



solve this problem, even for quantitative traits (Zhao et al., 2015b). In genome-wide prediction, many markers are used, and their effects are estimated in populations that have been genotyped and phenotyped. The estimated marker effects are then applied to predict the performance of non-phenotyped hybrids based on their molecular marker profiles.

The potential of genome-wide prediction was investigated in rapeseed with a focus on general combining ability (GCA) effects as the additive component of the hybrid performance (Jan et al., 2016). The study was based on two testcross series of 475 spring-type lines and revealed moderate-to-high prediction accuracies for a number of important agronomic traits. The potential of genome-wide prediction of hybrid performance, i.e., additive/general and non-additive/specific combining ability (SCA) effects, has not been examined in rapeseed. Simulation studies revealed that the prediction accuracy of the hybrid performance can be increased by modeling dominance as one type of non-additive effect, but the magnitude of improvement strongly depended on the relevance of the variance of specific vs. combining ability effects (Technow et al., 2012). Analyses of experimental data in maize (Bernardo, 1994), rice (Wang et al., 2017), wheat (Zhao et al., 2015a), triticale (Gowda et al., 2013), and sunflower (Reif et al., 2013) corroborated this finding, with hybrid prediction accuracies being either similar or higher when fitting additive and dominance effects.

Genome-wide prediction approaches of the hybrid performance can also accommodate epistasis, i.e., interaction effects between genes (Xu et al., 2014). In particular, semiparametric reproducing kernel Hilbert space (RKHS) regression models or extended genomic best linear unbiased predictions (EG-BLUP) are computational efficient approaches to capture epistasis in genome-wide predictions (Jiang and Reif, 2015). Experimental studies in rice (

pr2419261(s)-30

following mixed model:

$$y_{in} = \mu + g_i + \varepsilon_{in},$$

where  $y_{in}$  was the seed dry weight of the  $n$ th observation for the  $i$ th genotype,  $\mu$  was the mean value,  $g_i$  was the genetic value of the  $i$ th genotype, and  $\varepsilon_{in}$  was the corresponding residual. The mean value and genetic value were modeled as fixed effects, and the residual was treated as a random effect.

The adjusted means of seed dry weight for each genotype across environments were estimated with the following model:

$$y_{ik} = \mu + g_i + l_k + \varepsilon_{ik},$$

where  $y_{ik}$  was the adjusted mean of the  $i$ th genotype in the  $k$ th environment,  $\mu$  was the overall mean,  $g_i$  was the genetic value of the  $i$ th genotype,  $l_k$  was the effect of the  $k$ th environment confounded with replicate effects, and  $\varepsilon_{ik}$  was the residual. The mean and genetic values were modeled as fixed effects, environment effects, and residuals were treated as random effects.

In addition, we estimated the genetic variance components of hybrids and parental lines, as well as the variance of genotype  $\times$  environment interactions using a one-step model:

$$y_{ijkn} = \mu + a + p_{ij} + (pl)_{ijk} + m_i + f_j + s_{ij} + (ml)_{ik} + (fl)_{jk} + (sl)_{ijk} + l_k + e_{ijkn},$$

where  $y_{ijkn}$  was the phenotypic performance of the  $n$ th observation for the  $ij$ th entry (line  $i = j$ , or hybrid  $i \neq j$ ) in the  $k$ th environment,  $\mu$  was the overall mean,  $a$  was the group effect for lines and hybrids,  $p_{ij}$  was the genetic effect of the parental lines,  $(pl)_{ijk}$  was the interaction effect of the  $ij$ th parental line with the  $k$ th environment,  $m_i$  was the GCA effect of the  $i$ th male line,  $f_j$  was the GCA effect of the  $j$ th female line,  $s_{ij}$  was the SCA effect of crosses between lines  $i$  and  $j$ ,  $(ml)_{ik}$  and  $(fl)_{jk}$  were GCA  $\times$  environment effects of female and male lines, and  $(sl)_{ijk}$  was SCA  $\times$  environment interaction effects,  $l_k$  was the effect of the  $k$ th environment, and  $e_{ijkn}$  was the residual. All effects were treated as random effects, except for the mean value and group effects.

The broad-sense heritability was calculated as the ratio of genotypic to phenotypic variance:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G \times E}^2}{N_E} + \frac{\sigma_E^2}{N_E \times N_R}},$$

where  $N_E$  referred to the number of environments,  $N_R$  was the average number of replications per environment,  $\sigma_G^2$  was the genotypic variance,  $\sigma_{G \times E}^2$  was the variance of genotype multiplied by environment interaction, and  $\sigma_E^2$  referred to the error variance.

## Genome-Wide QTL Mapping

Design matrices for additive and dominance effects were specified for the hybrids and their parental lines according to the  $F_\infty$  metric (Falconer and Mackay, 1996). Data from each environment were used in QTL mapping. To correct for potential population stratification, the one minus the Rogers'

distance matrix was used as a kinship matrix in the genome-wide QTL mapping scan (Zhao et al., 2013). Genome-wide scans for marker-trait associations were conducted to detect main-effect QTL, as well as all first-order epistasis effect QTL.

