Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize

Jinliang Yang[1,¤a, ❂*], Sofiane Mezmouk[1, ¤b, ❂], Andy Baumgarten[2], Edward S. Buckler[3], Katherine E. Guill[4], Michael D. McMullen[4,5], Rita H. Mumm[6], and Jeffrey Ross-Ibarra[1,7 *]

1 Department of Plant Sciences, University of California, Davis, CA 95616, USA,
2 DuPont Pioneer, Johnston, IA 50131, USA,
3 US Department of Agriculture, Agricultural Research Service, Ithaca, NY 14853, USA,
4 US Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA,
5 Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA,
6 Department of Crop Sciences and the Illinois Plant Breeding Center, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA,
7 Center for Population Biology and Genome Center, University of California, Davis, CA 95616, USA.

❂These authors contributed equally to this work.
¤a Current Address: Department of Agriculture and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68588, USA.
¤b Current Address: KWS SAAT SE, Grimsehlstr. 31, 37555 Einbeck, Germany.
*Correspondence should be addressed to J.Y. (`jolyang@ucdavis.edu`) and J.R.-I. (`rossibarra@ucdavis.edu`).

Abstract

Deleterious alleles have long been proposed to play an important role in patterning phenotypic variation and are central to commonly held ideas explaining the hybrid vigor observed in the offspring by crossing two inbred parents. We test these ideas using evolutionary measures of sequence conservation to ask whether incorporating information about putatively deleterious alleles can inform genomic selection (GS) models and improve phenotypic prediction. We measured a number of agronomic traits in both the inbred parents and hybrids of an elite maize partial diallel population and re-sequenced the parents of the population. Inbred elite maize lines vary for more than 350,000 putatively deleterious sites, but show a lower burden of such sites than a comparable set of traditional landraces. Our modeling reveals widespread evidence for incomplete dominance at these loci, and supports theoretical models that more damaging variants are usually more recessive. We identify haplotype blocks using an identity-by-decent (IBD) analysis and perform genomic prediction analyses in which we weigh blocks on the basis of segregating putatively deleterious variants. Cross-validation results show that incorporating sequence conservation in genomic selection improves prediction accuracy for grain yield and other fitness-related traits as well as heterosis for those traits. Our results provide empirical support for an important role for incomplete dominance of deleterious alleles in explaining heterosis and demonstrate the utility of incorporating functional annotation in phenotypic prediction and plant breeding.

1

Author Summary

A key long-term goal of biology is understanding the genetic basis of phenotypic variation. Although most new mutations are likely disadvantageous, their prevalence and importance in explaining patterns of phenotypic variation is controversial and not well understood. In this study we combine whole genome-sequencing and field evaluation of a maize mapping population to investigate the contribution of deleterious mutations to phenotype. We show that *a priori* prediction of deleterious alleles correlates well with effect sizes for grain yield and that variants predicted to be more damaging are on average more recessive. We develop a simple model allowing for variation in the heterozygous effects of deleterious mutations and demonstrate its improved ability to predict both phenotypes and hybrid vigor. Our results help reconcile alternative explanations for hybrid vigor and highlight the use of leveraging evolutionary history to facilitate breeding for crop improvement.

Introduction

Understanding the genetic basis of phenotypic variation is critical to many biological endeavors from human health to conservation and agriculture. Although most new mutations are likely deleterious [1], their importance in patterning phenotypic variation is controversial and not well understood [2]. Empirical work suggests that, although the long-term burden of deleterious variants is relatively insensitive to demography [3], population bottlenecks and expansion may lead to an increased abundance of deleterious alleles over shorter time scales such as those associated with domestication [4], postglacial colonization [5] or recent human migration [6]. Even when the impacts on total load are minimal, demographic change may have important consequences for the contribution of deleterious variants to phenotypic variation [3, 7–9]. Together, these considerations point to a potentially important role for deleterious variants in determining patterns of phenotypic variation, especially for traits closely related to fitness.

Maize (*Zea mays*) is an ideal system in which to study the impacts of deleterious variants. In addition to its global agricultural importance, maize has long been an important genetic model system [10] and central to debates about the basis of hybrid vigor and the role of deleterious alleles [11, 12]. The maize domestication bottleneck has lead to an increased burden of deleterious alleles in maize compared to its wild ancestor teosinte [13], and rapid expansion following domestication likely lead to an increase in new mutations and stronger purifying selection [4]. More recently, modern maize breeding has lead to dramatic reductions in effective population size [14], but inbreeding during the development of modern inbred lines may have decreased load by purging recessive deleterious alleles [15]. Nonetheless, substantial evidence suggests an abundance of deleterious alleles present in modern germplasm, from changes in heterozygosity during the process of inbreeding [16, 17] and selection [18] to genome-wide association results that reveal an excess of associations with genes segregating for damaging protein-coding variants [19].

Modern maize agriculture takes advantage of hybrid maize plants that result from the cross between two parental inbred lines [12]. These crosses result in a phenomenon known as hybrid vigor or heterosis, in which the hybrid plant shows improved agronomic qualities compared to its parents. Heterosis cannot be easily predicted from parental phenotype alone, and the genetic underpinnings of heterosis remain largely unknown. The most straightforward explanation for heterosis has been simple complementation of recessive deleterious alleles homozygous in one of the inbred parents [20, 21]. While this model is supported by considerable empirical evidence [22, 23], it fails in its simplest

form to explain a number of observations, especially relating to heterosis and inbreeding depression in polyploid plants [11, 24, 25]. Other explanations, such as single-gene heterozygote advantage, clearly may play an important role in some cases [e.g. 26, 27], but mapping studies suggest such models are not easily generalizable [28].

In this study, we set out to investigate the contribution of deleterious alleles to phenotypic variation and hybrid vigor in maize. We created a partial diallel population from 12 maize inbred lines which together represent much of the ancestry of present-day commercial U.S. corn hybrids [29, 30]. We measured a number of agronomically relevant phenotypes in both parents and hybrids, including flowering time (days to 50% pollen shed, DTP; days to 50% silking, DTS; anthesis-silking interval, ASI), plant size (plant height, PHT; height of primary ear, EHT), grain quality (test weight which is a measure of grain density, TW), and grain yield (GY). We conducted whole genome sequencing of the parental lines and characterized genome-wide deleterious variants using genomic evolutionary rate profiling (GERP) [31]. We then test models of additivity and dominance for each phenotype using putatively deleterious variants and investigate the relationship between dominance and phenotypic effect size and the long-term fitness consequences of a mutation as measured by GERP. Finally, we take advantage of a Bayesian genomic selection framework [32] approach to explicitly test the utility of including GERP scores in phenotypic prediction for hybrid traits and heterosis.

Materials and Methods

Plant materials and phenotypic data.

We formed a partial diallel population from the F1 progeny of 12 inbred maize lines (**Table S1**, **Figure S1**). Field performance of the 66 F1 hybrids and 12 inbred parents were evaluated along with two current commercial check hybrids in Urbana, IL over three years (2009-2011) in a resolvable incomplete block design with three replicates. To avoid competition effects, inbreds and hybrids were grown in different blocks within the field. Plots consisted of four rows (5.3 m long with row spacing of 0.76 m at plant density of 74,000 plants ha^{-1}), with all observations taken from the inside two rows to minimize effects of shading and maturity differences from adjacent plots. We measured plant height (PHT, in cm), height of primary ear (EHT, in cm), days to 50% pollen shed (DTP), days to 50% silking (DTS), anthesis-silking interval (ASI, in days), grain yield adjusted to 15.5% moisture (GY, in bu/A), and test weight (TW, weight of 1 bushel of grain in pounds).

We estimated Best Linear Unbiased Estimates (BLUEs) of the genetic effects in ASReml-R (VSN International) with the following linear mixed model:

$$Y_{ijkl} = \mu + \varsigma_i + \delta_{ij} + \beta_{kij} + \alpha_l + \varsigma_i \cdot \alpha_l + \varepsilon$$

where Y_{ijkl} is the phenotypic value of the lth genotype evaluated in the kth block of the jth replicate within the ith year; μ, the overall mean; ς_i, the fixed effect of the ith year; δ_{ij}, the random effect of the jth replicate nested within the ith year; β_{kij}, the random effect of the kth block nested within the ith year and j^{th} replicate; α_l, the fixed genetic effect of the lth individual; $\varsigma_i \cdot \alpha_l$, the random interaction effect of the lth individual with the ith year; and ε, the model residuals. We calculated the broad sense heritability (H^2) of traits based on the analysis of all individuals (inbred parents, hybrid progeny, and checks) following the equation:

$$H^2 = V_G/(V_G + V_{G \times E}/i + V_E/(i \times j))$$

where $i = 3$ (number of years) and $j = 3$ (number of replicates per year).

3

The BLUE values for each cross can be found in **Table S1**; values across all hybrids were relatively normally distributed for all traits (Shapiro-Wilk normality tests P values > 0.05, **Figure S1**), though some traits were highly correlated (e.g. Spearman correlation $r = 0.98$ for DTS and DTP, **Figure S2**).

We estimated mid-parent heterosis (MPH) as:

$$MPH_{ij} = \hat{G}_{ij} - mean(\hat{G}_i, \hat{G}_j)$$

where $\hat{G}_{ij}$, $\hat{G}_i$ and $\hat{G}_j$ are the BLUE values of the hybrid and its two parents i and j. Note that for ASI, lower trait values are considered superior. General combining ability (GCA) was estimated following Falconer and Mackay [33], and the estimated values can be found in **Table S2**.

Sequencing and Genotyping.

We extracted DNA from the 12 inbred lines following [34] and sheared the DNA on a Covaris (Woburn, Massachusetts) for library preparation. Libraries were prepared using an Illumina paired-end protocol with 180 bp fragments and sequenced using 100 bp paired-end reads on a HiSeq 2000. Raw sequencing data are available at NCBI SRA (PRJNA381642).

We trimmed raw sequence reads for adapter contamination with Scythe (`https://github.com/vsbuffalo/scythe`) and for quality and sequence length (≥ 20 nucleotides) with Sickle (`https://github.com/najoshi/sickle`). We mapped filtered reads to the maize B73 reference genome (AGPv2) with bwa-mem [35], keeping reads with mapping quality higher than 10 and with a best alignment score higher than the second best one for further analyses.

We called single nucleotide polymorphisms (SNPs) using the *mpileup* function from samtools [36]. To deal with known issues with paralogy in maize [15], SNPs were filtered to be heterozygous in fewer than 3 inbred lines, have a mean minor allele depth of at least 4, have a mean depth across all individuals less than 30 and have missing alleles in fewer than 6 inbred lines. Data on total number of SNPs called and the rate of missing data per line are shown in **Table S3**. We estimated the allelic error rate using three independent data sets: for all individuals using 41,292 overlapping SNPs from the maize SNP50k bead chip [14]; for all individuals using 180,313 overlapping SNPs identified through genotyping-by-sequencing (GBS) [37]; and for B73 and Mo17 using 10,426,715 SNP from the HapMap2 project [15]. Alignments and genotypes for each of the 12 inbreds are available at CyVerse (`https://de.cyverse.org/de/?type=data& folder=/iplant/home/yangjl/pvp_diallel_data`). Because these parents are highly inbred, knowing their homozygous genotype also allows us to know the genotype of the F1 derived from any two of the parents.

