Improving genomic selection in hexaploid wheat with sub-genome additive and epistatic models

Augusto Tessele[1], David O. González-Diéguez, Jose Crossa[2], Blaine E. Johnson[3], Geoffrey P. Morris[4], Allan K. Fritz[1]

[1]Department of Agronomy, Kansas State University, Manhattan, Kansas, USA

[2]Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico City, Mexico

[3]Department of Agronomy and Horticulture, University of Nebraska, Lincoln, Nebraska, USA

[4]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, USA

Abstract

The goal of wheat breeding is the development of superior cultivars tailored to specific environments, and the identification of promising crosses is crucial for the success of breeding programs. Although genomic estimated breeding values were developed to estimate additive effects of genotypes before testing as parents, application has focused on predicting performance of candidate lines, ignoring non-additive genetic effects. However, non-additive genetic effects are hypothesized to be especially importance in allopolyploid species due to the interaction between homeologous genes. The objectives of this study were to model additive and additive-by-additive epistatic effects to better delineate the genetic architecture of grain yield in wheat and to the improve accuracy of genomewide predictions. The dataset utilized consisted of 3740 $F_{5:6}$ experimental lines tested in the K-State wheat breeding program across the years 2016 and 2018. Covariance matrices were calculated based on whole and sub-genome marker data and the natural and orthogonal interaction approach (NOIA) was used to estimate variance components for additive and additive-by-additive epistatic effects. Incorporating epistatic effects in additive models resulted in non-orthogonal partitioning of genetic effects but increased total genetic variance and reduced DIC. Estimation of sub-genome effects indicated that genotypes with the greatest whole genome effects often combine sub-genomes with intermediate to high effects, suggesting potential for crossing parental lines which have complementary sub-genome effects. Modeling epistasis in either whole-genome or sub-genome models led to a marginal (3%) but significant improvement in genomic prediction accuracy, which could result in significant genetic gains across multiple cycles of breeding.

Introduction

In wheat breeding, selection of experimental lines has historically been based on phenotypic selection, which encapsulates all forms of genetic effects. However, utilization of genomic selection (Meuwissen et al. 2001) to predict the performance of candidate lines overlooks the potential contribution of non-additive effects, especially epistasis, which potentially could play an important role in the genetic expression of agronomic traits (Chapman and McNeal, 1971; Sun et al., 1972; Goldringer et al., 1997; Zhang et al., 2008; Jiang et al., 2017). Epistasis is defined as the interaction between alleles of different genes and is normally classified into additive-by-additive, additive-by-dominance, and dominance-be-dominance, with higher order interactions normally ignored.

Historically, complex mating designs were used to create the progeny that could be used to estimate epistatic effects, e.g., the triple test cross (Bauman, 1959). Current breeding programs have an abundance of experimental lines genotyped with phenotypic records that could be leveraged to estimate genetic effects. With large number of markers covering the genome and quantitative trait loci assumed to have a normal distribution, genome-wide markers can be used to estimate the additive covariance between individuals and predict their additive genetic value (Nejati-Javaremi et al. 1997; VanRaden 2008). To obtain an approximation of the additive-by-additive relationship matrix between individuals, Henderson (1985) proposed the use of the Hadamard product of the additive covariance matrix based on pedigree information, which was later expanded to accommodate genomic-based additive relationship matrices (Su et al., 2012). Recently, the natural and orthogonal interaction approach (Alvarez-Castro and Carlborg, 2007) was expanded to build genomic relationship matrices using genotypic frequencies and to include genome-wide epistasis (Vitezica et al 2017), resulting in an orthogonal partitioning of genetic variance components, even if not in Hardy-Weinberg Equilibrium. Thus, there is the possibility of utilizing existing data sets to estimate and better delineate the genetic architecture of important agronomic traits in crop species.

Common wheat (*Triticum aestivum* spp *aestivum*) is a hexaploid species originated from the interspecific hybridization between tetraploid wheat, *Triticum turgidum* (Sax 1922; Kihara 1924), the source of the AABB genome, and *Aegilops tauschii* (Kihara 1944; McFadden and Sears, 1946), the source of the DD genome. The presence of homeologous genomes could create positive or negative epistasis through subfunctionalization (Lynch and Force, 2000) or substrate competition (Qian et al., 2010), while also impacting gene expression patterns (Leach et al., 2014; Akhunova et al., 2010). Intra-genomic interactions have also been reported in wheat (Tranquilli and Dubcovsky, 2000; Reif et al., 2011; Sehgal et al., 2020), underscoring the potential importance of epistasis underpinning agronomic traits. In addition, the inbreeding nature of wheat leads to a fast fixation of alleles (Charlesworth, 2003), revealing epistatic interactions whose favorable combinations are perpetuated with selection of superior genotypes in

breeding programs. Although hexaploid, the three sub-genome of common wheat undergo disomic segregation (Feldman and Levy, 2012) and are normally treated as diploids in breeding programs. For that reason, partitioning whole genome genetic effects into sub-genome opens up the possibility to attribute biological importance to whole genome effects and variance estimates (Santantonio et al., 2019). Thus, objectives of this study were to: (i) partition the genetic variance of wheat grain yield into additive and epistatic effects; (ii) partition whole genome effects into sub-genome; (iii) incorporate epistasis in genomic selection model.

Material and Methods

Estimation of Epistatic Variance

Phenotypic Data

Phenotypic data consisted of grain yield records for 3740 $F_{5:6}$ wheat genotypes grown in the early yield trail stage (IPSR) of the K-State hard red winter wheat breeding program across the years of 2016 and 2018. The IPSR is the first stage of the breeding program where selection is based on grain yield data (previous stages consisted of visual single plant selection). The IPRS represents the foundational genetic diversity within each breeding cycle. The experimental design used for the phenotypic evaluation consisted of unreplicated trials and was a modified augmented design (Federer and Raghavarao, 1975) type II with one replication of each experimental line. However, to quantify and remove spatial variability of these large experiments, two-stage analyses (Smith et al., 2001; Welham et al., 2010) were used, and included fitting a spatial model to account for field patterns associated with environmental factors (Cullis and Gleeson, 1991; Cullis et al., 1998; Smith et al., 2001). The first stage of a two-stage analysis is a within-location analysis while the second stage is across locations analysis. To accommodate spatial adjustments an extra step was added to the first-stage analysis, as reported in Arief et al. (2019). In the first step, genotypes were fitted as random effects to provide estimation of variance components. In the second step of the first-stage, genotypes were fitted as fixed effects and adjusted for spatial trends to calculate the best linear unbiased estimates (BLUEs) and two-stage weights for each location (Smith et al., 2001). BLUEs and weights from the first-stage analysis were used for the second-stage analysis to obtain the BLUPs for each genotype. The phenotypic analyses were executed using the Echidna Mixed Model Software (Gilmour, 2018).

Genotypic Data

The lines analyzed were in the $F_{5:6}$ generation and were genotyped using genotype-by-sequencing (GBS) (Poland et al. 2012). To ensure data quality, markers with a minor allele frequency (MAF) lower than

0.01, more than 20% missing values, and more than 20% heterozygotes were filtered out. The marker imputation process was carried out using the 'rrBLUP' package (Endelman, 2011), resulting in a dataset containing a total of 55,148 single nucleotide polymorphisms (SNPs).