For main effect QTL, the model was defined as the following Yu et al. (2006):

$$Y = X\beta + Ss + Zu + e.$$

$Y$  stands for the adjusted entry means of the 498 genotypes, i.e., 180 DH lines and 318 TNRC-F<sub>2</sub> hybrids, of each environment,  $\beta$  was a vector of environment effects,  $s$  was a vector of SNP effects,  $u$  was a vector of polygene background effects, and  $e$  was a vector of residual effects.  $X$ ,  $S$ , and  $Z$  were incidence matrices relating  $Y$  to  $\beta$ ,  $s$ , and  $u$ .  $\beta$  and  $s$  were treated as fixed effects, while  $u$  and  $e$  were treated as random effects. A Quantile-Quantile plot was used to test for proper control of population stratification (Yu et al., 2006). The Bonferroni-Holm procedure (Holm, 1979) was applied to correct for multiple tests at a significance level of  $P < 0.1$ .

We performed a full two-dimensional scan to detect epistatic QTL using the following model:

$$Y = X\beta + Z_{A1}a_1 + Z_{A2}a_2 + Z_{D1}d_1 + Z_{D2}d_2 + Z_{AA}i_{11} + Z_{A1D2}i_{12} + Z_{A2D1}i_{21} + Z_{DD}i_{22} + Zu + e,$$

where  $a_1$ ,  $a_2$ ,  $d_1$ , and  $d_2$ , are additive and dominance effects of the two loci and  $i_{11}$ ,  $i_{12}$ ,  $i_{21}$ , and  $i_{22}$  correspond to all four different epistatic interaction effects.  $Z_{A1}$ ,  $Z_{A2}$ ,  $Z_{D1}$ ,  $Z_{D2}$ ,  $Z_{AA}$ ,  $Z_{A1D2}$ ,  $Z_{A2D1}$ , and  $Z_{DD}$  were incidence matrices for the effects defined above. In this model, all effects were treated as fixed effects except for  $u$  and  $e$ , which were treated as random effects. The epistasis model was implemented using the efficient mixed-model association (EMMA) approach, which significantly reduced the computation time (Kang et al., 2008). A permutation analysis with 1,000 repetitions was employed to correct for multiple testing of epistatic effects at a significance level  $P < 0.05$  (Churchill and Doerge, 1994). The proportion of the phenotypic variance explained by single QTL was estimated using multiple regression with QTL ordered according to their  $P$ -values (Utz et al., 2000). The proportion of explained genotypic variance was determined as a proportion of explained phenotypic variance standardized by broad-sense heritability.

We applied five-fold cross validation to evaluate the accuracy to predict the genotypic values from the marker effects. The data set was randomly divided into an estimation set (100% of inbred lines and 80% of the hybrids) and a test set (20% of the remaining hybrids). QTL detection and estimation of marker effects were performed in the estimation set. The estimated marker effects were then used to predict the performance of the genotypes in the test set. The prediction accuracy was estimated as a Pearson correlation coefficient between predicted and observed values standardized with the square root of the heritability.

## Genome-Wide Prediction

Based on the adjusted entry means of the 498 genotypes, we applied genomic best linear unbiased prediction (Vanraden,

2008; Zhao et al., 2015a) that considered additive, dominance, and epistatic effects (Zhao et al., 2015a). The model including only additive effects was:

$$y = 1_n\mu + g_a + e.$$

The model including additive and dominance effects was:

$$y = 1_n\mu + g_a + g_d + e.$$

$Y$  were the adjusted entry means of the 498 genotypes across the three environments,  $1_n$  was a vector of ones and  $n$  was the number of genotypes,  $\mu$  referred to the overall mean across all three environments, and  $g_a$  and  $g_d$  represented the additive and dominance effects, respectively. In the model,  $\mu$  was a fixed effect, and the remaining effects were all random effects following normal distributions  $g_a \sim N(0, G_a\sigma_a^2)$ ,  $g_d \sim N(0, G_d\sigma_d^2)$ , and  $e \sim N(0, I\sigma_e^2)$ , where  $G_a$  and  $G_d$  were the relationship matrices corresponding to additive and dominance genetic effects, and  $\sigma_a^2$ ,  $\sigma_d^2$ , and  $\sigma_e^2$  were the variance of additive effects, dominance effects, and the residuals. Details on the implementation of these relationship matrices can be found in Zhao et al. (2015a). We used a shrinkage method to calculate the relationship matrices (Ledoit and Wolf, 2004; Endelman and Jannink, 2012). All of the above GBLUP models were implemented using the R package BGLR (Perez and Campos, 2014). In addition, we employed a G-BLUP approach using a Gaussian kernel to evaluate the relevance of epistasis for prediction accuracies. The Gaussian kernel method has the benefit of considering all higher-order epistatic interact effects. The G-BLUP model using the Gaussian kernel was implemented by using the R package rrBLUP (Endelman, 2011). The prediction accuracies were evaluated using the cross-validation scenarios outlined above. We also evaluated the prediction accuracy for the independent validation set using the same training set.