To test whether alignment to the B73 reference introduces a bias in relatedness estimation, we computed kinship matrices using both our SNP data as well as genotyping-by-sequencing data (version AllZeaGBSv2.7 downloaded from (`www.panzea.org`)) obtained from alignments to a set of sequencing reads ascertained from a broad germplasm base [38]. The two matrices were nearly identical (Pearson's correlation coefficient $r = 0.995$), suggesting the degree of relatedness among lines is not sensitive to using B73 as the reference genome.

Identifying putatively deleterious alleles.

We used genomic evolutionary rate profiling (GERP) [39] estimated from a multi-species whole-genome alignment of 13 plant genomes [40] including *Zea mays*, *Coelorachis tuberculosa*, *Vossia cuspidata*, *Sorghum bicolor*, *Oryza sativa*, *Setaria italica*,

Brachypodium distachyon, *Hordeum vulgare*, *Musa acuminata*, *Populus trichocarpa*, *Vitis vinifera*, *Arabidopsis thaliana*, and *Panicum virgatum*; the alignment and estimated GERP scores are available at CyVerse (`https://de.cyverse.org/de/?type=data&folder=/iplant/home/yangjl/pvp_diallel_data`). We define "GERP-SNPs" as the subset of SNPs with GERP score > 0, and at each SNP we assign the minor allele in the multi-species alignment as the likely deleterious allele. Finally, we predicted the functional consequences of GERP-SNPs based on genome annotation information obtained from SnpEff [41]. The multi-species alignment made use of the B73 AGPv3 assembly, and to ensure consistent coordinates, we ported our SNP coordinates from AGPv2 to AGPv3 using the Gramene assembly converter (`http://ensembl.gramene.org/Zea_mays/Tools/AssemblyConverter?db=core`).

To compare GERP scores (for all SNPs with GERP > 0) to recombination rate and allele frequencies, we obtained the NAM genetic map [42] from the Panzea website (`http://www.panzea.org/`) and allele frequencies from the $> 1{,}200$ maize lines sequenced as part of HapMap3.2 [43]. To compare the burden of deleterious alleles in modern inbred lines to landraces, we extracted genotypic data of 23 specially-inbred traditional landrace cultivars (see [15] for more details) from HapMap3.2. For each line, we calculated burden as the count of minor alleles present across all GERP-SNPs divided by the total number of non-missing sites. We separated sites into fixed (present in all individuals of a group) and segregating sites for landrace and modern maize samples separately.

Estimating effect sizes and dominance of GERP-SNPs.

We estimated the additive and dominant effects of individual GERP-SNPs using a GBLUP model [44] implemented in GVCBLUP [45]:

$$Y_i = \mu + \sum_{j=1}^{n} X_{ij} a_j + \sum_{j=1}^{n} W_{ij} d_j + \varepsilon$$

where Y_i is the BLUE value of the ith hybrid, a_j and d_j are the additive and dominant effects of the jth GERP-SNP, $X_{ij} = \{2p, 2p-1, 2p-2\}$, and $W_{ij} = \{2p^2, 2p(1-p), -2(1-p)^2\}$ are the genotype encodings for genotypes A_1A_1, A_1A_2, and A_2A_2 of the jth SNP in the ith hybrid, respectively, and ε is the model residuals. The additive and dominance SNP encoding ensures that the effects are independent for a given GERP-SNP. We first estimated the total variance explained under models of complete additivity ($d_j = 0$) or complete dominance ($a_j = 0$). Then, to assess correlations between SNP effects and GERP scores, we calculated the degree of dominance ($k = d/a$) [46] for SNPs that each explained greater than the genome-wide mean per-SNP variance (total variance explained divided by total number of GERP-SNPs). Because this approach can lead to very large absolute values of k, we truncated GERP-SNPs with $|k = d/a| > 2$ for all further analyses.

To compare the variance explained by our model to that explained by random SNPs, we used a 2-dimensional sampling approach to create 10 equal-sized datasets of randomly sampled SNPs (including SNPs with GERP score $<= 0$) matched for allele frequency (in bins of 10%) and recombination rate (in quartiles of cM/Mb). For each dataset we fit the above model separately and estimated SNP effects and phenotypic variance explained by each SNP.

To test the relationship between GERP score and dominance under a simple model of mutation-selection equilibrium, we estimated the selection coefficient s by assuming that yield is a measure of fitness. We assigned the yield-increasing allele at each GERP-SNP a random dominance value in the range of $0 \geq k \geq 1$ and calculated its equilibrium allele frequency p under mutation-selection balance using $p = \sqrt{\frac{\mu}{s}}$ for values

5

of $k > 0.98$ and $p = \frac{2\mu}{k+1}$ for $k \leq 0.98$. We then simulated datasets using binomial sampling to choose SNPs in a sample of size $n = 12$ inbreds.

Haplotype Analysis.

We imputed missing data and identified regions of identity by descent (IBD) between the 12 inbred lines using the fastIBD method implemented in BEAGLE [47]. We then defined haplotype blocks as contiguous regions within which there were no IBD break points across all pairwise comparisons of the parental lines (**Figure S3**). Haplotype blocks at least 1 Kb in size were kept for further analyses.

Because there is no recombination in an inbred parent, this allows us to project the diploid genotype of each F1 based on the haplotypes of the two parents. In the projected diploid genotype of each F1, haplotype blocks were weighted by the summed GERP scores of all GERP-SNPs (python script 'gerpIBD.py' available at https://github.com/yangjl/zmSNPtools); blocks with no SNPs with positive GERP scores were excluded from further analysis. For a particular SNP with a GERP score g, the homozygote for the conserved (major) allele was assigned a value of 0, the homozygote for the putatively deleterious allele a value of $2g$, and the heterozygote a value of $(1 + k) \times g$, where k is the dominance estimated from the GBLUP model above.

Genomic Selection.

We used the BayesC option from GenSel4 [32] for genomic selection model training with 41,000 iterations and removing the first 1,000 as burn-in. We used the model

$$Y_i = \mu + \sum_{j=1}^{n} r_j I_{ij} + \varepsilon$$

where Y_i is the BLUE value of the ith hybrid, r_j is the regression coefficient for the jth haplotype block, and I_{ij} is the sum of GERP scores under an additive, dominance or incompletely dominance model for the ith hybrid in the jth haplotype block.

To conduct prediction, we used a 5-fold cross-validation method, dividing the diallel population randomly into training (80%) and validation sets (20%) 100 times. After model training, we obtained prediction accuracies by comparing the predicted breeding values with the observed phenotypes in the corresponding validation sets. For comparison, we permuted GERP scores using 50k SNP ($\approx$ 100Mb or larger) windows which were circularly shuffled 10 times to estimate a null conservation score for each IBD block. We conducted permutations on all GERP-SNPs as well as on a restricted set of GERP-SNPs only in genic regions to control for GERP differences between genic (N $= 221,960$) and intergenic regions (N $= 123,216$). We conducted permutation cross-validation experiments using the same training and validation sets.

We estimated the posterior phenotypic variance explained using all of the data to derive correlations between breeding values estimated from the prediction model and observed BLUE values. Note that the correlation used here is different from the prediction accuracy (r) used for the cross-validation experiments, where the latter is defined as the correlation between real and estimated values; the two statistics will converge to the same value when there is no error in SNP/haplotype effect estimation [48].

Finally, to compare our genomic prediction model to a classical model of general combining ability, we used the following equations:

$$Y_{ij} = \mu + GCA_i + GCA_j + \varepsilon$$
$$Y_{ij} = \mu + GCA_i + GCA_j + G_{ij} + \varepsilon$$

where Y_{ij} is the BLUE value of the hybrid of the ith and j^{th} inbreds, μ is the overall mean, GCA_i and GCA_j are the general combining abilities of the ith and j^{th} inbreds, G_{ij} is the breeding value of the hybrid of the ith and j^{th} inbreds as estimated by our genomic prediction model, and ε, the model residuals.

Data and code accessibility

Sequencing data have been deposited in NCBI SRA (SRP103329) database, and code for all analyses are available in the public GitHub repository (`https://github.com/yangjl/GERP-diallel`).

Results

Heterosis in a partial diallel population

We created a partial diallel population from 12 maize inbred lines which together represent much of the ancestry of present-day commercial U.S. corn hybrids (**Table S1**) [29, 30]. We measured a number of agronomically relevant phenotypes in both parents and hybrids, including flowering time (days to 50% pollen shed, DTP; days to 50% silking, DTS; anthesis-silking interval, ASI), plant size (plant height, PHT; height of primary ear, EHT), test weight (TW; a measure of quality based on grain density), and grain yield (GY). In an agronomic setting GY — a measure of seed production per unit area — is the primary trait selected by breeders and thus analogous to fitness. Plant height and ear height, both common measures of plant health or viability, were significantly correlated to GY (**Figure S2**).

For each genotype we derived best linear unbiased estimators (BLUEs) of its phenotype from mixed linear models (**Table S1**) to control for spatial and environmental variation (see **Methods**). We estimated mid-parent heterosis (MPH, **Figure 1a**) for each trait as the percent difference between the hybrid compared to the mean value of its two parents (see **Methods**, **Table S1**). Consistent with previous work [28], we find that grain yield (GY) showed the highest level of heterosis (MPH of $182\% \pm 60\%$). While flowering time (DTS and DTP) is an important adaptive phenotype globally [49], it showed relatively little heterosis in this study, likely due to the relatively narrow geographic range represented by the parental lines.

