advances.sciencemag.org/cgi/content/full/6/24/eaay4897/DC1

Supplementary Materials for

Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat

Philipp H. G. Boeven, Yusheng Zhao, Patrick Thorwarth, Fang Liu, Hans Peter Maurer, Mario Gils, Ralf Schachschneider, Johannes Schacht, Erhard Ebmeyer, Ebrahim Kazman, Vilson Mirdita, Jost Dörnte, Stefan Kontowski, Ralf Horbach, Hilmar Cöster, Josef Holzapfel, Andreas Jacobi, Ludwig Ramgraber, Carsten Reinbrecht, Norbert Starck, Pierrick Varenne, Anne Starke, Friederike Schürmann, Martin Ganal, Andreas Polley, Jens Hartung, Sebastian Beier, Uwe Scholz, C. Friedrich H. Longin, Jochen C. Reif*, Yong Jiang, Tobias Würschum

*Corresponding author. Email: reif@ipk-gatersleben.de

Published 12 June 2020, *Sci. Adv.* **6**, eaay4897 (2020)
DOI: 10.1126/sciadv.aay4897

The PDF file includes:

Supplementary Note
Figs. S1 to S13
Tables S1 to S7

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/24/eaay4897/DC1)

Data file S1
Supplementary R code

Supplementary Materials

Supplementary Note

Quantitative genetic theory underlying the heterotic genetic distance

In this study, we devised a novel genetic distance measure that allows to incorporate information on the genetic architecture of heterosis. First, dominance effects were estimated in a Bayesian genome-wide prediction framework (*3, 10, 13*) by the model

$$Y = \mu + Z_A a + Z_D d + e \tag{1}$$

where Z_A is the design matrix for the additive marker effect $a = (a_1, \dots, a_L)$, the elements of Z_A are -1,0,1. Z_D is the design matrix for the dominance marker effect $d = (d_1, \dots, d_L)$, the elements of Z_D are 0,1, while L is the number of markers. These dominance effects d_u , $u = 1,2, \dots , L$, were then used to derive the weight for each marker u as

$w_u = \frac{d_u}{|\widehat{d}|}$, where $|\widehat{d}| = \frac{1}{L}\sum_{u=1}^{L}|d_u|$ is the mean of the absolute values of all dominance effects.

The heterotic genetic distance f_{RD} extends the Rogers' distance (*34*) by incorporating the weight w_u for each marker as

$$f_{RD}(X,Y) = \frac{1}{L}\sum_{u=1}^{L} w_u \sqrt{\frac{\sum_{j=1}^{n_u}\left(X_{uj} - Y_{uj}\right)^2}{2}} = \frac{1}{L}\sum_{u=1}^{L} w_u \left|X_{uj} - Y_{uj}\right| = \frac{\sum_{u=1}^{L} d_u \left|X_{uj} - Y_{uj}\right|}{\sum_{u=1}^{L}|d_u|}$$
$$= \frac{\sum_{u=1}^{L} d_u Z_{D_u}}{\sum_{u=1}^{L}|d_u|} \tag{2}$$

where Z_{D_u} is the element for *u*-th marker in the Z_D matrix for a hybrid with the two parents X and Y, and $n_u = 2$ is the number of alleles per marker. Notably, positive dominance effects increase the heterotic genetic distance, whereas negative dominance effects decrease it.

So $f_{RD}(X,Y)$ for all pairs of hybrids is equivalent to $\frac{Z_D d}{\sum_{u=1}^{L}|d_u|}$, where $Z_D d$ is derived from the model (1).

If we define the genetic values of the hybrids as G_{F_1}, then we have

$$Z_D d = G_{F_1} - Z_A a = G_{F_1} - \frac{1}{2}\left(Z_{P_1} + Z_{P_2}\right)a = G_{F_1} - \frac{1}{2}\left(G_{P_1} + G_{P_2}\right) \tag{3}$$

where G_{P_1} and G_{P_2} are genetic values of the two parents P_1 and P_2. From this it follows that $Z_D d$ is the genetic value of the midparent heterosis, when ignoring epistatic effects.

Here, the midparent value of a hybrid stems from the additive effects $MP = \mathrm{Z_A a}$, and the midparent heterosis from the dominance effects $MPH = Z_D d$. Thus, $\sum_{u=1}^{L}|d_u|$ is the range of the midparent heterosis $Z_D d$, because $\sum_{u=1}^{L}|d_u| = \sum_{u \in S1} d_u - \sum_{u \in S2} d_u$, where S_1 is the set of markers with positive dominance effect d_u and S_2 is the set of markers with negative d_u. Consequently, $\sum_{u \in S1} d_u$ is the maximum value of $Z_D d$ in this population and $\sum_{u \in S2} d_u$ is the minimum value of $Z_D d$ in this population.

From formula (2) we can see that if we assume all dominance effects d_u are none zero and equal, then the heterotic Rogers' distance is equal to the Rogers' distance between X and Y:

$$f_{RD}(X, Y) = \frac{1}{L} \sum_{u=1}^{L} \sqrt{\frac{\sum_{j=1}^{n_u} \left(X_{uj} - Y_{uj}\right)^2}{2}} = RD(X, Y) \qquad (4)$$

Table S1. Summary statistics for evaluated traits and heterosis. Shown for the traits grain yield, heading time, and plant height

Source	Grain yield (Mg ha^{-1})	Heading (days)	Plant height (cm)
Elite hybrids (*n* = 1655)			
h²	0.81	0.94	0.94
Mean (Min; Max)	9.93 (7.48; 10.89)	152.74 (148.19; 158.31)	90.01 (72.68; 105.18)
Mean MPH (Min, Max)	0.83 (-0.75; 2.39)	-0.75 (-3.62; 2.77)	6.21 (-7.54; 18.83)
Mean MPH% (Min; Max)	9.12 (-8.21; 28.30)	-0.49 (-2.36; 1.82)	7.44 (-8.96; 23.49)
Mean BPH (Min; Max)	0.60 (-1.63; 1.91)	-1.60 (-7.49; 1.35)	2.32 (-11.55; 14.56)
Mean BPH% (Min; Max)	6.45 (-17.41; 21.38)	-1.04 (-4.80; 0.89)	2.74 (-13.71; 18.30)
Historic×Elite hybrids (*n* = 96)			
h²	0.83	0.79	0.76
Mean (Min; Max)	9.35 (8.30; 10.72)	152.97 (148.71; 157.75)	91.02 (83.00; 98.73)
Mean MPH (Min; Max)	0.81 (-0.12; 2.57)	-1.30 (-4.79; 1.39)	5.96 (-1.64; 13.30)
Mean MPH% (Min; Max)	9.48 (-1.32; 31.51)	-0.84 (-3.06; 0.91)	6.33 (-1.73; 14.03)
Mean BPH (Min; Max)	0.47 (-0.62; 2.03)	-2.42 (-6.51; 0.28)	3.03 (-4.02; 11.46)
Mean BPH% (Min; Max)	5.27 (-6.16; 23.34)	-1.56 (-4.20; 0.18)	3.46 (-4.62; 13.14)
Exotic×Elite hybrids (*n* = 152)			
h²	0.82	0.64	0.84
Mean (Min; Max)	8.76 (7.59; 9.63)	155.22 (151.93; 158.40)	112.71 (94.07; 124.03)
Mean MPH (Min; Max)	0.72 (-0.77; 1.79)	-1.41 (-5.37; 6.13)	9.02 (-1.22; 32.81)
Mean MPH% (Min; Max)	9.17 (-9.18; 24.44)	-0.88 (-3.23; 4.09)	9.15 (-1.03; 37.59)
Mean BPH (Min; Max)	-0.66 (-1.49; 0.14)	-3.30 (-10.09; 1.36)	-13.33 (-38.11; 31.05)
Mean BPH% (Min; Max)	-7.00 (-16.29; 1.46)	-2.07 (-6.12; 0.88)	-9.19 (-25.00; 34.88)

