

## Article

# Genomic Prediction from Multi-Environment Trials of Wheat Breeding

Guillermo García-Barrios <sup>1</sup>, Leonardo Crespo-Herrera <sup>2</sup>, Serafín Cruz-Izquierdo <sup>1</sup>, Paolo Vitale <sup>2</sup>, José Sergio Sandoval-Islas <sup>3</sup>, Guillermo Sebastián Gerard <sup>2</sup>, Víctor Heber Aguilar-Rincón <sup>1</sup>, Tarsicio Corona-Torres <sup>1</sup>, José Crossa <sup>2,4,\*</sup> and Rosa Angela Pacheco-Gil <sup>2,\*</sup>

- <sup>1</sup> Postgrado en Recursos Genéticos y Productividad-Genética, Colegio de Postgraduados, Texcoco 56264, Estado de México, Mexico; garcia.guillermo@colpos.mx (G.G.-B.); sercruz@colpos.mx (S.C.-I.); aheber@colpos.mx (V.H.A.-R.); tcoro-nat@colpos.mx (T.C.-T.)
- <sup>2</sup> International Maize and Wheat Improvement Center (CIMMYT), Km 35 Carretera México-Veracruz, Texcoco 56237, Estado de México, Mexico; l.crespo@cgiar.org (L.C.-H.); p.vitale@cgiar.org (P.V.); guillermosgerard@gmail.com (G.S.G.)
- <sup>3</sup> Postgrado en Fitosanidad, Colegio de Postgraduados, Texcoco 56264, Estado de México, Mexico; sandoval@colpos.mx
- <sup>4</sup> Posgrado en Socioeconomía Estadística e Informática, Colegio de Postgraduados, Texcoco 56264, Estado de México, Mexico
- \* Correspondence: j.crossa@cgiar.org (J.C.); r.a.pacheco@cgiar.org (R.A.P.-G.)

**Abstract:** Genomic prediction relates a set of markers to variability in observed phenotypes of cultivars and allows for the prediction of phenotypes or breeding values of genotypes on unobserved individuals. Most genomic prediction approaches predict breeding values based solely on additive effects. However, the economic value of wheat lines is not only influenced by their additive component but also encompasses a non-additive part (e.g., additive  $\times$  additive epistasis interaction). In this study, genomic prediction models were implemented in three target populations of environments (TPE) in South Asia. Four models that incorporate genotype  $\times$  environment interaction ( $G \times E$ ) and genotype  $\times$  genotype (GG) were tested: Factor Analytic (FA), FA with genomic relationship matrix (FA + G), FA with epistatic relationship matrix (FA + GG), and FA with both genomic and epistatic relationship matrices (FA + G + GG). Results show that the FA + G and FA + G + GG models displayed the best and a similar performance across all tests, leading us to infer that the FA + G model effectively captures certain epistatic effects. The wheat lines tested in sites in different TPE were predicted with different precisions depending on the cross-validation employed. In general, the best prediction accuracy was obtained when some lines were observed in some sites of particular TPEs and the worse genomic prediction was observed when wheat lines were never observed in any site of one TPE.



**Citation:** García-Barrios, G.; Crespo-Herrera, L.; Cruz-Izquierdo, S.; Vitale, P.; Sandoval-Islas, J.S.; Gerard, G.S.; Aguilar-Rincón, V.H.; Corona-Torres, T.; Crossa, J.; Pacheco-Gil, R.A. Genomic Prediction from Multi-Environment Trials of Wheat Breeding. *Genes* **2024**, *15*, 417. <https://doi.org/10.3390/genes15040417>

Academic Editor: Qinghu Ma

Received: 24 February 2024

Revised: 24 March 2024

Accepted: 26 March 2024

Published: 27 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** bread wheat; genomic selection; factor analytic; epistatic variance

## 1. Introduction

Genomic selection is an alternative in breeding programs to address complex traits that are regulated by many genes that display minor effects [1]. Breeding value predictions are based on a covariance matrix that describes the additive relationship between the individuals considered [2]. However, an individual's total genetic value is determined through a combination of additive and non-additive effects (dominance and epistasis) [3,4].

In quantitative genetics, epistasis refers to any interaction between genotypes in two or more loci [5]. If genotypes for the different loci display epistatic interactions, these result in a variance component, which is the variance of the epistatic interactions [6].

Epistatic variance is subdivided into components. First, a subdivision is carried out according to the number of loci involved: the two-factor interaction comes from the

interaction of both loci, the three-factor interaction, from three loci, etc. The interactions involving a greater number of loci contribute so little to the variance that they can be ignored [6].

The subsequent subdivision of epistatic variance is based on whether the interaction involves breeding values or dominance deviations; thus, three types of two-factor interactions emerge. The interaction between the two breeding values results in additive  $\times$  additive variance ( $V_{AA}$ ), the interaction between the breeding values of a locus and the dominance deviation of the other results in additive  $\times$  dominance variance ( $V_{AD}$ ), and the interaction between the two dominance deviations results in the dominance  $\times$  dominance variance ( $V_{DD}$ ). Thus, the epistatic variance splits into components in the following way:  $V_I = V_{AA} + V_{AD} + V_{DD} + etc.$ , where the additional terms are similar components that rise from the epistatic interactions between over two loci [6].

Modeling additive  $\times$  additive epistatic effects in genomic prediction can be restrictive because of the high computational burden caused by the large number of interactions among markers when all the interactions are considered [7,8]. An alternative to modeling epistasis and reducing the computational burden is to use models that include genomic relationship matrices as covariance structures for individuals [7].

Commonly, interactions between pairs of loci are modeled through regression coefficients of the Hadamard products of the genomic relationship matrix with itself. This approach has been implemented in several genomic prediction studies, covering wheat [7,9], sugar beet, rye [10], potato [11], *Pinus taeda* [12] and Japanese larch [13].

Initially, epistatic effects can be exploited in the theory of quantitative genetic selection [14]. However, estimating epistatic effects and their inclusion in breeding values have produced variable results in prediction accuracy.

In barley, Abed et al. [15] reported that modeling epistasis resulted in slight increases in the prediction accuracy of days to heading, days to maturity and plant height, but no advantages were evident for fusarium resistance traits, weight and grain yield.

Lorenzana and Bernardo [16] pointed out that including epistatic effects in prediction models resulted in a lower prediction accuracy compared to the models that only considered additive effects. This was observed in multiple *Arabidopsis*, maize and barley traits.