Statistical Models

Four models were used to quantify additive and additive-by-additive epistatic (herein described as epistasis) effects across sub-genomes. The first two models consisted of analyses at the whole genome level while the last two models partitioned genetic effects, both additive and epistatic, by sub-genome. Variance components, for all models, were estimated using the natural and orthogonal interaction approach (NOIA) parametrizations outlined by Alvarez-Castro and Carlborg (2007) and Vitezica et al. (2017).

Whole Genome Models

Model 1 – Additive Effects

The linear model for grain yield that includes only an additive term is given by:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Zg} + \boldsymbol{\varepsilon} \quad (1)$$

where $\boldsymbol{y}$ is the response vector (adjusted phenotypic data), μ is the intercept, $\boldsymbol{\varepsilon}$ is a vector of random residuals with $\boldsymbol{\varepsilon} \sim N(0, \sigma_e^2)$, where σ_e^2 is the residual variance, $\boldsymbol{Z}$ is a design matrix of random effects and $\boldsymbol{g}$ is the vector of additive genomic breeding values with $\boldsymbol{g} \sim N(0, \sigma_A^2\boldsymbol{A})$, where σ_A^2 is the genomic additive genetic variance and the $\boldsymbol{A}$ is an additive genomic relationship constructed according to Vitezica et al. (2017) (see Raffo et al. 2022).

Model 2–- Additive and Epistatic Effects

Following Henderson (1985), the epistatic covariance of individuals can be calculated as the Hadamard product of the component covariance matrices. Hence, the second model used in this study extends the first model by including an additive-by-additive epistatic term:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Zg} + \boldsymbol{Zh} + \boldsymbol{\varepsilon} \quad (2)$$

where $\boldsymbol{y}, \mu, \boldsymbol{Z}, \boldsymbol{g}$ and $\boldsymbol{\varepsilon}$ are the same as in the first model, $\boldsymbol{h}$ is a vector of additive-by-additive epistatic genomic effects with $\boldsymbol{h} \sim N(0, \sigma_H^2\boldsymbol{H})$, where σ_H^2 is the epistatic variance and $\boldsymbol{H}$ is the additive-by-additive relationship matrix, built following Vitezica et al. (2017).

Sub-Genome Models

Model 3 – Additive Effects

As described in Santantonio et al (2019), G_j can be decomposed into individual additive effects for each sub-genome, such that $\mathrm{G_j} = \mathrm{A_j} + \mathrm{B_j} + \mathrm{D_j}$, resulting in the following model:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}_A\boldsymbol{g}_A + \boldsymbol{Z}_B\boldsymbol{g}_B + \boldsymbol{Z}_D\boldsymbol{g}_D + \boldsymbol{\varepsilon} \quad (3)$$

Decomposition of the whole genome into sub-genomes permits each sub-genome to have an individual additive genetic variance and covariance among individuals, such that $\boldsymbol{g}_A \sim N(\mathbf{0}, \sigma_A^2\boldsymbol{A})$, $\boldsymbol{g}_B \sim N(\mathbf{0}, \sigma_B^2\boldsymbol{B})$ and $\boldsymbol{g}_D \sim N(\mathbf{0}, \sigma_D^2\boldsymbol{D})$. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{D}$ are the realized additive genetic variance-covariance for each sub-genome and are calculated using the method used for calculating the whole genome additive variance-covariance matrix.

Model 4 – Additive and Epistatic Effects

The whole genome epistatic component, $\boldsymbol{H}$, can also be decomposed into intra and inter subgenome epistatic interactions, such that $H_i = AA_i + BB_i + DD_i + AB_i + AD_i + BD_i$. The new additive plus epistatic sub-genome model is then expressed as:

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}_A\boldsymbol{g}_A + \boldsymbol{Z}_B\boldsymbol{g}_B + \boldsymbol{Z}_D\boldsymbol{g}_D + \boldsymbol{Z}_{AA}\boldsymbol{g}_{AA} + \boldsymbol{Z}_{BB}\boldsymbol{g}_{BB} + \boldsymbol{Z}_{DD}\boldsymbol{g}_{DD} + \boldsymbol{Z}_{AB}\boldsymbol{g}_{AB} + \boldsymbol{Z}_{AD}\boldsymbol{g}_{AD} + \boldsymbol{Z}_{BD}\boldsymbol{g}_{BD} + \boldsymbol{\varepsilon} \quad (4)$$

Decomposition of the whole genome epistatic term into sub-genomes results in expression of additive-by-additive genetic variance-covariance among individuals for each pair of sub-genomes. This decomposition is generalized as $\boldsymbol{g}_{AA} \sim N(0, \sigma_{AA}^2\boldsymbol{AA})$, and $\boldsymbol{g}_{\boldsymbol{AB}} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{AB}}^{\mathbf{2}}\boldsymbol{AB})$, where the matrices $\boldsymbol{AA}$, and $\boldsymbol{AB}$, are the realized intra and inter sub-genome epistatic variance-covariance matrices, respectively, calculated using the previously described procedures.

Variance Components, Covariance, and Heritability

Variance components were estimated using the Bayesian BGLR R-package (Pérez and de los Campos, 2014). To enhance computational efficiency, eigenvalue decomposition of the variance-covariance matrices described by Acosta-Pech et al. (2017), was used and modeled as Bayesian Ridge Regression (BRR). For each model, a total of 50,000 iterations were executed, with the first 5,000 iterations discarded as burn-in. Subsequent iterations were thinned by selecting one of every ten samples, resulting in 4,500 samples used for estimating the variance components. Based upon a Bayesian Markov Chain Monte Carlo (MCMC) framework, genetic variance components were calculated using the genetic values derived from each

MCMC sample, per the methods of Sorensen, Fernando, and Gianola (2001). Estimation of genomic variance quantified the covariance between loci, i.e., linkage disequilibrium (Lehermeier et al., 2017). Variance components from each MCMC sample were used to calculate the correlation between additive and epistatic terms in both whole genome (Model 2) and sub-genome models (Model 3 and 4). Heritability was calculated for each model as the ratio of total genetic variance divided by total phenotypic variance.

Genomic prediction

Five-fold cross validation, with ten replicates was used to test predictive ability of the models. For each replicate, the set of 3740 lines was randomly divided into five groups, with four used to train the model and the remaining used to predict genetic values. Using that structure, all four above-described models were fitted, resulting in predicted genetic values for each line. Prediction of the five folds was correlated with the phenotypic value, with the resulting value representing the prediction accuracy of the model. Correlation between predicted genetic values and phenotypic values provided comparisons of accuracy amongst models. To test statistical significance between the prediction accuracies of the models fitted, Tukey's honestly significant difference was calculated using the 'agricolae' package on R (de Mendiburu, 2019).