## RESULTS

### Phenotypic Data Analyses of Seed Yield

The Best Linear Unbiased Estimations of seed yield for the single environments were significantly ( $P < 0.01$ ) correlated, with Pearson moment correlation coefficients ranging from 0.48 to 0.59 (Supplementary Figure 2). The analyses across environments revealed genetic variance components, which were significantly ( $P < 0.05$ ) larger than zero (Table 1). The broad-sense heritability estimates were 0.47 for the TNRC-F<sub>2</sub> hybrid population and 0.48 for the TNDH parental population. We further decomposed the genetic variance for seed yield of the TNRC-F<sub>2</sub> hybrid population into variances due to general ( $\sigma_{GCA}^2$ ) and specific combining ability effects ( $\sigma_{SCA}^2$ ). The  $\sigma_{SCA}^2$  was 1.4 times larger than  $\sigma_{GCA}^2$ , and the variance of interaction effects between environments and SCA effects  $\sigma_{Environment \times SCA}^2$  was 2.3 times larger than  $\sigma_{Environment \times GCA}^2$ .

Substantial transgressive variation was observed when comparing the seed yield of the TNDH parental population with the performance of the two founder lines, Tapidor and Ningyou7 (Figure 1). The average seed yield of the 318 single-cross hybrids

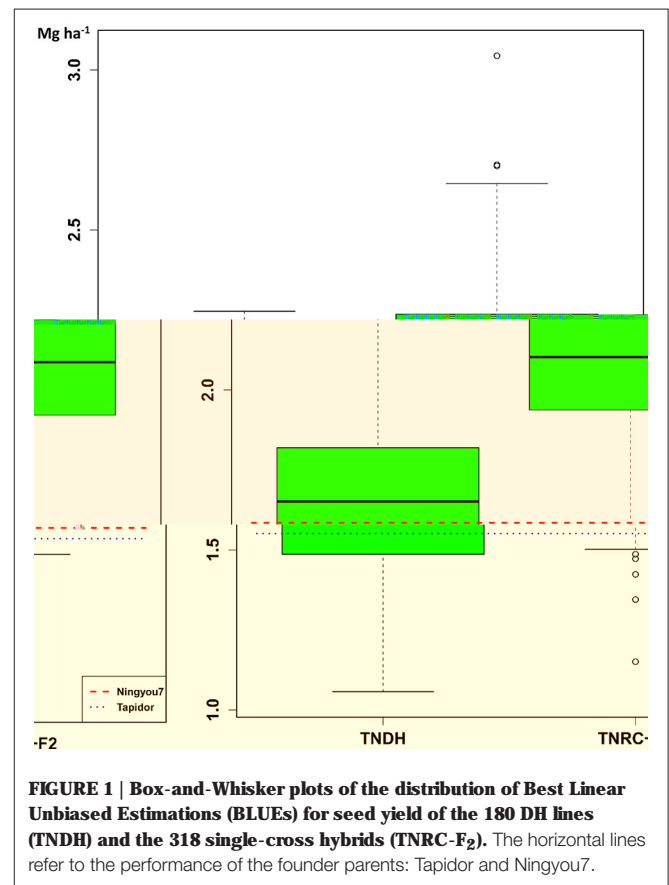
**TABLE 1 | Estimates of variance components ( $\sigma^2$ ) and broad-sense heritability of 318 hybrids and 180 parents evaluated for seed yield (Mg ha<sup>-1</sup>) across three environments.**

Source	Hybrids	Parents
$\sigma_{Genotype}^2$	0.0267***	0.0266***
$\sigma_{GCA}^2$	0.0110*	–
$\sigma_{SCA}^2$	0.0156**	–
$\sigma_{Genotype \times Environment}^2$	0.0805***	0.0770***
$\sigma_{Environment \times GCA}^2$ <sup>a</sup>	0.0243***	–
$\sigma_{Environment \times SCA}^2$ <sup>b</sup>	0.0561***	–
$\sigma_{Residual}^2$	0.0211	0.0211
Heritability	0.47	0.48

\*, \*\*, and \*\*\* means significantly different from zero at  $P < 0.1$ ,  $P < 0.01$ , and  $P < 0.001$ , respectively.

<sup>a</sup>General combining ability effects.

<sup>b</sup>Specific combining ability effects.



**FIGURE 1 | Box-and-Whisker plots of the distribution of Best Linear Unbiased Estimations (BLUEs) for seed yield of the 180 DH lines (TNDH) and the 318 single-cross hybrids (TNRC-F<sub>2</sub>). The horizontal lines refer to the performance of the founder parents: Tapidor and Ningyou7.**

of the TNRC-F<sub>2</sub> population was 2.09 Mg ha<sup>-1</sup> and was 1.27 times larger than the average seed yield of the 180 lines of the TNDH parental population.