Annotation of deleterious alleles

We resequenced the 12 inbred parents to an average depth of $\approx 10\times$, resulting in a filtered set of 13.8M SNPs. Compared to corresponding SNPs identified by previous studies (see **Methods**), we observed a mean genotypic concordance rate of 99.1%. In order to quantify the deleterious consequences of variants *a priori*, we made use of Genomic Evolutionary Rate Profiling (GERP) [39] scores of the maize genome [50]. GERP scores provide a quantitative measure of the evolutionary conservation of a site across a phylogeny that allows characterization of the long-term fitness consequences of both coding and noncoding positions in the genome [51]. Sites with more positive GERP scores are inferred to be under stronger purifying selection, and SNPs observed at such sites are thus inferred to be more deleterious. At each site with GERP scores > 0 (hereafter called GERP-SNPs), we designated the minor allele from the multispecies alignment as putatively deleterious. Of the 350k total segregating GERP-SNPs in our parental lines, 14% are detected in coding regions, equally split between synonymous (N = 64,439) and non-synonymous (N = 65,376) sites (**Table S4**). Each line carries, on average, 139k potential deleterious SNPs (**Table S5**). The
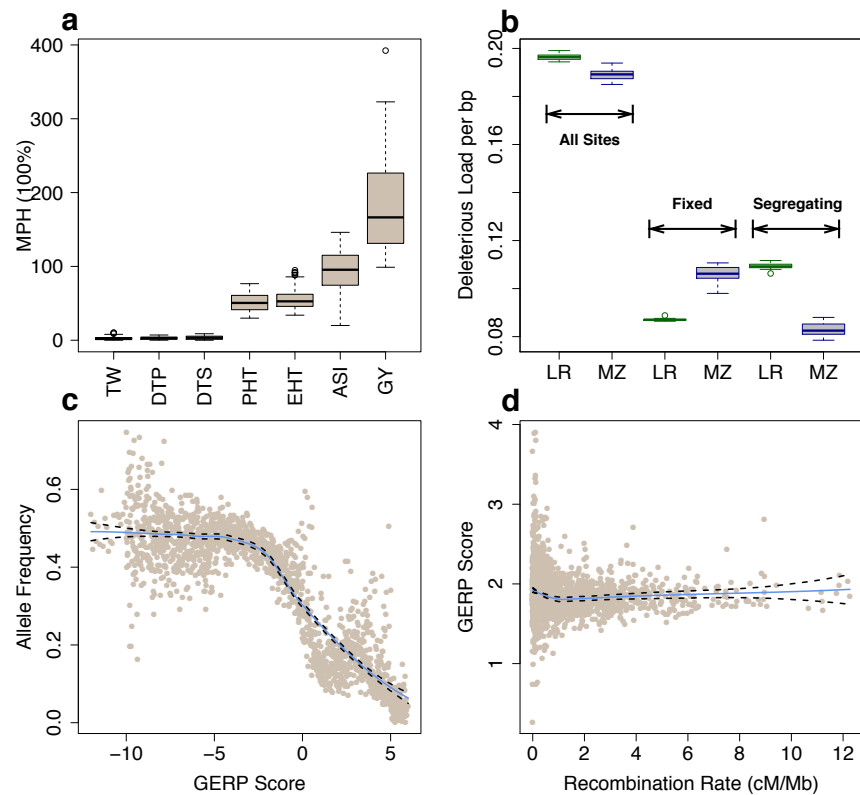
7

Fig 1. Heterosis and deleterious variants. (**a**) Boxplots (median and interquartile range) of percent mid-parent heterosis (MPH). (**b**) Proportion of deleterious alleles in landraces (LR, green) and elite maize (MZ, blue) lines. (**c**) The allele frequency of the minor alleles in the multi-species alignment in bins of 0.01 GERP score (including GERP <= 0 sites). (**d**) The mean GERP score for putatively deleterious sites (GERP > 0). Each point represents a 1 Mb window. In (**c**) and (**d**) the solid blue and dashed black lines define the best-fit regression line and its 95% confidence interval.

reference genome B73 contains only $\approx 1/3$ of the deleterious SNPs of the other parents, likely due to reference bias in identifying deleterious variants. The F1 hybrids of the diallel each contain an average of $\approx 56,000$ homozygous deleterious SNPs, ranging from 47,219 (PH207 x PHG35) to 77,210 (PHG84 x PHZ51) (**Table S6**).

To compare the burden of deleterious variants between our elite maize lines and traditionally cultivated landraces, we used genotypes from the maize HapMap3.2 [43] for our diallel parents and 23 specially-inbred landrace lines [15] (**Table S5**). Compared to landraces, the parents of our diallel exhibited a greater burden of fixed (allele frequency of 1) deleterious variants but a much smaller burden of segregating SNPs, resulting in a slightly lower overall proportion of deleterious sites (mean of 1.3M deleterious alleles out of 6.5M total sites vs. 0.6/3.3M; **Figure 1b**).

Population genetic theory predicts that deleterious variants should be at low overall frequencies, and that such variants should be enriched in regions of the genome with extremely low recombination [52]. Using data from more than 1,200 lines in maize HapMap3.2 [43], we find that allele frequency of the minor alleles in the multi-species alignment shows a strong negative correlation with GERP score (**Figure 1c**). This negative correlation holds using allele frequency derived from our 12 parental lines (**Figure S4**), though as expected is less significant given the smaller sample size. SNPs

8

found in regions of the genome with low recombination also show higher overall GERP scores (**Figure 1d**), a trend particularly noticeable around centromeres (**Figure S5**). These results match previous empirical findings in maize that deleterious alleles are rare [19] and most abundant in the lowest recombination regions [17, 40, 53], and support the use of GERP scores as a quantitative measure of the long-term fitness effects of an observed variant.
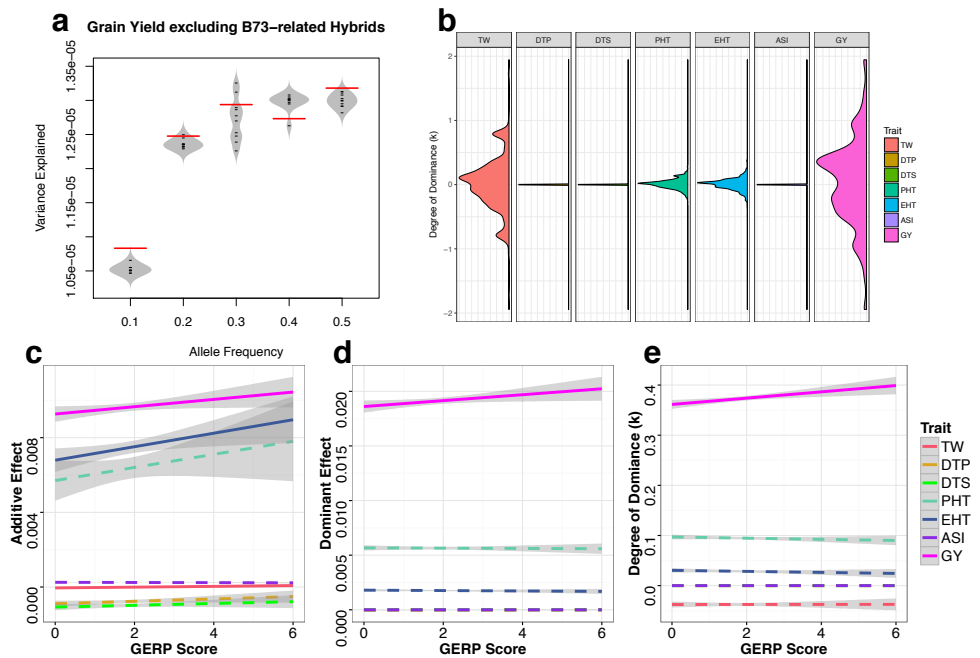

Fig 2. Variance explained and degree of dominance (k) of GERP-SNPs for traits *per se*. **(a)** Total per-SNP variance explained for grain yield trait *per se* by GERP-SNPs (red lines) and randomly sampled SNPs (grey beanplots). **(b)** Density plots of the degree of dominance (k). Extreme values of k were truncated at 2 and -2. **(c-e)** Linear regressions of additive effects **(c)**, dominance effects **(d)**, and degree of dominance **(e)** of seven traits *per se* against SNP GERP scores. Solid and dashed lines represent significant and nonsignificant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for SNPs that explain more than the mean genome-wide per-SNP variance (see **Methods** for details).

Phenotypic effects of deleterious SNPs

We first investigate the impacts of deleterious variants on phenotype using simple linear regressions. Across all hybrids, the number of homozygote GERP-SNPs was negatively correlated with grain yield, plant height, and ear-height *per se* (see **Table S6** for complementation data and **Table S7** for correlations with all traits).

We next applied a genomic best linear unbiased prediction (GBLUP) [44] modeling approach to estimate the effect sizes and variance explained by GERP-SNPs for each of the phenotypes *per se* across our diallel (see **Methods**). GERP-SNPs had larger average effects and explained more phenotypic variance than the same number of randomly sampled SNPs (including SNPs with GERP score $<= 0$) matched for allele frequency and recombination (**Figure 2a**). We found the cumulative proportion of

9

dominance variance explained by GERP-SNPs was higher for traits showing high heterosis (Spearman correlation P value < 0.01, $r = 0.9$), from ≈ 0 for flowering time traits to as much as 24% for grain yield (**Figure S6**). Distributions of per-SNP dominance $k = \frac{d}{a}$ (see **Methods**) across traits were consistent with the cumulative partitioning of variance components (**Figure 2b**) and matched well with expectations from previous studies showing a predominantly additive basis for flowering time [54] and plant height [55] but meaningful contributions of dominance to test weight and grain yield [28, 30]. Although our diallel population is relatively small, our estimated values explain as much (for traits with low dominance variance like flowering time) or more variance (for traits with substantial dominance variance like grain yield) than sets of data with randomly shuffled values of dominance (n=10 randomizations of k per trait; **Figure S7**).

We then evaluated the relationship between GERP score and SNP effect size, dominance, and contribution to phenotypic variance. We found weak or negligible correlations between effect size and GERP score for flowering time and grain quality, but a strong positive correlation for fitness-related traits (**Figure 2c-d**). The variance explained by individual SNPs, however, was largely independent of GERP score (**Figure S8**), likely due to the observed negative correlation between allele frequency and GERP score (**Figure 1c**). Finally, we observed a positive relationship between GERP score and the degree of dominance (k) for grain yield (**Figure 2e**), such that the putatively deleterious allele at SNPs with higher GERP scores are also estimated to be more recessive for their phenotypic effects on grain yield (larger k for the major allele).

We investigated a number of possible caveats to the results presented in **Figure 2**. First, to control for the potential inflation of SNP effect sizes in regions of high linkage disequilibrium, we removed SNPs from regions of the genome in the lowest quartile of recombination. While some individual correlations changed significance, our overall results appear robust to the removal of low recombination regions (**Figure S9**). Second, we tested the impact of reference bias caused by inclusion of the B73 genome in the multi-species alignment used to estimate GERP scores. To do so, we removed the 11 hybrids which include as one parent the reference genome line B73 and repeated the above analyses. Doing so dramatically reduces the size of our dataset, but we nonetheless find significant correlations between complementation and phenotype (**Table S7**), that GERP-SNPs explain a greater proportion of overall variation than randomly sampled SNPs (**Figure S10a**), and that the relative pattern of dominance among traits remains the same (**Figure S10b**). While most of the correlations between effect size and GERP score lose significance (**Figure S10c-d**), likely due to the decreased sample size, the positive correlation between dominance and GERP score remains significant even in the absence of B73-derived hybrids (**Figure S10e**). Finally, because natural selection will maintain dominant deleterious alleles at lower frequencies than their recessive counterparts, we investigated whether the ascertainment bias against rare alleles present in our small sample would lead to the observed correlation between GERP and dominance. Simulations of SNPs with random dominance at mutation-selection balance (see **Methods**), however, failed to find any relationship between dominance and GERP score (**Figure S11**), though we caution that the dramatic demographic shifts involved in the recent history of maize [4] make such a simulation approximate at best.