Results

Variance Components and Heritability

Whole Genome Models 1 and 2

Estimated posterior means of the whole genome additive Model 1 resulted in additive genetic effects accounting for 27% of the observed phenotypic variance for wheat grain yield. However, in whole genome additive plus epistatic Model 2, the contribution of additive variance decreased to only 11%, while epistatic variance became prominent, explaining 29% of the observed phenotypic variance. By including an epistatic term in the whole genome additive plus epistatic Model 2, total genetic variance increased while residual variance decreased compared to whole genome additive Model 1 (Figure 1). These results suggest that the epistatic term, although not orthogonal to additive effects, captures a portion of the genetic variance that is not captured by the additive term alone. Consequently, heritability estimated using the whole genome additive plus epistatic Model 2 (0.39) exceeded the estimate of the fully additive model (0.27 – see Supplemental Table 1). Interestingly, while epistatic effects did not result in orthogonal partitioning of genetic effects, the correlation between additive and epistatic terms was virtually zero (Supplemental Table 5). The deviance information criteria value was smaller for whole genome additive plus epistatic Model 2

than for whole genome additive Model 1 (Supplemental Table 1), suggesting that the incorporation of an epistatic term captured epistatic interactions and provide a better fit of data.

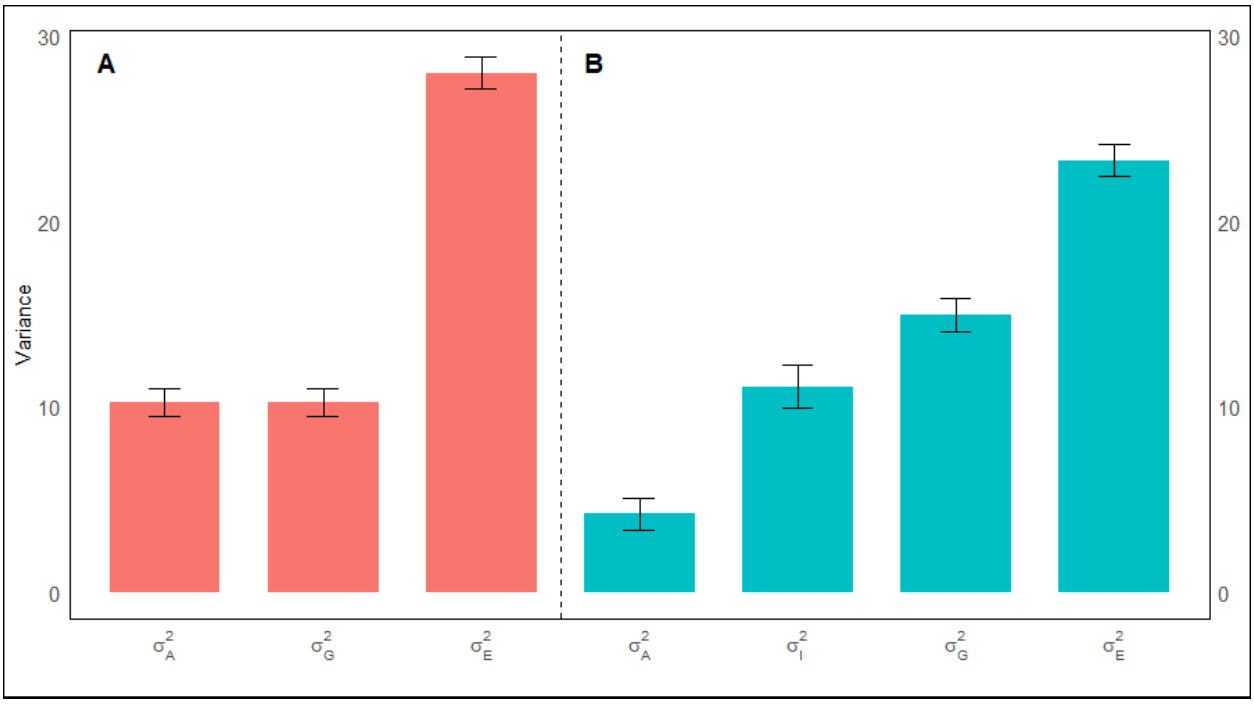


Figure 1. Substantial additive-by-additive epistatic contribution to variance of grain yield. Estimated posterior means and standard deviations of genetic variance components of Model 1 (panel A) and Model 2 (panel B). σ_A^2 refers to additive variance, σ_I^2 to additive-by-additive epistatic variance, σ_G^2 to total genetic variance, and σ_E^2 is the residual variance of the model.

Sub-Genome Models

The sum of additive genetic variances across genomes for the sub-genome additive Model 3 (10.88) was similar to the total genetic variance estimated using the whole genome additive Model 1, which was expected since breaking the genome into sub-genomes should not imply different additive effects (Figure 1 and 2). For the sub-genome additive plus epistatic Model 4, the sum of sub-genome additive effects added to 4.6 and the epistatic variance was 12.7, indicating a slight increase in epistatic variance with little change in additive variance compared to corresponding values from the whole genome additive plus epistatic Model 2 (Supplemental Table 2). The sum of the additive and epistatic variance deviated from the observed total genotypic variance because of a negative covariance between the terms in the sub-genome additive plus epistatic Model 4 (Supplemental Table 6). Using sub-genome additive Model 3, the largest additive variance was estimated for the A sub-genome, followed by the corresponding estimates for the D and B

sub-genomes (Figure 2). Inclusion of the intra- and inter- sub-genome interaction terms reduced the amount of additive variance associated with each sub-genome, and the D subgenome displayed the highest additive effects among the three sub-genomes. As with whole genome additive plus epistatic Model 2, the addition of intra and inter sub-genome epistatic terms resulted in non-orthogonal partitioning of genetic effects. Likewise, the correlation within and between additive and epistatic terms was nearly zero in all instances (Supplemental Table 6).