In research conducted by Wang et al. [17] in winter wheat, models incorporating marker and epistasis information were found to outperform models using marker information alone.

An aspect that has been scarcely explored is the epistatic interaction across environments. Multi-environment trials have been widely used in genetic plant breeding to evaluate genotype by environment interactions ( $G \times E$ ) and play an important role when selecting stable and high-yielding lines in different environments or lines adapted to local environmental constraints [18].

The aim of modeling epistasis and including multiple environments in genomic prediction is to achieve a greater prediction accuracy in comparison with the conventional models, which only consider additive effects. Hence, the intention of partially replacing costly experiments and/or save time, increasing the selection gain [9,19].

Multi-environment trials can be conducted across various sites and years, constituting a target population of environments (TPE). During evaluations, genotypes are grown under environmental conditions that differ among farms and years. These trials provide information to predict how a genotype will perform across the TPE [20,21].

Factor Analytic (FA) is a statistical method to model genotype  $\times$  environment ( $G \times E$ ) interaction, adjusted within the framework of mixed models. Its implementation requires calculating the  $G \times E$  covariance matrix (which represents the relationships between genotypes and environments). The FA model efficiently captures common patterns of  $G \times E$  interaction [22–25].

The aims of this investigation were: (1) to predict the grain yield of bread wheat in three TPEs in South Asia and (2) to evaluate the predictive ability of four genomic prediction models: Factor Analytic (FA), Factor Analytic with a genomic relationship matrix (FA + G),

Factor Analytic with an epistatic relationship matrix (FA + GG) and Factor Analytic with a genomic and epistatic relationship matrix (FA + G + GG).

## 2. Materials and Methods

### 2.1. Phenotypic Data

A total of 485 bread wheat genotypes, developed by the International Maize and Wheat Improvement Center (CIMMYT) were used, and distributed in 162 wheat cultivars, which were included in the target population of environment 1 (TPE1). 161 wheat cultivars were in tested TPE2 and 162 wheat lines were included in sites of TPE3. The TPEs in question are found in South Asia in India, Nepal and Bangladesh.

TPE1 contains six locations in India and one in Nepal. The sites in which the experiments were conducted were the Borlaug Institute for South Asia (Ludhiana, Punjab, India), Punjab Agricultural University (Ludhiana, Punjab, India), Indian Institute of Wheat and Barley Research (Karnal, Haryana, India), Indian Agricultural Research Institute (New Delhi, India), Chaudhary Charan Singh Haryana Agricultural University (Hisar, Haryana, India) and Khumaltar, Nepal.

TPE2 had three locations in India, one in Nepal and one in Bangladesh. The sites in which the experiments were conducted were Bihar Agriculture University (Bhagalpur, Bihar, India), (Kalyani, Bengala, India), (Siddharthanagar, Lumbini, Nepal), the Borlaug Institute for South Asia (Pusa, Bihar, India) and (Jamalpur, Bangladesh).

TPE3 was represented in three sites in India: The Agharkar Research Institute (Pune, Maharashtra, India), the Borlaug Institute for South Asia (Jabalpur, Madhya Pradesh, India) and the Indian Agricultural Research Institute (Indore, Madhya Pradesh, India).

Yield evaluations were conducted between 2021 and 2023, using an alpha lattice experimental design with two repetitions. For more details of the grain yield evaluation methodology, see Pietragalla and Pask [26]. Usually sites and wheat breeding lines for each TPE change from year to year.

The wheat lines were genotyped using the genotyping-by-sequencing method [27], and libraries were sequenced using the Illumina NovaSeq 6000 (Illumina, CA, USA). SNP calling was conducted using the system patented by DArT P/L (Canberra, Australia).

SNP marker matrix was transformed to the numeric format 1, 0, −1, where 1 indicates the individual is homozygous for the most frequent allele, 0 denotes heterozygosity and −1 indicates homozygosity for the least frequent allele. The matrix marker was further edited and updated applying the standard norms of quality control by removing markers with over 40% missing data and minor allele frequencies lower than 5%. The missing data were imputed using the expectation–maximization algorithm. From this marker data the genomic similarity matrix (G) matrix was developed and used for the genomic prediction.

### 2.2. Statistical Models

Implementing the mixed model was carried out following Crossa et al. [28]:

$$Y = Xb + Z_r r + Z_g g + Z_{ge} ge + e$$

where  $Y$  is the vector of phenotypic observations,  $X$ ,  $Z_r$ ,  $Z_g$  and  $Z_{ge}$  are the design matrices for the fixed effects of sites, random effects of replications within sites, genotypes and  $G \times E$  interaction, respectively. Vector  $b$  denotes the fixed effects of sites and vectors  $r$ ,  $g$ ,  $ge$  and  $e$  contain the random effects of replications within sites, genotypes, genotype  $\times$  environment interaction, and residuals, respectively.

The assumption is made that  $r$ ,  $g$ ,  $ge$  and  $e$  are random and normally distributed, with zero-mean vectors and variance-covariance matrices  $R$ ,  $G$ ,  $GE$  and  $E$ , respectively, such that:

$$\begin{bmatrix} r \\ g \\ ge \\ e \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R & 0 & 0 & 0 \\ 0 & G & 0 & 0 \\ 0 & 0 & GE & 0 \\ 0 & 0 & 0 & E \end{bmatrix} \right)$$

The variance in  $Y$  is  $V(Y) = Z_r R Z_r' + Z_g G Z_g' + Z_{ge} G E Z_{ge}' + E$ . Under the assumption, the  $R$  and  $E$  variance-covariance matrices have a simple variance component structure.

$$R = \Sigma_r \otimes I_r = \begin{bmatrix} \sigma_{r_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{r_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{r_s}^2 \end{bmatrix} \otimes I_{r'}$$

$$E = \Sigma_e \otimes I_{rg} = \begin{bmatrix} \sigma_{e_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{e_s}^2 \end{bmatrix} \otimes I_{rg}$$

where  $r$  is the number of repetitions,  $g$  is the number of genotypes,  $I_{r'}$  and  $I_{rg}$  are the identity matrices of orders  $r$  and  $rg$ , respectively.  $\Sigma_r$  and  $\Sigma_e$  are the repetition and residual variance matrices, respectively.  $\otimes$  is the Kronecker product of both matrices. The residuals are assumed to have a normal multivariate distribution normal with a mean of zero and an  $E$  covariance matrix.

