNPs to train the genomic prediction models. In more detail, GWAS using the total SNP marker panel was performed in the training set of the designated CV scenarios. First, the TA-SNPs with *p* values less than 0.01 were recorded. Then each chromosome was divided into bins spanning 300 kb (the bin size is decided by the LD decay, with the physical distance between pairs of SNPs based on the total population). Finally, the most trait-associated SNPs with the lowest *p*-value in each bin was picked together with the TA-SNPs ( $p < 0.01$ ) and recorded to constitute the selective SNP marker panel of each repeat of CV. The MT-UN genotype-based model was implemented using the genotypic data of the selective SNPs. Contrastingly, the adjacent selective SNPs located within 300 kb (highly possible as the position of the selective SNP from each bin is unfixed) were combined to compile the haplotypes using the phased genotypic data as the LD decay implied a non-negligible LD among them. The remaining SNPs without close neighbors within 300 kb were maintained, and their genotypic profiles were used. Therefore, the MT-UN haplotype-based model took advantage of both haplotypes and genotypes. The GWAS was implemented using GCTA software (Yang et al., 2011; Yang et al., 2014). The additive genomic relationship matrix was exclusively used in the GWAS model to account for the relatedness between accessions.



**TABLE 2 |** Genomic heritabilities (diagonal and bold) and genetic correlations (upper triangle) of the trace elements traits studied.

Traits	Cd	Fe	Mn	Cu	Zn
Cd	<b>0.62</b>	0.95	0.39	0.71	0.67
Fe		<b>0.5</b>	0.07	0.79	0.76
Mn			<b>0.35</b>	0.59	0.7
Cu				<b>0.21</b>	0.95
Zn					<b>0.14</b>

Diagonal and bold are genomic heritabilities as indicated.

## RESULTS

### Linkage Disequilibrium Decay, Kinship, and Population Structure

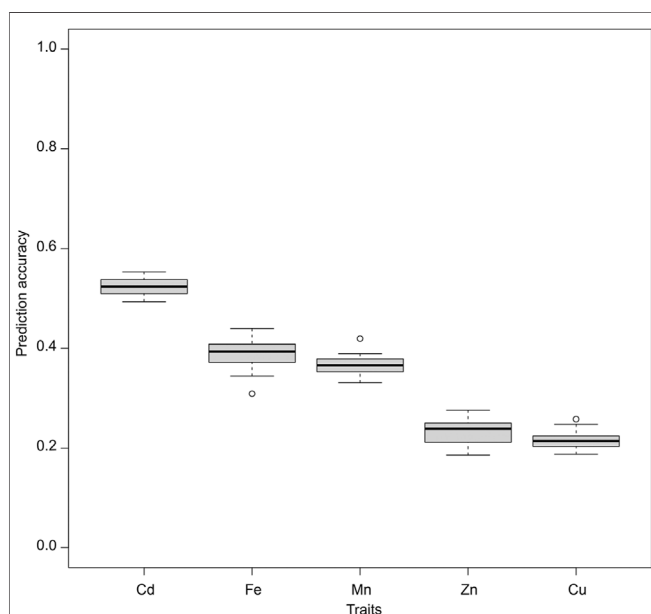
The LD decay distance between all SNP markers for the 250 accessions was ~250–300 kb when the cut-off value ( $r^2$ ) was set at 0.1, assuming non-negligible SNP pairwise correlation ( $r = 0.3$ ) (Figure 1A). The kinship between accessions was determined based on pairwise Euclidean distances. Pairwise distances among accessions ranging from 0 to 0.2 accounted for less than 5% of all the pairwise distances. Pairwise distances from 0.6 to 0.8 were the most frequent and accounted for 12%–23% of all the pairwise distances (Figure 1B). Further, no genetically structured sub-populations were observed among the 250 varieties used in this study. However, several families were detected (Supplementary Figure S1).

### Distribution of Phenotypes, Genomic Heritability, and Genetic Correlation

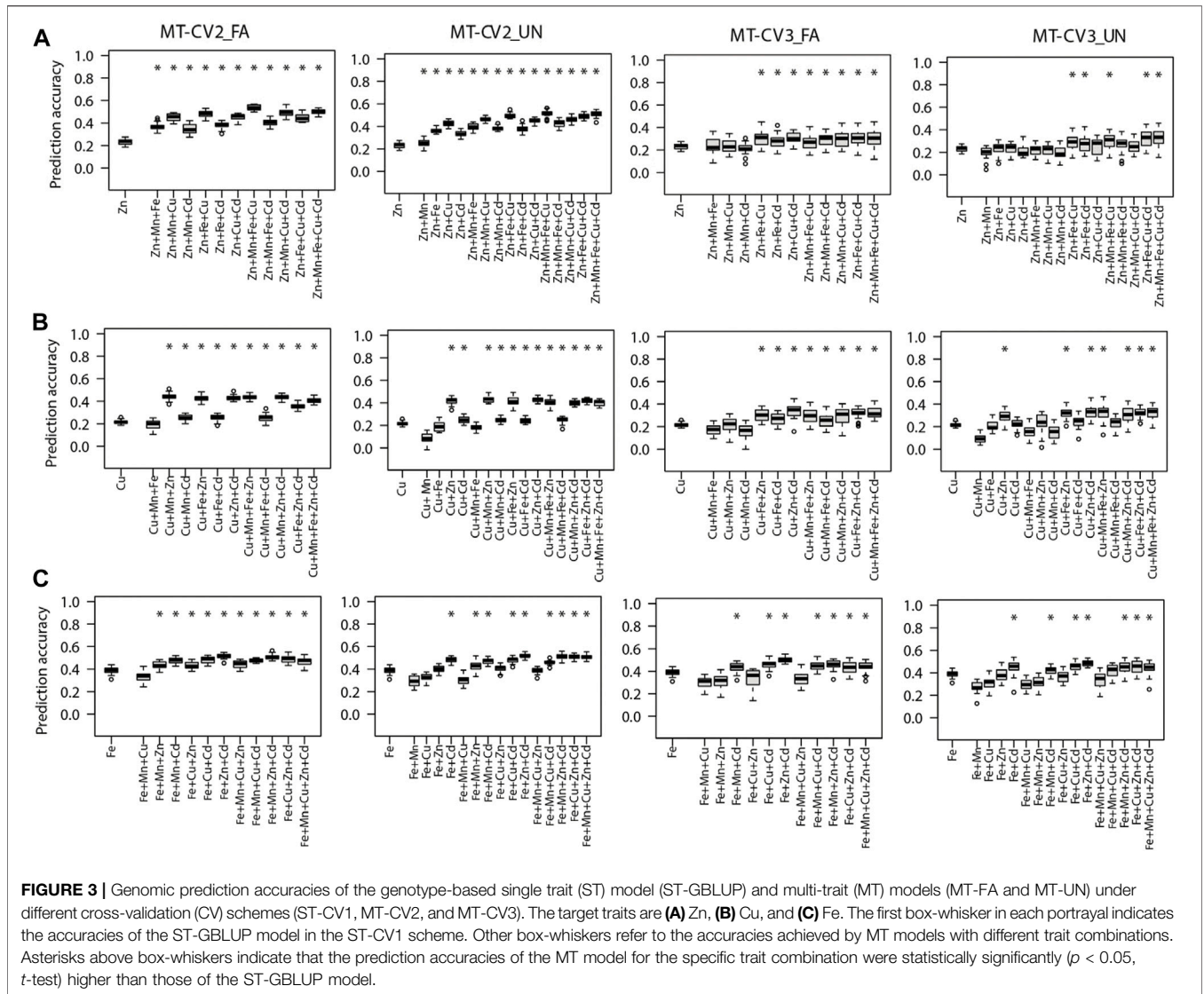
The distributions of phenotypes (adjusted phenotypic means) based on the 250 accessions varied among the five traits studied. The distribution of Zn was almost symmetrical. The skewness was high and negative for Cu, and moderate and negative for Fe, Cd, and Mn (Supplementary Figure S2). The genomic heritability for all studied traits ranged from low (Zn: 0.14 and Cu: 0.21) to medium (Mn: 0.35) and high (Fe: 0.5 and Cd: 0.62) (Table 2). The genetic correlation estimated with the MTM model was highest between Fe and Cd (0.95) and Cu and Zn (0.95) and lowest between Mn and Cd (0.39) and Mn and Fe (0.44). Zn had the highest genetic correlations with all the other studied traits, ranging from 0.67 to 0.95 (Table 2).

### Prediction Accuracy of Single-Trait Model Versus Multi-Trait Model Using Whole-Genome Markers

The average prediction accuracy with the traditional ST-GBLUP model under the CV1 scheme was the highest for



**FIGURE 2 |** Genomic prediction accuracies of the studied traits were assessed using a single trait GBLUP (ST-GBLUP) model.