Genomic prediction by incoporating GERP information

To explicitly test the informativeness of alleles identified *a priori* as putatively deleterious, we implemented a haplotype-based genomic prediction model that incorporates GERP scores as weights (see **Methods**). We explored the explanatory power with several different models and found that a model which incorporates both
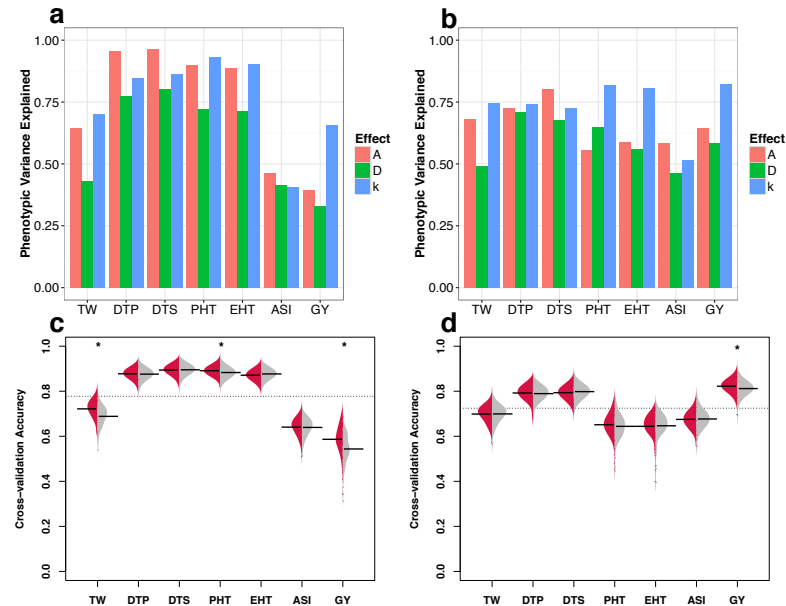
10

Fig 3. Genomic prediction models incorporating GERP. (a-b) Total phenotypic variance explained for traits *per se* (**a**) and heterosis (MPH) (**b**) under models of additivity (red), dominance (green), and incomplete dominance (blue). **(c-d)** Beanplots represent prediction accuracy estimated from cross-validation experiments for traits *per se* (**c**) and heterosis (MPH) (**d**) under a model of incomplete dominance. Prediction accuracy using estimated values for each GERP-SNP under an incomplete dominance model is shown on the left (red) and permutated values on the right (grey). Horizontal bars indicate mean accuracy for each trait and the grey dashed lines indicate the overall mean accuracy. Stars above the beans indicate prediction accuracies significantly (FDR < 0.05) higher than permutations. Results for pure additive and dominance models are shown in **Figure S13**.

GERP scores and dominance (k) estimated from our GBLUP model explained a greater amount of the posterior phenotypic variance for most traits *per se* (**Figure 3a**) and heterosis (MPH) (**Figure 3b**). A simple additive model showed superior explanatory power for flowering time, however, consistent with previous association mapping results that flowering time traits are predominantly controlled by a large number of additive effect loci [54].

To explicitly test the utility of incorporating GERP information in prediction models, we compared cross-validation prediction accuracies of the observed GERP scores to those from datasets in which GERP scores were circularly shuffled along the genome (see **Methods**). Models incorporating our observed GERP scores out-performed permutations (**Figure 3c-d**), even when considering only SNPs in genes (**Figure S12**). Our model improved prediction accuracy of grain yield by more than 4.3%, and improvements were also seen for plant height (0.8%) and testing weight (3.3%). While our model showed no improvement in predicting heterosis for traits showing low levels of heterosis (**Figure 1a**), including GERP scores significantly improved prediction accuracy for heterosis of grain yield (by 1%). Finally, our approach also significantly improved model fit for phenotypes of all traits *per se* as well as heterosis for GY and PHT compared to traditional models of genomic selection that use general combining ability (see **Methods**, **Table S2**) calculated directly from the pedigree of the hybrid

11

population [56] (ANOVA FDR < 0.01 and difference in AIC < 0, **Table S8**).   358

Discussion   359

We combine *a priori* prediction of deleterious alleles from whole genome sequence data   360
with multi-year field evaluation of important agronomic phenotypes to test the role of   361
incomplete dominance in determining hybrid phenotypes and heterosis in maize.   362

We first show that GERP scores are meaningful quantitative estimates of the fitness   363
consequences of individual alleles, as SNPs with higher GERP scores are found at lower   364
allele frequencies (**Figure 1c**), enriched in regions of low recombination (**Figure 1d**),   365
and associated with larger effect sizes on grain yield (**Figure 2c,d**). Although a   366
number of other methods exist to identify deleterious alleles from sequence data, GERP   367
scores include both coding and noncoding sequence, do not require additional functional   368
annotation, and show higher sensitivity and specificity than other related approaches   369
[51]. While the GERP scores used here reflect conservation of across relatively deep   370
phylogenetic time, future efforts may be able to increase power by incorporating   371
information from within-species polymorphism data [57] as well as other types of   372
annotations that have been shown to contribute substantially to phenotypic variation   373
(e.g. Wallace *et al.,* [58] and Rodgers-Melnick *et al.,* [50]).   374

Using GERP scores as a proxy for deleterious alleles, we then ask whether our elite   375
maize inbreds show an increased burden of deleterious alleles compared to a set of   376
traditional landrace varieties. We find that modern inbreds are characterized by an   377
increase in the proportion of deleterious variants fixed within the population   378
(**Figure 1b**), consistent with the strong impact of drift associated with rapid decreases   379
in effective population size during modern breeding [14]. In contrast, modern maize   380
inbreds exhibit a much smaller proportion of segregating deleterious variants than   381
landraces. This latter result is likely due to increased inbreeding in smaller populations,   382
an effect exacerbated by the transition from traditional open-pollinated maize to the   383
intentional formation of inbred lines. Inbreeding facilitates the removal of deleterious   384
variants by selection, as evidenced by the striking inbreeding depression exhibited by   385
open-pollinated maize [59]. Supporting our interpretation of these results, our observed   386
differences in the burden of deleterious variants closely mimic results from simulations   387
of partially recessive deleterious alleles in populations that have recently undergone   388
demographic bottlenecks [3].   389

We next use the set of SNPs with GERP > 0 scores (or GERP-SNPs) to investigate   390
the phenotypic effects of deleterious variants. Across phenotypes, our results largely   391
mirror previous work, finding that dominance contributes substantially to grain yield   392
[28], while traits such as flowering time appear to be largely additive [54]. At the level   393
of individual SNPs we find correlations between GERP score and phenotypic effect size   394
for yield and ear height, suggesting that long-term evolutionary constraint as measured   395
by GERP is a useful predictor of the phenotypic effects of variants on traits related to   396
fitness. Both traits are well explained by a model allowing for incomplete dominance   397
(**Figure 3a**), as is plant height, which shows a positive but not significant correlation   398
between effect size and GERP score. For grain yield, we also find that more deleterious   399
alleles (those with higher GERP score) are more likely to be recessive. We are unaware   400
of previous demonstrations of the genome-wide relationship between dominance and   401
fitness in other multicellular organisms, but this result follows predictions based on   402
models of metabolic pathways [60] and supports previous empirical evidence from gene   403
knockouts in yeast [61]. Though our population size is small, our partial diallel crossing   404
design and genomic selection model circumvent some of the problems with standard   405
genome-wide association analyses, including genome-wide multiple testing thresholds   406
and difficulties in assessing the effects of rare alleles due to limited replication. And   407

12

while there is likely substantial error in individual SNP estimates, permutation analyses show our overall results nonetheless produce meaningful results (**Figure 2a**, **Figure S7**).

After showing that GERP-SNPs explain a substantial portion of the observed phenotypic variation when combined with our estimates of dominance and effect size, we more rigorously test the direct utility of GERP scores using cross-validation prediction methods. We show that for both plant height and grain yield, our GERP-enabled prediction model has significantly improved accuracy compared to randomized data, even when only considering SNPs within genes (which have higher on average GERP scores; **Figure S12**). As genotyping costs continue to decline, genomic prediction models are increasing in popularity [62]. Most previous work on genomic prediction, however, focuses exclusively on statistical properties of the models, ignoring potentially useful biological information (but see Edwards *et al.,* [63] for a recent example). Our results suggest that incorporating additional annotations — in particular information on evolutionary constraint — can provide additional, inexpensive benefits to existing genomic prediction frameworks.

Finally, our results also have implications for understanding the genetic basis of heterosis. Heterosis has been observed across many species, from yeast [64] to plants [65] and vertebrates [66], and a number of hypotheses have been put forth to explain the phenomenon [11, 21]. Of all these explanations, complementation of recessive deleterious alleles [12, 21] remains the simplest genetic explanation and is supported by considerable empirical evidence [22, 23, 67]. It remains controversial, however, because complementation of completely recessive mutations cannot fully explain a number of empirical observations including unexpected differences in heterosis and inbreeding depression among polyploids [11, 68]. For example, a model of simple complementation of purely recessive alleles is unable to explain differing levels of heterosis between triploid hybrids with different numbers of parental genomes (e.g. AAB vs ABB) [24] or why the cross of two tetraploid F1 hybrids shows greater heterosis than the original F1 [25]. Our results, however, indicate that most deleterious SNPs show incomplete dominance (**Figure 2b**) for traits with high levels of heterosis, and our genomic prediction models find improvement in predictions of heterosis when incorporating GERP scores under such a model (**Figure 3d**). These results are in line with other empirical evidence suggesting that new mutations tend to be partially recessive [69] and that GWAS hits exhibit incomplete dominance for phenotypes *per se* among hybrids [27]. We argue that allowing for incomplete dominance effectively unifies models of simple complementation with those of gene dosage [24]. Combined with observations that deleterious SNPs are enriched in low-recombination pericentromeric regions [40] (**Figure 1d**), such a model can satisfactorily explain changes in heterozygosity during breeding [18, 53], enrichment of yield QTL and apparent overdominance in centromeric regions [28], and even observed patterns of heterosis in polyploids (**Figure S14**). It is unlikely of course that any single explanation is sufficient for a phenomenon as complex as heterosis, and other processes such as overdominance likely make important contributions (e.g. Guo *et al.,* [70] and Huang *et al.,* [27]), but we argue here that a simple model of incompletely dominant deleterious alleles may provide substantial explanatory power not only for fitness-related phenotypic traits but for hybrid vigor as well.