In order to model the terms  $G$  and  $GE$  in terms of some hypothetical factors, the factor analytic structure was used, which consisted in: suppose that the effect of the unobservable factor of the  $i$ -th genotype in the  $j$ -th environment can be expressed as  $\sum_{k=1}^t \delta_{ik} x_{jk} + d_{ij}$ , where  $\delta_{ik}$  is the  $k$ -th random regression coefficient of the  $i$ -th genotype;  $x_{jk}$  is the  $k$ -th latent variable related to the  $j$ -th site and  $d_{ij}$  is the residual of the end of the interaction. In matrix terms, the vector of genotypic effects is represented by  $g = \Delta x + d$ , so the variance-covariance matrix of  $g$  is  $V(g) = \Delta V(x) \Delta' + D$  and given that  $V(g) = I$ , then  $V(g) = \Delta \Delta' + D$ . With this, one can define that the structure of the analysis of factors with  $q \leq e$  factors (or components) (FA( $q$ )) is in the form  $\Delta \Delta' + D$ , where  $\Delta$  is an  $e \times q$  matrix of the  $\delta$ 's and  $D$  is the positive diagonal matrix  $e \times e$  [29].

The implemented models were named FA, FA + G, FA + GG and FA + G + GG. FA model was the simplest and was built with grain yield data at different sites and considering the  $G \times E$  interaction. The additive model FA + G included the genomic relationship matrix derived from markers in the form:

$$G_A = \frac{(X - \mu_E)(X - \mu_E)^T}{2 \sum_{j=1}^p p_j (1 - p_j)}$$

where  $X$  is the SNP marker matrix of dimensions  $j \times p$ .  $p_j$  is the minor allele frequency of SNP  $j = 1, \dots, p$  and  $\mu_E$  is the expected value of matrix  $X$  under the Hardy–Weinberg equilibrium.

FA + GG model considers the additive  $\times$  additive epistatic effects. The first order epistatic relationship matrix was computed using the Hadamard product (cell-by-cell multiplication)  $G_A \# G_A$ .

The most comprehensive model was FA + G + GG, which includes the genomic relationship matrices and the additive  $\times$  additive epistatic relationship matrices. These two matrices were constructed using the sommer package [30] in R [31]. The models were adjusted using the software ASReml-R 4.1 [32].

### 2.3. Cross-Validation Schemes

Three cross-validation schemes (CV) were implemented. CV1 comprised predicting the breeding value of genotypes that have not been evaluated; 80% of the genotypes were used for training and 20% for testing. In CV2, the breeding value of genotypes that have been evaluated in some sites but not in others was predicted. The training and testing sets consisted of disjoint sets of genotypes. CV3 comprised predicting the breeding value only at one site, with 10% of genotypes for training and the remaining 90% for testing; at

other sites, all genotypes were included in the training set. Table 1 contains an example of the number of genotypes used for training and testing under the validation schemes in TPE1 sites.

**Table 1.** Example of the number of genotypes used in the training and testing set for 1-fold. In CV1, the breeding value of 32 genotypes was predicted across six locations. CV2 used prior performance data from Ludhiana (PAU), New Delhi, and Khumaltar to predict the performance of 32 genotypes at Ludhiana (BISA), Karnal, and Hisar. In CV3, the breeding value of 142 genotypes was predicted in Ludhiana (BISA).

Cross-Validation Schemes	Set	Ludhiana (BISA)	Ludhiana (PAU)	Khumaltar	Karnal	New Delhi	Hisar
CV1	Training	130	130	130	130	130	130
	Testing	32	32	32	32	32	32
CV2	Training	130	162	162	130	162	130
	Testing	32	0	0	32	0	32
CV3	Training	20	162	162	162	162	162
	Testing	142	0	0	0	0	0

Each model's accuracy was determined by the average of the Pearson correlation between the predicted and observed breeding values of the test set in the 5-fold. The results were plotted using the R package ggpubr [33].

### 3. Results

#### 3.1. Descriptive Yield Statistics in the TPEs

The average grain yield in TPE1 was  $5.88 \pm 0.4 \text{ t ha}^{-1}$ . The yield correlations between sites were variable, from 0.9 (between Karnal and New Delhi) down to an absence of correlation (0.12) between Ludhiana (PAU) and Ludhiana (BISA) (Table 2).

**Table 2.** Yield correlation between sites of TPE1.

	Ludhiana (BISA)	Ludhiana (PAU)	Khumaltar	Karnal	New Delhi	Hisar
Ludhiana (BISA)	1					
Ludhiana (PAU)	0.12	1				
Khumaltar	0.20	0.00	1			
Karnal	0.59	0.14	0.21	1		
New Delhi	0.55	0.18	0.23	0.90	1	
Hisar	0.35	0.18	0.48	0.42	0.48	1

In TPE2, the average grain yield obtained was  $4.47 \pm 0.27 \text{ t ha}^{-1}$ . The yield correlations between sites were positive and moderate, fluctuating between 0.5 and 0.69 (Table 3).

**Table 3.** Yield correlation between sites of TPE2.

	Bhagalpur	Jamalpur	Pusa	Bhairahawa	Kalyani
Bhagalpur	1				
Jamalpur	0.69	1			
Pusa	0.63	0.64	1		
Bhairahawa	0.57	0.59	0.50	1	
Kalyani	0.69	0.64	0.60	0.62	1