MPH, absolute midparent heterosis; *MPH%*, relative midparent heterosis; *BPH*, absolute better parent heterosis; *BPH%*, relative better parent heterosis

Table S2. Genome-wide prediction accuracies for mid-parent heterosis in the Elite and Exotic×Elite sets considering different genetic effects

Model	Elite	Exotic×Elite
D[a]	0.59	0.65
AA[c]	0.77	0.68
D+AA	0.75	0.69
D+AA+AD[c]	0.76	0.69
D+AA+AD+DD[d]	0.75	0.69

[a] Dominance
[b] Additive-by-additive
[c] Additive-by-dominance
[d] Dominance-by-dominance

Table S3. Correlations among marker-derived kinship matrices of dominance (D), and respective digenic epistatic (AA, AD, DD) effects in the genome-wide prediction model partitioning genetic variance components (VC) within the entire hybrid panel

VC	D	AA	AD	DD
D	1.00	0.06	0.92	0.93
AA		1.00	0.20	0.16
AD			1.00	0.89
DD				1.00

Table S4. Correlations among marker-derived kinship matrices of dominance (D), and respective digenic epistatic (AA, AD, DD) effects in the genome-wide prediction model partitioning genetic variance components (VC) within the Elite set

VC	D	AA	AD	DD
D	1.00	0.46	0.73	0.99
AA		1.00	0.76	0.53
AD			1.00	0.75
DD				1.00

Table S5. Correlations among marker-derived kinship matrices of dominance (D), and respective digenic epistatic (AA, AD, DD) effects in the genome-wide prediction model partitioning genetic variance components (VC) within the Exotic×Elite set

VC	D	AA	AD	DD
D	1.00	0.84	0.82	0.99
AA		1.00	0.97	0.86
AD			1.00	0.83
DD				1.00

Table S6. Number of significant genetic effects and heterotic effects detected in the Elite and Exotic×Elite sets

Genetic effect	Elite [a]	Exotic×Elite [b]
D	1	3
AA	111	7
AD	71	27
DD	30	106
Heterotic effect	50	21

[a] Elite set: Treshold $P < 0.05$ after Bonferroni correction, [b] Exotic×Elite set: Treshold $P < 0.05$ after modified Bonferroni correction

Table S7. Location information. Locations used for the field trials in the growing-season 2015/16

Location	Latitude	Longitude	Altitude	Mean annual precipitation	Mean annual temperature	Set
Asendorf	52°44'17.934'' N	9°0'24.105'' E	45 m	751 mm	9.3 °C	Elite
Biendorf	51°45'0'' N	11°50'59'' E	79 m	470 mm	8.8 °C	Elite
Gatersleben	51°50'35.7'' N	11°18'1.512'' E	125 m	519 mm	9.8 °C	Elite, Historic/Exotic
Hadmersleben	51°58'37.916'' N	11°18'10.414'' E	91 m	500 mm	8.7 °C	Elite
Rosenthal	52°18'18.89'' N	10°10'52.88'' E	70 m	700 mm	9.8 °C	Elite
Seligenstadt	49°51'16.30'' N	10°06'2.30'' E	280 m	606 mm	9.2 °C	Elite
Stuttgart-Hohenheim	49°43'2.646'' N	9°11'12.699'' E	406 m	697 mm	8.8 °C	Historic/Exotic
Renningen	48°44'29.583'' N	8°55'15.35'' E	484 m	690 mm	7.9 °C	Historic/Exotic
Schackstedt	51°43'12'' N	11°37'11.999'' E	134 m	504 mm	8.8 °C	Historic/Exotic
Böhnshausen	51°51'31.401'' N	10°57'44.669'' E	130 m	580 mm	9.2 °C	Historic/Exotic