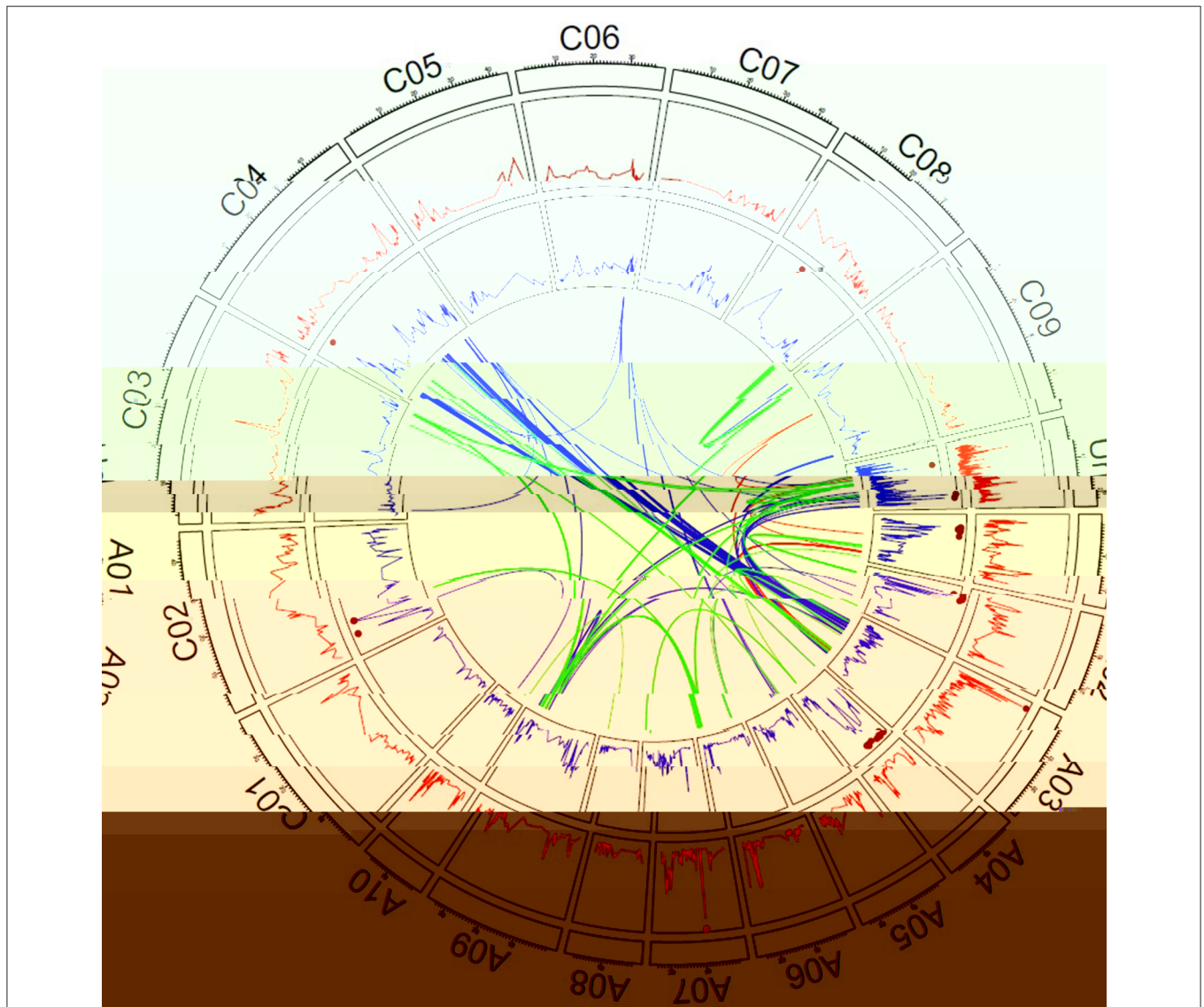
### Genome-Wide QTL Mapping

Genome-wide QTL mapping was performed in the combined TNRC-F<sub>2</sub> hybrid and TNDH parental population. The Quantile-Quantile plots for the additive and dominance

effects revealed that population stratification was properly controlled with a model including the kinship matrix (**Supplementary Figure 3**). In the genome-wide QTL mapping scan, there were 2 SNPs located on chromosomes A03 and A07, which significantly contributed to the additive genetic variation for seed yield (**Figure 2**). The SNP located on chromosome A03 explained 7% and the SNP on chromosome A07 explained 4% of the total genotypic variance (**Supplementary Table 2**). Moreover, 32 SNPs exhibited significant dominance effects and were mainly located on chromosomes A01, A02, A04, and C02. The SNPs with significant dominance effects explained 44% of the genotypic variance.

A full two-dimensional scan for epistatic effects was performed, which revealed 18 significant additive by additive effects, 129 significant additive by dominant effects, and 104 significant dominance by dominance effects (**Figure 2**). The epistatic interactions mainly involved SNPs, which were located on chromosome A03, A09, and C04. All 251 epistatic effects explained 31% of genetic total genetic variance (**Supplementary Table 3**).

Five-fold cross-validation was performed for the genome-wide QTL mapping study to obtain unbiased estimates of the total genetic variation explained by additive, dominance, and epistatic effects. The prediction accuracy was measured as a Pearson moment correlation between predicted and observed



**FIGURE 2 | Genetic architecture of hybrid performance in rapeseed.** The red lines refer to the  $-\log_{10}$  ( $P$ -values) of the additive effects, and the blue lines refer to the  $-\log_{10}$  ( $P$ -values) of the dominance effects. The brown dots mark the significant effects of  $P < 0.1$  adjusted to apply the Bonferroni–Holm procedure (Holm, 1979). Links in the center of the circle represent significant digenic interactions between SNP markers; red lines reflect additive-by-additive, blue lines additive-by-dominance, and green lines dominance-by-dominance interactions.



values of the test population standardized with the square root of the broad-sense heritability. The prediction accuracy of additive effects amounted to 0.18 and increased to 0.26 when modeling additional dominance effects. Adding epistatic interaction effects did not alter the prediction accuracy of the hybrid performance.