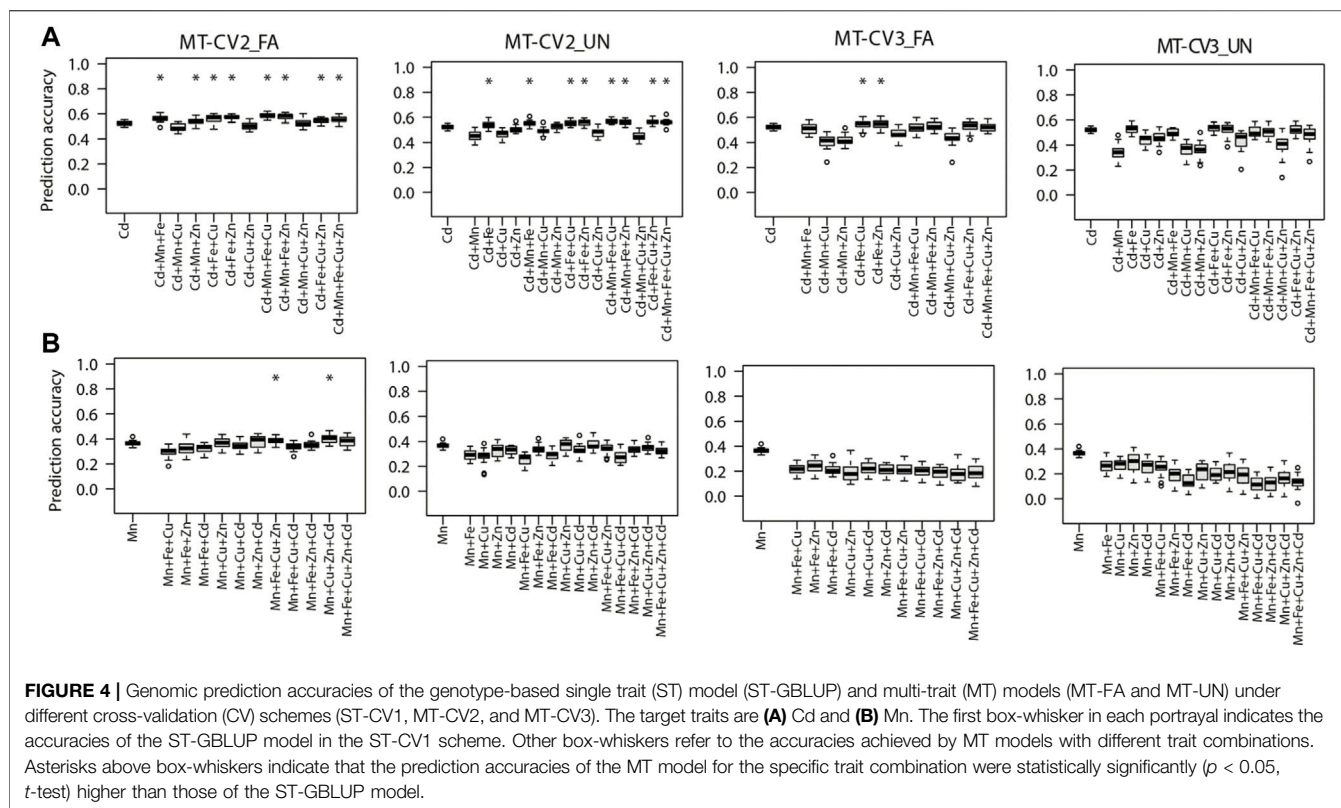
Cd (0.52), followed by Fe (0.39), Mn (0.36), Zn (0.23), and Cu (0.21) (Figure 2). Also, under the CV1 scheme, prediction accuracies of MT models were not statistically significantly ( $p < 0.05$ ) superior to those of ST-GBLUP irrespective of the models and traits studied (Supplementary Tables S2–S6).

As compared, when phenotypes of the auxiliary traits were made available in both training and test sets (MT-CV2) or merely in the test set (MT-CV3), the MT models, namely FA or UN, significantly ( $p < 0.05$ ) outperformed the ST-GBLUP model (Figures 3, 4). For most of the studied traits, the highest performance of MT models was observed under the MT-CV2 scheme (Figures 3, 4). However, MT-GBLUP and BMORS MT models were not significantly ( $p < 0.05$ ) superior to the ST-GBLUP model for all the CV schemes studied (Supplementary Tables S7–S11).

We further compared scenarios where the prediction of the target traits was assisted with a single auxiliary trait or a

combination of multiple auxiliary traits in MT models. Supporting the prediction of Zn with one of its correlated traits (Cu, Fe, Mn, or Cd) was sufficient to significantly ( $p < 0.05$ ) increase the prediction accuracy MT-UN model relative to ST-GBLUP model in MT-CV2 (Figure 3A). Cu was the best single auxiliary trait for predicting Zn. Incorporating observations from Cu in MT-UN model (under MT-CV2) significantly ( $p < 0.05$ ) increased the prediction accuracy of Zn by 82.6% (0.23–0.42) relative to the ST-GBLUP model. However, the highest increase in prediction accuracy (126% or 0.23–0.52) of MT models was observed when observations from Mn, Fe, and Cu were combined as supporting traits for Zn under the MT-CV2 scheme (Figure 3A). Under MT-CV3, the MT-UN model outperformed the ST-GBLUP model only after multiple auxiliary traits were used to support the prediction of Zn (Figure 3A).

Similarly, compared to ST-GBLUP, the prediction accuracy of Cu by the MT-UN model significantly ( $p < 0.05$ ) increased by



95.2% (0.21–0.41) and 38% (0.21–0.29) in MT-CV2 and MT-CV3, respectively, when Zn was used as a single supporting trait (**Figure 3B**). Yet, after including other traits in MT-CV2 (Mn and Zn) and MT-CV3 (Zn and Cd), the prediction accuracy improved by 109.5% (0.21–0.44) and 57.1% (0.21–0.33) relative to ST-GBLUP, respectively.

Similarly, when Fe was the target trait, MT-UN model accounting information from Cd significantly ( $p < 0.05$ ) outperformed ST-GBLUP by 23% (0.39–0.48) and 12.8% (0.39–0.44) under MT-CV2 and MT-CV3 schemes, respectively (**Figure 3C**). Nevertheless, considering phenotypes from more auxiliary traits in MT-CV2 (Zn and Cd) and MT-CV3 (Zn and Cd) provided 33.3% (0.39–0.52) and 23% (0.39–0.48) gains in the prediction accuracy of Fe with the MT-UN model, respectively (**Figure 3C**).

Furthermore, the prediction accuracy of Cd (the most heritable trait) with the MT-UN model was significantly ( $p < 0.05$ ) improved by 3.8% (0.52–0.54) when Fe, its strongly correlated trait, was used as a single auxiliary trait under the MT-CV2 scheme (**Figure 4A**). As observed for the other traits, 7.6% (0.52–0.56) and 11.5% (0.52–0.58) gains in prediction accuracy were attained after using combined information from multiple auxiliary traits (Mn, Fe, and Cu) in MT-UN and MT-FA models under MT-CV2, respectively (**Figure 4A**). Similarly, under MT-CV3, MT models did not significantly outperform ST-GBLUP models in scenarios where a single auxiliary trait was used. However, when information from Fe and Zn or Fe and Cu was considered, an improvement of 5.7% (0.52–0.55) in the prediction

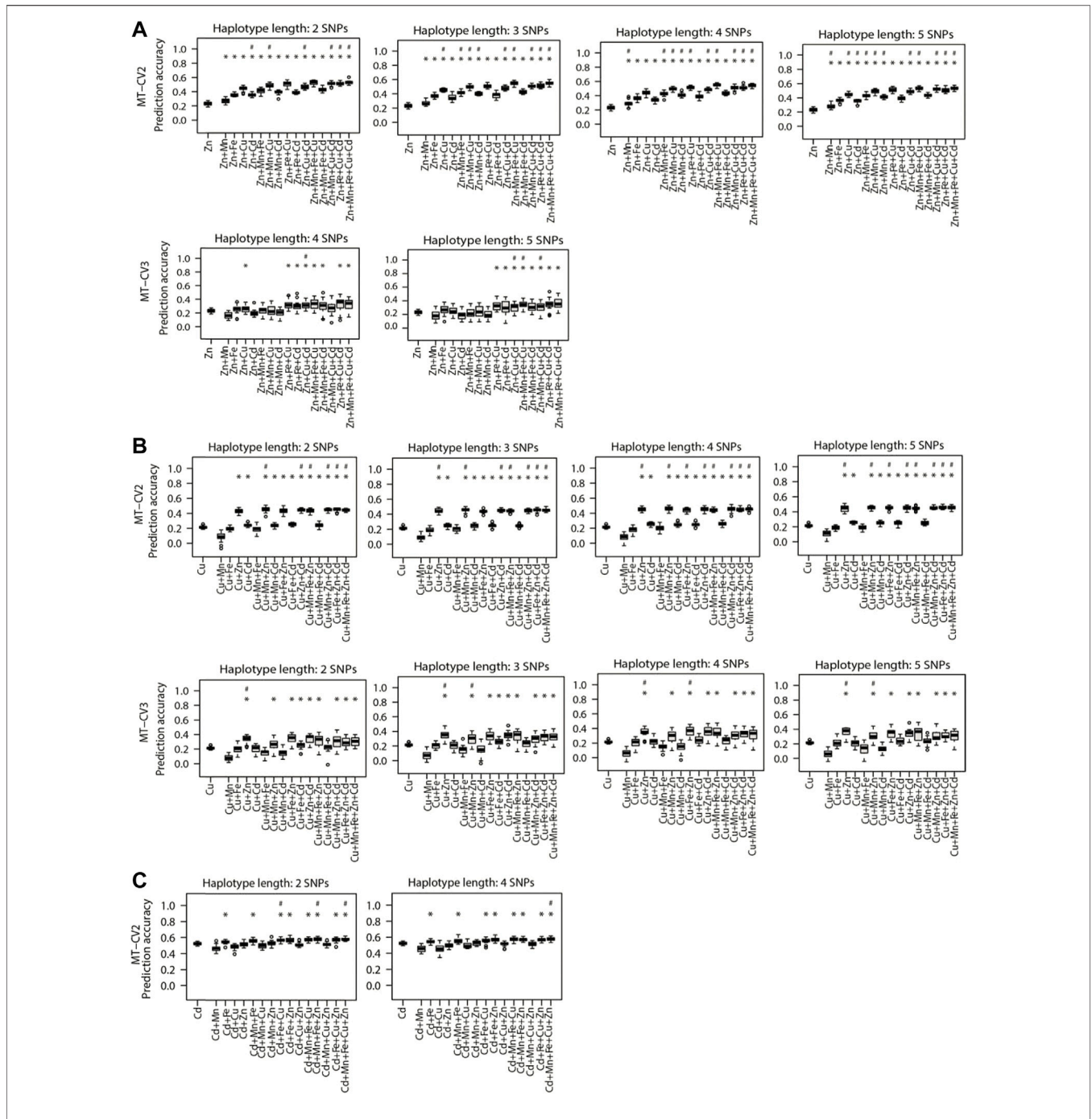
accuracy of the MT-FA model over the ST-GBLUP model was observed (**Figure 4A**).