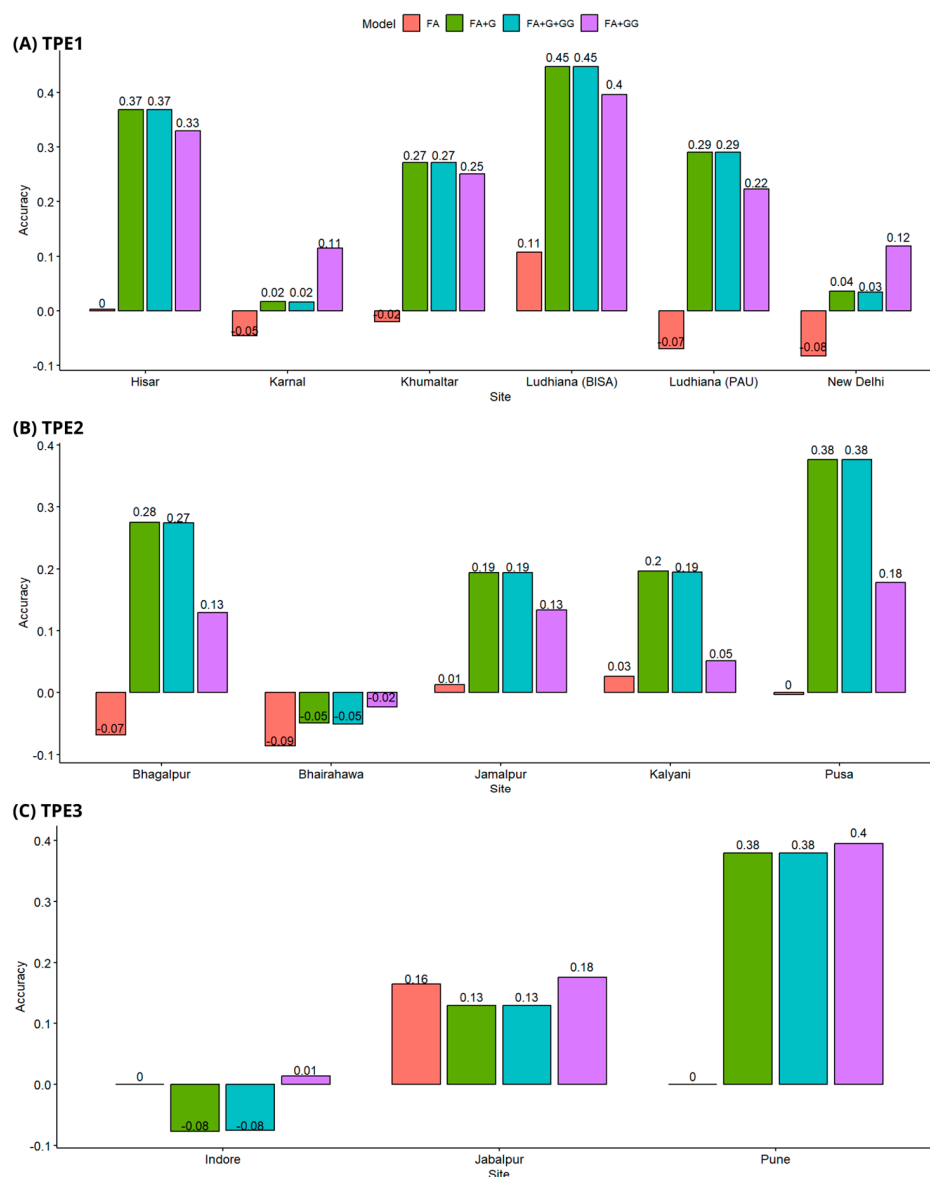
In TPE3, the average grain yield was  $5.74 \pm 0.29 \text{ t ha}^{-1}$ . The locations of Indore and Jabalpur had a correlation of 0.24, whereas there was an absence of correlation between the locations of Pune and Indore, Indore and Jabalpur (Table 4).

**Table 4.** Yield correlation between sites of TPE3.

	Pune	Indore	Jabalpur
Pune	1		
Indore	0.09	1	
Jabalpur	−0.04	0.24	1

**3.2. Prediction of the Breeding Value of Non-Evaluated Genotypes (Cross-Validation CV1)**

In TPE1, the sites with the best prediction accuracy were Ludhiana (BISA) and Hisar. The models with the best performance for these sites were FA + G y FA + G + GG, with prediction accuracies ranging from  $0.45 \pm 0.06$  to  $0.37 \pm 0.12$ . The sites with the lowest accuracies were New Delhi and Karnal. For these two sites, the epistatic model FA + GG had the best performance. The model with the worst average performance in TPE1 was FA (Figure 1A).



**Figure 1.** Prediction accuracy scheme CV1, (A) Predictive accuracy in TPE1, (B) Predictive accuracy in TPE2, (C) Predictive accuracy in TPE3. Models Tested: FA (Factor Analytic), FA + G (FA with genomic relationship matrix), FA + GG (FA with epistatic relationship matrix), FA + G + GG (FA with genomic and epistatic relationship matrices).

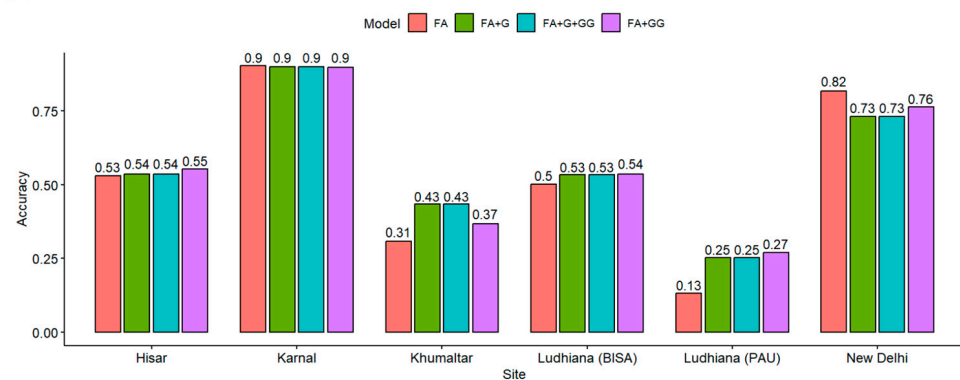
In TPE2 the models with the best performance were also FA + G y FA + G + GG, with accuracies that ranged from  $0.19 \pm 0.14$  for Kalyani to  $0.19 \pm 0.17$  for Jamalpur to  $0.38 \pm 0.13$  for Pusa. The site of Bhairahawa had negative predictions for all models (Figure 1B).

In TPE3, the location of Pune had the best predictions, with epistatic model FA + GG having the best performance with an accuracy of  $0.4 \pm 0.06$ . For the location of Indore, no model achieved acceptable predictions. The model with the lowest performance in TPE3 was FA, which did not adjust due to having insufficient degrees of freedom (Figure 1C).