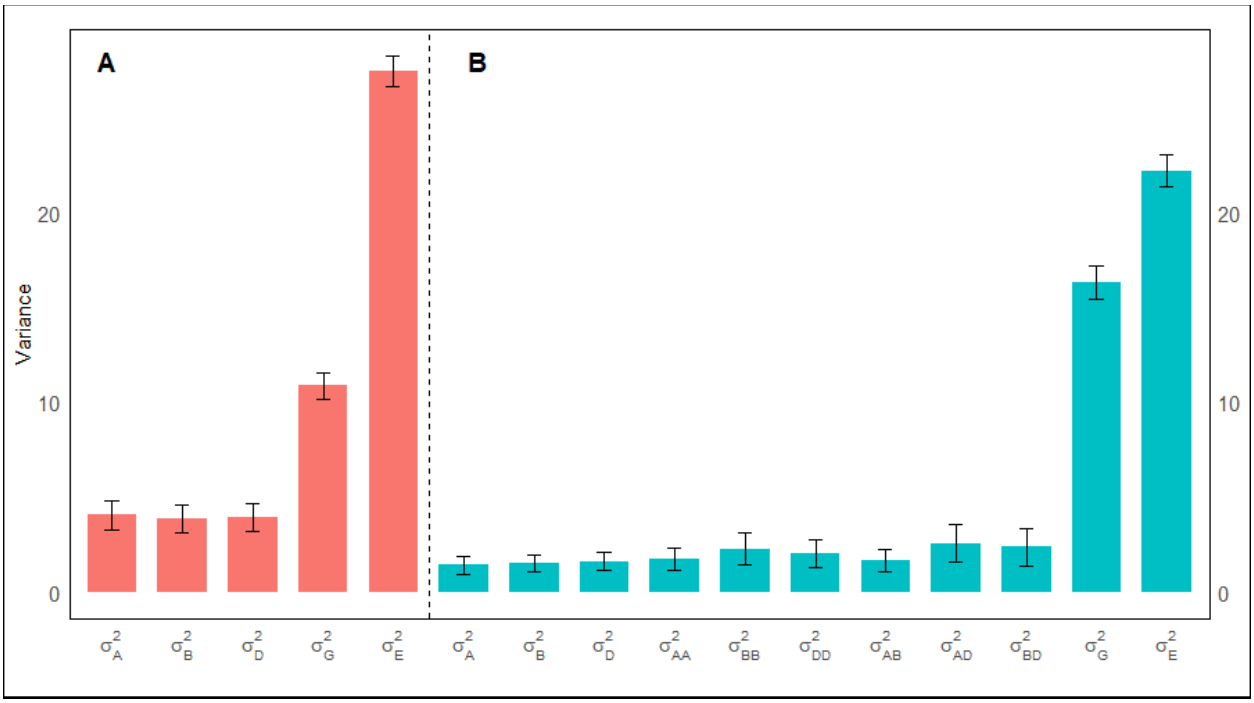
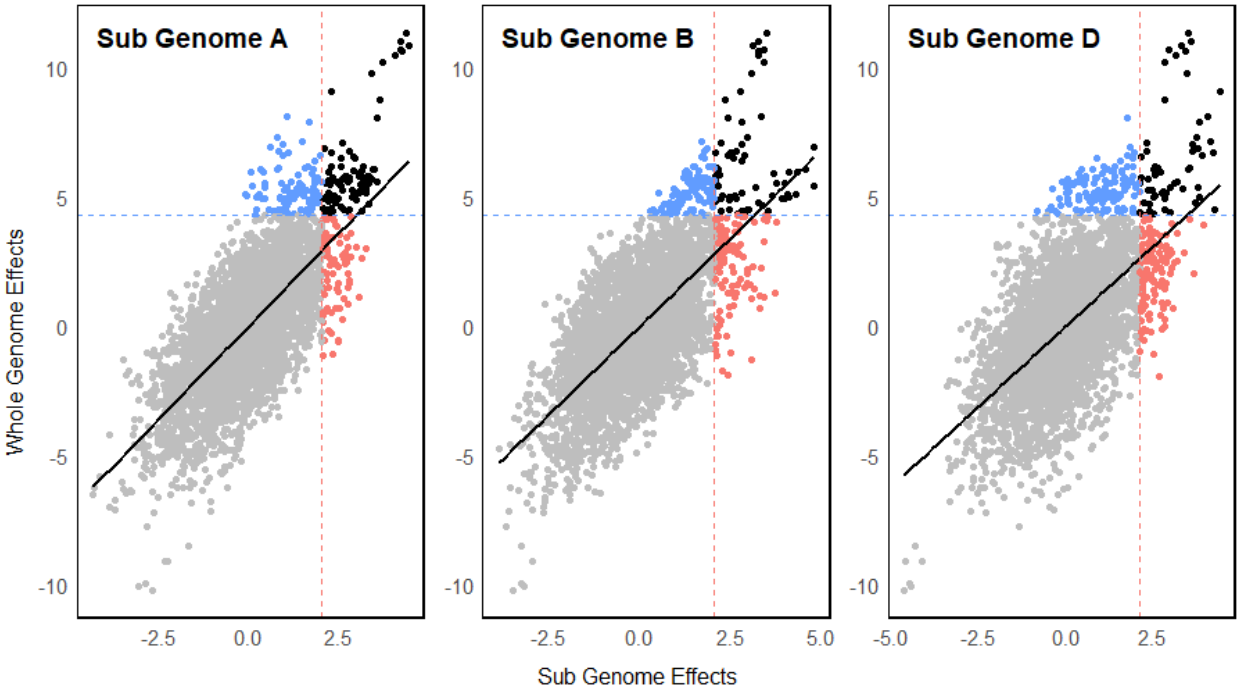


Figure 2. Wheat sub-genomes additive and epistatic variance components in grain yield. Estimated posterior means and standard deviations of genetic variance components of Model 3 (panel A) and Model 4 (panel B). σ^2_A refers to additive variance in subgenome A, σ^2_B to subgenome B and σ^2_D to D. Intra and inter sub-genomes additive-by-additive epistatic variance are depicted by σ^2_{AA}, σ^2_{BB}, σ^2_{DD}, and σ^2_{AB}, σ^2_{AD}, σ^2_{BD}, respectively. σ^2_G to total genetic variance, and σ^2_E is the residual variance of the model.

The larger total genetic variance estimated using Model 4 increased the magnitude of estimated heritability compared to estimates from the whole genome Models 1 and 2 (Supplemental Table 2). In addition, the deviance information criteria values were smaller than those obtained using the whole genome models, indicating a better fit of data.

Sub Genome Additive Effects

232 Correlations between whole and sub-genomes effects were estimated to characterize the
233 relationships between sub-genome additive effects and whole genome effects (Figure 3). The smallest
234 correlation was observed between the effects of sub-genome D and the whole genome effects (0.61 –
235 Supplemental Table 3), while the correlations between the effects of sub-genome A and B with the whole
236 genome effects were higher (0.72 and 0.69, respectively). The correlations amongst sub genome effects
237 was very low. Interestingly, when analyzing the sub-genome additive effects from Model 4, which includes
238 epistasis, the correlations between sub-genome effects and whole genome effects, as well as between sub-
239 genomes, generally decreased, except for the correlation between sub-genomes A and D. Although
240 correlations between sub-genomes and whole genome varied, all three sub-genomes resulted in similar
241 maximum and minimum additive effects, and standard deviations (Supplemental Table 4).
242



243

244 **Figure 3. Sub-genome additive effects correlates with whole genome effects for grain yield in wheat**. The
245 dotted lines indicate the 95 percentile for whole or subgenome additive effects. The solid black lines indicate
246 the correlation between whole genome and sub genome effects. Blue dots indicate genotypes with whole
247 genome additive effects above the 95-percentile. Pink dots indicate genotypes with sub genome additive
248 effects above the 95-percentile. Black dots indicates genotypes above the 95-percentile for both thresholds.