Conclusion

In this study, we use genomic and phenotypic data from a partial diallel population of maize to show that an incomplete dominance model of deleterious mutation both fits predictions of population genetic theory and explains phenotypic variation for fitness-related phenotypes and hybrid vigor. We find genome-wide support for

13

hypotheses predicting that more damaging variants are more recessive. Finally, we show that leveraging evolutionary annotation information *in silico* enables us to predict grain yield and other traits, including heterosis, with greater accuracy. Together, these results help reconcile alternative explanations for hybrid vigor and point to the utility of leveraging evolutionary history to facilitate breeding for crop improvement.

Supporting Information

S1 Fig.   A half-diallel population and distributions of phenotypes.

S2 Fig.   Pairwise correlation plots of seven phenotypic traits.

S3 Fig.   Haplotype block identification using an IBD approach.

S4 Fig.   The minor allele frequency estimated from 12 parental lines in bins of 0.01 GERP score.

S5 Fig.   Segregating genetic load across ten maize chromosomes.

S6 Fig.   Cumulative variance explained by GERP-SNPs.

S7 Fig.   Phenotypic variance explained for observed data and for randomly shuffled data using the genomic selection model.

S8 Fig.   Linear regressions of GERP-SNPs' additive variance, dominance variance and total variance of seven traits *per se* against their GERP scores.

S9 Fig.   Linear regressions after filtering out GERP-SNPs located in regions in the lowest quartiles of recombination.

S10 Fig.   Phenotypic variance explained for grain yield and degree of dominance (k) of GERP-SNPs after removing 11 hybrids that B73 as one parent.

S11 Fig.   Regression of degree of dominance (k) on GERP scores for simulated data.

S12 Fig.   Cross-validation accuracy using GERP-SNPs in genic regions.

S13 Fig.   Cross-validation prediction accuracy for trait *per se* and heterosis.

S14 Fig.   Breeding values of diploid and simulated triploid hybrids.

S1 Table.   BLUE values and levels of heterosis of the seven phenotypic traits for the 66 hybrids.

S2 Table.   General combining ability and specific combining ability of the seven phenotypic traits.

14

S3 Table.   SNP missing rate in our diallel parental lines.   491

S4 Table.   Summary statistics of SNP annotation results.   492

S5 Table.   Number of deleterious SNPs carried per line.   493

S6 Table.   Number of complementation and homozygote deleterious load   494
for GERP-SNPs in hybrids.   495

S7 Table.   The correlation between the number of homozygote   496
GERP-SNPs and the hybrid phenotypes.   497

S8 Table.   Model comparisons P values and AICs.   498

Acknowledgements   499

Author Contributions   506

J.Y., S.M., and J.R.-I. designed the work. E.S.B., K.E.G., M.D.M., J.R.-I. and R.H.M.   507
generated the data. J.Y., S.M. and J.R.-I. analyzed data. A.B. provided statistical   508
advice. J.Y. and J.R.-I. wrote the manuscript.   509

Competing Interests Statement   510

The authors declare no competing financial interests.   511

References

1. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new
   mutations. Nature Reviews Genetics. 2007;8(8):610–618.

2. Mitchell-Olds T, Willis JH, Goldstein DB. Which evolutionary processes
   influence natural genetic variation for phenotypic traits? Nature Reviews
   Genetics. 2007;8(11):845–856.

3. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is
   insensitive to recent population history. Nature genetics. 2014;46(3):220.

4. Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J.
   Recent demography drives changes in linked selection across the maize genome.
   Nature Plants. 2016;2(7):16084. doi:10.1038/nplants.2016.84.

5. Laenen B, Tedder A, Nowak MD, Toräng P, Wunder J, Wötzel S, et al. Demography And Mating System Shape The Genome-Wide Impact Of Purifying Selection In Arabis alpina. bioRxiv. 2017; p. 127209.

6. Peischl S, Dupanloup I, Foucal A, Jomphe M, Bruat V, Grenier JC, et al. Relaxed selection during a recent human expansion. bioRxiv. 2016; p. 064691.

7. Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 2014;10(5):e1004379.

8. Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome research. 2016;26:863–873.

9. Sanjak J, Long AD, Thornton KR. The Genetic Architecture of a Complex Trait is more Sensitive to Genetic Model than Population Growth. bioRxiv. 2016;doi:10.1101/048819.

10. Nannas NJ, Dawe RK. Genetic and genomic toolbox of Zea mays. Genetics. 2015;199(3):655–669.

11. Birchler JA, Auger DL, Riddle NC. In search of the molecular basis of heterosis. The Plant Cell. 2003;15(10):2236–2239.

12. Crow JF. 90 years ago: the beginning of hybrid maize. Genetics. 1998;148(3):923–928.

13. Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford M. The interplay of demography and selection during maize domestication and expansion. bioRxiv. 2017; p. 114579.

14. van Heerwaarden J, Hufford MB, Ross-Ibarra J. Historical genomics of North American maize. Proceedings of the National Academy of Sciences. 2012;109(31):12420–12425.

15. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, et al. Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet. 2012;44(7):803–7. doi:10.1038/ng.2313.

16. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, et al. Genetic properties of the maize nested association mapping population. Science. 2009;325(5941):737–740.

17. Gore Ma, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, et al. A first-generation haplotype map of maize. Science (New York, NY). 2009;326(5956):1115–7. doi:10.1126/science.1177837.

18. Gerke JP, Edwards JW, Guill KE, Ross-Ibarra J, McMullen MD. The genomic impacts of drift and selection for hybrid performance in maize. Genetics. 2015;201(3):1201–1211.

19. Mezmouk S, Ross-Ibarra J. The pattern and distribution of deleterious mutations in maize. G3 (Bethesda, Md). 2014;4(January):163–71. doi:10.1534/g3.113.008870.

20. Crow JF. 90 years ago: the beginning of hybrid maize. Genetics. 1998;148(3):923–928.

21. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nature reviews Genetics. 2009;10(11):783–96. doi:10.1038/nrg2664.

22. Garcia AAF, Wang S, Melchinger AE, Zeng ZB. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. Genetics. 2008;180(3):1707–1724.

23. Xiao J, Li J, Yuan L, Tanksley SD. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics. 1995;140(2):745–754.

24. Yao H, Gray AD, Auger DL, Birchler JA. Genomic dosage effects on heterosis in triploid maize. Proceedings of the National Academy of Sciences. 2013;110(7):2665–2669.

25. Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA. Heterosis. The Plant Cell. 2010;22(7):2105–2112.

26. Krieger U, Lippman ZB, Zamir D. The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. Nature genetics. 2010;42(5):459–463.

27. Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, et al. Genomic architecture of heterosis for yield traits in rice. Nature. 2016;537(7622):629–633.

28. Larièpe a, Mangin B, Jasson S, Combes V, Dumas F, Jamin P, et al. The genetic basis of heterosis: multiparental quantitative trait loci mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (Zea mays L.). Genetics. 2012;190(2):795–811. doi:10.1534/genetics.111.133447.

29. Mikel MA, Dudley JW. Evolution of North American dent corn from public to proprietary germplasm. Crop science. 2006;46(3):1193–1205.

30. Macke JA, Bohn MO, Rausch KD, Mumm RH. Genetic Factors Underlying Dry-Milling Efficiency and Flaking-Grit Yield Examined in US Maize Germplasm. Crop Science. 2016;56(5):2516–2526.

31. Cooper GM, Stone Ea, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome research. 2005;15(7):901–13. doi:10.1101/gr.3577405.

32. Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the Bayesian alphabet for genomic selection. BMC bioinformatics. 2011;12(1):186.

33. Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. Longman; 1996. Available from: http://books.google.de/books?id=7ASZNAEACAAJ.

34. Doyle J, Doyle J. Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochem Bull. 1987;19(11):11–15.

35. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324.

36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. doi:10.1093/bioinformatics/btp352.

37. Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, et al. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 2013;14(6):R55. doi:10.1186/gb-2013-14-6-r55.

38. Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, et al. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS one. 2014;9(2):e90346.

39. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS computational biology. 2010;6(12):e1001025. doi:10.1371/journal.pcbi.1001025.

40. Rodgers-Melnick E, Bradbury PJ, Elshire RJ, Glaubitz JC, Acharya CB, Mitchell SE, et al. Recombination in diverse maize is stable, predictable, and associated with genetic load. Proceedings of the National Academy of Sciences. 2015;112(12):3823–3828.

41. Cingolani P, Platts A, Coon M, Nguyen T, Wang L, Land SJ, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80–92.

42. Ogut F, Bian Y, Bradbury PJ, Holland JB. Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. Heredity. 2015;114(6):552.

43. Bukowski R, Guo X, Lu Y, Zou C, He B, Rong Z, et al. Construction of the third generation Zea mays haplotype map. bioRxiv. 2015; p. 026963.

44. Da Y, Wang C, Wang S, Hu G. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using SNP markers. PloS one. 2014;9(1).

45. Wang C, Prakapenka D, Wang S, Pulugurta S, Runesha HB, Da Y. GVCBLUP: a computer package for genomic prediction and variance component estimation of additive and dominance effects. BMC Bioinformatics. 2014;15(1):1–9. doi:10.1186/1471-2105-15-270.

46. Lynch M, Walsh B, et al. Genetics and analysis of quantitative traits. vol. 1. Sinauer Sunderland, MA; 1998.

47. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84(2):210–23. doi:10.1016/j.ajhg.2009.01.005.

48. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. Nature Reviews Genetics. 2013;14(7):507–515.

49. Navarro JAR, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nature genetics. 2017;49(3):476–480.

50. Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. Proceedings of the National Academy of Sciences. 2016;113(22):E3177–E3184. doi:10.1073/pnas.1525244113.

51. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310–315.

52. Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. Reduced efficacy of selection in regions of the Drosophila genome that lack crossing over. Genome biology. 2007;8(2):R18. doi:10.1186/gb-2007-8-2-r18.

53. McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, et al. Genetic properties of the maize nested association mapping population. Science (New York, NY). 2009;325(5941):737–40. doi:10.1126/science.1174320.

54. Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, et al. The genetic architecture of maize flowering time. Science. 2009;325(5941):714–718.

55. Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, et al. The genetic architecture of maize height. Genetics. 2014;196(4):1337–1356.

56. Greenberg AJ, Hackett SR, Harshman LG, Clark AG. A hierarchical Bayesian model for a novel sparse partial diallel crossing design. Genetics. 2010;185(1):361–373.

57. Joly-Lopez Z, Flowers JM, Purugganan MD. Developing maps of fitness consequences for plant genomes. Current opinion in plant biology. 2016;30:101–107.

58. Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES. Association mapping across numerous traits reveals patterns of functional variation in maize. PLoS Genet. 2014;10(12):e1004845.

59. East EM. Inbreeding in corn. Rep Conn Agric Exp Stn. 1908;1907:419–428.

60. Kacser H, Burns JA. The molecular basis of dominance. Genetics. 1981;97(3-4):639–666.

61. Phadnis N, Fry JD. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. Genetics. 2005;171(1):385–392.

62. Desta ZA, Ortiz R. Genomic selection: genome-wide prediction in plant improvement. Trends in plant science. 2014;19(9):592–601.