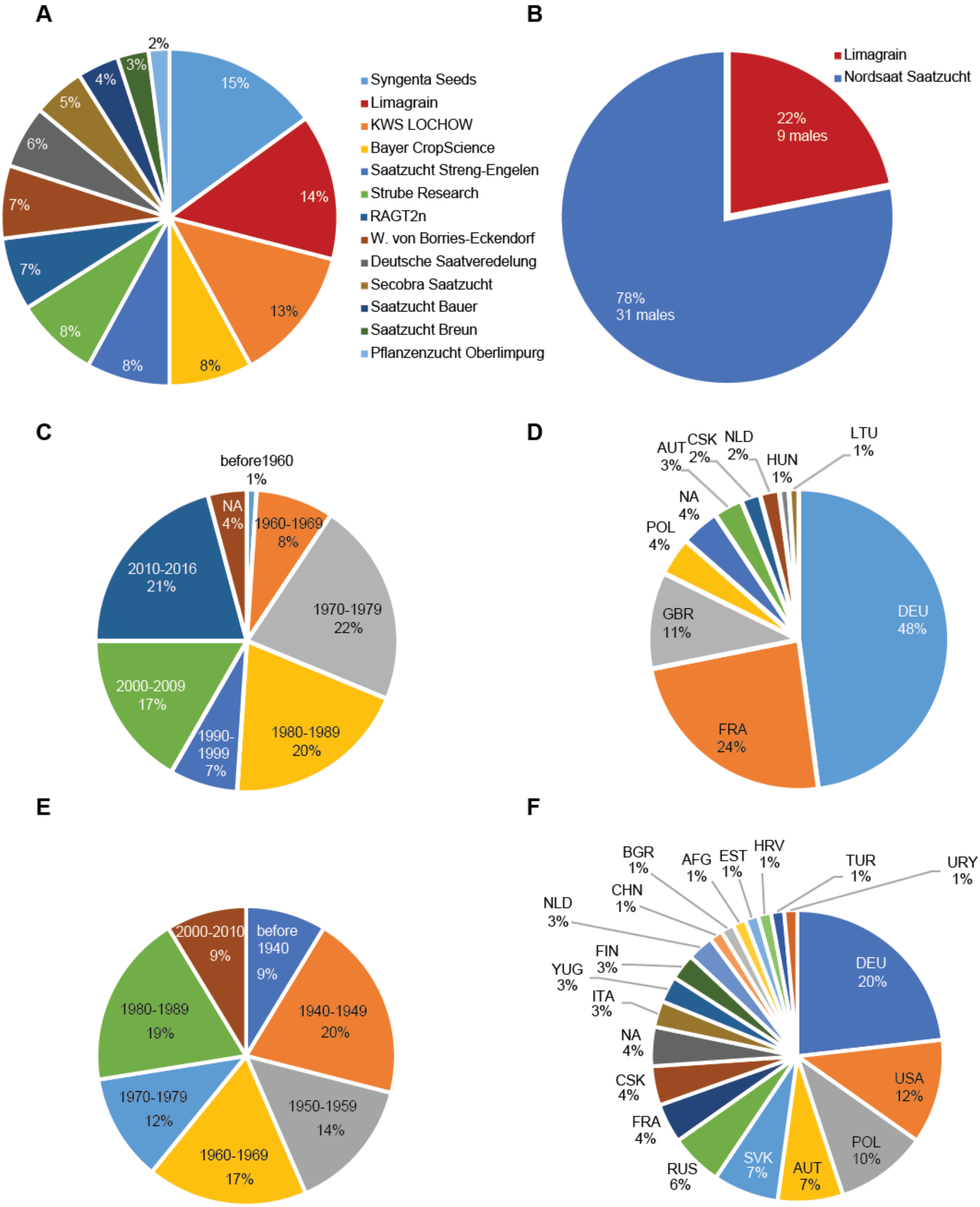

Fig. S1. Plant material information of parental lines. Breeders' affiliation of (**A**) 177 female lines and (**B**) 40 male lines and used within the Elite set. (**C**) Decade of release and (**D**) origin of 96 female lines within the Historic×Elite set. (**E**) Decade of release and (**F**) origin of 69 male lines within the Exotic×Elite set. *AFG*, Afghanistan; *AUT*, Austria; *BGR*, Bulgaria; *CHN*, China, *CSK*, former Czechoslovakia; *DEU*, Germany; *EST*, Estonia; *FIN*, Finland; *FRA*, France; *GBR*; Great Britain; *HRV*, Croatia; *HUN*, Hungary; *ITA*, Italy; *JPN*, Japan; *LTU*, Lithuania; *NA*, no data available; *NLD*, Netherlands; *NPL*, Nepal; *POL*, Poland; *RUS*, Russia; *SVK*, Slovak Republic; *TUR*, Turkey; *URY*, Uruguay; *USA*, United States of America; *YUG*, former Yugoslavia.

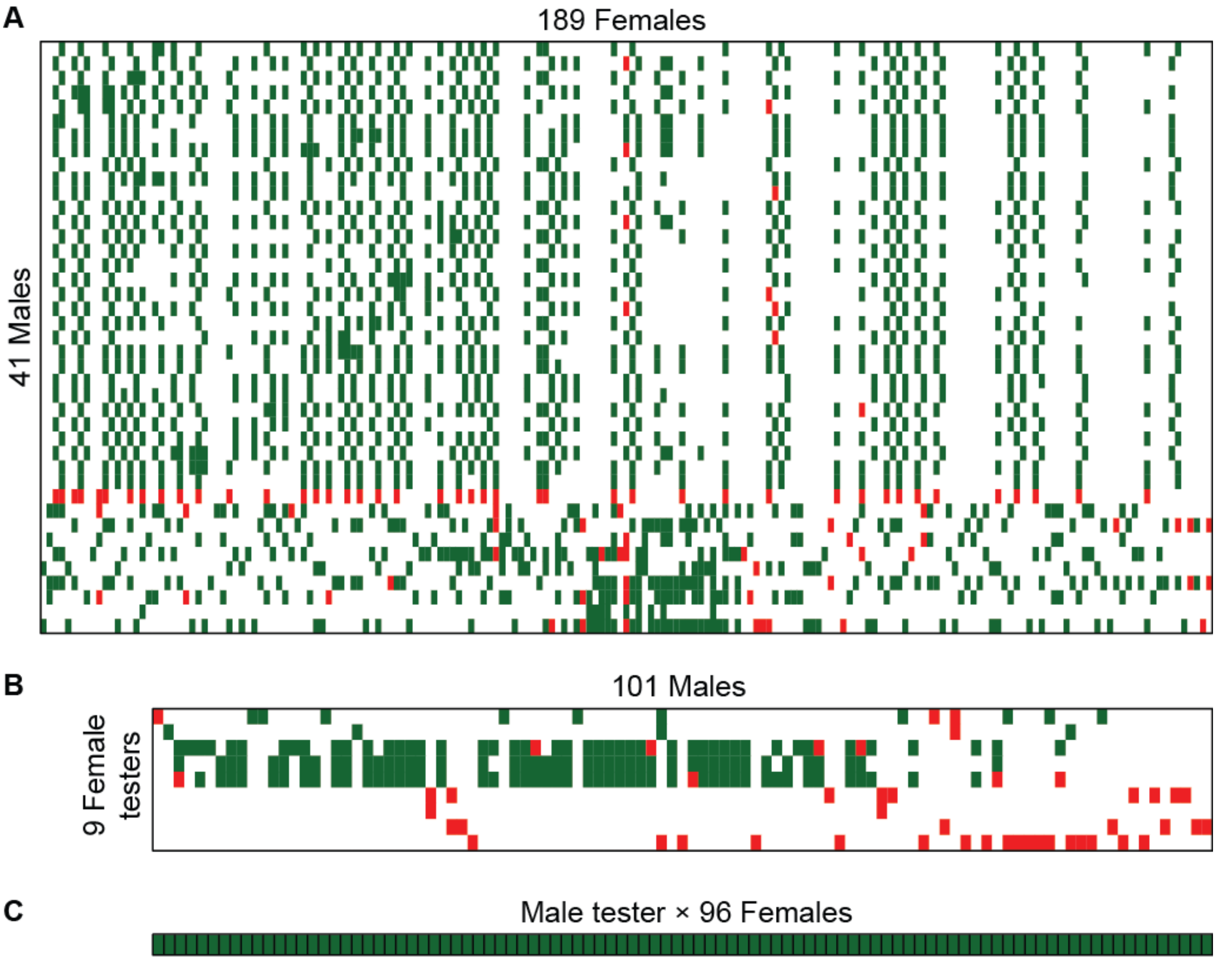

Fig. S2. Schematic illustration of the mating designs. Produced and tested hybrids that were used for the analyses are indicated by green boxes. Hybrids that were excluded for different reasons described in the Methods section are shown in red. (**A**) Elite; incomplete factorial mating design with 189 females, 41 males, resulting in 1,750 hybrids of which 1,655 were used for analyses. These Elite hybrids were evaluated at six locations in 2015/2016. (**B**) Exotic×Elite; incomplete factorial mating design with 101 males, 9 female testers, resulting in 200 hybrids of which 152 were used for analyses. These Exotic×Elite hybrids were evaluated at five locations in 2015/2016. (**C**) Historic×Elite; topcross mating design resulting in 96 hybrids that were tested at the same five locations as the Exotic×Elite hybrids.