249

250 In total, 98 lines were found above the 95-percentile threshold for both whole and sub-genome A

251 additive effects (Figure 3). Of those, 28 and 22 genotypes were also above the 95-percentile threshold for
252 sub-genome B and D, respectively. In sub-genome B and D, 75 and 57 lines were found above the 95-
253 percentile threshold for both sub-genomes and whole genome additive effects, with 22 lines above the 95-
254 percentile threshold for both sub genomes B and D. In total, 11 genotypes were above the 95-percentile for
255 all sub-genomes, 9 of which were sister lines. Hence, in general, genotypes with the highest whole genome
256 effects tended to also exhibit intermediate to high additive effects in more than one sub-genome. Of the
257 top-100 genotypes with highest whole genome additive effects, 48 were advanced to the preliminary yield
258 trials (PYT). Among these, 6 were selected for the advanced yield trials (AYT), but none of these genotypes
259 were advanced to the elite trials (EYT). One possible reason for the discarding of genotypes with high whole
260 genome additive effects could be selection for or against other important agronomic traits not considered
261 in this prediction model, traits such as disease resistance, maturity, and lodging. For example, from this pool
262 of 3740 genotypes evaluated across three breeding cycles, only eight reached the EYT stage, five of which
263 were above the top-90 percentile. It is necessary to acknowledge that the total genetic value of each
264 genotype across many traits is used to make advancement decisions, with those genetic values being a
265 function of both additive and epistatic effects. In that regard, the sub-genome additive plus epistatic Model
266 4 estimated higher genotypic values for the eight lines that reached the EYT stage, with seven of those eight
267 falling withing the top-10 percentile. Compared to the sub-genome additive Model 3, these eight lines had
268 an average rank change of 127 positions towards the top when epistasis was modeled. This result of
269 selection highlights the importance of epistasis when making selection decisions based upon the genetic
270 value of candidate genotypes.
271
272 *Genomic Predictive Ability*
273 The prediction accuracy between the predicted total genetic values and the phenotypic values was
274 evaluated using the five-fold cross validation method (Figure 4). For both whole and sub-genome models,
275 inclusion of epistatic genetic effects significantly improved the prediction accuracy compared to fully
276 additive models, based upon Tukey's honestly significant difference test. In contrast, partitioning whole
277 genome effects into sub-genome effects did not improve the accuracy of predictions. The correlations
278 amongst total genetic values of the four models fitted showed very high values (Supplemental Table 7). As
279 expected, the smallest correlation was observed between whole genome additive Model 1 and sub-genome
280 additive plus epistatic Model 4. Similarly, correlations amongst genomic estimated breeding values (sum of
281 additive effects only) included a smaller correlation between the purely additive and the additive plus
282 epistatic models (Supplemental Table 8).

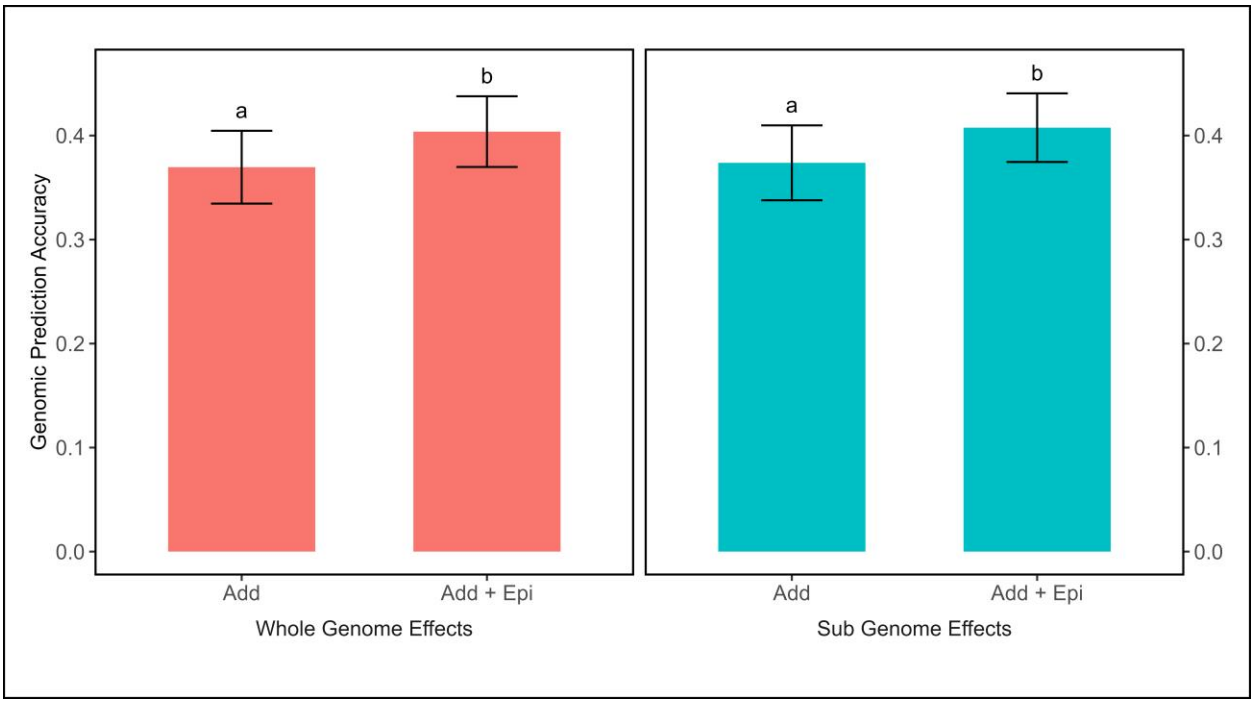


Figure 4. Increased prediction accuracy using models that include epistasis. Bar plot of predictive abilities of additive whole (panel A) and sub-genome (panel B) models with the inclusion or not of an epistatic term using a five-fold cross validation appraoch with 10 replications.

Discussion

Modeling Epistasis Improves the Total Genetic Variance and Broad Sense Heritability

The role of epistasis is expected to be especially relevant in polyploid species due to the interaction between homologous genes, but epistasis has been commonly ignored in wheat. In this study, based upon modeling variance components, inclusion of epistatic effects in whole and sub-genome additive models reduced the additive variance, indicating a lack of statistical independence between additive and epistatic effects. The lack of independence due to non-orthogonal partitioning of genetic effects could be attributed to linkage disequilibrium, which introduces non-independence between loci (Vitezica et al. 2017). Under the presence of linkage disequilibrium, covariance between genetic effects can be introduced and partitioning genetic effects is not possible (Hill and Mäki-Tanila 2015; Zeng et al. 2005). In simulated and experimental studies, lack of orthogonality was evidenced through a high correlation between additive and epistatic variance component estimates (Vitezica et al., 2017; Raffo et al., 2022), with similar correlations being near zero in this study (Supplemental Table 5 and 6).

303 In this study inclusion of epistasis effects in additive models increased the total genetic variance
304 and reduced the DIC. In contrast, a study of the Nordic Seed A/S program reported a slightly reduction in
305 total genetic variance when epistasis was modeled (Raffo et al., 2022), while a study of Cornell's and
306 CIMMYT's breeding populations found that partitioning whole genome effects into sub-genome additive
307 and epistatic effects only marginally increased the AIC for wheat grain yield (Santantonio et al., 2019). One
308 possible explanation for differing results could relate to the large genetic diversity of the K-State wheat
309 breeding program, which is expected to have relatively high minor allele frequencies that increase the
310 statistical power to capture epistatic interactions.
311
312 *Sub-Genome Additive Effects as a Tool for Parental Selection if Epistasis is Irrelevant*
313 Estimation of sub-genome additive effects has the potential of attributing biological importance to
314 genetic effects commonly assigned to the whole genome (Santantonio et al 2019) and could be leveraged
315 for selecting parental lines with complementary sub-genome effects when designing breeding crosses. A
316 study of Cornell's and CIMMYT's wheat breeding populations found that genotypes with the highest sub-
317 genome additive effects were normally not among the top genotypes having high whole genome effects,
318 and a low correlation between sub-genomes was regarded as a potential indicator of independence
319 between sub-genomes (Santantonio et al., 2019). In this study, making breeding crosses using parental lines
320 (genotypes) having highest additive effects within each sub-genome could produce theoretical maximum
321 whole genome effects of 13.7, 21% higher than the observed highest whole genome effect (11.3). The
322 success of this approach is tightly linked to the interplay of epistatic effects between and within sub-
323 genomes. Although low pairwise correlations amongst sub-genomes were observed, high epistatic variance
324 resulted from both intra- and inter- sub-genome interactions, variances which were nearly zero in Cornell's
325 and CIMMYT's breeding populations (Santantonio et al., 2019). In addition, estimation of sub-genome
326 additive effects were less accurate than sub-genome additive plus epistatic effects in predicting which lines
327 reached the elite stages of evaluation in the wheat breeding pipeline. Although the magnitude of epistasis
328 is hard to estimate, epistasis has been reported to affect several traits in wheat, including stem rust (Singh
329 et al., 2013; Rouse et al., 2014), stripe rust (Vazques et al., 2015), plant height (Zhang et al., 2008;
330 Sannemann et al., 2018), photoperiod response (Shaw et al., 2020), vernalization (Kippes et al., 2014), and
331 grain yield heterosis (Boeven et al., 2020; Jiang et al., 2017). These reports suggest epistatic effects should
332 not be overlooked when estimating or predicting the genetic value of wheat genotypes.
333 In the germplasm used in this study, all sub-genomes exhibited similar additive effects for grain
334 yield in wheat. These results are surprising as the D sub-genome is known for having limited genetic

