Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize

Jinliang Yang[1,†,§], Sofiane Mezmouk[1,2,†], Andy Baumgarten[3], Edward S. Buckler[4], Katherine E. Guill[5], Michael D. McMullen[5,6], Rita H. Mumm[7], and Jeffrey Ross-Ibarra[1,8,§]

1 Department of Plant Sciences, University of California, Davis, CA 95616, USA,
2 Current address: KWS SAAT SE, Grimsehlstr. 31, 37555 Einbeck, Germany,
3 DuPont Pioneer, Johnston, IA 50131, USA,
4 US Department of Agriculture, Agricultural Research Service, Ithaca, NY 14853, USA,
5 US Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA,
6 Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA,
7 Department of Crop Sciences and the Illinois Plant Breeding Center, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA,
8 Center for Population Biology and Genome Center, University of California, Davis, CA 95616, USA.

† These authors contributed equally to this work.
§ Correspondence should be addressed to J.Y. (jolyang@ucdavis.edu) and J.R.-I. (rossibarra@ucdavis.edu).

Abstract

Complementation of deleterious alleles has long been proposed as a major contributor to the hybrid vigor observed in the offspring of inbred parents. We test this hypothesis 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 500,000 putatively deleterious sites, but show a lower burden of such sites than a comparable set of inbred 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 weight blocks on the basis of segregating putatively deleterious variants.
Cross-validation results show that incorporating sequence conservation in genomic selection improves prediction accuracy for yield and several other traits as well as heterosis for those traits. Our results provide strong 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.

Results

A key long-term goal of biology is understanding the genetic basis of phenotypic variation, which 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, demographic changes common to many species may lead to an increased abundance of deleterious alleles over shorter time scales such as those associated with domestication [4] or recent human migration [5]. 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, 6–8]. 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.

In addition to its global agricultural importance, maize has long been an important genetic model system [9] and central to debates about the basis of hybrid vigor and the role of deleterious alleles [10, 11]. Rapid expansion following maize domestication likely lead to an increase in new mutations and stronger purifying selection [4], but inbreeding during the development of modern inbred lines may have decreased load by purging recessive deleterious alleles [12]. Nonetheless, patterns of heterozygosity during inbreeding [13, 14] and selection [15] suggest an abundance of deleterious alleles in modern germplasm, and genome-wide association reveals an excess of associations with genes segregating for deleterious protein-coding variants [16]. We set out to investigate the contribution of deleterious alleles to phenotypic variation and hybrid vigor in maize, creating a partial diallel population from 12 maize inbred lines (**Supplementary Figure S1**) which together represent much of the ancestry of present-day commercial U.S. corn hybrids [17, 18]. 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). 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 variations (see **Methods**).

We estimated best-parent heterosis (BPH, **Figure 1a**) for each trait as the percent difference between the hybrid compared to the better of its two parents (see **Methods**, **Table S2**.). Consistent with previous work [19], we find that grain yield (GY) — analogous to fitness in an agronomic setting — showed the highest heterosis (BPH of 127% ± 35%). Flowering time (DTS and DTP), although important to fitness globally across wide latitudinal ranges, showed relatively little heterosis in this study as the parental lines vary within a narrow range in our field trials.

In order to quantify the deleterious consequences of variants *a priori*, we resequenced the 12 inbred parents to an average depth of ≈ 10× resulting in a filtered set of 13.8M SNPs with a mean concordance rate of 99.1% to SNPs genotyped in previous studies (**Supplementary Note**). We annotated variable sites in our sample using Genomic Evolutionary Rate Profiling (GERP) [20] scores, 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. 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 of the 506,898 sites with GERP scores > 0, we designated the minor allele from the multispecies alignment as putatively deleterious. Compared to a panel of specially inbred landrace (traditionally outcrossing) lines [12], the parents of our diallel exhibited

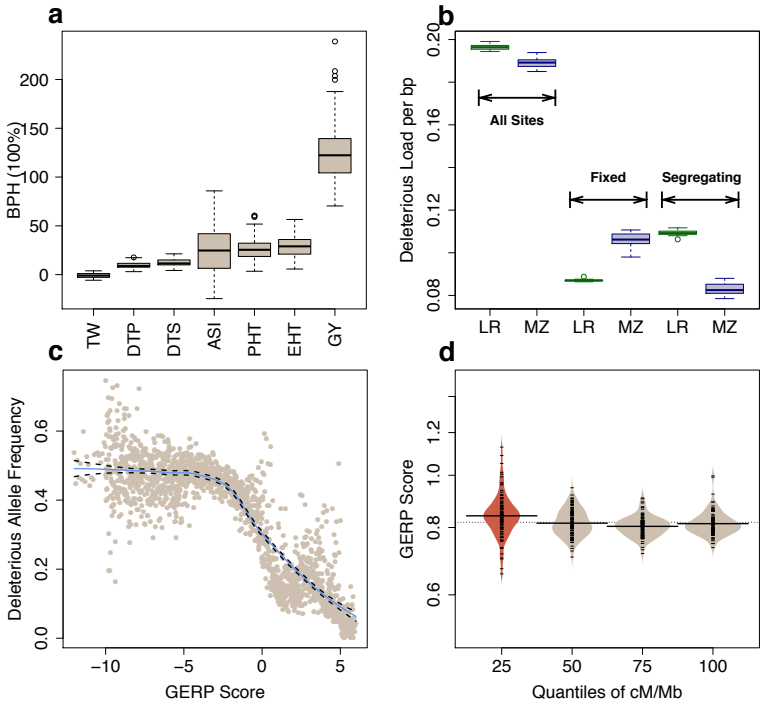


Figure 1. Heterosis and deleleterious variants. (**a**) Boxplots (median and interquartile range) of percent best-parent heterosis (BPH). (**b**) Burden of deleterious alleles in landrace (LR, green boxplots) and elite (MZ, blue boxplots) maize lines. (**c**) The mean frequency of the putatively deleterious allele in bins of 0.01 GERP score. Blue solid and black dashed lines define the best-fit regression line and its 95% confidence interval. (**d**) Density plots of mean GERP scores in quartiles of recombination rates (cM/Mb). The dashed line indicates the overall mean; means for each quartile are shown as solid lines and colors represent quartiles that are significantly different.