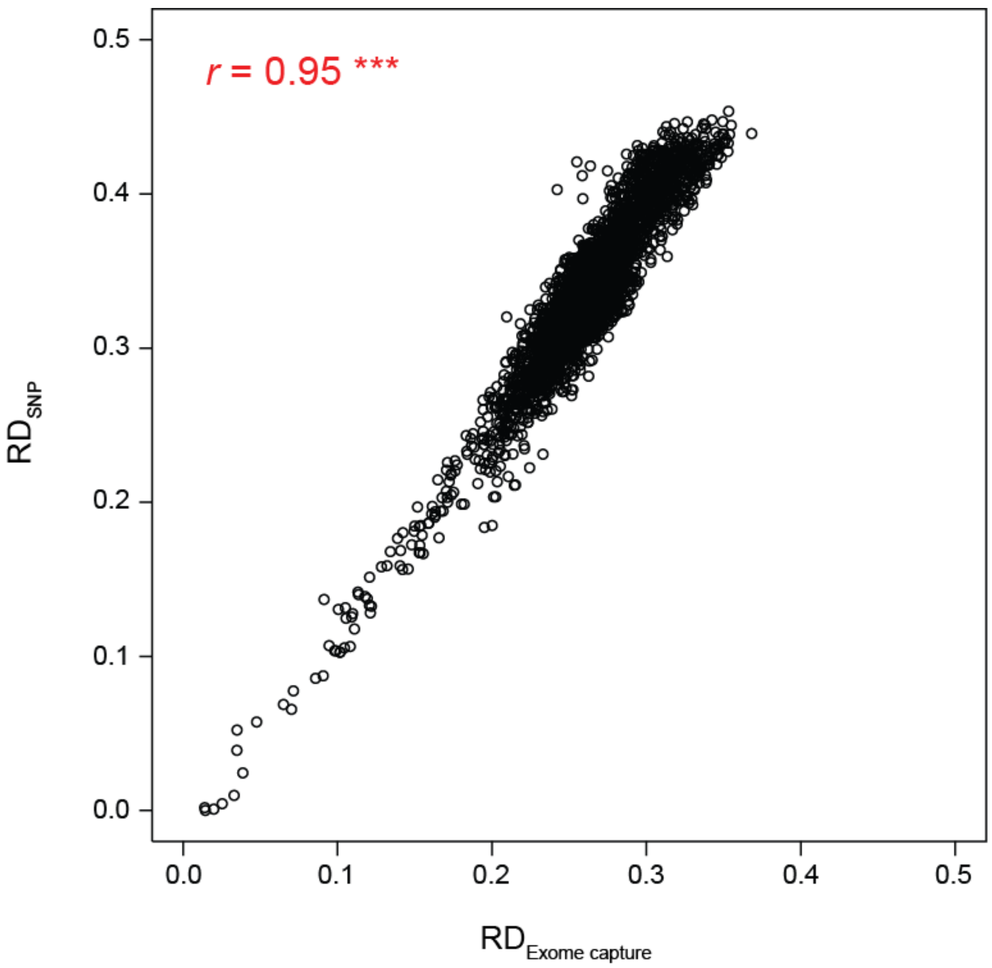

Fig. S3. Absence of ascertainment bias. All lines were genotyped by a 15,000 SNP array and a subset of 77 lines (40 Elite lines, 20 Historic and 17 Exotic genetic resources) were additionally fingerprinted using exome capture. The correlation between the genetic distance matrixes (Rogers' distance, RD) based on the SNP array data (RD_{SNP}) and exome capture ($RD_{Exome\ capture}$) was high with $r = 0.95$ ($P < 0.001$) and shows absence of ascertainment bias.