335 variability. The high additive effects could be in part attributed to major QTLs conferring resistance to barley
336 yellow dwarf virus (BYD). In 2017, a severe BYD infection affected the experimental farm, but a large
337 proportion of the advanced genotypes had parental lines carrying *Bdv1* or *Bdv2* resistance genes. *Bdv1*
338 reportedly confers partial tolerance to BYD and is located on chromosome 7D (Singh et al. 1993; Ayala et al.
339 2002), while *Bdv2*, an introgression from *Thinopyrum intermedium*, is also found on chromosome 7D in
340 hexaploid wheat (Banks et al.,1995; Hohmann et al., 1996; Larkin et al., 2002). Hence, the presence of *Bdv1*
341 and *Bdv2* alleles in many advanced families could have drastically increased the genetic effects associated
342 with the D sub-genome, as the high disease pressure greatly affected grain yield performance.
343
344 *Prediction Accuracy Improves When Epistasis is Modeled*
345 As epistasis is potentially an important component of genetic effects impacting grain yield, inclusion
346 of epistatic effects in genomic prediction models could significantly improve prediction accuracies. In this
347 study, modeling epistasis in whole or sub-genome models marginally but significantly improved prediction
348 accuracies, contrasting with the results from Cornell's and CIMMYT's wheat breeding populations, where
349 no improvement in prediction accuracy was observed for grain yield, even though inclusion of epistatic
350 effects improved predictability of other traits (Santantonio et al., 2019). In a study of CIMMYT's Global
351 Wheat and Semiarid wheat breeding populations, an average improvement of 6% in prediction accuracy
352 was observed when modeling additive-by-additive epistatic interactions (Jiang and Reif, 2015), while an
353 increase of 5% in prediction accuracy was achieved in KWS European wheat breeding program (He et al.,
354 2016). These consistent but small improvement in prediction accuracy reported do not support the
355 expectation that epistasis has a major role in wheat given the allopolyploid nature of wheat. In allopolyploid
356 species, sub-functionalization could create positive or negative epistasis (Lynch and Force, 2000),
357 competition for the same substrate (Qian et al., 2010), or affect the patterns of gene expression (Leach et
358 al., 2014; Akhunova et al., 2010), and directly influence the expression of agronomic traits. Furthermore,
359 homeologous interaction effects could be minor relative to the total epistatic genetic effect in wheat, since
360 epistasis has been reported in diploid wheat, *Triticum monccum* (Tranquilli and Dubcovsky, 2000), and also
361 within sub-genomes in hexaploid wheat (Reif et al., 2011; Sehgal et al., 2020). One potential reason for the
362 marginal improvement in prediction accuracy when modeling epistasis lies in the highly quantitative genetic
363 architecture of grain yield. Every stage of crop development is influenced by environmental stresses, and all
364 of which add to total expression of grain yield at the end of the breeding cycle. As various genetic
365 interactions are potentially triggered during crop development, the contribution of each component to
366 grain yield may be so small that capturing epistatic effects is extremely difficult.

Conclusion

To delineate the genetic architecture of wheat, as expressed for grain yield, whole genome additive and epistatic effects were estimated at the whole genome level and at the sub-genome level. Although estimated phenotypic variance matched the observed phenotypic variance, partitioning of genetic variance components into additive and epistatic variance components was not orthogonal. Incorporating epistasis in additive models did augment the total genetic variance and reduced the DIC, indicating existence of epistatic effects controlling the expression of the trait grain yield. Using a fully additive model, additive effects were estimated for each sub-genome with results indicating that genotypes with largest whole genome additive effects also tended towards intermediate to large sub-genome additive effects. Estimating sub-genome additive effects has potential as a tool during parental selection, as it could permit pairing lines with complementary sub-genome effects. However, the success of this approach is tightly linked to the extent of inter and intra sub-genome epistatic effects. Modeling epistasis in genomic selection models marginally but significantly improved prediction accuracy for both whole-genome and sub-genome models, which can potentially contribute to significant improvements in genetic gain over multiple cycles of selection and advancement.

Data Availability

Supplemental files are available at the Figshare portal. The files Sub-Genome-A.hmp.txt, Sub-Genome-B.hmp.txt and Sub-Genome-D.hmp.txt refer to the molecular marker data partitioned by each sub-genome of hexaploid wheat. The BLUPs.txt file refers to the grain yield BLUPs. The Var.R, GS.R and Rank.R are the R codes used to estimate the variance components, to run the genomic prediction models, and to rank lines according to sub-genome additive effects, respectively.

Conflict of Interest

Node declared.

References

Acosta-Pech, R., Crossa, J., de Los Campos, G., Teyssèdre, S., Claustres, B., Pérez-Elizalde, S., & Pérez-Rodríguez, P. (2017). Genomic models with genotype× environment interaction for predicting hybrid performance: an application in maize hybrids. Theoretical and Applied Genetics, 130, 1431-1440.

Akhunova, A. R., Matniyazov, R. T., Liang, H., & Akhunov, E. D. (2010). Homoeolog-specific transcriptional bias in allopolyploid wheat. BMC genomics, 11, 1-16.

Alvarez-Castro, J. M., & Carlborg, O. (2007). A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. Genetics, 176(2), 1151-1167.

Arief, V. N., Desmae, H., Hardner, C., DeLacy, I. H., Gilmour, A., Bull, J. K., & Basford, K. E. (2019). Utilization of multiyear plant breeding data to better predict genotype performance. Crop Science, 59(2), 480-490.

Ayala, L., Henry, M., van Ginkel, M., Singh, R., Keller, B., & Khairallah, M. (2002). Identification of QTLs for BYDV tolerance in bread wheat. Euphytica, 128, 249-259.

Banks, P. M., Larkin, P. J., Bariana, H. S., Lagudah, E. S., Appels, R., Waterhouse, P. M., ... & Zhou, G. H. (1995). The use of cell culture for subchromosomal introgressions of barley yellow dwarf virus resistance from Thinopyrum intermedium to wheat. Genome, 38(2), 395-405.