Finally, when Mn was the target trait, MT-UN with a single auxiliary trait failed to improve its prediction accuracy under both MT-CV2 and MT-CV3 schemes (**Figure 4B**). However, considering information from additional traits (Cu, Zn, and Cd), using the MT-FA model significantly ( $p < 0.05$ ) improved the prediction accuracy of Mn up to 11.1% (0.36–0.40) over ST-GBLUP under MT-CV2 (**Figure 4B**). On the other hand, under the MT-CV3 scheme, MT-UN or MT-FA did not significantly outperform ST-BLUP even after using multiple auxiliary traits.

## Prediction Accuracy of Haplotype-Based Model Versus Genotype-Based Models

We further investigated the benefits of accommodating local epistatic effects on the prediction accuracy of MT models by using haplotypes instead of genotypes in the UN model. Comparing to the genotype-based UN model, the observed largest and significant ( $p < 0.05$ ) increment of prediction accuracies using haplotype-based models was 3.8% for Zn (0.52–0.54), 4.6% for Cu (0.43–0.45), and 3.5% (0.56–0.58) for Cd under MT-CV2. For Zn, the above improvement in prediction accuracy was achieved with a haplotype length of 3 SNPs, and when Mn, Fe, Cu, and Cd were collectively used as auxiliary traits (**Figure 5A**). For Cu, the observed gains were realized with a haplotype length of 4 SNPs and when auxiliary traits Mn and Zn were used together (**Figure 5B**). For Cd, the gains were from the

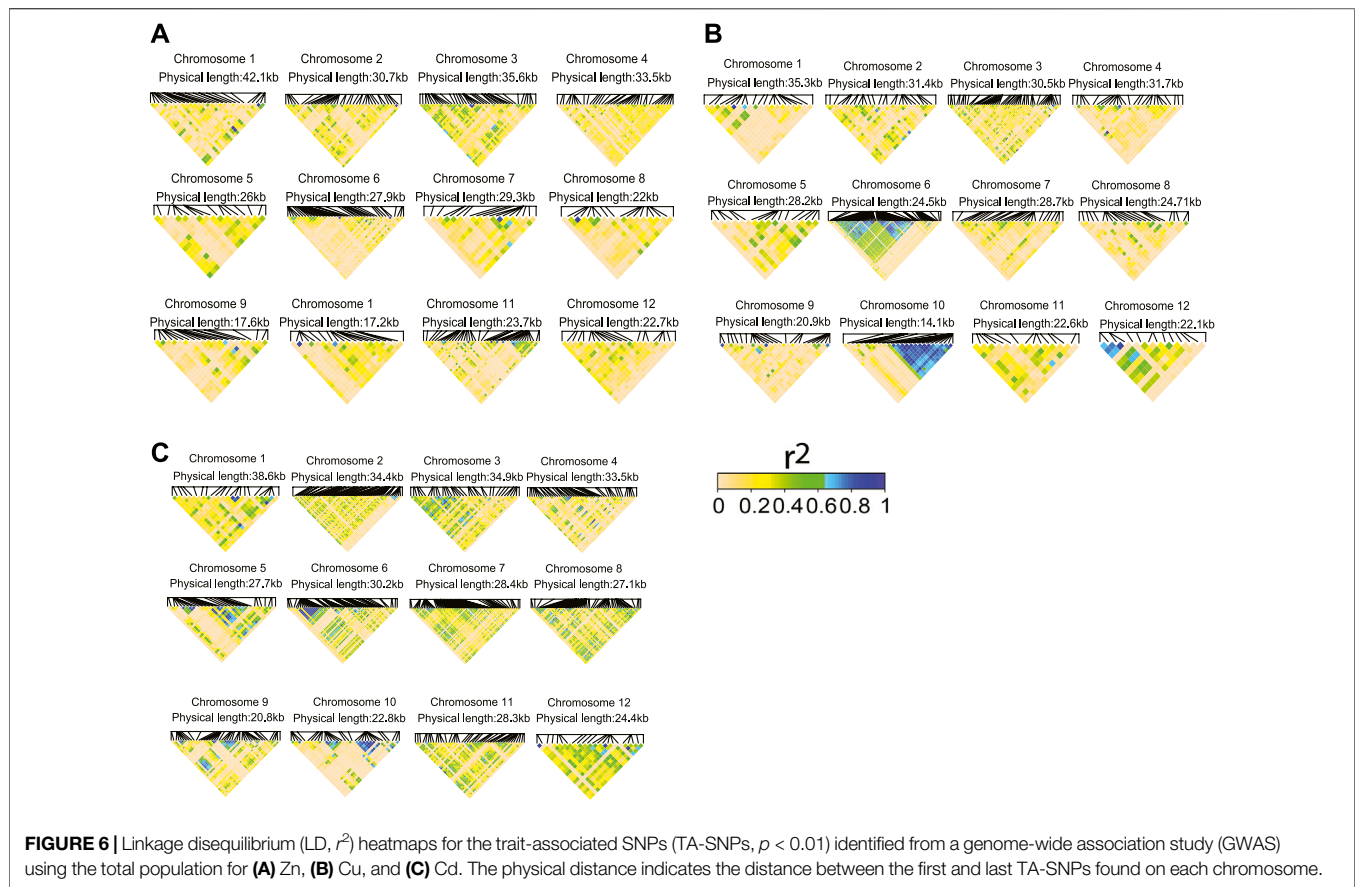




**FIGURE 5 |** Genomic prediction accuracies of the genotype-based single trait (ST) model (ST-GBLUP) and haplotype-based multi-trait (MT) model (MT-UN) under different cross-validation (CV) schemes (ST-CV1, MT-CV2, and MT-CV3). The target traits are **(A)** Zn, **(B)** Cu, and **(C)** Cd. The number of SNPs contained in haplotype blocks ranged from three to five. The first box-whisker in each portrayal indicates the accuracies of the ST-GBLUP model in the ST-CV1 scheme. Other box-whiskers refer to the accuracies achieved by MT models with different trait combinations. Asterisks above box-whiskers indicate that the prediction accuracies of the haplotype-based MT-UN model for the specific trait combination were statistically significantly ( $p < 0.05$ ,  $t$ -test) higher than those of the ST-GBLUP approach. Pounds above box-whiskers indicate that the prediction accuracies of the haplotype-based MT-UN model were statistically significantly ( $p < 0.05$ ,  $t$ -test) higher than those of its corresponding genotype-based counterparts. Only the scenarios where the haplotype-based MT-UN model were statistically significantly outperformed ( $p < 0.05$ ,  $t$ -test) the genotype-based MT-UN model are presented.

MT-UN model with a haplotype length of 2 SNPs, and when Mn, Fe, Cu, and Zn were combined as the auxiliary traits (**Figure 5C**). Under MT-CV3, the haplotype-based UN model was

significantly ( $p < 0.05$ ) superior to the genotype-based UN model by 12.5% (0.32–0.36) for Zn (**Figure 5A**) and 6% (0.33–0.35) for Cu (**Figure 5B**).



Compared to the ST-GBLUP model, the haplotype-based UN models were significantly ( $p < 0.05$ ) superior with an increment of prediction accuracy of 134.7% (0.23–0.54), 114.2% (0.21–0.45), 23% (0.39–0.48), and 11.5% (0.52–0.58) for Zn (Figure 5A), Cu (Figure 5B), Fe (Supplementary Figure S3), and Cd (Figure 5C), respectively.

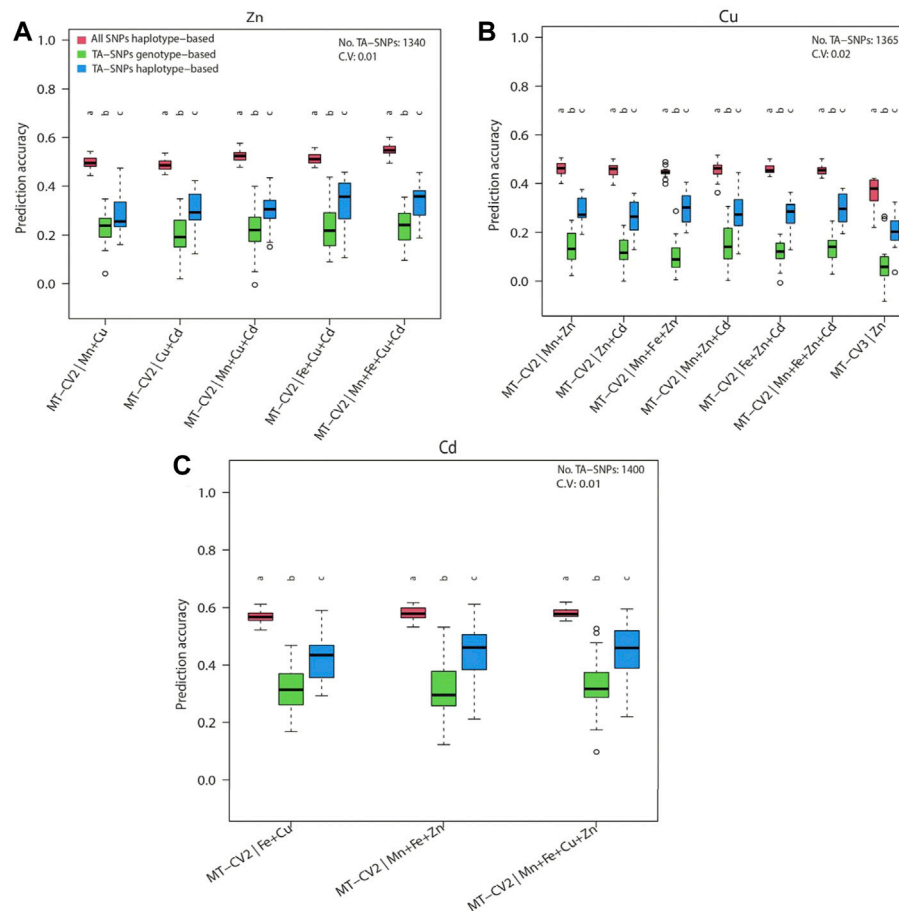
## Prediction Accuracy of a Haplotype-Based Model Capitalizing on Trait-Associated SNPs