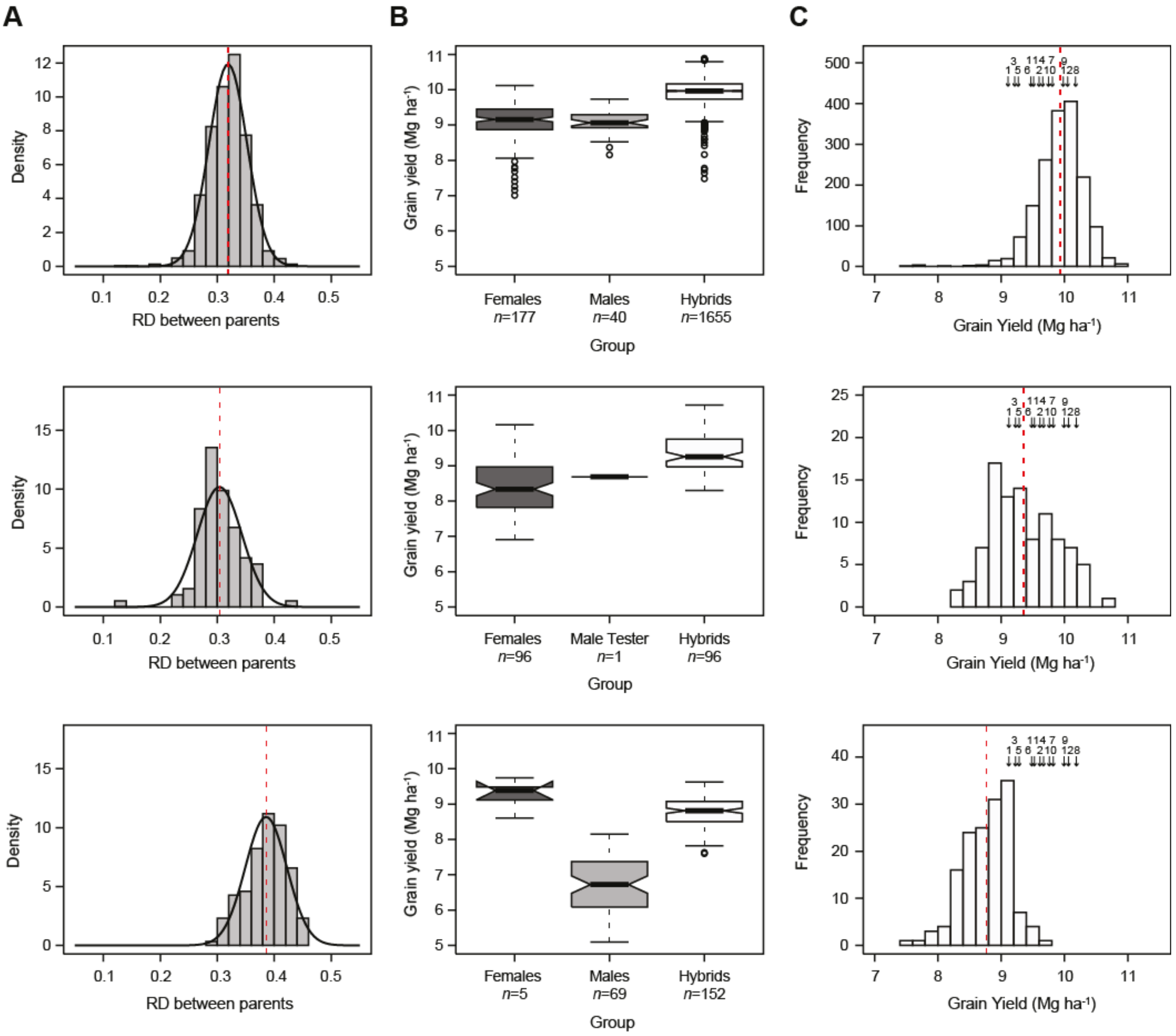

Fig. S4. Distributions of grain yield and Rogers' distance between parents in the three different sets of hybrids. (**A**) Distributions of Rogers' distances (RD) between 217 Elite parents, between 97 Historic parents including the male tester, and between 74 Exotic parents. (**B**) Boxplots showing grain yield of female parents, male parents, and hybrids for Elite, Historic×Elite and Exotic×Elite sets. (**C**) Distributions of grain yield for 1,655 Elite hybrids, 96 Historic×Elite hybrids and 152 Exotic×Elite hybrids. Grain yield of eleven commercial varieties used as checks is also indicated (1=Colonia, 2=Elixer, 3=Hybred, 4=Hystar, 5=JB Asano, 6=Julius, 7=KWS Loft, 8=KWS Smart, 9=LG Alpha, 10=RGT Reform, 11=Rumor and 12=Tobak).

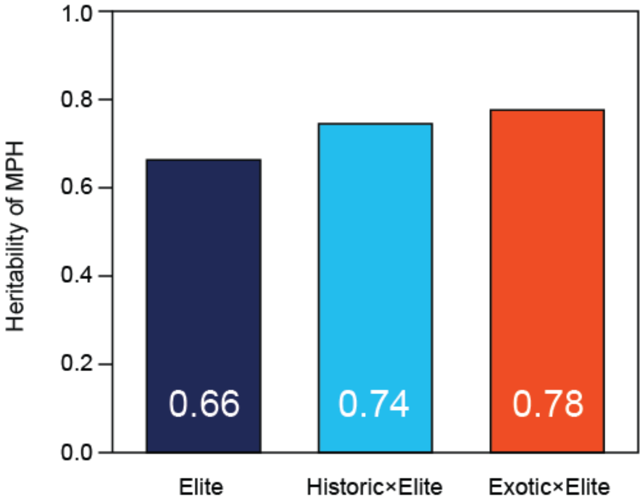

Fig. S5. Heritability estimates for midparent heterosis of grain yield.

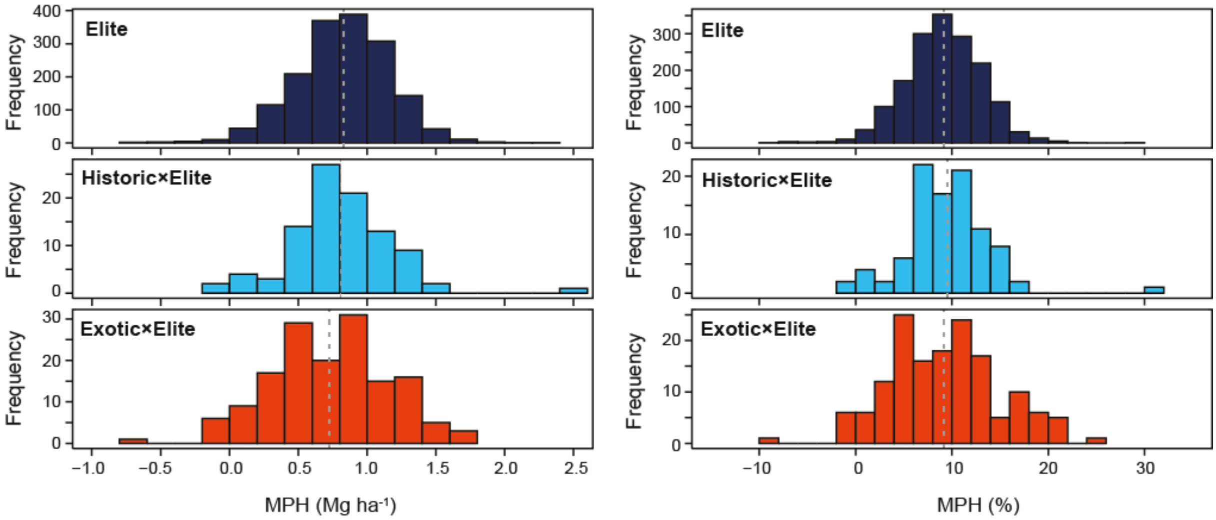

Fig. S6. Heterosis for grain yield. Distribution of absolute and relative midparent heterosis (MPH) for grain yield for the Elite, Historic×Elite and Exotic×Elite sets of hybrids. The dashed vertical line indicates the mean.