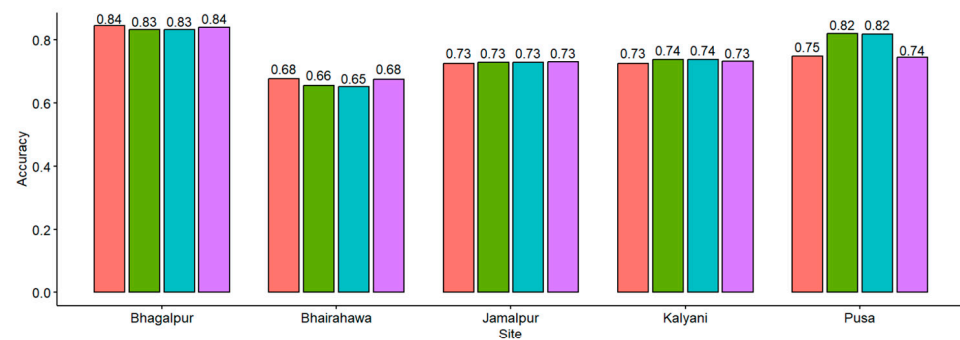
### 3.3. Prediction of Breeding Value in Genotypes Evaluated in Some Environments, but Not in All (Cross-Validation CV2)

In predicting the breeding values of genotypes that have been evaluated in some environments, but not in all, the FA + G and FA + G + GG models had similar performances across all TPEs (Figure 2).

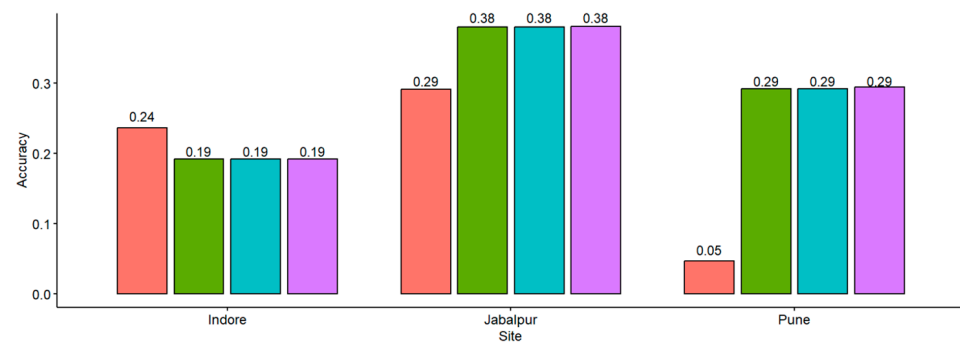
#### (A) TPE1



#### (B) TPE2



#### (C) TPE3



**Figure 2.** Prediction accuracy scheme CV2, (A) Predictive accuracy in TPE1, (B) Predictive accuracy in TPE2, (C) Predictive accuracy in TPE3. Models Tested: FA (Factor Analytic), FA + G (FA with genomic relationship matrix), FA + G + GG (FA with genomic and epistatic relationship matrices), FA + GG (FA with epistatic relationship matrix).

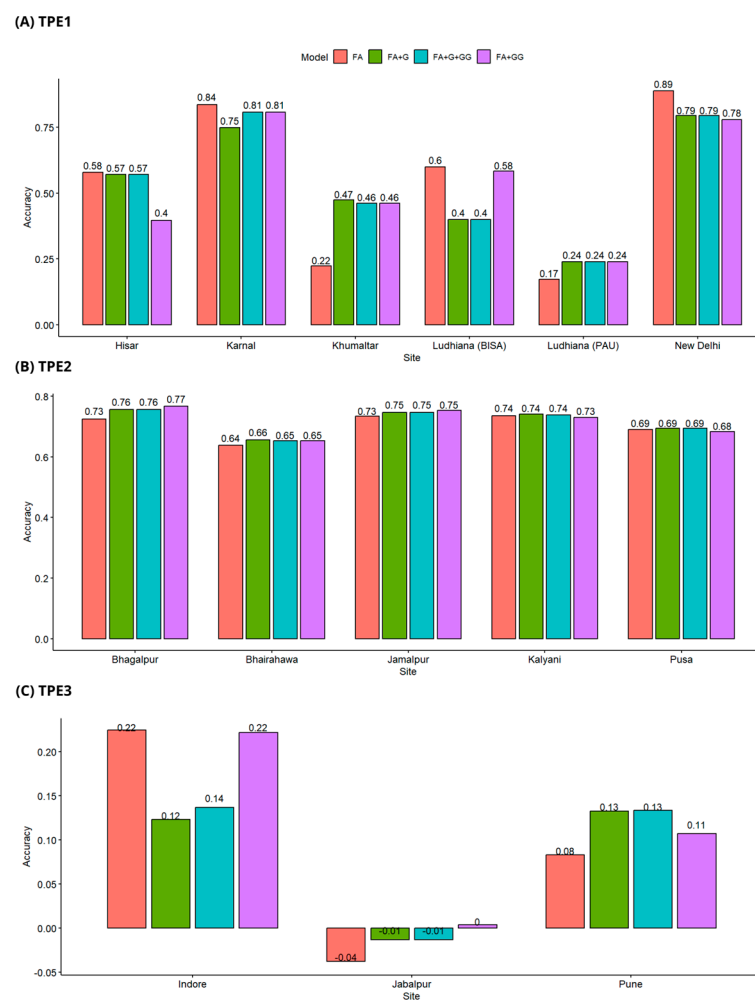
In TPE1, the site of Karnal had the best predictions, reaching the highest accuracy of  $0.90 \pm 0.06$  with the four models tested. The site with the worst predictions was Ludhiana (PAU), with a maximum accuracy for this site being  $0.27 \pm 0.19$  (Model FA + GG), and the lowest being  $0.13 \pm 0.14$  (Model FA) (Figure 2A).

TPE2 had a better prediction accuracy than TPE1 and TPE3 under the CV2 validation scheme. The four models tested in TPE2 had a consistent performance across all sites, with the highest precision being  $0.84 \pm 0.06$ , corresponding to the FA and FA + GG models in the site of Bhagalpur, and the lowest precision  $0.65 \pm 0.11$  for models FA + G and FA + G + GG in the site of Bhairahawa (Figure 2B).

In TPE3, models FA + G, FA + GG and FA + G + GG had a consistent performance, obtaining the same prediction accuracy within sites. FA had a variable performance, being the worst model for the site of Pune ( $0.05 \pm 0.18$  accuracy), yet this model achieved the highest precision for the site of Indore ( $0.24 \pm 0.24$ ) (Figure 2C).