With this study, we sought to investigate whether the prediction accuracy of Zn, Cu, and Cd with the haplotype-based model can be improved by using the selective marker panel made by TA-SNPs derived from GWAS (Supplementary Table S12). We purposely selected Zn and Cu because for both traits the haplotype-based MT model performed superiorly for several scenarios irrespective of the lengths of haplotypes (2–5 SNPs) (Figure 5). We also investigated Cd in addition to Zn and Cu since successive significant ( $p < 0.2$ ) TA-SNPs were observed in GWAS based on the total population (Supplementary Figures 4A–C). The non-negligible LD ( $r^2 \geq 0.1$ ) observed between the TA-SNPs, especially in Cu

and Cd, underpinned the necessity of modelling local epistatic effects among TA-SNPs (Figures 6A–C). The haplotype-based UN model accounting for TA-SNPs significantly ( $p < 0.05$ ) outperformed their genotype-based counterparts; however, it was significantly ( $p < 0.05$ ) inferior to the model using all genome-wide markers for all the traits and scenarios evaluated (Figures 7A–C).

## DISCUSSION

Quantifying trace element content in food crops is labor- and time-intensive. As a result, trace element traits have been the subject of few genomic prediction studies (Owens et al., 2014; Leplat et al., 2016; Frouin et al., 2019; Guo et al., 2020; Ahmadi et al., 2021) compared to agronomic or physiological traits. This study demonstrates how MT models with appropriate CV strategies can be useful in saving phenotyping resources for trace element traits in diverse rice collections without compromising the prediction accuracy. It also provides the first proof of concept in diverse rice for incorporating local epistatic effects and trait-associated SNPs into MT genomic prediction models.



**FIGURE 7** | Genomic prediction accuracies of the haplotype-based multi-trait model (MT-UN) with uninformative genomic markers (all SNPs haplotype-based) and haplotype- or genotype-based multi-trait model (MT-UN) with trait-associated SNPs (TA-SNPs) under MT-CV2 and MT-CV3. The target traits were **(A)** Zn, **(B)** Cu, and **(C)** Cd. The size of the haplotype blocks containing the TA-SNPs is maximally 300 kb. Different letters above box whiskers indicate statistically significant ( $p < 0.05$ ,  $t$ -test) differences among compared groups. The average number and the coefficient of variation (CV) of the used TA-SNPs for the observed predictions are shown for each trait.

## Multi-Trait Models Improved the Prediction Accuracy of Trace Elements in Rice Grain

In this study, MT models did not significantly outperform ST-GBLUP under the CV1 scheme for all the scenarios evaluated (**Supplementary Tables S2–S6**). Earlier studies also reported insignificant differences in the prediction accuracies of MT-CV1 and ST-CV1 (Calus and Veerkamp, 2011; dos Santos et al., 2016; Bhatta et al., 2020), implying that MT models are not always robust over ST models, especially when information on auxiliary traits is only available in the training set and the unobserved accessions are predicted only based on genotypic data. In contrast, when phenotypes of the auxiliary traits were present in the training and test set (MT-CV2) or merely in the test set (MT-CV3), the prediction accuracy of MT models (MT-UN and MT-FA) for the unobserved target traits (Zn, Cu, Fe, and Cd) was significantly improved relative to ST-GBLUP (**Figures 3, 4**). Previous studies attributed the predictive performance of MT models to both higher heritability of the auxiliary trait and strong

genetic correlation between the target and auxiliary traits (Sun et al., 2017; Fernandes et al., 2018).

## Accounting for the Information From Multiple Auxiliary Traits Boosted the Predictive Ability of Multi-Trait Models

Using a single auxiliary trait in the MT-UN model significantly ( $p < 0.05$ ) improved the prediction accuracy of target traits Zn, Cu, Fe, and Cd relative to the ST-GBLUP model (**Figures 3, 4**). When a strong genetic correlation exists between target and auxiliary trait, the prediction accuracy of MT models could still be improved under MT-CV2 or MT-CV3 regardless of trait heritability. For instance, supporting the prediction of Cu with Zn, its strongly correlated trait ( $cor_{Zn, Cu} = 0.95$ ) but with lower heritability ( $h^2_{Zn} = 0.14$ ;  $h^2_{Cu} = 0.21$ ) significantly improved the prediction accuracy of Cu with the MT-UN model (**Figure 3B**). Also, supporting Cd with Fe, its strongly correlated trait ( $cor_{Cd, Fe} = 0.95$ ) but with lower heritability

( $h^2_{Cd} = 0.62$ ;  $h^2_{Fe} = 0.50$ ), improved the prediction accuracy of Cd with the MT-UN model (**Figure 4A**). Arojju et al. (2020) also indicated that the genetic correlation was the main cause of the observed gain in prediction accuracy of MT models. The same study further showed that when a trait in strong genetic correlation with the target trait is used in the MT model, the predictive performance of the MT model was still superior to the ST model even after reducing the training population size by 50%.

Collectively accounting for phenotypes of multiple auxiliary traits further improved the predictive ability of the MT models compared to the MT models with a single auxiliary trait. For example, the highest increase in the prediction of Zn was 82.6% when a single auxiliary trait was used in MT-UN models. Yet, using multiple traits collectively in the same model improved the prediction of Zn by 126% compared to the ST-GBLUP model (**Figure 3A**). Also, MT models with one auxiliary trait showed no benefit over the ST-GBLUP model when predicting Mn, with relative medium heritability and no strong genetic correlation with any other studied trait. However, when auxiliary traits were collectively used in the MT-FA model, significant improvements in the prediction accuracy over the ST-GBLUP model were observed (**Figure 4B**). Multiple auxiliary traits would optimize MT models, though the assisting trait *per se* is neither strongly genetically correlated with the target trait nor highly heritable. Therefore, when no single auxiliary trait meets the criterion of heritability or genetic correlation, combining multiple auxiliary traits in the MT model could be an effective approach to enhance the predictive ability of MT models. These findings are concurrent with previous findings by Wang et al. (2017), indicating that the prediction accuracy of MT models was highest when eight different traits were used as auxiliary traits to predict grain yield in rice.

### Modeling Local Epistatic Effects is Beneficial in Multi-Trait Models Irrespective of Using Total or Selective Marker Panel

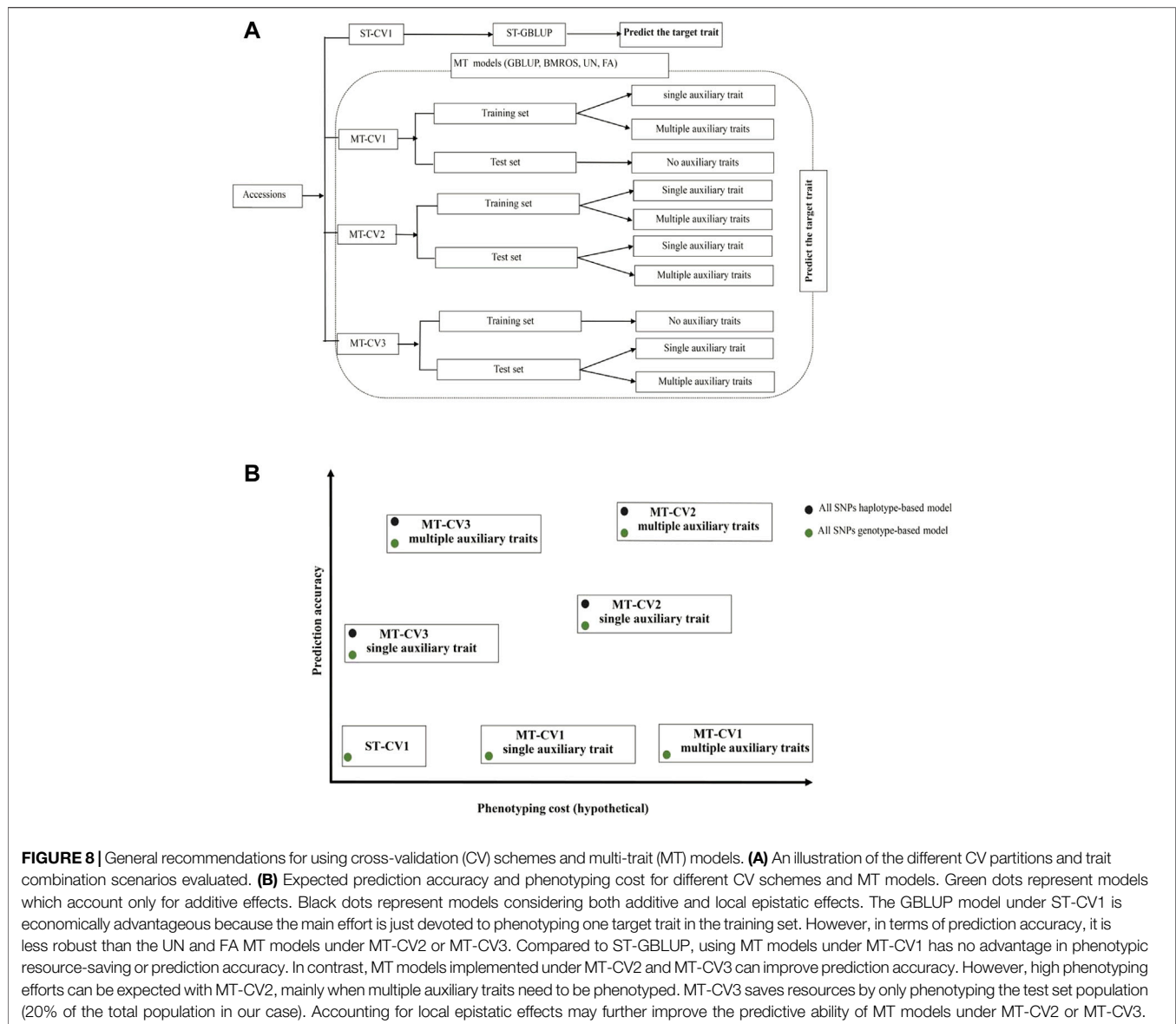
Previous studies demonstrated that accounting for local epistatic effects besides the additive effect in genomic prediction could improve the prediction accuracy of ST models (Akdemir and Jannink, 2015; Akdemir et al., 2017; Jiang et al., 2018; He et al., 2019). Here, we are the first to attempt to model the local epistatic effects in the context of MT genomic prediction. Accounting for the local epistatic effects in haplotype-based MT models significantly improved the prediction accuracy of Zn, Cu, and Cd relative to genotype-based MT models, only capturing additive effects (**Figures 5A–C**). Relative to ST-GLUP, the highest increase in prediction accuracy, 134.7% for Zn, was observed after incorporating the local epistatic effects into the MT-UN model (**Figure 5A**). These findings imply that the potential of MT models can be maximized by accounting for local epistatic effects besides additive effects in the model.