63. Edwards SM, Sørensen IF, Sarup P, Mackay TF, Sørensen P. Genomic Prediction for Quantitative Traits Is Improved by Mapping Variants to Gene Ontology Categories in Drosophila melanogaster. Genetics. 2016; p. 110–116.

64. Shapira R, Levy T, Shaked S, Fridman E, David L. Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. Heredity. 2014;113(4):1–11. doi:10.1038/hdy.2014.33.

65. Shull GH. The composition of a field of maize. Journal of Heredity. 1908;1(1):296–301.

66. Gama LT, Bressan MC, Rodrigues EC, Rossato LV, Moreira OC, Alves SP, et al. Heterosis for meat quality and fatty acid profiles in crosses among Bos indicus and Bos taurus finished on pasture or grain. Meat Science. 2013;93(1):98–104. doi:10.1016/j.meatsci.2012.08.005.

67. Wang L, Greaves IK, Groszmann M, Wu LM, Dennis ES, Peacock WJ. Hybrid mimics and hybrid vigor in Arabidopsis. Proceedings of the National Academy of Sciences. 2015;112(35):E4959–E4967.

68. Li Z, Li B, Tong Y. The contribution of distant hybridization with decaploid Agropyron elongatum to wheat improvement in China. Journal of Genetics and Genomics. 2008;35(8):451–456.

69. Halligan DL, Keightley PD. Spontaneous mutation accumulation studies in evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics. 2009;40:151–172.

70. Guo M, Rupe MA, Wei J, Winkler C, Goncalves-Butruille M, Weers BP, et al. Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize yield. Journal of experimental botany. 2014;65(1):249–260.

1 Supporting Tables

Table S1. BLUE values and levels of heterosis of the seven phenotypic traits for the 66 hybrids. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_trait_heterosis.csv`)

Table S2. General combining ability and specific combining ability of the seven phenotypic traits.
(`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_CA.csv`)

Table S3. SNP missing rate in our diallel parental lines. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_SNP_missing_rate.csv`)

Table S4. Summary statistics of SNP annotation results. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_S_snpeff_results.xlsx`)

Table S5. Number of deleterious SNPs carried per line. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_S_del_per_line.csv`)

Table S6. Number of complementation and homozygote deleterious load for GERP-SNPs in hybrids. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_S_del_complemenation.csv`)

Table S7. The correlation between the number of homozygote GERP-SNPs and the hybrid phenotypes. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_hyb_load_pheno.csv`)

Table S8. Model comparisons P values and AICs. (`https://github.com/yangjl/GERP-diallel/blob/master/table/Table_model_comp.csv`)

21

2 Supporting Figures
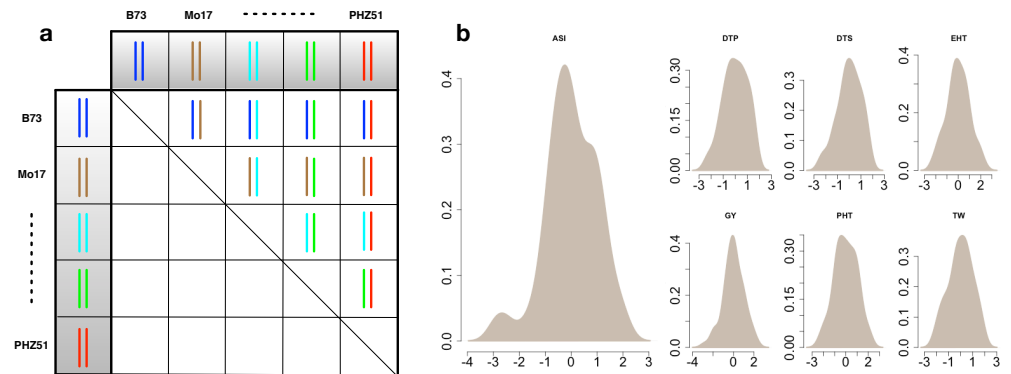
Fig S1. A half-diallel population and distributions of phenotypes. (**a**) Twelve maize inbred lines were selected and crossed in a half-diallel fashion. Each inbred lines was used as both male and female and the resulting F1 seed was bulked. (**b**) Density plots of normalized BLUE values for the seven phenotypic traits. We used "scale" function in R to normalize the BLUE values by first centering on zero and then dividing the numbers by their standard deviation. The seven phenotypic traits are plant height (PHT), height of primary ear (EHT), days to 50% pollen shed (DTP), days to 50% silking (DTS), anthesis-silking interval (ASI), grain yield adjusted to 15.5% moisture (GY), and test weight (TW).
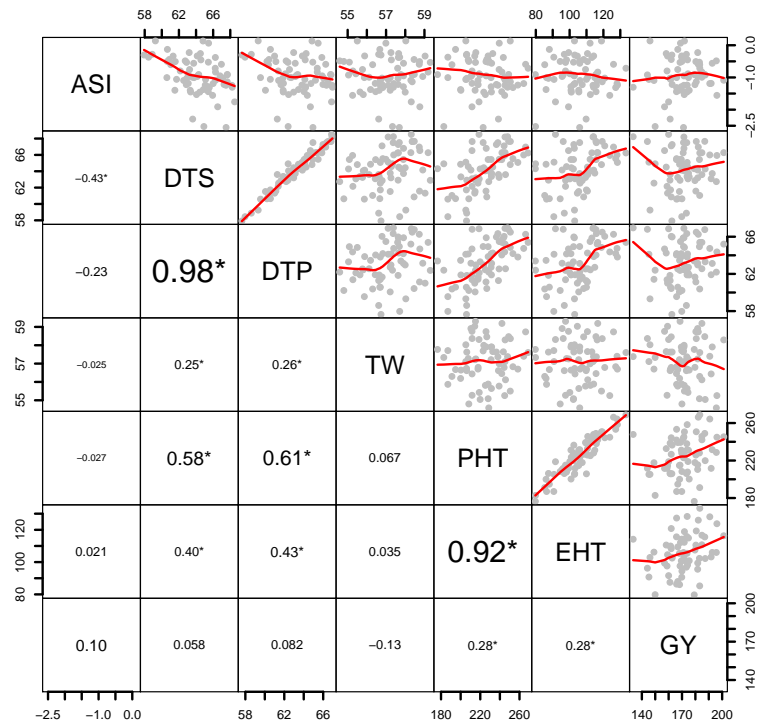
Fig S2. Pairwise correlation plots of seven phenotypic traits. The upper right panels show the scatter plots of all possible pairwise comparisons of two traits. Red line is a fitted smooth curve using "loess" method. In the lower left panels, the numbers are the Spearman correlation coefficients (r) and the asterisks (*) indicate the correlation coefficients are statistically significant (Spearman correlation test P value < 0.05). Units for various traits are plant height (PHT, in cm), height of primary ear (EHT, in cm), days to 50% pollen shed (DTP), days to 50% silking (DTS), anthesis-silking interval (ASI, in days), grain yield adjusted to 15.5% moisture (GY, in bu/A), and test weight (TW, weight of 1 bushel of grain in pounds).