### 3.4. Predicting the Breeding Value Using 90% of Genotypes for Training (Cross-Validation CV3)

In the validation scheme CV3, models FA + G and FA + G + GG had a similar performance in the three TPEs, with similar prediction precisions. FA had an acceptable performance in TPE1 and TPE2, which contain six and five sites, respectively, although this same model had a bad performance in TPE3, which is composed of three sites (Figure 3).



**Figure 3.** Prediction accuracy scheme CV3, (A) Predictive accuracy in TPE1, (B) Predictive accuracy in TPE2, (C) Predictive accuracy in TPE3. Models Tested: FA (Factor Analytic), FA + G (FA with genomic relationship matrix), FA + G + GG (FA with genomic and epistatic relationship matrices), FA + GG (FA with epistatic relationship matrix).



In TPE1, the sites that had the best predictions were New Delhi and Karnal, with the highest precisions being 0.89 and 0.84, respectively, with the model FA. The site of Ludhiana (PAU) had the worst prediction accuracies, with values of 0.17 for model FA and 0.24 for the rest of the models (Figure 3A).

In TPE2, the four models tested had an acceptable performance and there was minor variation between them. The highest prediction accuracy was 0.77, which corresponds to the site of Bhagalpur and the epistatic model FA + GG. The lowest accuracy was 0.64, which corresponds to the location of Bhairahawa and the model FA (Figure 3B).

Out of the three environments evaluated, TPE3 displayed the worst accuracies. The site of Indore reached the highest accuracies (0.22) with models FA and FA + GG (Figure 3C).

## 4. Discussion

### 4.1. Including Epistasis in the Genomic Prediction Models of Bread Wheat Does Not Increase Prediction Accuracy

In this study, we evaluated four prediction models that considered the additive and epistatic effects, as well as their combination. The model that included the additive and epistatic effects (FA + G + GG) was expected to have a higher prediction accuracy than the rest of the models, although FA + G + GG had a similar performance to the additive model FA + G in almost all tests.

Mackay [5] establishes a difference in the additive genetic variance, considering it “real” when most of the loci that affect the trait have additive genetic action, or “apparent” when the main non-zero effects arise from the action of the epistatic gene in many loci. Under this distinction, we can suggest that our additive model could capture some of the epistatic effects; therefore, the additive (FA + G) and additive + epistatic (FA + G + GG) models displayed similar performances.

The average effects of individual loci may capture a portion of epistasis, implying that interaction effects between loci become “apparent” additive effects; this is more likely to happen when allele frequencies approach 0 or 1 due to genetic drift or selection [34–36].

In a study conducted on *P. taeda* by Muñoz et al. [12] modest variations in the Akaike information criterion were obtained when comparing additive models and models that simultaneously considered additive and non-additive effects. The authors suggest additive models capture part of the genetic variation, due to the dominance and epistasis and that the additive model should not suffer much if these components are omitted.

Dong et al. [13] implemented genomic prediction models related to growth, physical and chemical properties in wood in *Larix kaempferi*. Their results showed that the traditional GLUP model performed similarly to GLUP with non-additive effects. These results coincide with those obtained in this study.

A study by Wang et al. [17] presented opposite results to ours. They reported that the consideration of epistatic effects on the genomic prediction models resulted in a higher prediction accuracy. However, this study was conducted considering only the effects of major markers and ignores that epistasis is not limited to interactions between loci associated with the trait, but interactions also occur with loci that have no significant effects on the traits when evaluated individually.

Wientjes et al. [37] conducted a simulation study on a livestock population where they evaluated genomic prediction models. In the first generation of selection, the accuracy of the additive and additive + dominance models was ~0.83, while it was ~0.72 for the model incorporating additive, dominance, and epistatic effects. These results are consistent with those observed in our study, wherein the inclusion of additional factors in the model did not lead to improved predictive accuracy.

### 4.2. Cross-Validation Schemes

Having information about performing a line across multiple sites makes it easier to predict its performance at a new site, while predicting the performance of a newly developed line (with no history) is more challenging and requires using the performance

of similar genotypes for predictions. This is supported by our results, where the CV2 validation scheme had an advantage over CV1.

In TPE2, under the CV2 and CV3 scheme, prediction accuracies ranged from 0.64 to 0.84. As a general rule, a correlation of 0.5 is acceptable, given that prediction accuracy is the correlation between the phenotypic value and the predicted breeding value, it is common for accuracy to be downward biased because the phenotype is an imperfect estimator of the breeding value [38].

The good accuracy accomplished in TPE2 with the CV2 and CV3 scheme is possibly an indicator that genomic selection can be implemented in this TPE, which would allow for the reduction in costly experiments and increase selection gains. Further studies in TPE2 may include evaluations over seasons or years.

The advantage of the CV3 validation scheme over CV1 may be attributed to the larger size of the training set in CV3, thus capable of capturing more variance than a small training set and therefore generating high prediction accuracies.

#### *4.3. Advantages of the Factor Analytic Model for Studying Genotype $\times$ Environment Interaction in the Context of Genomic Prediction*

The FA offers flexibility in capturing the underlying structure of  $G \times E$  interaction. It allows for the estimation of latent factors that represent the main patterns of  $G \times E$  interaction, which can be beneficial for understanding the genetic basis of complex traits and designing breeding strategies. Also, FA identifies latent factors associated with  $G \times E$  interaction. This can help breeders understand the genetic architecture of traits and make informed decisions regarding genotype selection and environmental management. However, FA may require more computational resources compared to simpler models. Efficient algorithms and parallel computing techniques can help improve the computational efficiency of FA.

## 5. Conclusions

It was confirmed that additive models capture part of the epistatic variance, showing the suitability of the FA + G model to include in genomic selection programs. Our research also corroborates the efficiency of the Factor Analytic model to capture  $G \times E$  interactions, particularly when integrating information from multiple sites. Furthermore, our results indicated that the FA model exhibited good performance when using data from six to five sites (TPE1 and TPE2, respectively); however, it performed poorly in TPE3, which contained only three sites.

**Author Contributions:** Conceptualization, J.C. and L.C.-H.; methodology, S.C.-I. and P.V.; software, R.A.P.-G.; validation, J.S.S.-I., G.S.G. and V.H.A.-R.; formal analysis, G.G.-B.; investigation, T.C.-T., P.V. and J.C.; resources, L.C.-H.; data curation, R.A.P.-G.; writing—original draft preparation, G.G.-B.; writing—review and editing, J.S.S.-I., S.C.-I. and G.S.G.; visualization, T.C.-T.; supervision, J.C. and R.A.P.-G.; project administration, L.C.-H.; funding acquisition, J.C. and L.C.-H. All authors have read and agreed to the published version of the manuscript.