Using a selective marker panel based on approaches exploiting the trait biological and genetic background knowledge such as GWAS has been proven effective to improve the predictive ability of GS models (Owens et al., 2014; Wang et al., 2019). Our study did not show any improvements of prediction accuracy by using the TA-SNPs instead of all genome-wide SNPs for Cu, Zn, and Cd (**Figures 7A–C**). These findings could be attributed to the complex genetic architectures of the trace elements we studied (**Supplementary Figures S4A–C**). Our approach was slightly similar to previous methods using GWAS-derived TA-SNPs to construct the trait-associated matrix (Zhang et al., 2014; Ahmadi et al., 2021), except that we did not assign weights to haplotype- or genotype-based genomic relationship matrices. Though numerous studies reported improved gains from using the above strategy (Bhandari et al., 2019; Ahmadi et al., 2021), Veerkamp et al. (2016) showed that the proportion of total variance explained by the TA-SNPs combined in a GRM was considerably smaller than that explained by all variants in Holstein-Friesian cattle population. A potentially more promising way to use TA-SNPs would be to fit them as fixed effects in the GP model along with all other SNPs as random effects (Spindel et al., 2016). However, the latter approach is best suited for features with a few large-effect QTLs in a polygenic context (Poland and Rutkoski, 2016; Bian and Holland, 2017; Rice and Lipka, 2019). Therefore, the genetic architecture of the target traits must be studied before applying this strategy to a breeding program. The marked advantage of the haplotype-based UN model over their genotype-based counterparts using the TA-SNPs (**Figures 7A–C**) substantiates the existence of local epistasis in trace element traits (Sharma V. et al., 2021) and the merit of modelling local epistatic effects in MT-GP program.

### Factors Affecting the Observed Prediction Accuracies: Trait Heritability, Genetic Correlation, and Population Relatedness

Various factors affect the predictive ability of GP models used in GS (Crossa et al., 2017; Xu et al., 2021). In this study, the genomic heritabilities spanned a wide range from 0.14 to 0.62 (**Table 2**), which enabled the evaluation of the performance of MT models under contrasting levels of genomic heritability. The genomic heritability of Zn was the lowest ( $h^2 = 0.14$ , **Table 2**), which contradicts several previous studies that reported moderate to high heritability of Zn (Norton et al., 2010; Pinson et al., 2015; Naik et al., 2020). The poor heritability estimate of Zn in this study could be due to potential environmental effects. Unfortunately, our study does not include multi-environment trials and therefore does not provide insight into environmental factors and GxE interactions on genomic prediction of trace element traits.

Strong genetic correlations ( $Cor > 0.75$ ) were observed between pairs of the studied trace element traits (**Table 2**). This was expected due to their overlapping genetic and physiological mechanisms (Sasaki et al., 2012; Cu et al., 2020). For example, transporter gene families like zinc-iron permease (ZIP), natural resistance-associated macrophage proteins (NRAMPs), and heavy metal transporting, ATPases (HMAs)



have been associated with uptake and translocation of several trace elements in plants (Fernández-Paz et al., 2021; Vanderschueren et al., 2021; Zhang et al., 2022). As a result, borrowing information from correlated traits overall improved the prediction accuracy of the MT models.

The success of GS is also highly dependent on the LD between markers and unknown causal variants. The genetically distant training and test sets would have different LD decay patterns and consequently impede the prediction (Snelling et al., 2013; Desta and Ortiz, 2014; Thistlethwaite et al., 2020). Such a problem is typically prominent in germplasm accessions, limiting the power of GP (Crossa et al., 2017). This is also the case for the diverse rice population used in this study, as most accession pairs were distantly unrelated (Figure 1B). As a result, the predictive

ability for ST-GBLUP was poor (Figure 2), particularly for traits with the lowest estimated heritabilities, namely Zn and Cu (Table 2). Adding related materials to the training population has been suggested to overcome the problem of low relatedness between training and test populations and improve the accuracy of genomic prediction (Arenas et al., 2021). Nevertheless, increasing relatedness will damage genetic gain in the long term because genetic variation will be limited or exhausted if related populations are overused (Jannink et al., 2010; Moeinizada et al., 2019). We show that MT models are powerful tools for predicting trace element traits in populations with diverse backgrounds. However, further studies with larger datasets are needed to elucidate the utility of different populations and marker optimization strategies in the context of MT genomic prediction.

## The Prospect of Multi-Trait Models for the Evaluation of Genetic Resources in Gene Banks

Expediting genomic selection in gene banks to predict the genetic merit of the unobserved accessions would enable accurate identification of promising donor accessions without a comprehensive phenotypic test of all the accessions in the field (Pace et al., 2015; Yu et al., 2016; Crossa et al., 2017; Tanaka et al., 2021; He et al., 2022). In fact, as the traits of breeders' interest are extensive, the genetic resources archived in gene banks would be evaluated for several traits. MT genomic prediction is an effective method to realize this comprehensive evaluation. Our study tested several MT models under three different CV schemes, implying different phenotyping layouts and costs. We found that high prediction accuracy of MT models can be achieved under prediction schemes MT-CV2 and MT-CV3. The MT-CV2 scheme requires more budget for phenotyping auxiliary traits in both training and test sets. Therefore, breeders may kindly MT-CV2 if phenotypes for the auxiliary traits can be inexpensively obtained. Otherwise, MT-CV3 is more cost-effective as phenotypes for the auxiliary traits are only required for the test set (i.e., 20% of the entire population). Since using multiple auxiliary traits collectively in the MT model can improve prediction even if the individual auxiliary traits do not fully meet the heritability and genetic correlation conditions, an ideal situation would be to phenotype less expensive and more manageable traits (e.g., root system architecture, 100-grain weight, data to heading, etc.) to support the prediction of expensive target traits with the MT models. Besides, accounting for local epistatic effects in MT models would help to improve the predictive ability. The different scenarios studied here and their respective potentials in terms of prediction accuracy and phenotyping cost are illustrated in **Figures 8A,B**.

To breed safe and nutritious crop varieties, further studies using the genomic selection index (Habyarimana et al., 2020), for

example, are desired to provide a comprehensive understanding of the strategies to optimize essential nutrients and toxic metals such as Cd in food crops.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; further inquiries can be directed to the corresponding authors.

## AUTHOR CONTRIBUTIONS

GY and SH conceived the project. WZ, JZ, and MB prepared the samples and collected the data. SH, JZ, and MB analyzed the data. WZ and MB drafted the manuscript. SH and GY revised the manuscript. All the authors have read, edited, and approved the current version of the manuscript.