### Genome-Wide Prediction of Seed Yield Evaluated By Applying Cross Validations

The accuracies of three different genome-wide prediction models were compared by applying five-fold cross validation (Figure 3). We observed an increase in the prediction accuracy from the model, which considered only additive effects (0.49) compared with the model with additive plus dominance effects (0.65) and with the highest accuracy of 0.72 observed when modeling additive, dominance, and epistatic effects. To summarize, the benefits were twice as much when including additionally dominance effects compared with epistatic effects.

### Genome-Wide Prediction of Seed Yield Evaluated By Applying Independent Validations

The accuracy of the prediction model developed based on the TNDH and TNRC-F<sub>2</sub> populations was further evaluated in an independent validation. Therefore, an additional set of 37 RC-F<sub>2</sub> single-cross hybrids were generated and evaluated at an independent environment. The prediction accuracy was assessed again by standardizing the Pearson moment correlation between predicted and observed values by the square root of the broad-sense heritability. The latter was estimated based on the variance components observed for the TNRC-F<sub>2</sub> population. The prediction accuracy was high and amounted to 0.49 for

the genome-wide prediction approach while also considering additive effects. Interestingly, the prediction accuracy dropped severely when additionally dominance (0.15) or epistatic effects (0.08) were included.

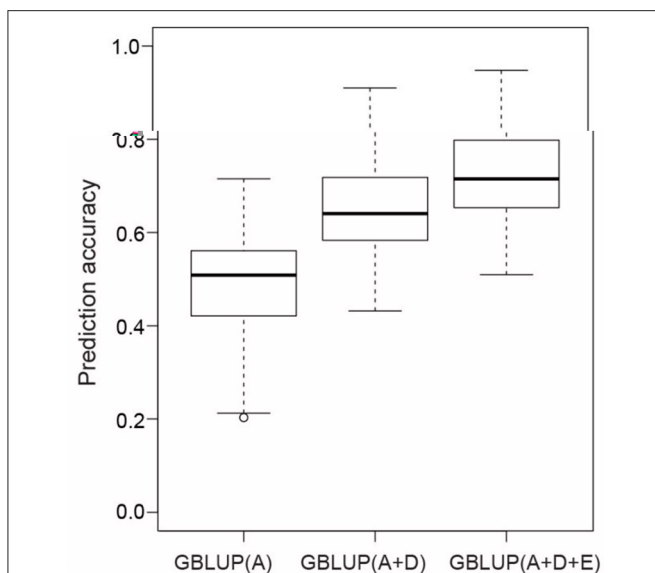
## DISCUSSION

The immortalized TNRC-F<sub>2</sub> population used in our study was initially designed to determine the genetic basis of midparent heterosis of hybrids between the European winter-type cultivar Tapidor and the Chinese semi-winter type cultivar Ningyou7 (Shi et al., 2011). Immortalized F<sub>2</sub> populations exhibited an expected allele frequency of one-half, which maximized the variance of dominance effects (Falconer and Mackay, 1996) and enhanced the variance of epistatic effects (Mackay, 2014). Because dominance and epistatic effects determine midparent heterosis (Melchinger et al., 2008), its genetic basis can be studied thoroughly in immortalized F<sub>2</sub> populations. The variance of additive effects is twice as high in doubled haploid populations compared with hybrid populations (Falconer and Mackay, 1996). Thus, an integrated analysis of doubled haploid and immortalized F<sub>2</sub> populations represents a powerful approach for studying the contribution of additive, dominance, and epistatic effects to hybrid performance. This approach encouraged us to implement a strategy that involved reanalyzing seed yield data of the TNDH and TNRC-F<sub>2</sub> populations (Shi et al., 2011) using the newly available high-throughput 60,000 SNP array (Zhang et al., 2016).

### Hybrid Performance Is Substantially Influenced By Dominance Effects

Distinguishing the variance of the hybrid performance into  $\sigma_{GCA}^2$  and  $\sigma_{SCA}^2$  provides the first insights into the role of additive and dominance effects. Assuming there is an absence of epistasis,  $\sigma_{GCA}^2$  is determined for immortalized F<sub>2</sub> populations through additive effects and  $\sigma_{SCA}^2$  was controlled by dominance effects (Lynch and Walsh, 1998). Thus, the predominance of  $\sigma_{SCA}^2$  was 1.4 times larger than  $\sigma_{GCA}^2$  (Table 1) and clearly points to the relevance of dominance effects in the TNRC-F<sub>2</sub> hybrid rapeseed population. Our findings agreed with previous results on grain yield reported for F<sub>2</sub> populations in maize (Wolf et al., 2000) and rice (Li et al., 2010) generated using the North Carolina Design III (Comstock and Robinson, 1952). By contrast, studies using factorial mating designs based on rapeseed (Brandle and McVetty, 1989), maize (Parisseau and Bernardo, 2004), wheat (Zhao et al., 2015a), or barley diversity panels (Philipp et al., 2016) observed that  $\sigma_{GCA}^2$  is the main component of the variance of the hybrid performance. These discrepancies can be explained by allele frequencies substantially deviating from one-half in factorial crosses among diverse inbred lines, which severely impacts the ratio of  $\sigma_{GCA}^2$  vs.  $\sigma_{SCA}^2$  (Falconer and Mackay, 1996).