**Funding:** We acknowledge the financial support provided by the Bill and Melinda Gates Foundation (INV-003439 BMGF/FCDO Accelerating Genetic Gains in Maize and Wheat for Improved Livelihoods (AGG)) as well as the USAID projects (Amend. No. 9 MTO 069033, USAID-CIMMYT Wheat/AGGMW, Genes 2023, 14, 927 14 of 18 AGG-Maize Supplementary Project, AGG (Stress Tolerant Maize for Africa)), which generated the CIMMYT data analyzed in this study.

**Data Availability Statement:** The data utilized in this study, encompassing both phenotypic and genomic information, are available for download from the following link. <https://github.com/guillermogarci4/MultiEnv-Wheat-Genomic-Prediction.git> (accessed on 1 March 2024).

**Acknowledgments:** The authors are thankful for the support of the technicians involved in field data collection. GGB expresses gratitude to CONAHCYT (Consejo Nacional de Humanidades, Ciencias y Tecnologías) for the doctoral fellowship.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Meuwissen, T.H.; Hayes, B.J.; Goddard, M.E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **2001**, *157*, 1819–1829. [[CrossRef](#)]
2. VanRaden, P.M. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)] [[PubMed](#)]
3. Alves, K.; Brito, L.F.; Baes, C.F.; Sargolzaei, M.; Robinson, J.A.B.; Schenkel, F.S. Estimation of Additive and Non-Additive Genetic Effects for Fertility and Reproduction Traits in North American Holstein Cattle Using Genomic Information. *J. Anim. Breed. Genet.* **2020**, *137*, 316–330. [[CrossRef](#)]
4. Su, G.; Christensen, O.F.; Ostersen, T.; Henryon, M.; Lund, M.S. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE* **2012**, *7*, e45293. [[CrossRef](#)]
5. Mackay, T.F.C. Epistasis and Quantitative Traits: Using Model Organisms to Study Gene–Gene Interactions. *Nat. Rev. Genet.* **2014**, *15*, 22–33. [[CrossRef](#)] [[PubMed](#)]
6. Falconer, D.S.; Mackay, T. *Introduction to Quantitative Genetics*, 4th ed.; Pearson, Prentice Hall: Harlow, UK, 1996.
7. Raffo, M.A.; Sarup, P.; Guo, X.; Liu, H.; Andersen, J.R.; Orabi, J.; Jahoor, A.; Jensen, J. Improvement of Genomic Prediction in Advanced Wheat Breeding Lines by Including Additive-by-Additive Epistasis. *Theor. Appl. Genet.* **2022**, *135*, 965–978. [[CrossRef](#)] [[PubMed](#)]
8. Vojgani, E.; Hölker, A.C.; Mayer, M.; Schön, C.-C.; Simianer, H.; Pook, T. Genomic Prediction Using Information across Years with Epistatic Models and Dimension Reduction via Haplotype Blocks. *PLoS ONE* **2023**, *18*, e0282288. [[CrossRef](#)] [[PubMed](#)]
9. Martini, J.W.R.; Wimmer, V.; Erbe, M.; Simianer, H. Epistasis and Covariance: How Gene Interaction Translates into Genomic Relationship. *Theor. Appl. Genet.* **2016**, *129*, 963–976. [[CrossRef](#)] [[PubMed](#)]
10. Kristensen, P.S.; Sarup, P.; Fé, D.; Orabi, J.; Snell, P.; Ripa, L.; Mohlfeld, M.; Chu, T.T.; Herrström, J.; Jahoor, A.; et al. Prediction of Additive, Epistatic, and Dominance Effects Using Models Accounting for Incomplete Inbreeding in Parental Lines of Hybrid Rye and Sugar Beet. *Front. Plant. Sci.* **2023**, *14*, 1193433. [[CrossRef](#)] [[PubMed](#)]
11. Endelman, J.B.; Carley, C.A.S.; Bethke, P.C.; Coombs, J.J.; Clough, M.E.; da Silva, W.L.; De Jong, W.S.; Douches, D.S.; Frederick, C.M.; Haynes, K.G.; et al. Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. *Genetics* **2018**, *209*, 77–87. [[CrossRef](#)] [[PubMed](#)]
12. Muñoz, P.R.; Resende, M.F.R.; Gezan, S.A.; Resende, M.D.V.; de los Campos, G.; Kirst, M.; Huber, D.; Peter, G.F. Unraveling Additive from Nonadditive Effects Using Genomic Relationship Matrices. *Genetics* **2014**, *198*, 1759–1768. [[CrossRef](#)] [[PubMed](#)]
13. Dong, L.; Xie, Y.; Zhang, Y.; Wang, R.; Sun, X. Genomic Dissection of Additive and Non-Additive Genetic Effects and Genomic Prediction in an Open-Pollinated Family Test of Japanese Larch. *BMC Genom.* **2024**, *25*, 11. [[CrossRef](#)] [[PubMed](#)]
14. Flury, C.; Täubert, H.; Simianer, H. Extension of the Concept of Kinship, Relationship, and Inbreeding to Account for Linked Epistatic Complexes. *Livest. Sci.* **2006**, *103*, 131–140. [[CrossRef](#)]
15. Abed, A.; Pérez-Rodríguez, P.; Crossa, J.; Belzile, F. When Less Can Be Better: How Can We Make Genomic Selection More Cost-Effective and Accurate in Barley? *Theor. Appl. Genet.* **2018**, *131*, 1873–1890. [[CrossRef](#)] [[PubMed](#)]
16. Lorenzana, R.E.; Bernardo, R. Accuracy of Genotypic Value Predictions for Marker-Based Selection in Biparental Plant Populations. *Theor. Appl. Genet.* **2009**, *120*, 151–161. [[CrossRef](#)]
17. Wang, D.; Salah El-Basyoni, I.; Stephen Baenziger, P.; Crossa, J.; Eskridge, K.M.; Dweikat, I. Prediction of Genetic Values of Quantitative Traits with Epistatic Effects in Plant Breeding Populations. *Heredity* **2012**, *109*, 313–319. [[CrossRef](#)] [[PubMed](#)]
18. Crossa, J.; Montesinos-López, O.A.; Pérez-Rodríguez, P.; Costa-Neto, G.; Fritsche-Neto, R.; Ortiz, R.; Martini, J.W.R.; Lillemo, M.; Montesinos-López, A.; Jarquin, D.; et al. Genome and Environment Based Prediction Models and Methods of Complex Traits Incorporating Genotype × Environment Interaction. In *Genomic Prediction of Complex Traits: Methods and Protocols*; Ahmadi, N., Bartholomé, J., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2022; pp. 245–283. [[CrossRef](#)]
19. Schrauf, M.F.; Martini, J.W.R.; Simianer, H.; de los Campos, G.; Cantet, R.; Freudenthal, J.; Korte, A.; Munilla, S. Phantom Epistasis in Genomic Selection: On the Predictive Ability of Epistatic Models. *G3 Genes Genomes Genet.* **2020**, *10*, 3137–3145. [[CrossRef](#)] [[PubMed](#)]
20. Braun, H.J.; Atlin, G.; Payne, T. Multi-Location Testing as a Tool to Identify Plant Response to Global Climate Change. In *Climate Change and Crop Production*; Reynolds, P.M., Ed.; CABI Climate Change Series; CAB International: Wallingford, UK, 2010; pp. 115–138.
21. Crespo-Herrera, L.A.; Crossa, J.; Huerta-Espino, J.; Mondal, S.; Velu, G.; Juliana, P.; Vargas, M.; Pérez-Rodríguez, P.; Joshi, A.K.; Braun, H.J.; et al. Target Population of Environments for Wheat Breeding in India: Definition, Prediction and Genetic Gains. *Front. Plant Sci.* **2021**, *12*, 638520. [[CrossRef](#)] [[PubMed](#)]
22. Coast, O.; Posch, B.C.; Rognoni, B.G.; Bramley, H.; Gaju, O.; Mackenzie, J.; Pickles, C.; Kelly, A.M.; Lu, M.; Ruan, Y.-L.; et al. Wheat Photosystem II Heat Tolerance: Evidence for Genotype-by-Environment Interactions. *Plant J.* **2022**, *111*, 1368–1382. [[CrossRef](#)] [[PubMed](#)]
23. Kelly, A.M.; Smith, A.B.; Eccleston, J.A.; Cullis, B.R. The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. *Crop. Sci.* **2007**, *47*, 1063–1070. [[CrossRef](#)]
24. Meyer, K. Factor-Analytic Models for Genotype × Environment Type Problems and Structured Covariance Matrices. *Genet. Sel. Evol.* **2009**, *41*, 21. [[CrossRef](#)] [[PubMed](#)]