Bauman, L. F. (1959). Evidence of Non-Allelic Gene Interaction in Determining Yield, Ear Height, and Kernel Row Number in Corn 1. Agronomy Journal, 51(9), 531-534.

Boeven, P. H., Zhao, Y., Thorwarth, P., Liu, F., Maurer, H. P., Gils, M., ... & Würschum, T. (2020). Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. Science Advances, 6(24), eaay4897.

Chapman, S. R., & McNeal, F. H. (1971). Gene Action for Yield Components and Plant Height in a Spring Wheat Cross 1. Crop science, 11(3), 384-386.

Charlesworth, D. (2003). Effects of inbreeding on the genetic diversity of populations. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358(1434), 1051-1070.

Cullis, B. R., & Gleeson, A. C. (1991). Spatial analysis of field experiments-an extension to two dimensions. Biometrics, 1449-1460.

Cullis, B., Gogel, B., Verbyla, A., & Thompson, R. (1998). Spatial analysis of multi-environment early generation variety trials. Biometrics, 1-18.

De Mendiburu, F., & De Mendiburu, M. F. (2019). Package ‘agricolae’. R Package, version 1.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package

rrBLUP. The plant genome, 4(3).

Federer, W. T., & Raghavarao, D. (1975). On augmented designs. Biometrics, 29-35.

Feldman, M., & Levy, A. A. (2012). Genome evolution due to allopolyploidization in wheat. Genetics, 192(3), 763-774.

Gilmour, A. R. (2018, February). Echidna mixed model software. In Proceedings of the World Congress Applied to Livestock Genetics (Vol. 995).

Goldringer, I., Brabant, P., & Gallais, A. (1997). Estimation of additive and epistatic genetic variances for agronomic traits in a population of doubled-haploid lines of wheat. Heredity, 79(1), 60-71.

He, S., Schulthess, A. W., Mirdita, V., Zhao, Y., Korzun, V., Bothe, R., ... & Jiang, Y. (2016). Genomic selection in a commercial winter wheat population. Theoretical and Applied Genetics, 129, 641-651.

Henderson, C. R. (1985). Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. Journal of animal science, 60(1), 111-117.

Hill, W. G., & Mäki-Tanila, A. (2015). Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. Journal of animal breeding and genetics, 132(2), 176-186.

Hohmann, U., Busch, W., Badaeva, K., Friebe, B., & Gill, B. S. (1996). Molecular cytogenetic analysis of Agropyron chromatin specifying resistance to barley yellow dwarf virus in wheat. Genome, 39(2), 336-347.

Jiang, Y., & Reif, J. C. (2015). Modeling epistasis in genomic selection. Genetics, 201(2), 759-768.

Jiang, Y., Schmidt, R. H., Zhao, Y., & Reif, J. C. (2017). A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nature genetics, 49(12), 1741-1746.

Kihara, H. (1924). Cytologische und genetische Studien bei wichtigen Getreidearten mit besonderer Rücksicht auf das Verhalten der Chromosomen und die Sterilität in den Bastarden. Memoirs of the College of Science, Kyoto Imperial University. Ser. B, 1(1), 1-200.

Kihara, H. (1944). Discovery of the DD-analyser, one of the ancestors of Triticum vulgare. Agric. Hort., 19, 13-14.

Kippes, N., Zhu, J., Chen, A., Vanzetti, L., Lukaszewski, A., Nishida, H., ... & Dubcovsky, J. (2014). Fine mapping and epistatic interactions of the vernalization gene VRN-D4 in hexaploid wheat. Molecular Genetics and Genomics, 289, 47-62.

Larkin, P. J., Kleven, S., & Banks, P. M. (2002). Utilizing Bdv2, the Thinopyrum intermedium source of BYDV resistance, to develop wheat cultivars. Recent Advances and Future Strategies, M. Henry and A. McNab eds.(Mexico: CIMMYT Texcoco), 60-63.

Leach, L. J., Belfield, E. J., Jiang, C., Brown, C., Mithani, A., & Harberd, N. P. (2014). Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. BMC genomics, 15, 1-19.

Lehermeier, C., de Los Campos, G., Wimmer, V., & Schön, C. C. (2017). Genomic variance estimates: With or without disequilibrium covariances?. Journal of Animal Breeding and Genetics, 134(3), 232-241.

Lynch, M., & Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics, 154(1), 459-473.

McFadden, E. S., & Sears, E. R. (1946). The origin of Triticum spelta and its free-threshing hexaploid relatives. Journal of heredity, 37(3), 81-89.

Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. genetics, 157(4), 1819-1829.

Nejati-Javaremi, A., Smith, C., & Gibson, J. P. (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. Journal of animal science, 75(7), 1738-1745.

Pérez, P., & de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. Genetics, 198(2), 483-495.

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PloS one, 7(2), e32253.

Qian, W., Liao, B. Y., Chang, A. Y. F., & Zhang, J. (2010). Maintenance of duplicate genes and their functional redundancy by reduced expression. Trends in Genetics, 26(10), 425-430.

Raffo, M. A., Sarup, P., Guo, X., Liu, H., Andersen, J. R., Orabi, J., ... & Jensen, J. (2022). Improvement of genomic prediction in advanced wheat breeding lines by including additive-by-additive epistasis. Theoretical and Applied Genetics, 135(3), 965-978.

Reif, J. C., Maurer, H. P., Korzun, V., Ebmeyer, E., Miedaner, T., & Würschum, T. (2011). Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theoretical and Applied genetics, 123, 283-292.

Rouse, M. N., Talbert, L. E., Singh, D., & Sherman, J. D. (2014). Complementary epistasis involving Sr12 explains adult plant resistance to stem rust in Thatcher wheat (Triticum aestivum L.). Theoretical and applied genetics, 127, 1549-1559.

Sannemann, W., Lisker, A., Maurer, A., Léon, J., Kazman, E., Cöster, H., ... & Pillen, K. (2018). Adaptive selection of founder segments and epistatic control of plant height in the MAGIC winter wheat population WM-800. BMC genomics, 19, 1-16.

Santantonio, N., Jannink, J. L., & Sorrells, M. (2019). Prediction of subgenome additive and interaction effects in allohexaploid wheat. G3: Genes, Genomes, Genetics, 9(3), 685-698.

Sax, K. (1922). Sterility in wheat hybrids. II. Chromosome behavior in partially sterile hybrids. Genetics, 7(6), 513.

Sehgal, D., Rosyara, U., Mondal, S., Singh, R., Poland, J., & Dreisigacker, S. (2020). Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. Frontiers in plant science, 11, 502915.

Shaw, L. M., Li, C., Woods, D. P., Alvarez, M. A., Lin, H., Lau, M. Y., ... & Dubcovsky, J. (2020). Epistatic interactions between *PHOTOPERIOD1*, *CONSTANS1* and *CONSTANS2* modulate the photoperiodic response in wheat. PLoS genetics, 16(7), e1008812.

Singh, A., Knox, R. E., DePauw, R. M., Singh, A. K., Cuthbert, R. D., Campbell, H. L., ... & Clarke, F. (2013). Identification and mapping in spring wheat of genetic factors controlling stem rust resistance and the study of their epistatic interactions across multiple environments. Theoretical and Applied Genetics, 126, 1951-1964.