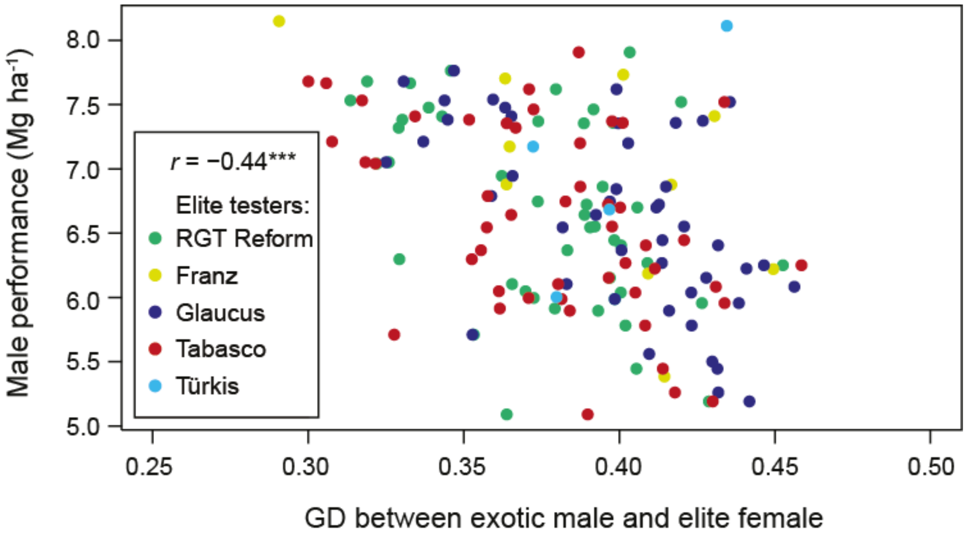

Fig. S7. Grain yield of the Exotic male lines decreases with genetic distance (GD) to the respective elite female tester.

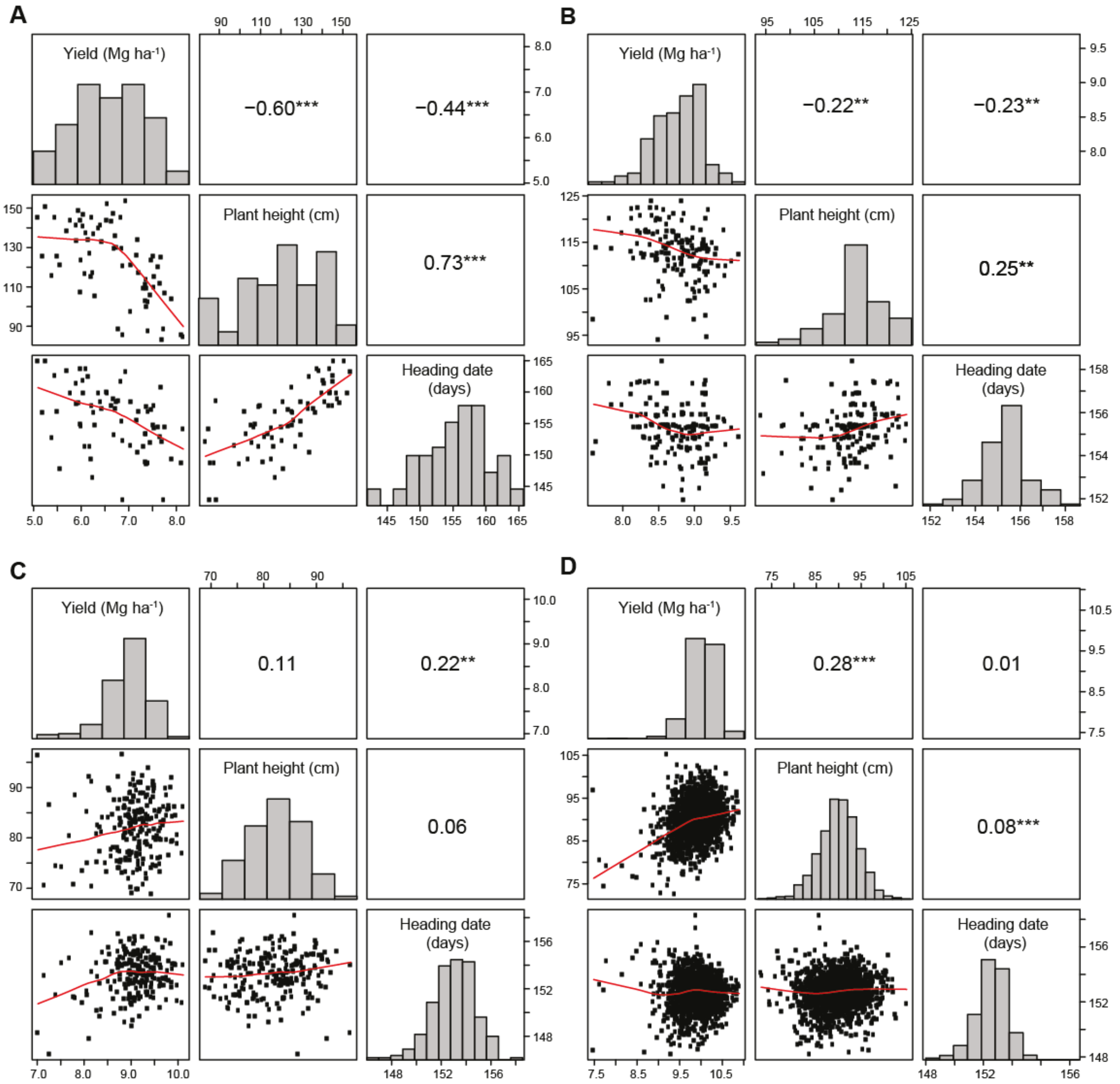

Fig. S8. Phenotypic trait correlations. (**A**) Exotic lines and (**B**) Exotic×Elite hybrids. (**C**) Elite lines and **d,** Elite hybrids. Grain yield was found to be strongly negatively correlated with plant height and heading time within the Exotic lines, but much less so in their hybrids from crosses with adapted elite material. In contrast, for the parental lines of the Elite hybrids weak positive correlations were observed.