25. Oliveira, I.C.M.; Guilhen, J.H.S.; Ribeiro, P.C.d.O.; Gezan, S.A.; Schaffert, R.E.; Simeone, M.L.F.; Damasceno, C.M.B.; Carneiro, J.E.d.S.; Carneiro, P.C.S.; Parrella, R.A.d.C.; et al. Genotype-by-Environment Interaction and Yield Stability Analysis of Biomass Sorghum Hybrids Using Factor Analytic Models and Environmental Covariates. *Field Crop. Res.* **2020**, *257*, 107929. [[CrossRef](#)]
26. Pietragalla, J.; Pask, A. Chapter 18. Grain Yield and Yield Components. In *Physiological Breeding II: A Field Guide to Wheat Phenotyping*; Pask, A., Pietragalla, J., Mullan, D., Chávez, P., Reynolds, M.P., Eds.; CIMMYT: México-Veracruz, México, 2013; pp. 95–103.
27. Poland, J.A.; Brown, P.J.; Sorrells, M.E.; Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE* **2012**, *7*, e32253. [[CrossRef](#)]
28. Crossa, J.; Burgueño, J.; Cornelius, P.L.; McLaren, G.; Trethowan, R.; Krishnamachari, A. Modeling Genotype  $\times$  Environment Interaction Using Additive Genetic Covariances of Relatives for Predicting Breeding Values of Wheat Genotypes. *Crop. Sci.* **2006**, *46*, 1722–1733. [[CrossRef](#)]
29. Smith, A.; Cullis, B.; Thompson, R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics* **2001**, *57*, 1138–1147. [[CrossRef](#)] [[PubMed](#)]
30. Covarrubias-Pazarán, G. Genome-Assisted Prediction of Quantitative Traits Using the R Package Sommer. *PLoS ONE* **2016**, *11*, e0156744. [[CrossRef](#)] [[PubMed](#)]
31. R Core Team. R: A Language and Environment for Statistical Computing. 2023. Available online: <https://www.R-project.org/> (accessed on 15 December 2023).
32. Gilmour, A.R.; Gogel, B.J.; Cullis, B.R.; Welham, S.J.; Thompson, R. *ASReml User Guide Release 4.1*; VSN International Ltd.: Hemel Hempstead, UK, 2015; Volume Release 4.1.
33. Kassambara, A. Ggpubr: “ggplot2” Based Publication Ready Plots. 2023. Available online: <https://rpkgs.datanovia.com/ggpubr/#ggpubr-ggplot2-based-publication-ready-plots> (accessed on 15 December 2023).
34. Duenk, P.; Bijma, P.; Calus, M.P.L.; Wientjes, Y.C.J.; van der Werf, J.H.J. The Impact of Non-Additive Effects on the Genetic Correlation between Populations. *G3 Genes Genomes Genet.* **2020**, *10*, 783–795. [[CrossRef](#)] [[PubMed](#)]
35. Hill, W.G.; Goddard, M.E.; Visscher, P.M. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genet.* **2008**, *4*, e1000008. [[CrossRef](#)] [[PubMed](#)]
36. Holland, J.B. Epistasis and Plant Breeding. In *Plant Breeding Reviews*; Janick, J., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001; Volume 21, pp. 27–92. [[CrossRef](#)]
37. Wientjes, Y.C.J.; Bijma, P.; Calus, M.P.L.; Zwaan, B.J.; Vitezica, Z.G.; van den Heuvel, J. The Long-Term Effects of Genomic Selection: 1. Response to Selection, Additive Genetic Variance, and Genetic Architecture. *Genet. Sel. Evol.* **2022**, *54*, 19. [[CrossRef](#)] [[PubMed](#)]
38. McCaugh, S.E.; Lorenz, A.J.; Flagel, L.E. The Utility of Genomic Prediction Models in Evolutionary Genetics. *Proc. R. Soc. B* **2021**, *288*, 20210693. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.