Singh, R. P., Burnett, P. A., Albarran, M., & Rajaram, S. (1993). Bdv1: a gene for tolerance to barley yellow dwarf virus in bread wheats. Crop Science, 33(2), 231-234.

Smith, A., Cullis, B., & Thompson, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics, 57(4), 1138-1147.

Sorensen, D., Fernando, R., & Gianola, D. (2001). Inferring the trajectory of genetic variance in the course of artificial selection. Genetics Research, 77(1), 83-94.

Su, G., Christensen, O. F., Ostersen, T., Henryon, M., & Lund, M. S. (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers.

Sun, P. L. F., Shands, H. L., & Forsberg, R. A. (1972). Inheritance of kernel weight in six spring wheat crosses 1. Crop Science, 12(1), 1-5.

Tranquilli, G., & Dubcovsky, J. (2000). Epistatic interaction between vernalization genes Vrn-Am1 and Vrn-Am2 in diploid wheat. Journal of Heredity, 91(4), 304-306.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. Journal of dairy science, 91(11), 4414-4423.

Vazquez, M. D., Zemetra, R., Peterson, C. J., Chen, X. M., Heesacker, A., & Mundt, C. C. (2015). Multi-location wheat stripe rust QTL analysis: genetic background and epistatic interactions. Theoretical and applied genetics, 128, 1307-1318.

Vitezica, Z. G., Legarra, A., Toro, M. A., & Varona, L. (2017). Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. Genetics, 206(3), 1297-1307.

Welham, S. J., Gogel, B. J., Smith, A. B., Thompson, R., & Cullis, B. R. (2010). A comparison of analysis methods for late-stage variety evaluation trials. Australian & New Zealand Journal of Statistics, 52(2), 125-149.

Zeng, Z. B., Wang, T., & Zou, W. (2005). Modeling quantitative trait loci and interpretation of models. Genetics, 169(3), 1711-1725.

Zhang, K., Tian, J., Zhao, L., & Wang, S. (2008). Mapping QTLs with epistatic effects and QTL× environment interactions for plant height using a doubled haploid population in cultivated wheat. Journal of Genetics and Genomics, 35(2), 119-127.

Supplemental Materials- Improving genomic selection in hexaploid wheat with sub-genome additive and epistatic models

Table 1. Posterior means (± posterior standard deviation) of additive variance (V_A), epistatic variance (V_I), total genetic variance (V_G), residual variance (V_E), phenotypic variance (V_P), genomic broad sense heritability (H^2) and deviance information criteria (DIC) estimated with models WG-A (whole genome additive) and WG-AE (whole genome additive plus epistatic).

Model	V_A	V_I	V_G	V_E	V_P	H^2	DIC
1	10.21 ± 0.74	-	10.21 ± 0.74	27.99 ± 0.85	38.19 ± 0.85	0.27 ± 0.02	23592.16
2	4.21 ± 0.83	11.09 ± 1.19	14.96 ± 0.9	23.29 ± 0.89	38.26 ± 0.81	0.39 ± 0.02	23346.74

Table 2. Posterior means (± posterior standard deviation) of additive variance in sub-genome A (V_A), B (V_B) and D (V_D), intra genomic epistatic variance in sub-genomes A (V_{AA}), B (V_{BB}) and D (V_{DD}), inter genomic epistatic variance in sub-genomes AB (V_{AB}), AD (V_{AD}) and BD (V_{BD}), total genetic variance (V_{TG}), residual variance (V_E), phenotypic variance (V_P), genomic broad sense heritability (H^2) and deviance information criteria (DIC) estimated with models SG-A (sub-genome additive) and SG-AE (sub-genome additive plus epistatic).

Model	V_A	V_B	V_D	V_{AA}	V_{BB}	V_{DD}	V_{AB}	V_{AD}	V_{BD}	V_G	V_E	V_P	H^2	DIC
3	4.05 ± 0.75	3.87 ± 0.71	3.94 ± 0.72	-	-	-	-	-	-	10.88 ± 0.72	27.45 ± 0.8	38.33 ± 0.84	0.28 ± 0.02	23569.99
4	1.42 ± 0.47	1.54 ± 0.44	1.62 ± 0.47	1.76 ± 0.57	2.29 ± 0.81	2.05 ± 0.74	1.67 ± 0.56	2.58 ± 0.97	2.39 ± 0.99	16.33 ± 0.86	22.22 ± 0.85	38.55 ± 0.8	0.42 ± 0.02	23289.3

Table 3. Correlation between whole genome (Model 1) and sub-genome additive effects. The above diagonal shows the correlation of sub-genome additive effects from Model 3 (purely additive) and below the diagonal from Model 4 (additive plus epistatic).

	A	B	D	Whole
A	-	0.18	0.27	0.72
B	0.14	-	0.14	0.69
D	0.29	0.10	-	0.61
Whole	0.63	0.63	0.54	-

Table 4. Mean (standard deviation), maximum and minimum additive effects associated with sub-genome A, B and D (Model 3) and the whole genome (Model 1).

	A	B	D	Whole
Mean	0 ± 1.27	0 ± 1.27	0 ± 1.25	0 ± 2.51
Maximum	4.47	4.81	4.43	11.34
Minimum	-4.33	-3.88	-4.64	-10.19

Table 5. Variance (diagonal), covariance (below diagonal) and correlation (above diagonal) of additive and epistatic effects from whole genome additive and epistatic Model 2.

	Additive	Epistatic
Additive	4.21	0.00
Epistatic	-0.17	11.09

Table 6. Variance (diagonal), covariance (below diagonal) and correlation (above diagonal) between additive and epistatic effects from sub-genome additive and epistatic Model 4.

	V_A	V_B	V_D	V_{AA}	V_{BB}	V_{DD}	V_{AB}	V_{AD}	V_{BD}
V_A	1.42	0	0	0	0	0	0	0	0
V_B	-0.05	1.54	0	0	0	0	0	0	0
V_D	0.02	-0.06	1.62	0	0	0	0	0	0
V_{AA}	-0.04	-0.01	-0.01	1.76	0	0	0	0	0
V_{BB}	-0.01	-0.02	-0.01	-0.01	2.29	0	0	0	0
V_{DD}	-0.01	-0.01	-0.03	-0.01	-0.01	2.05	0	0	0
V_{AB}	0.00	0.00	-0.01	-0.02	-0.01	-0.01	1.67	0	0
V_{AD}	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01	-0.01	2.58	0
V_{BD}	-0.01	-0.01	-0.02	-0.01	-0.02	-0.01	-0.01	0.00	2.39

Table 7. Correlation between total genetic effects (that is the sum of additive and epistatic effects, if epistasis was modeled) of Models 1-4.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Model 1	-	0.92	0.98	0.90

Model 2	-	0.91	0.99
Model 3		-	0.91
Model 4			-

Table 8. Correlation between additive effects of Models 1-4.

	Model 1	**Model 2**	**Model 3**	**Model 4**
Model 1	-	0.91	0.99	0.89
Model 2		-	0.89	0.97
Model 3			-	0.90
Model 4				-