## FUNDING

This work was supported by the Agricultural Science and Technology Innovation Program, National Key R&D Program of China (2020YFE0202300), Shenzhen Science and Technology Projects (JCYJ20200109150650397), and China Postdoctoral Science Foundation (2020M682998).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.883853/full#supplementary-material>

## REFERENCES

- Ahmadi, N., Cao, T.-V., Frouin, J., Norton, G. J., and Price, A. H. (2021). Genomic Prediction of Arsenic Tolerance and Grain Yield in Rice: Contribution of Trait-specific Markers and Multi-Environment Models. *Rice Sci.* 28 (3), 268–278. doi:10.1016/j.rsci.2021.04.006
- Akdemir, D., Jannink, J.-L., and Isidro-Sánchez, J. (2017). Locally Epistatic Models for Genome-wide Prediction and Association by Importance Sampling. *Genet. Sel. Evol.* 49 (1), 74. doi:10.1186/s12711-017-0348-8
- Akdemir, D., and Jannink, J.-L. (2015). Locally Epistatic Genomic Relationship Matrices for Genomic Association and Prediction. *Genetics* 199 (3), 857–871. doi:10.1534/genetics.114.173658
- Alemu, A., Brazauskas, G., Gaikpa, D. S., Henriksson, T., Islamov, B., Jørgensen, L. N., et al. (2021). Genome-Wide Association Analysis and Genomic Prediction for Adult-Plant Resistance to Septoria Tritici Blotch and Powdery Mildew in Winter Wheat. *Front. Genet.* 12, 661742. doi:10.3389/fgene.2021.661742
- Arao, T., and Ae, N. (2003). Genotypic Variations in Cadmium Levels of Rice Grain. *Soil Sci. Plant Nutr.* 49 (4), 473–479. doi:10.1080/00380768.2003.10410035
- Ardlie, K. G., Kruglyak, L., and Seielstad, M. (2002). Patterns of Linkage Disequilibrium in the Human Genome. *Nat. Rev. Genet.* 3 (4), 299–309. doi:10.1038/nrg777
- Arenas, S., Cortés, A. J., Mastretta-Yanes, A., and Jaramillo-Correa, J. P. (2021). Evaluating the Accuracy of Genomic Prediction for the Management and Conservation of Relictual Natural Tree Populations. *Tree Genet. Genomes* 17. doi:10.1007/s11295-020-01489-1
- Arojju, S. K., Cao, M., Trolove, M., Barrett, B. A., Inch, C., Eady, C., et al. (2020). Multi-Trait Genomic Prediction Improves Predictive Ability for Dry Matter Yield and Water-Soluble Carbohydrates in Perennial Ryegrass. *Front. Plant Sci.* 11, 1197. doi:10.3389/fpls.2020.01197
- Aschner, M., and Erikson, K. (2017). Manganese. *Adv. Nutr.* 8 (3), 520–521. doi:10.3945/an.117.015305
- Bandumula, N. (2018). Rice Production in Asia: Key to Global Food Security. *PNAS* 88 (4), 1323–1328. doi:10.1007/s40011-017-0867-7
- Bernardo, R. (2014). Genomewide Selection when Major Genes Are Known. *Crop Sci.* 54 (1), 68–75. doi:10.2135/cropsci2013.05.0315
- Bhandari, A., Bartholomé, J., Cao-Hamadoun, T.-V., Kumari, N., Frouin, J., Kumar, A., et al. (2019). Selection of Trait-specific Markers and Multi-Environment Models Improve Genomic Predictive Ability in Rice. *PLoS One* 14 (5), e0208871. doi:10.1371/journal.pone.0208871
- Bhatta, M., Gutierrez, L., Cammarota, L., Cardozo, F., Germán, S., Gómez-Guerrero, B., et al. (2020). Multi-trait Genomic Prediction Model Increased the Predictive Ability for Agronomic and Malting Quality Traits in Barley (*Hordeum Vulgare* L.). *G3 (Bethesda)* 10 (3), 1113–1124. doi:10.1534/g3.119.400968
- Bian, Y., and Holland, J. B. (2017). Enhancing Genomic Prediction with Genome-wide Association Studies in Multiparental Maize Populations. *Heredity* 118 (6), 585–593. doi:10.1038/hdy.2017.4

- Breseghele, F., and Sorrells, M. E. (2006). Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics* 172 (2), 1165–1177. doi:10.1534/genetics.105.044586
- Burgueño, J., de los Campos, G., Weigel, K., and Crossa, J. (2012). Genomic Prediction of Breeding Values When Modeling Genotype × Environment Interaction Using Pedigree and Dense Molecular Markers. *Crop Sci.* 52 (2), 707–719. doi:10.2135/cropsci2011.06.0299
- Calus, M. P., and Veerkamp, R. F. (2011). Accuracy of Multi-Trait Genomic Selection Using Different Methods. *Genet. Sel. Evol.* 43 (1), 26. doi:10.1186/1297-9686-43-26
- 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. *GigaSci.* 4 (1), s13742–0047. doi:10.1186/s13742-015-0047-8
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., De Los Campos, G., et al. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant Sci.* 22 (11), 961–975. doi:10.1016/j.tplants.2017.08.011
- Cu, S. T., Guild, G., Nicolson, A., Velu, G., Singh, R., and Stangoulis, J. (2020). Genetic Dissection of Zinc, Iron, Copper, Manganese and Phosphorus in Wheat (*Triticum aestivum* L.) Grain and Rachis at Two Developmental Stages. *Plant Sci.* 291, 110338. doi:10.1016/j.plantsci.2019.110338
- Cuevas, J., Crossa, J., Montesinos-López, O. A., Burgueño, J., Pérez-Rodríguez, P., de los Campos, G., et al. (2017). Bayesian Genomic Prediction with Genotype × Environment Interaction Kernel Models. *G3 (Bethesda)* 7 (1), 41–53. doi:10.1534/g3.116.035584
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27 (15), 2156–2158. doi:10.1016/j.tplants.2017.08.01110.1093/bioinformatics/btr330
- De los Campos, G., and Pérez-Rodríguez, P. (2015). “BGLR: Bayesian generalized linear regression. R package version 1.0.4,” in *R Foundation for Statistical Computing*. Available at: <http://cran.r-project.org/package=BGLR> (Accessed June 3, 2022).
- De los Campos, G., and Grüneberg, A. (2016). “MTM Package,” in *QuantGen Group, Michigan State Univ. East Lansing*. Available at: <http://quantgen.github.io/MTM/vignette.html>.
- De los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting Quantitative Traits with Regression Models for Dense Molecular Markers and Pedigree. *Genetics* 182 (1), 375–385. doi:10.1534/genetics.109.101501
- Delaneau, O., Marchini, J., and Zagury, J.-F. (2012). A Linear Complexity Phasing Method for Thousands of Genomes. *Nat. Methods* 9 (2), 179–181. doi:10.1038/nmeth.1785
- Descalsota-Empleo, G. I., Amparado, A., Inabangan-Asilo, M. A., Tesoro, F., Stangoulis, J., Reinke, R., et al. (2019). Genetic Mapping of QTL for Agronomic Traits and Grain Mineral Elements in Rice. *Crop J.* 7 (4), 560–572. doi:10.1016/j.cj.2019.03.002
- Desta, Z. A., and Ortiz, R. (2014). Genomic Selection: Genome-wide Prediction in Plant Improvement. *Trends Plant Sci.* 19 (9), 592–601. doi:10.1016/j.tplants.2014.05.006
- Dos Santos, J. P. R., Vasconcellos, R. C. d. C., Pires, L. P. M., Balestre, M., and Von Pinho, R. G. (2016). Inclusion of Dominance Effects in the Multivariate GBLUP Model. *PLoS One* 11 (4), e0152045. doi:10.1371/journal.pone.0152045
- Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., et al. (2017). Sequencing and De Novo Assembly of a Near Complete Indica Rice Genome. *Nat. Commun.* 8 (1), 15324. doi:10.1038/ncomms15324
- Du, J., Zeng, D., Wang, B., Qian, Q., Zheng, S., and Ling, H.-Q. (2013). Environmental Effects on Mineral Accumulation in Rice Grains and Identification of Ecological Specific QTLs. *Environ. Geochem Health* 35 (2), 161–170. doi:10.1007/s10653-012-9473-z
- Falconer, D. S. (1996). *Introduction to Quantitative Genetics*. 4th Edition. Essex: Pearson Education Ltd.
- Fernandes, S. B., Dias, K. O. G., Ferreira, D. F., and Brown, P. J. (2018). Efficiency of Multi-Trait, Indirect, and Trait-Assisted Genomic Selection for Improvement of Biomass Sorghum. *Theor. Appl. Genet.* 131 (3), 747–755. doi:10.1007/s00122-017-3033-y
- Fernández-Paz, J., Cortés, A. J., Hernández-Varela, C. A., Mejía-de-Tafur, M. S., Rodríguez-Medina, C., and Baligar, V. C. (2021). Rootstock-Mediated Genetic Variance in Cadmium Uptake by Juvenile Cacao (*Theobroma Cacao* L.) Genotypes, and its Effect on Growth and Physiology. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.777842
- Frouin, J., Labeyrie, A., Boissard, A., Sacchi, G. A., and Ahmadi, N. (2019). Genomic Prediction Offers the Most Effective Marker Assisted Breeding Approach for Ability to Prevent Arsenic Accumulation in Rice Grains. *PLoS One* 14 (6), e0217516. doi:10.1371/journal.pone.0217516
- Gao, L., and Xiong, J. (2018). Improving Rice Grain Quality by Enhancing Accumulation of Iron and Zinc while Minimizing Cadmium and Lead. *IntechOpen*. doi:10.5772/intechopen.72826
- García-Oliveira, A. L., Tan, L., Fu, Y., and Sun, C. (2009). Genetic Identification of Quantitative Trait Loci for Contents of Mineral Nutrients in Rice Grain. *J. Integr. Plant Biol.* 51 (1), 84–92. doi:10.1111/j.1744-7909.2008.00730.x
- Guo, R., Dhaliwayo, T., Mageto, E. K., Palacios-Rojas, N., Lee, M., Yu, D., et al. (2020). Genomic Prediction of Kernel Zinc Concentration in Multiple Maize Populations Using Genotyping-By-Sequencing and Repeat Amplification Sequencing Markers. *Front. Plant Sci.* 11 (534), 00534. doi:10.3389/fpls.2020.00534
- Habyarimana, E., Lopez-Cruz, M., and Baloch, F. S. (2020). Genomic Selection for Optimum Index with Dry Biomass Yield, Dry Mass Fraction of Fresh Material, and Plant Height in Biomass Sorghum. *Genes* 11 (1), 61. doi:10.3390/genes11010061
- Han, Y., Ling, Q., Dong, F., de Dios, V. R., Li, Z., Zhang, W., et al. (2021). Iron and Copper Micronutrients Influences Cadmium Accumulation in Rice Grains by Altering its Transport and Allocation. *Sci. Total Environ.* 777, 146118. doi:10.1016/j.scitotenv.2021.146118
- Hao, X., Zeng, M., Wang, J., Zeng, Z., Dai, J., Xie, Z., et al. (2018). A Node-Expressed Transporter OsCCX2 Is Involved in Grain Cadmium Accumulation of Rice. *Front. Plant Sci.* 9 (476), 00476. doi:10.3389/fpls.2018.00476
- He, S., Liu, H., Zhan, J., Meng, Y., Wang, Y., Wang, F., et al. (2022). Genomic Prediction Using Composite Training Sets Is an Effective Method for Exploiting Germplasm Conserved in Rice Gene Banks. *Crop J.* doi:10.1016/j.cj.2021.11.011
- He, S., Reif, J. C., Korzun, V., Bothe, R., Ebmeyer, E., Jiang, Y., et al. (2017). Genome-wide Mapping and Prediction Suggests Presence of Local Epistasis in a Vast Elite Winter Wheat Populations Adapted to Central Europe. *Theor. Appl. Genet.* 130 (4), 635–647. doi:10.1007/s00122-016-2840-x
- He, S., Thistlethwaite, R., Forrest, K., Shi, F., Hayden, M. J., Trethowan, R., et al. (2019). Extension of a Haplotype-Based Genomic Prediction Model to Manage Multi-Environment Wheat Data Using Environmental Covariates. *Theor. Appl. Genet.* 132 (11), 3143–3154. doi:10.1007/s00122-019-03413-1
- Henderson, C. R., and Quaas, R. L. (1976). Multiple Trait Evaluation Using Relatives’ Records. *J. Anim. Sci.* 43 (6), 1188–1197. doi:10.2527/jas1976.4361188x
- Houk, R. S., Fassel, V. A., Flesch, G. D., Svec, H. J., Gray, A. L., and Taylor, C. E. (1980). Inductively Coupled Argon Plasma as an Ion Source for Mass Spectrometric Determination of Trace Elements. *Anal. Chem.* 52 (14), 2283–2289. doi:10.1021/ac50064a012
- Huang, W., and Mackay, T. F. C. (2016). The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS Genet.* 12 (11), e1006421. doi:10.1371/journal.pgen.1006421
- Huang, X.-Y., Deng, F., Yamaji, N., Pinson, S. R. M., Fujii-Kashino, M., Danku, J., et al. (2016). A Heavy Metal P-type ATPase OsHMA4 Prevents Copper Accumulation in Rice Grain. *Nat. Commun.* 7 (1), 12138. doi:10.1038/ncomms12138
- Jannink, J.-L., Lorenz, A. J., and Iwata, H. (2010). Genomic Selection in Plant Breeding: from Theory to Practice. *Briefings Funct. Genomics* 9 (2), 166–177. doi:10.1093/bfpg/eq001
- Jarquín, D., Specht, J., and Lorenz, A. (2016). Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical Data Creates Robust Models for Enhancing Selection of Accessions. *G3 (Bethesda)* 6 (8), 2329–2341. doi:10.1534/g3.116.031443
- Jiang, Y., Schmidt, R. H., and Reif, J. C. (2018). Haplotype-Based Genome-wide Prediction Models Exploit Local Epistatic Interactions Among Markers. *G3 (Bethesda)* 8 (5), 1687–1699. doi:10.1534/g3.117.300548
- Lado, B., Vázquez, D., Quincke, M., Silva, P., Aguilar, I., and Gutiérrez, L. (2018). Resource Allocation Optimization with Multi-Trait Genomic Prediction for Bread Wheat (*Triticum aestivum* L.) Baking Quality. *Theor. Appl. Genet.* 131 (12), 2719–2731. doi:10.1007/s00122-018-3186-3