	HB1	HB2	HB3	HB4	HB5	HB6	HB7	HB8	HB9	HB10	HB11	HB12	HB13	HB14
Hyb1	0	0	0	1	1	1	1	1	0	0	1	1	1	0
Hyb2	1	1	0	0	0	1	1	1	1	1	1	0	0	0
Hyb3	0	1	1	1	0	1	1	1	1	1	1	0	0	0
Hyb4	0	0	1	1	1	1	0	1	1	0	0	0	1	1
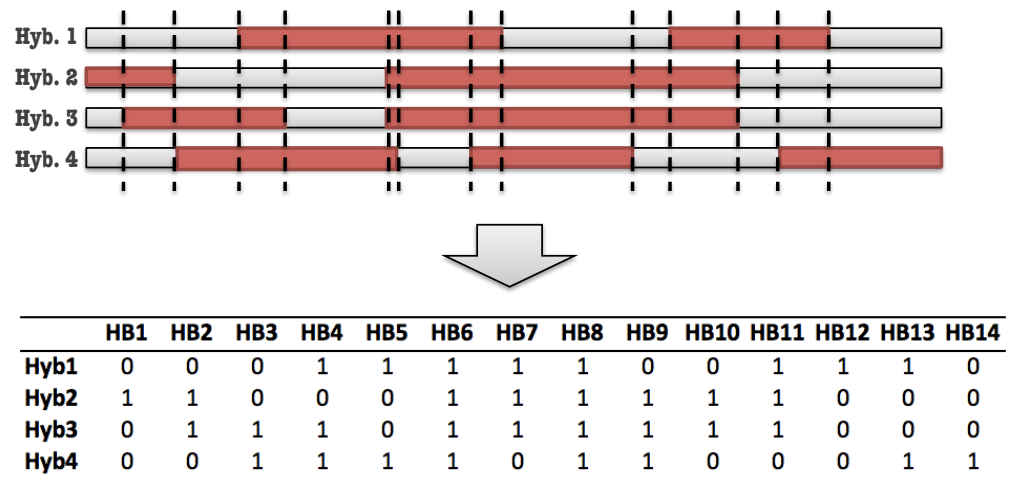
Fig S3. Haplotype block identification using an IBD approach. In the upper panel, regions in red are IBD blocks identified by pairwise comparison of the two parental lines of a hybrid. The vertical dashed lines define haplotype blocks. In the lower panel, hybrid genotype in each block are coded as heterozygotes (0) or homozygotes (1).
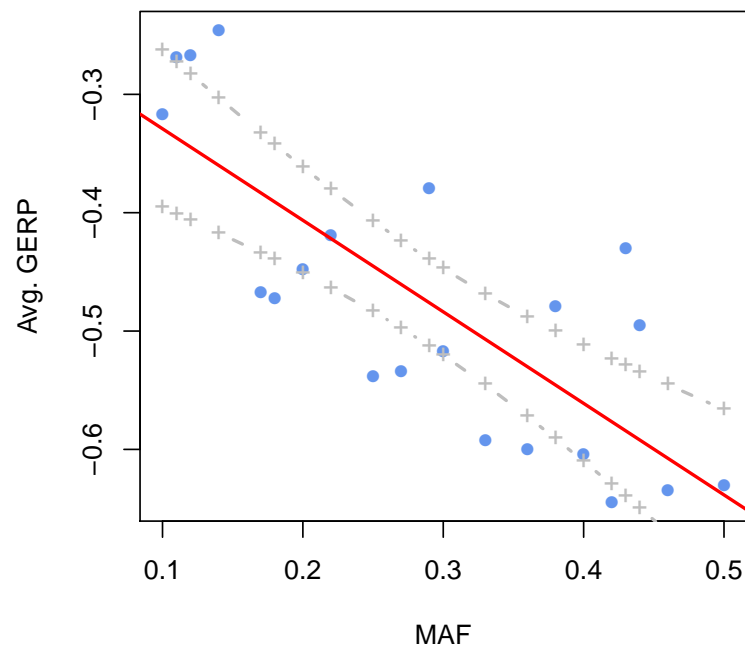
Fig S4. The minor allele frequency estimated from 12 parental lines in bins of 0.01 GERP score. Red solid and grey dashed lines define the best-fit regression line and its 95% confidence interval.
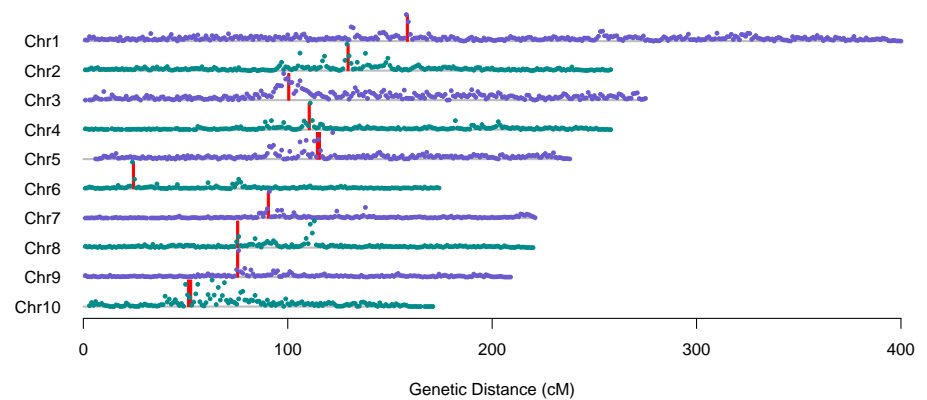
Fig S5. Segregating genetic load across ten maize chromosomes. Dots indicate mean GERP scores of putatively deleterious SNPs (GERP scores > 0) carried by our 12 elite maize lines (bin size = 1 cM). Vertical red lines indicate centromeres.
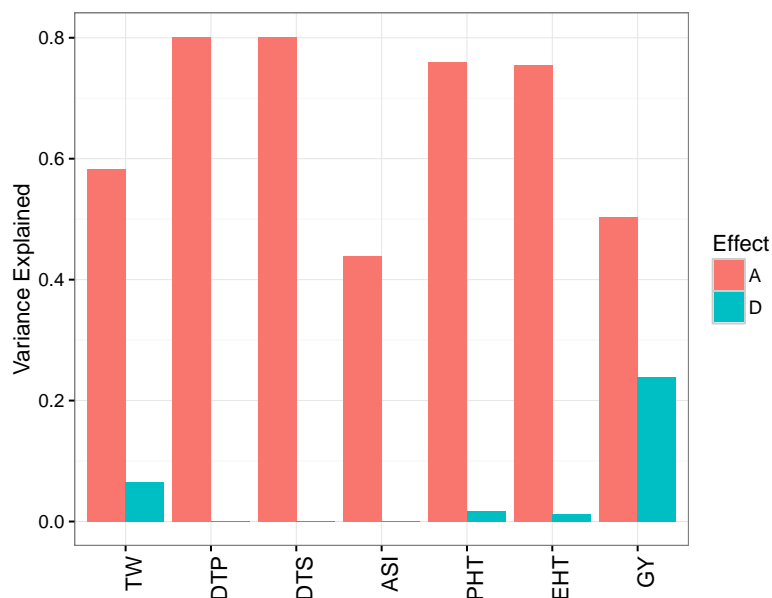
Fig S6. Cumulative variance explained by GERP-SNPs. Additive and dominance effects are indicated by red and blue colors respectively.
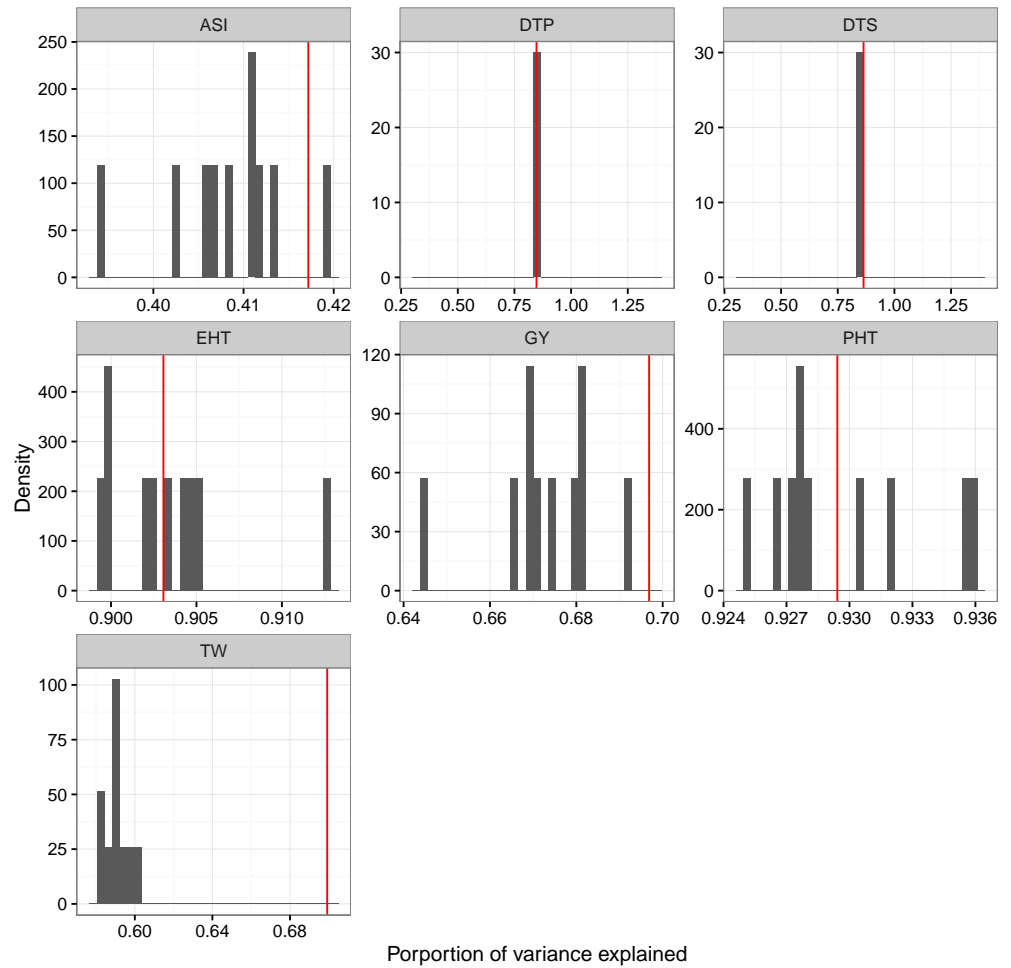
Fig S7. Phenotypic variance explained for observed data and for randomly shuffled data using the genomic selection model. Histograms show the results for the randomly shuffled (10 times) degrees of dominance (k) in each trait. Red lines are phenotypic variance explained using the observed k.
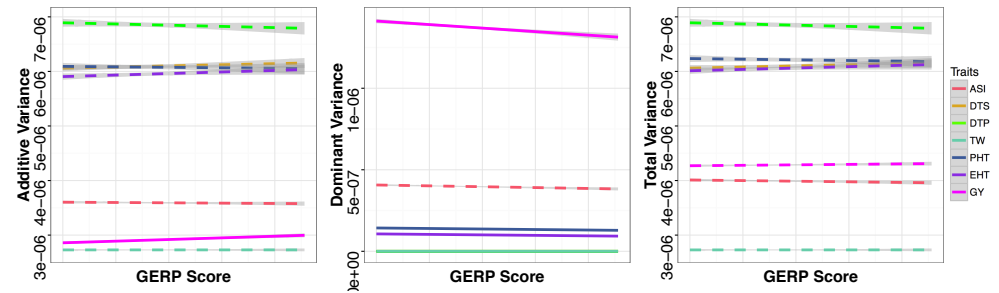
Fig S8. Linear regressions of GERP-SNPs' additive variance, dominance variance and total variance of seven traits *per se* against their GERP scores. Solid and dashed lines represent significant and non-significant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for SNPs with $> 1\times$ of the mean genome-wide phenotypic variance explained.
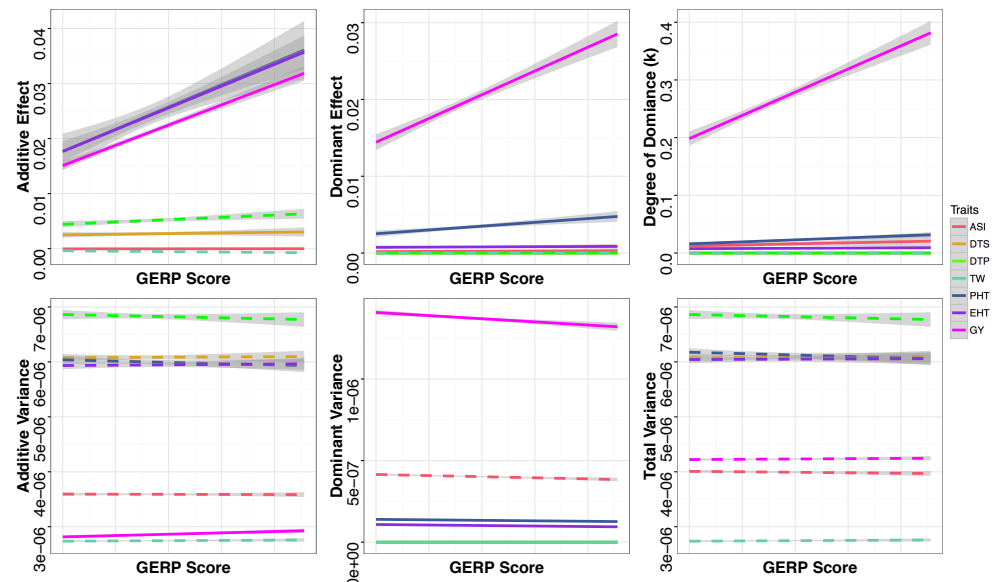
Fig S9. Linear regressions after filtering out GERP-SNPs located in regions in the lowest quartiles of recombination. Solid and dashed lines represent significant and non-significant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for GERP-SNPs with $> 1\times$ of the mean genome-wide variance explained and with > 1st quantile of the recombination rate (cM/Mb).
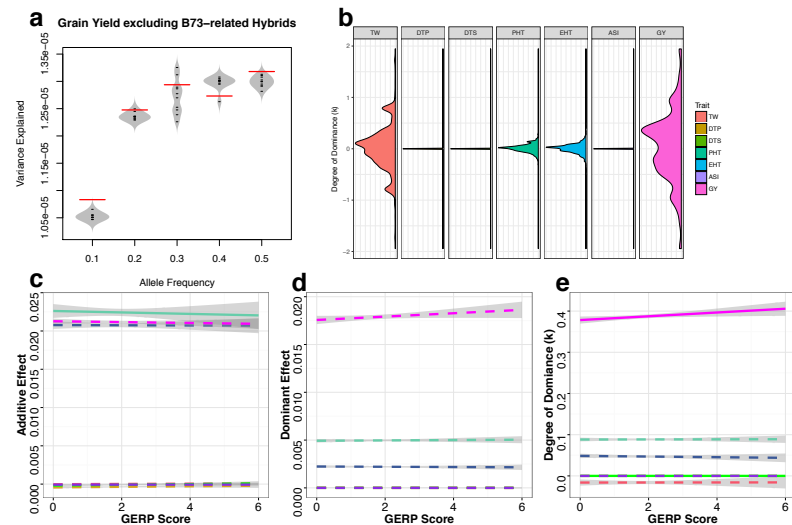