The findings of the genome-wide QTL mapping study (Supplementary Table 2) complement the picture of the relevance of dominance effects at the molecular level. We detected 16 times as many marker-trait associations that contributed to the dominance more than to the additive



**FIGURE 3 |** Box-Whisker plots of whole-genome best linear unbiased prediction (GBLUP) accuracies of seed yield evaluated with five-fold cross validation. Additive (A), dominance (D), and epistatic effects (E) were gradually included in the GBLUP models.

variation (Figure 2). The findings were in line with earlier studies reporting that dominance effects are key factors of heterosis in rapeseed (Radoev et al., 2008; Basunanda et al., 2010; Shi et al., 2011; Bu et al., 2015; Wen et al., 2015). Nevertheless, it is important to note that these results were not cross-validated, which was recommended to obtain an unbiased picture of the contribution of genetic effects to phenotypic variation (Utz et al., 2000). When we applied cross validation, the variance explained by all main effect QTL decreased from 55 to 7%, but the advantage for prediction accuracy when fitting data besides the additive, as well as the dominance effects, remains substantial (44%).

The small proportion of genetic variance explained by all QTL indicates a complex genetic architecture of seed yield in rapeseed, with the presence of many QTL each contributing only a little to the phenotypic variation. Genome-wide predictions are more suitable to tackle such complex traits (Riedelsheimer et al., 2012). Similarly, we observed a 2.5 times higher accuracy for the genome-wide predictions based on main effects (Figure 3) compared with the approach using QTL only. Nevertheless, the benefits of modeling the effects of the additive along with the dominance effects was comparable (33%, Figure 3) to the results observed for QTL mapping (44%). In summary, the phenotypic data analyses, the QTL mapping, and the genome-wide prediction study suggested that dominance effects were substantially contributing to the phenotypic variation of seed yield for the Tapidor × Ningyou7 rapeseed hybrid.

## Genome-Wide Prediction Revealed a Prominent Role of Epistasis Contributing to the Hybrid Performance

According to previous results (Radoev et al., 2008; Basunanda et al., 2010; Shi et al., 2011; Bu et al., 2015; Wen et al., 2015), a large number of 251 significant digenic epistatic effects were detected in the genome-wide QTL mapping study (Supplementary Table 3). Around 31.5% (79) of the digenic interaction were additive by dominance effects involving SNPs on chromosome A03 and C04 (Supplementary Table 3). These 79 epistatic pairs trace back to eight clusters of markers, five on chromosome A03, and three on chromosome C04. Some of the SNPs of the eight clusters were also associated with main effect QTL for flowering time, seed weight, seed number, and oil content (Supplementary Table 3). Marker Bn-scaff\_16534\_1-p2320270 located on C04 showed significant dominance effects (Supplementary Table 2) and was together with closely linked markers involved in 26 of the epistatic interactions between chromosomes A03 and C04. Thus, Bn-scaff165341-p2320270 is an interesting candidate for further fine mapping studies.

Despite the large number of 251 significant digenic epistatic effects, the cross-validated prediction accuracy did not increase when modeling epistatic effects beyond main effects. This can either point to the irrelevance of epistasis for seed yield in rapeseed or can be explained by the challenge to detect epistatic effects for the complex trait seed yield (Mackay, 2014). The results of the genome-wide prediction study support the

latter explanation. The prediction accuracy increased by 11% compared with the prediction approach based on additive and dominance effects only (Figure 3). The observed increase is much higher than that reported in previous studies on genome-wide prediction of hybrid performance. Zhao et al. (2015a) observed a 2% increase in prediction ability when modeling epistatic effects in a large population of 1,604 wheat hybrids and their 135 parental inbred lines. The lower benefit observed for wheat vs. rapeseed can be explained by the more marginal allele frequencies in factorial mating designs compared with immortalized F<sub>2</sub> design. Surprisingly, Xu et al. (2014) observed no benefits when modeling epistatic effects in a large immortalized F<sub>2</sub> population comprised of 240 rice inbred lines and 360 F<sub>2</sub> genotypes. The discrepancy between our findings and that of Xu et al. (2014) might point to differences in the genetic architecture among crop species, as has been observed for the genetic basis of heterosis in the self-pollinating species rice vs. the outcrossing species of maize (Garcia et al., 2008) and deserves further research.