- Leplat, F., Jensen, J., and Madsen, P. (2016). Genomic Prediction of Manganese Efficiency in Winter Barley. *Plant Genome* 9 (2). doi:10.3835/plantgenome2015.09.0085
- Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. Available at: <http://arxiv.org/pdf/1303.3997pp>.
- Lien, K.-W., Pan, M.-H., and Ling, M.-P. (2021). Levels of Heavy Metal Cadmium in Rice (*Oryza Sativa* L.) Produced in Taiwan and Probabilistic Risk Assessment for the Taiwanese Population. *Environ. Sci. Pollut. Res.* 28 (22), 28381–28390. doi:10.1007/s11356-020-11902-w
- Liu, J., Zhan, J., Chen, J., Lu, X., Zhi, S., and Ye, G. (2021a). Validation of Genes Affecting Rice Grain Zinc Content through Candidate Gene-Based Association Analysis. *Front. Genet.* 12 (1354). doi:10.3389/fgene.2021.701658
- Liu, S., Zou, W., Lu, X., Bian, J., He, H., Chen, J., et al. (2021b). Genome-Wide Association Study Using a Multiparent Advanced Generation Intercross (MAGIC) Population Identified QTLs and Candidate Genes to Predict Shoot and Grain Zinc Contents in Rice. *Agriculture* 11 (1), 70. doi:10.3390/agriculture11010070
- Long, X., Xiang, X., Xu, Y., Su, W., and Kang, C. (2014). Absorption, Transfer and Distribution of Cd in Indica and Japonica Rice under Cd Stress. *Chin. J. Rice Sci.* 28 (2), 177–184.
- Lu, K., Li, L., Zheng, X., Zhang, Z., Mou, T., and Hu, Z. (2008). Quantitative Trait Loci Controlling Cu, Ca, Zn, Mn and Fe Content in Rice Grains. *J. Genet.* 87 (3), 305–310. doi:10.1007/s12041-008-0049-8
- Lyra, D. H., de Freitas Mendonça, L., Galli, G., Alves, F. C., Granato, Í. S. C., and Fritsche-Neto, R. (2017). Multi-trait Genomic Prediction for Nitrogen Response Indices in Tropical Maize Hybrids. *Mol. Breed.* 37 (6), 80. doi:10.1007/s11032-017-0681-1
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. doi:10.1101/gr.107524.110
- Meng, L., Wang, B., Zhao, X., Ponce, K., Qian, Q., and Ye, G. (2017). Association Mapping of Ferrous, Zinc, and Aluminum Tolerance at the Seedling Stage in Indica Rice Using MAGIC Populations. *Front. Plant Sci.* 8. doi:10.3389/fpls.2017.01822
- 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. doi:10.1093/genetics/157.4.1819
- Miller, W. J. (1970). Zinc Nutrition of Cattle: A Review. *J. Dairy Sci.* 53 (8), 1123–1135. doi:10.3168/jds.S0022-0302(70)86355-X
- Moeiniazade, S., Hu, G., Wang, L., and Schnable, P. S. (2019). Optimizing Selection and Mating in Genomic Selection with a Look-Ahead Approach: An Operations Research Framework. *G3 (Bethesda)* 9 (7), 2123–2133. doi:10.1534/g3.118.200842
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Toledo, F. H., Pérez-Hernández, A. K., Eskridge, K. M., et al. (2016). A Genomic Bayesian Multi-Trait and Multi-Environment Model. *G3 (Bethesda)* 6 (9), 2725–2744. doi:10.1534/g3.116.032359
- Montesinos-López, O. A., Montesinos-López, A., Crossa, J., Cuevas, J., Montesinos-López, J. C., Gutiérrez, Z. S., et al. (2019). A Bayesian Genomic Multi-Output Regressor Stacking Model for Predicting Multi-Trait Multi-Environment Plant Breeding Data. *G3 (Bethesda)* 9 (10), 3381–3393. doi:10.1534/g3.119.400336
- Naik, S. M., Raman, A. K., Nagamallika, M., Venkateshwarlu, C., Singh, S. P., Kumar, S., et al. (2020). Genotype × Environment Interactions for Grain Iron and Zinc Content in Rice. *J. Sci. Food Agric.* 100 (11), 4150–4164. doi:10.1002/jsfa.10454
- Norton, G. J., Deacon, C. M., Xiong, L., Huang, S., Meharg, A. A., and Price, A. H. (2010). Genetic Mapping of the Rice Ironome in Leaves and Grain: Identification of QTLs for 17 Elements Including Arsenic, Cadmium, Iron and Selenium. *Plant Soil* 329 (1), 139–153. doi:10.1007/s11104-009-0141-8
- Olivares, M., and Uauy, R. (1996). Copper as an Essential Nutrient. *Am. J. Clin. Nutr.* 63 (5), 791S–796S. doi:10.1093/ajcn/63.5.791
- Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., et al. (2014). A Foundation for Provitamin A Biofortification of Maize: Genome-wide Association and Genomic Prediction Models of Carotenoid Levels. *Genetics* 198 (4), 1699–1716. doi:10.1534/genetics.114.169979
- Pace, J., Yu, X., and Lübberstedt, T. (2015). Genomic Prediction of Seedling Root Length in Maize (*Zea mays* L.). *Plant J.* 83 (5), 903–912. doi:10.1111/tpj.12937
- Pinson, S. R. M., Tarpley, L., Yan, W., Yeater, K., Lahner, B., Yakubova, E., et al. (2015). Worldwide Genetic Diversity for Mineral Element Concentrations in Rice Grain. *Crop Sci.* 55 (1), 294–311. doi:10.2135/cropsci2013.10.0656
- Poland, J., and Rutkoski, J. (2016). Advances and Challenges in Genomic Selection for Disease Resistance. *Annu. Rev. Phytopathol.* 54, 79–98. doi:10.1146/annurev-phyto-080615-100056
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Available at: <https://www.R-project.org/>.
- Rice, B., and Lipka, A. E. (2019). Evaluation of RR-BLUP Genomic Selection Models that Incorporate Peak Genome-Wide Association Study Signals in Maize and Sorghum. *Plant Genome* 12 (1), 180052. doi:10.3835/plantgenome2018.07.0052
- Rutkoski, J., Poland, J., Mondal, S., Autrique, E., Pérez, L. G., Crossa, J., et al. (2016). Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. *G3 (Bethesda)* 6 (9), 2799–2808. doi:10.1534/g3.116.032888
- Sasaki, A., Yamaji, N., Yokosho, K., and Ma, J. F. (2012). Nramp5 Is a Major Transporter Responsible for Manganese and Cadmium Uptake in Rice. *Plant Cell.* 24 (5), 2155–2167. doi:10.1105/tpc.112.096925
- Sharma, S., Pinson, S. R. M., Gealy, D. R., and Edwards, J. D. (2021a). Genomic Prediction and QTL Mapping of Root System Architecture and Above-Ground Agronomic Traits in Rice (*Oryza Sativa* L.) with a Multitrait Index and Bayesian Networks. *G3 (Bethesda)* 11 (10). doi:10.1093/g3journal/jkab178
- Sharma, V., Choudhary, M., Kumar, P., Choudhary, J. R., Khokhar, J. S., Kaushik, P., et al. (2021b). Harnessing the Wild Relatives and Landraces for Fe and Zn Biofortification in Wheat through Genetic Interventions—A Review. *Sustainability* 13 (23), 12975. doi:10.3390/su132312975
- Snelling, W. M., Cushman, R. A., Keele, J. W., Maltecca, C., Thomas, M. G., Fortes, M. R. S., et al. (2013). Breeding and Genetics Symposium: Networks and Pathways to Guide Genomic Selection 1–3. *J. Anim. Sci.* 91 (2), 537–552. doi:10.2527/jas.2012-5784
- Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redoña, E., Jannink, J.-L., et al. (2016). Genome-wide Prediction Models that Incorporate De Novo GWAS Are a Powerful New Tool for Tropical Rice Improvement. *Heredity* 116 (4), 395–408. doi:10.1038/hdy.2015.113
- Spyromitros-Xioufous, E., Tsoumakas, G., Groves, W., and Vlahavas, I. (2012). Multi-label classification methods for multi-target regression. ArXiv e-prints. Available at: <https://arxiv.org/abs/1211.6581v1>. (Accessed June 4, 2022).
- Spyromitros-Xioufous, E., Tsoumakas, G., Groves, W., and Vlahavas, I. (2016). Multi-Target Regression Via Input Space Expansion: Treating Targets as Inputs. *Mach. Learn.* 104, 55–98. doi:10.1007/s10994-016-5546-z
- Sun, J., Rutkoski, J., Poland, J. A., Crossa, J., Jannink, J.-L., and Sorrells, M. E. (2017). Multitrait, Random Regression, or Simple Repeatability Model in High-Throughput Phenotyping Data Improve Genomic Prediction for Wheat Grain Yield. *Plant Genome* (102), 1940–3372. doi:10.3835/plantgenome2016.11.0111
- Swamy, B. P. M., Descalsota, G. I. L., Nha, C. T., Amparado, A., Inabangan-Asilo, M. A., Manito, C., et al. (2018). Identification of Genomic Regions Associated with Agronomic and Biofortification Traits in DH Populations of Rice. *PLoS One* 13 (8), e0201756. doi:10.1371/journal.pone.0201756
- Tanaka, R., Mandaharisoa, S. T., Rakotondramanana, M., Ranaivo, H. N., Pariasca-Tanaka, J., Kajiya-Kanegae, H., et al. (2021). From Gene Banks to Farmer's Fields: Using Genomic Selection to Identify Donors for a Breeding Program in Rice to Close the Yield Gap on Smallholder Farms. *Theor. Appl. Genet.* 134 (10), 3397–3410. doi:10.1007/s00122-021-03909-9
- Thistlethwaite, F. R., Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Porth, I., Chen, C., et al. (2020). Linkage Disequilibrium vs. Pedigree: Genomic Selection Prediction Accuracy in Conifer Species. *PLoS one* 15 (6), e0232201. doi:10.1371/journal.pone.0232201
- Ueno, D., Yamaji, N., Kono, I., Huang, C. F., Ando, T., Yano, M., et al. (2010). Gene Limiting Cadmium Accumulation in Rice. *Proc. Natl. Acad. Sci. U.S.A.* 107 (38), 16500–16505. doi:10.1073/pnas.1005396107
- Uraguchi, S., and Fujiwara, T. (2012). Cadmium Transport and Tolerance in Rice: Perspectives for Reducing Grain Cadmium Accumulation. *Rice* 5 (1), 5. doi:10.1186/1939-8433-5-5