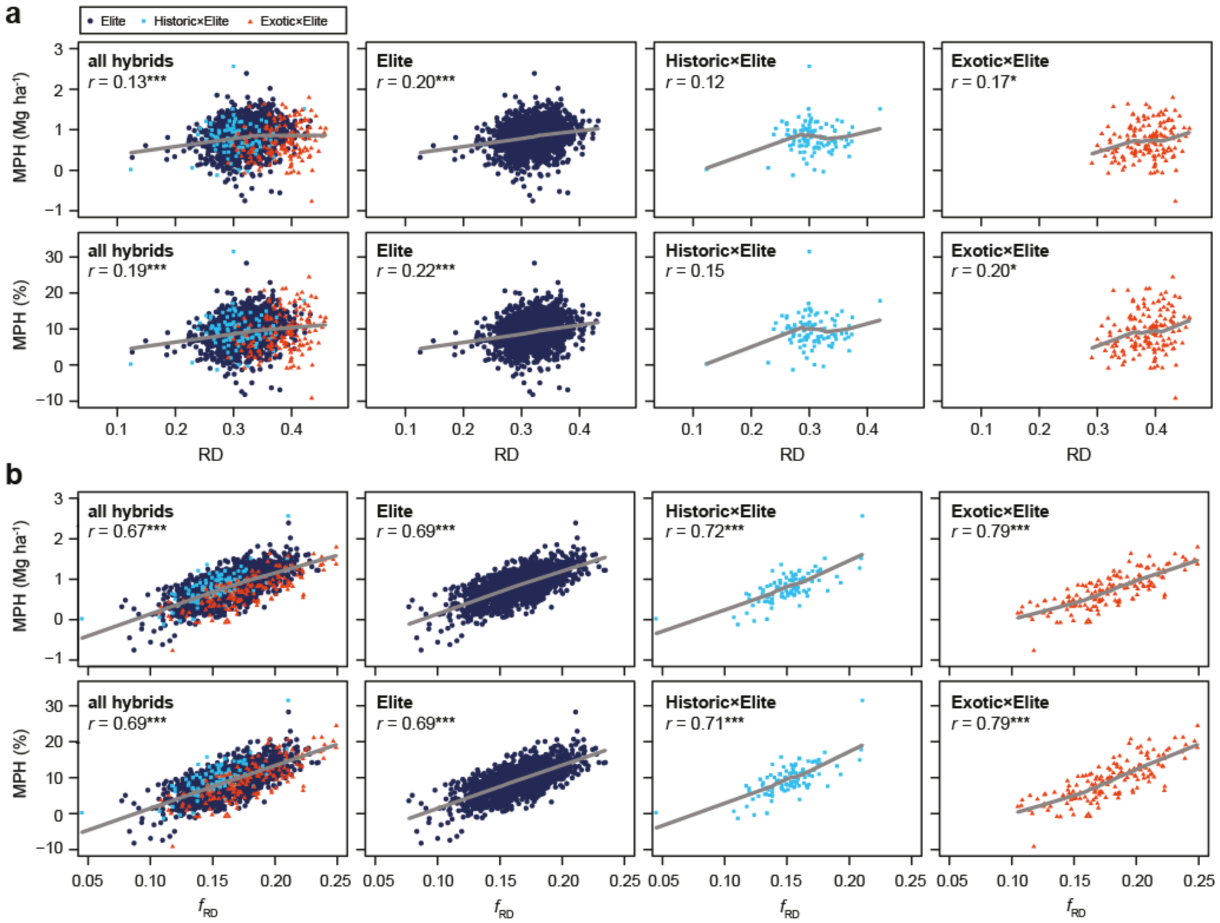

Fig. S9. Association between grain yield heterosis and genetic distance. Association between absolute or relative midparent heterosis of grain yield and (**A**) Rogers' distance (RD) or (**B**) heterotic genetic distance (f_{RD}), shown for the different sets of hybrids. The gray lines are locally weighted regression lines.

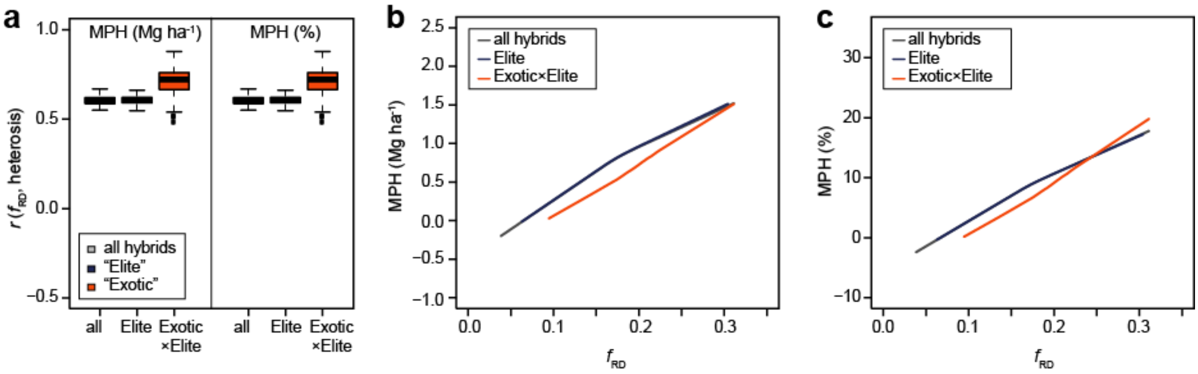

Fig. S10. Cross-validated correlation between heterosis and the heterotic genetic distance. (**A**) Boxplots showing correlations between heterosis and the heterotic Rogers' distance (f_{RD}) for dominance effects estimated in 80% of the hybrids and parental lines, while the correlation was assessed in the remaining 20%. (**B,C**) Natural smoothing splines to the cross-validated values between f_{RD} and (**B**) absolute and (**C**) relative midparent heterosis.

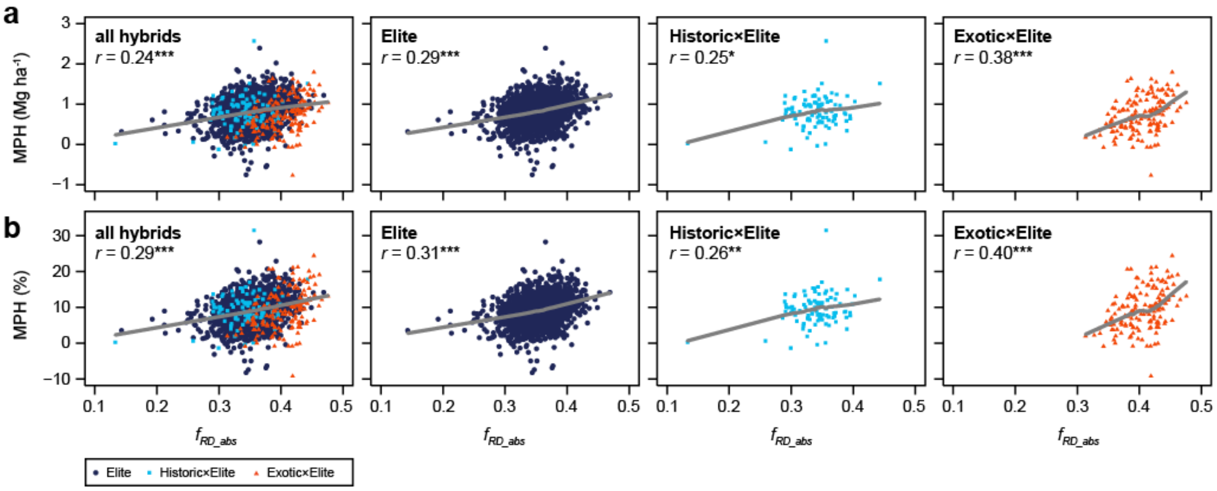

Fig. S11. Association between grain yield heterosis and heterotic genetic distance based on absolute dominance effects (f_{RD_abs}). (**A**) Midparent heterosis (**B**) relative midparent heterosis, shown for the different sets of hybrids. The grey lines are locally weighted regression lines.