The total population was used to estimate the genetic components of variance with a Bayesian generalized linear regression based on the design matrices of additive, dominance, and all types of digenic epistatic effects. We observed a prominent role of the variance of dominance effects contributing to 49% of the genetic variance (Supplementary Figure 4). Additive (27%) and the sum of all digenic epistatic effects (24%) contributed nearly equally to the genetic variance of the hybrid performance. This suggests that the Tapidor × Ningyou7 rapeseed hybrid successfully exploits all types of genetic effects. Whether our findings are also valid for further rapeseed hybrids or not, deserve further research in other genetic backgrounds.

## Additive Effects Were Less Affected By Environment Interactions than Non-additive Effects

The phenotypic data analyses revealed that  $\sigma_{GCA}^2$  was less affected by varying environmental conditions than  $\sigma_{SCA}^2$  (Table 1). This result contrasted with previous findings based on an F<sub>2</sub> population derived from the B73 × Mo17 maize hybrid (Wolf et al., 2000), which again points to differences in the genetic architecture among crop species. Validating the developed prediction model in an independent sample of genotypes and environments revealed that the observation at the phenotypic level was also reflected at the molecular level. The prediction accuracy is stable for additive effects (0.49) but collapsed when adding dominance (0.15) or epistatic effects (0.08) (Figure 3). Thus, the development of robust hybrid prediction models, including dominance and epistatic effects, requires a much deeper phenotyping analysis in multi-environmental trials. The flip side of our finding is that prediction models focusing exclusively on additive effects yields already stable and high prediction accuracies. Additive models can be easier implemented simplifying the application of hybrid prediction in plant breeding programs.

## CONCLUSIONS

Our study was based on a doubled haploid and immortalized  $F_2$  population derived from the single-cross hybrid Tapidor  $\times$  Ningyou7, and the conclusions are restricted to this gene space. The phenotypic data analyses, the QTL mapping, and the genome-wide predictions revealed that hybrid performance is driven by a mix of additive, dominance, and epistatic effects. Prediction accuracies substantially profited when integrating dominance and epistatic effects, which is most likely to result from using a mapping population with expected allele frequencies of one-half. Transferring the results to a broader diversity involves the challenge that this entails to move to a gene space with less balanced allele frequencies, which led to a low power to exploit dominance and epistatic effects. Further research is required to search for an optimum compromise between the inference space of the results and precision to predict additive as well as dominance and epistatic effects.

## AUTHOR CONTRIBUTIONS

PL and YZ performed the research, analyzed the data, and wrote the manuscript; GL, MW, DH, and JH analyzed the data; MW helped the genotyping and phenotyping of new hybrids; JM provided plant materials, the previous published data and suggestions; JZ and JR designed the research, analyzed the data, and wrote the manuscript. All authors revised, read, and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (No. 2016YFD0101300), and

## REFERENCES

- Anscombe, F. J., and Tukey, J. W. (1963). The examination and analysis of residuals. *Technometrics* 5, 141–160. doi: 10.1080/00401706.1963.10490071
- Basunanda, P., Radoev, M., Ecke, W., Friedt, W., Becker, H. C., and Snowdon, R. J. (2010). Comparative mapping of quantitative trait loci involved in heterosis for seedling and yield traits in oilseed rape (

the National Basic Research Program of China (Grant No. 2015CB150200).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00815/full#supplementary-material>

**Supplementary Table 1 | The environments used for seed yield field trial of TNRC- $F_2$  population.**

**Supplementary Table 2 | Significant (Bonferroni-Holm significant level  $P < 0.1$ ) marker-trait associations and the proportion of explained genotypic variance (GVE) detected in a genome-wide association mapping approach for seed yield.**

**Supplementary Table 3 | Significant (Bonferroni-Holm significant level  $P < 0.1$ ) epistasis interaction the proportion of explained genotypic variance (GVE) detected in a genome-wide-association mapping approach for seed yield.**

**Supplementary Figure 1 | The schematic diagram of the crossing design for 318 TN RC- $F_2$  hybrids used in this study.** The corresponding DH parents involved in each of the crossings are shown on the x-axis and y-axis, respectively. The red boxes indicate the crosses.

**Supplementary Figure 2 | The distribution and correlation among best linear unbiased estimates (BLUEs) of each environment.**

**Supplementary Figure 3 | Quantile-Quantile plots of association mapping for determining the models of QTL detection for hybrid seed yield. (A,B)** showed the expected  $P$ -value of association mapping using different models for detecting the QTLs with additive effects and dominance effects, respectively. The blue and red lines showed  $P$ -values of association mapping using the model without kinship and with kinship, respectively.