- Van Der Straeten, D., Bhullar, N. K., De Steur, H., Gruijssem, W., MacKenzie, D., Pfeiffer, W., et al. (2020). Multiplying the Efficiency and Impact of Biofortification through Metabolic Engineering. *Nat. Commun.* 11 (1), 5203. doi:10.1038/s41467-020-19020-4
- Vanderschueren, R., Argüello, D., Blommaert, H., Montalvo, D., Barraza, F., Maurice, L., et al. (2021). Mitigating the Level of Cadmium in Cacao Products: Reviewing the Transfer of Cadmium from Soil to Chocolate Bar. *Sci. Total Environ.* 781, 146779. doi:10.1016/j.scitotenv.2021.146779
- Veerkamp, R. F., Bouwman, A. C., Schrooten, C., and Calus, M. P. L. (2016). Genomic Prediction Using Preselected DNA Variants from a GWAS with Whole-Genome Sequence Data in Holstein-Friesian Cattle. *Genet. Sel. Evol.* 48 (1), 95. doi:10.1186/s12711-016-0274-1
- Wang, N., Liu, B., Liang, X., Zhou, Y., Song, J., Yang, J., et al. (2019). Genome-wide Association Study and Genomic Prediction Analyses of Drought Stress Tolerance in China in a Collection of Off-PVP Maize Inbred Lines. *Mol. Breed.* 39 (8), 113. doi:10.1007/s11032-019-1013-4
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic Variation in 3,010 Diverse Accessions of Asian Cultivated Rice. *Nature* 557 (7703), 43–49. doi:10.1038/s41586-018-0063-9
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. (2017). Predicting Rice Hybrid Performance Using Univariate and Multivariate GBLUP Models Based on North Carolina Mating Design II. *Heredity* 118 (3), 302–310. doi:10.1038/hdy.2016.87
- White, P. J., and Broadley, M. R. (2009). Biofortification of Crops with Seven Mineral Elements Often Lacking in Human Diets - Iron, Zinc, Copper, Calcium, Magnesium, Selenium and Iodine. *New Phytol.* 182 (1), 49–84. doi:10.1111/j.1469-8137.2008.02738.x
- Wu, X., Islam, A. S. M. F., Limpot, N., Mackasmiel, L., Mierzwa, J., Cortés, A. J., et al. (2020). Genome-wide SNP Identification and Association Mapping for Seed Mineral Concentration in Mung Bean (*Vigna radiata* L.). *Front. Genet.* 11, 1664–8021. doi:10.3389/fgene.2020.00656
- Xu, Y., Ma, K., Zhao, Y., Wang, X., Zhou, K., Yu, G., et al. (2021). Genomic Selection: A Breakthrough Technology in Rice Breeding. *Crop J.* (93), 669–677. doi:10.1016/j.cj.2021.03.008
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs Explain a Large Proportion of the Heritability for Human Height. *Nat. Genet.* 42 (7), 565–569. doi:10.1038/ng.608
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* 88 (1), 76–82. doi:10.1016/j.ajhg.2010.11.011
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and Pitfalls in the Application of Mixed-Model Association Methods. *Nat. Genet.* 46 (2), 100–106. doi:10.1038/ng.2876
- Yang, M., Lu, K., Zhao, F.-J., Xie, W., Ramakrishna, P., Wang, G., et al. (2018). Genome-Wide Association Studies Reveal the Genetic Basis of Ionomics Variation in Rice. *Plant Cell.* 30 (11), 2720–2740. doi:10.1105/tpc.18.00375
- Yu, X., Li, X., Guo, T., Zhu, C., Wu, Y., Mitchell, S. E., et al. (2016). Genomic Prediction Contributing to a Promising Global Strategy to Turbocharge Gene Banks. *Nat. Plants* 2 (10), 16150. doi:10.1038/nplants.2016.150
- Zhang, J., Zhu, Y., Yu, L., Yang, M., Zou, X., Yin, C., et al. (2022). Research Advances in Cadmium Uptake, Transport and Resistance in Rice (*Oryza sativa* L.). *Cells.* 11 (3), 569. doi:10.3390/cells11030569
- Zhang, Z., Ober, U., Erbe, M., Zhang, H., Gao, N., He, J., et al. (2014). Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. *PLoS one* 9 (3), e93017. doi:10.1371/journal.pone.0093017
- Zhou, G., Zhu, Q., Mao, Y., Chen, G., Xue, L., Lu, H., et al. (2021). Multi-Locus Genome-wide Association Study and Genomic Selection of Kernel Moisture Content at the Harvest Stage in Maize. *Front. Plant Sci.* 12, 1664–462X. doi:10.3389/fpls.2021.697688

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Muvunyi, Zou, Zhan, He and Ye. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.