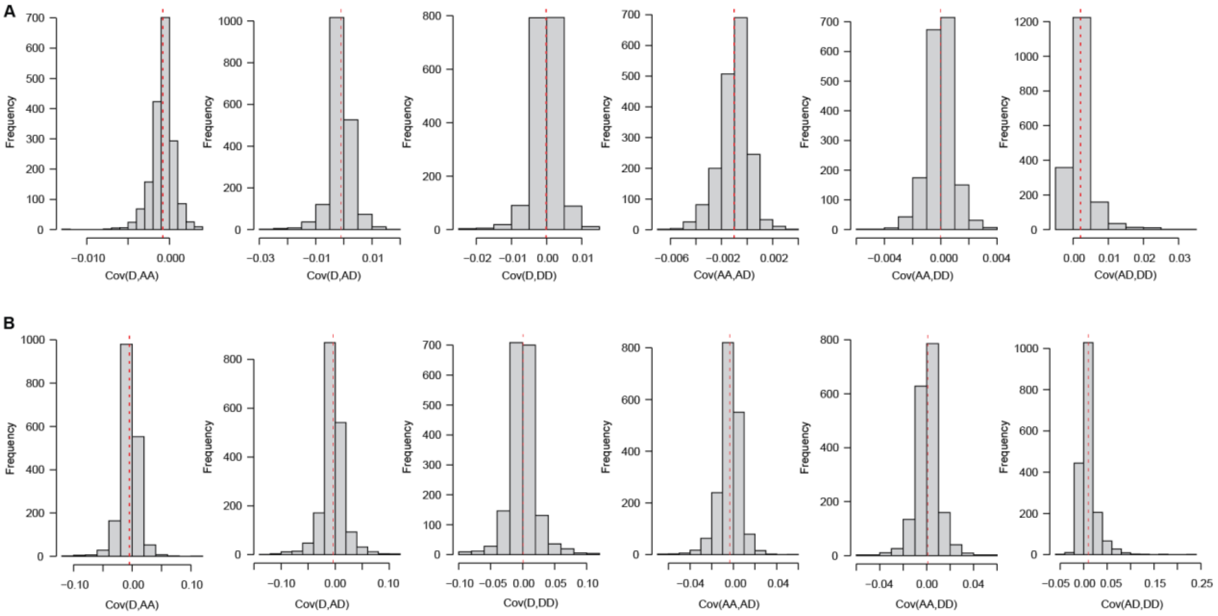

Fig. S12. Covariances between genetic effects. Posterior distribution of the covariances between dominance and the three types of digenic epistatic effects (AA, AD and DD) shown for (**A**) Elite hybrids and (**B**) Exotic×Elite hybrids. Results were obtained from a genomic prediction model (Eq. 7) where the estimated genetic effects were extracted in each sampling round of the Bayesian iteration procedure. This illustrates that although the model is not orthogonal and the kinship matrices are correlated, the different types of genetic effects are basically independent as assumed in the model.

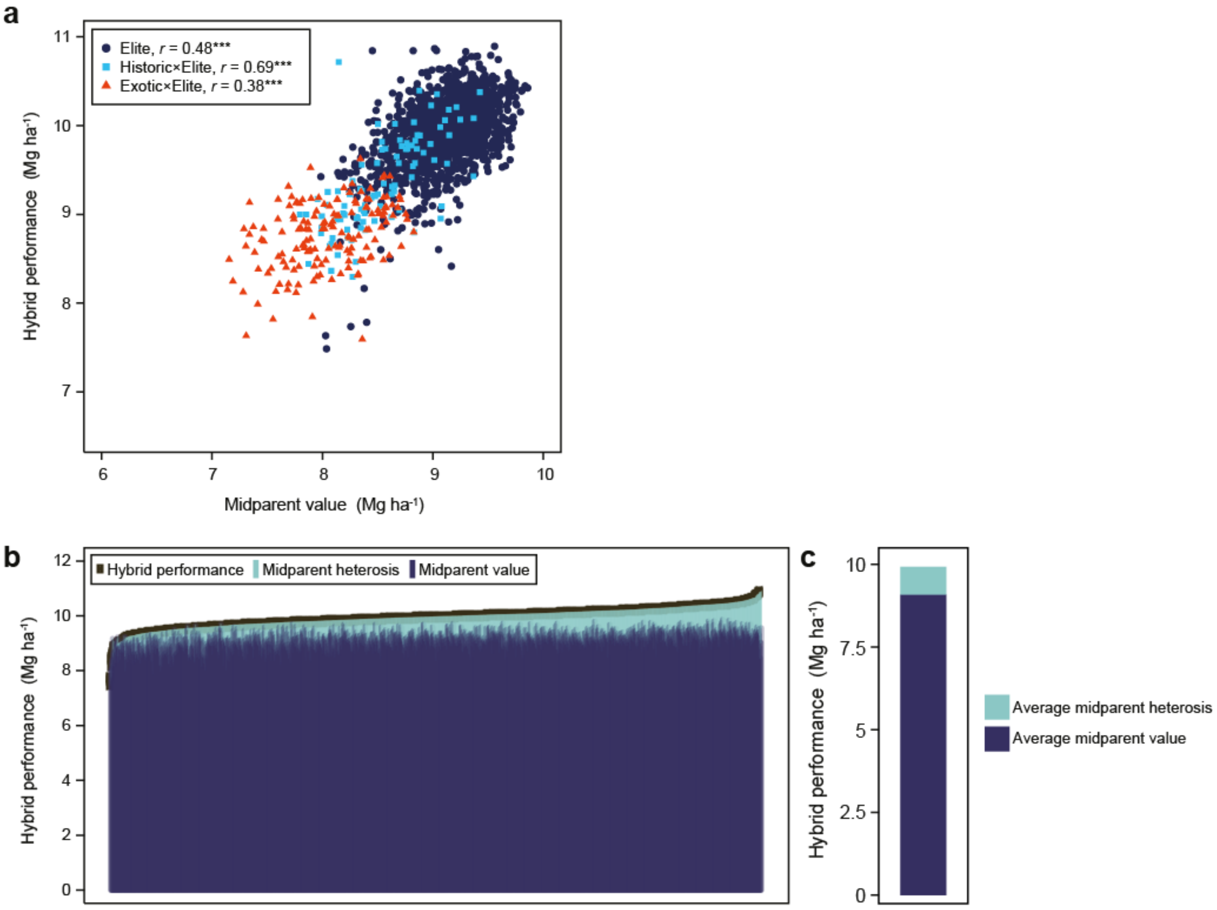

Fig. S13. Midparent value and hybrid performance for grain yield. (**A**) Positive association between midparent value and hybrid performance. (**B**) Hybrids of the Elite set ordered for increasing hybrid performance, illustrating the contribution of midparent value and heterosis. (**C**) Average midparent value and heterosis in the Elite set. Midparent value contributed to 91.6% of the variance of the hybrid performance.