**Supplementary Figure 4 | Pie chart of genetic components of variance (additive variance  $\sigma_A^2$ , dominance variance  $\sigma_D^2$ , and respective epistatic variance components) estimated with Bayesian generalized linear regression.**



- He, S., Reif, J. C., Korzun, V., Bothe, R., Ebmeyer, E., and Jiang, Y. (2016a). Genome-wide mapping and prediction suggests presence of local epistasis in a vast elite winter wheat populations adapted to Central Europe. *Theor. Appl. Genet.* 130, 635–647. doi: 10.1007/s00122-016-2840-x
- He, S., Schulthess, A. W., Mirdita, V., Zhao, Y., Korzun, V., Bothe, R., et al. (2016b). Genomic selection in a commercial winter wheat population. *Theor. Appl. Genet.* 129, 641–651. doi: 10.1007/s00122-015-2655-1
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Jan, H. U., Abbadi, A., Lucke, S., Nichols, R. A., and Snowdon, R. J. (2016). Genomic prediction of testcross performance in canola (*Brassica napus*). *PLoS ONE* 11:e0147769. doi: 10.1371/journal.pone.0147769
- Jiang, Y., and Reif, J. C. (2015). Modeling epistasis in genomic selection. *Genetics* 201, 759–768. doi: 10.1534/genetics.115.177907
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101
- Kempe, K., and Gils, M. (2011). Pollination control technologies for hybrid breeding. *Mol. Breed.* 27, 417–437. doi: 10.1007/s11032-011-9555-0
- Kumar, S., Molloy, C., Munoz, P., Daetwyler, H., Chagne, D., and Volz, R. (2015). Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. *G3* 5, 2711–2718. doi: 10.1534/g3.115.021105
- Ledoit, O., and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* 88, 365–411. doi: 10.1016/S0047-259X(03)00096-4
- Liu, G., Wang, Y., Wang, Y., Wang, Y., Wang, Y., Wang, Y., et al. (2015). Epistatic actions at loci underlying several quantitative traits in two elite rice hybrids. *Mol. Genet. Genomics* 284, 383–397. doi: 10.1007/s00438-010-0575-y
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits [M]*. Sunderland: Sinauer.
- Mackay, T. F. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat. Rev. Genet.* 15, 22–33. doi: 10.1038/nrg3627
- Melchinger, A. E., Utz, H. F., and Schon, C. C. (2008). Genetic expectations of quantitative trait loci main and interaction effects obtained with the triple testcross design and their relevance for the analysis of heterosis. *Genetics* 178, 2265–2274. doi: 10.1534/genetics.107.084871
- Parisseaux, B., and Bernardo, R. (2004). *In silico* mapping of quantitative trait loci in maize. *Theor. Appl. Genet.* 109, 508–514. doi: 10.1007/s00122-004-1666-0
- Perez, P., and Campos, G. D. L. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Philipp, N., Liu, G., Zhao, Y., He, S., Spiller, M., Stiewe, G., et al. (2016). Genomic prediction of barley hybrid performance. *Plant Genome* 9:2. doi: 10.3835/plantgenome2016.02.0016
- Qiu, D., Morgan, C., Shi, J., Long, Y., Liu, J., Li, R., et al. (2006). A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor. Appl. Genet.* 114, 67–80. doi: 10.1007/s00122-006-0411-2
- Radoev, M., Becker, H. C., and Ecke, W. (2008). Genetic analysis of heterosis for yield and yield components in rapeseed (*Brassica napus* L.) by quantitative trait locus mapping. *Genetics* 179, 1547–1558. doi: 10.1534/genetics.108.089680
- Reif, J. C., Zhao, Y. S., Wurschum, T., Gowda, M., and Hahn, V. (2013). Genomic prediction of sunflower hybrid performance. *Plant Breed.* 132, 107–114. doi: 10.1111/pbr.12007
- Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220. doi: 10.1038/ng.1033
- Shi, J. Q., Li, R. Y., Zou, J., Long, Y., and Meng, J. L. (2011). A dynamic and complex network regulates the heterosis of yield-correlated traits in rapeseed (*Brassica napus* L.). *PLoS ONE* 6:e21645. doi: 10.1371/journal.pone.0021645
- Technow, F., Riedelsheimer, C., Schrag, T. A., and Melchinger, A. E. (2012). Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. *Theor. Appl. Genet.* 125, 1181–1194. doi: 10.1007/s00122-012-1905-8
- Utz, H. F., Melchinger, A. E., and Schon, C. C. (2000). Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. *Genetics* 154, 1839–1849.
- Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980
- Wang, X., Li, L., Yang, Z., Zheng, X., Yu, S., Xu, C., et al. (2017). Predicting rice hybrid performance using univariate and multivariate GBLUP models based on North Carolina mating design II. *Heredity* 118, 302–310. doi: 10.1038/hdy.2016.87
- Wen, J., Zhao, X., Wu, G., Xiang, D., Liu, Q., Bu, S. H., et al. (2015). Genetic dissection of heterosis using epistatic association mapping in a partial NCII mating design. *Sci. Rep.* 5:18376. doi: 10.1038/srep18376
- Wolf, D. P., Peternelli, L. A., and Hallauer, A. R. (2000). Estimates of genetic variance in an F2 maize population. *J. Hered.* 91, 384–391. doi: 10.1093/jhered/91.5.384
- Xu, S. H., Zhu, D., and Zhang, Q. F. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111
- Yan, X., Zeng, X., Wang, S., Li, K., Yuan, R., Gao, H., et al. (2016). Aberrant meiotic prophase I leads to genic male sterility in the novel TE5A mutant of *Brassica napus*. *Sci. Rep.* 6:33955. doi: 10.1038/srep33955
- Yu, C. Y., Guo, Y. F., Ge, J., Hu, Y. M., Dong, J. G., and Dong, Z. (2015). Genomic prediction of hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111