Fig S10. Phenotypic variance explained for grain yield and degree of dominance (k) of GERP-SNPs after removing 11 hybrids that B73 as one parent. (a) Total per-SNP variance explained for grain yield trait *per se* by deleterious (red lines) and randomly sampled SNPs (grey beanplots). **(b)** Density plots of the degree of dominance (k). Extreme values of k were truncated at 2 and -2 for visualization. **(c-e)** Linear regressions of additive effects **(c)**, dominance effects **(d)**, and degree of dominance **(e)** of seven traits *per se* against SNP GERP scores. Colors in **(c-e)** are the same as the legend for **(b)**. Solid and dashed lines represent significant and nonsignificant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for deleterious alleles with the above mean genome-wide variance explained.
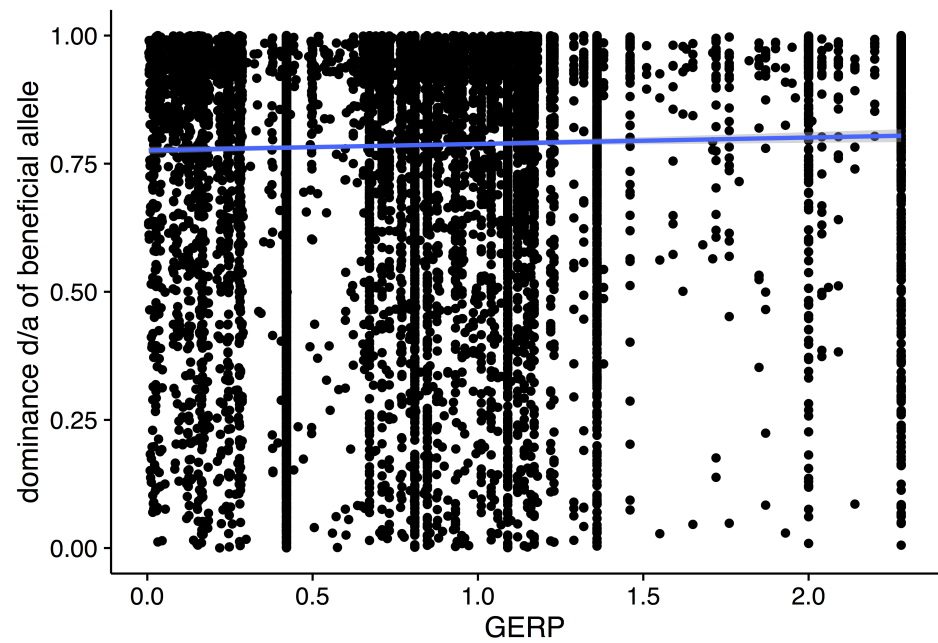
Fig S11. Regression of degree of dominance (k) on GERP scores for simulated data. Solid blue line indicates the regression line fitted to data simulated under mutation-selection balance (see **Methods** for details).
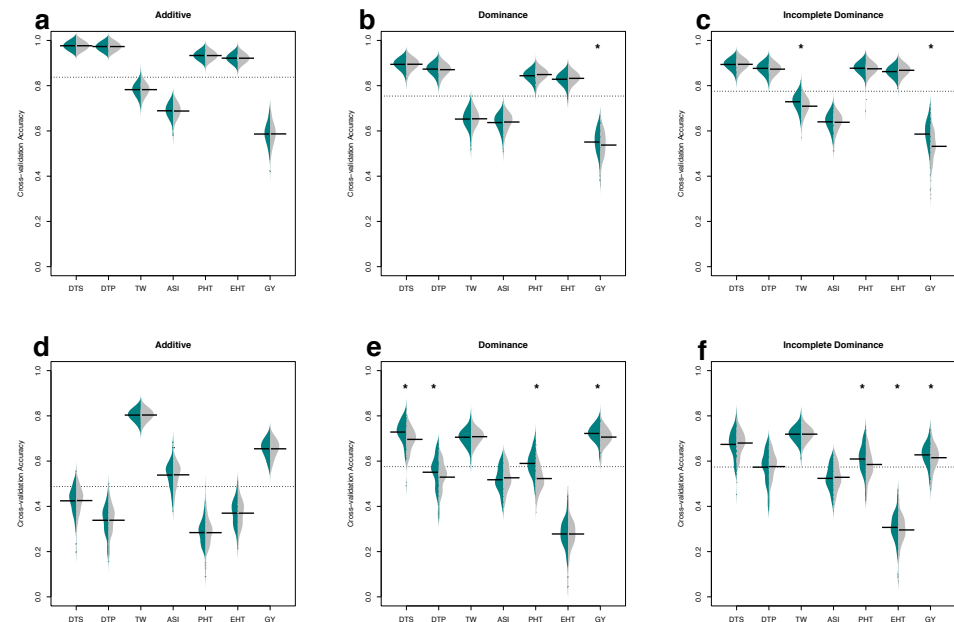
Fig S12. Cross-validation accuracy using GERP-SNPs in genic regions.
Beanplots represent prediction accuracy estimated from cross-validation experiments for traits *per se* (**a, b, c**) and heterosis (**d, e, f**) under additive (**a, d**), dominance (**b, e**), and incomplete dominance (**c, f**) models. Prediction accuracy using real data is shown on the left (green) and permutation results on the right (grey). Horizontal bars indicate mean accuracy and the grey dashed lines indicate the overall mean accuracy. Stars indicate significantly (permutation FDR < 0.05) higher than cross-validation accuracy.
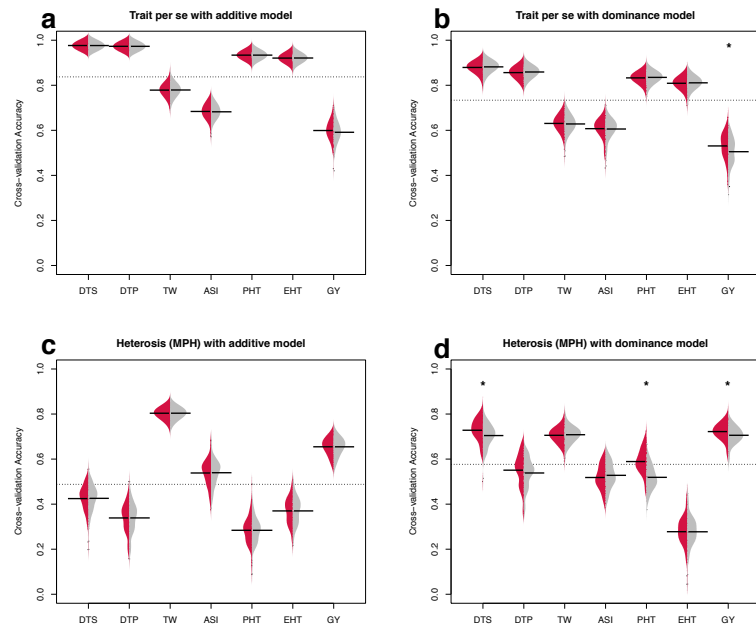
Fig S13. Cross-validation prediction accuracy for trait *per se* and heterosis.
Beanplots represent prediction accuracy estimated from cross-validation experiments for
traits *per se* (**a, b**) and heterosis (**c, d**) under additive (**a, c**) and dominance (**b, d**)
models. Prediction accuracy using real data is shown on the left (red) and permutation
results on the right (grey). Horizontal bars indicate mean accuracy of each trait and the
grey dashed lines indicate the mean accuracy of all traits. Stars indicate real data
having significantly (t-test P value < 0.05) higher cross-validation accuracy than
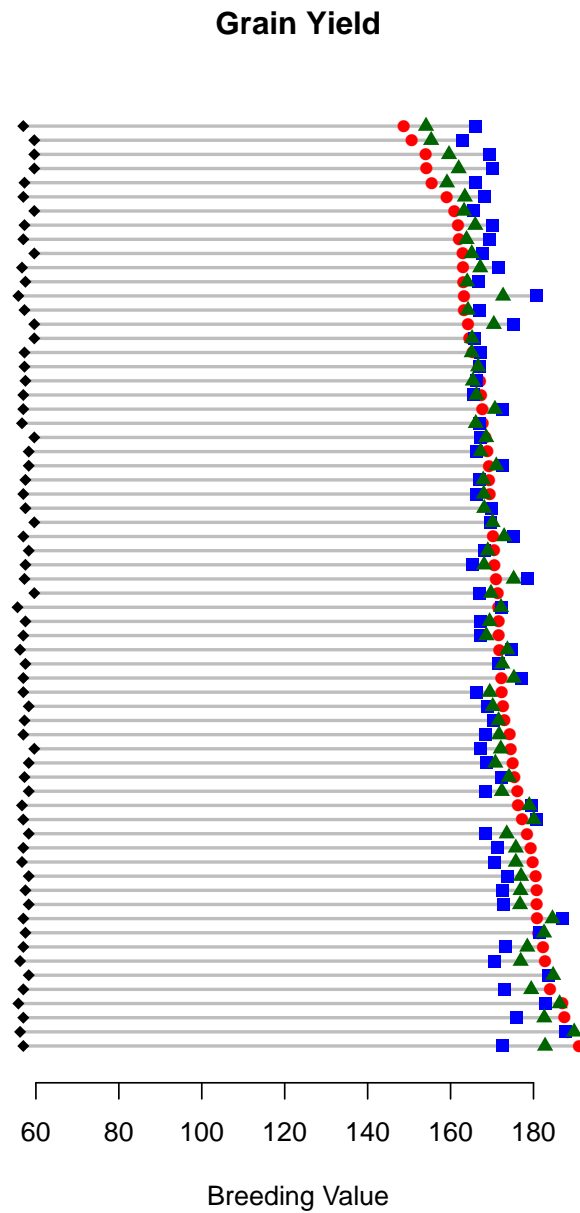permuted data.

Fig S14. Breeding values of grain yield for diploid and simulated triploid hybrids. Each line represents the posterior breeding values of a diploid hybrid (red circle), its best parent (black diamond), and predicted breeding values of AAB triploid (blue square) and ABB triploid (green triangle) based on estimated effect sizes and dominance values for each SNP.