a slightly decreased burden of deleterious alleles (23,000 fewer per genome, **Figure 1b**); Deleterious allele frequencies showed a strong negative correlation with GERP score across the more than 1,200 lines in maize HapMap 3.2 [21] (**Figure 1c** and **Supplementary Note**) and GERP scores were highest in regions of the genome with low recombination (Student's t-test P value < 0.05; **Figure 1d**), especially in pericentromeric regions (**Supplementary Figure S2**). These results match well with predictions from population genetic theory [22] and previous empirical work [14, 16, 23, 24], supporting the use of GERP scores as a quantitative measure of the fitness effects of an observed variant.

We estimated the effect sizes and variance explained by deleterious (GERP > 0) SNPs across hybrid phenotypes *per se* using a genomic best linear unbiased prediction (GBLUP) [25] approach. Deleterious SNPs had larger average effects and explained more phenotypic variance than randomly sampled SNPs matched for allele frequency and recombination (**Figure 2a**). We found the cumulative proportion of dominance variance explained by deleterious 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 (**Supplementary Figure S3**). 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 previous estimates [18, 26, 27]. We then evaluated the relationship between the strength of

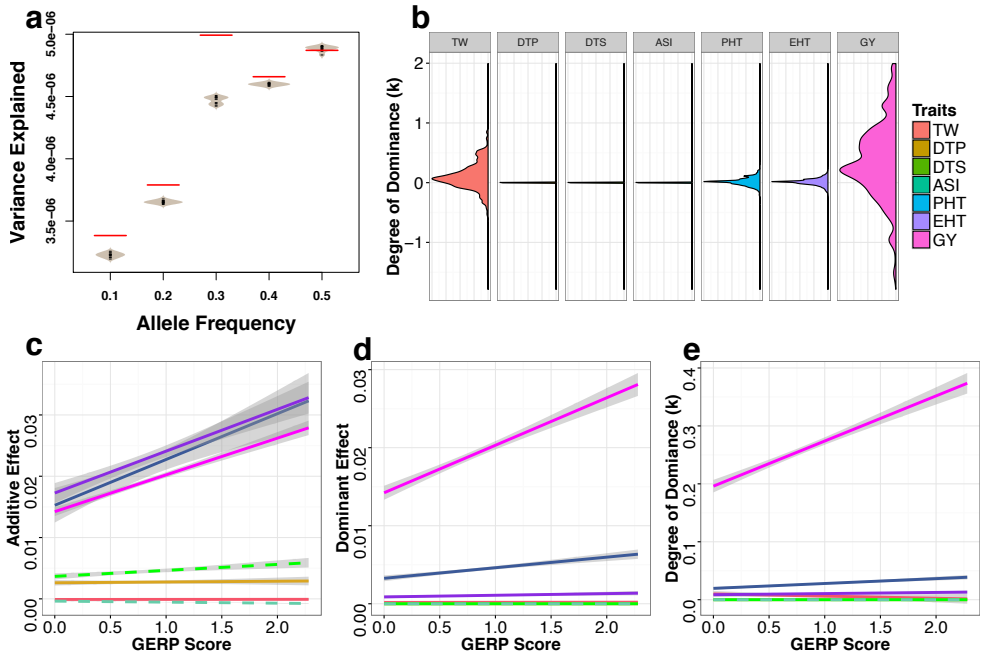


Figure 2. Variance explained and degree of dominance (k) of GERP-SNPs for traits *per se*. **(a)** Total per-SNP variance explained by deleterious (red lines) and randomly sampled SNPs (grey beanplots). **(b)** Density plots of the degree of dominance (k). Extreme values of k were truncated at 2 and -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 mean genome-wide variance explained (see **Methods**).

selection against SNPs and their effect size, contribution to phenotypic variance, and dominance. We found weak or negligible correlations between effect size and GERP score for flowering time and grain quality traits, but a strong positive correlation for fitness-related traits (**Figure 2c-d**). The variance explained by individual SNPs, however is largely independent of how deleterious they are (**Supplemetary Figure S4**), likely due to the negative correlation between frequency and effect size for fitness-related traits. We see similar patterns for both analyses even removing SNPs in low recombination regions where effect size might be inflated due to high levels of linkage disequilibrium (**Supplementary Figure S5**). Finally, we observed a positive relationship between GERP score and degree of dominance (k) (**Figure 2e**) such that more deleterious alleles (larger GERP score) were more recessive (larger k for the beneficial allele), simulations suggest ascertainment bias alone is unlikely to explain this result (**Supplementary Figure S6**). Though we are unaware of previous demonstrations of this realtionship across multiple traits in other multicellular organisms, these finding closely follow predicted patterns of dominance in models of metabolic pathways [28] and support previous empirical evidence from gene knockouts in yeast [29].

As genotyping costs continue to decline, genomic prediction models are increasing in

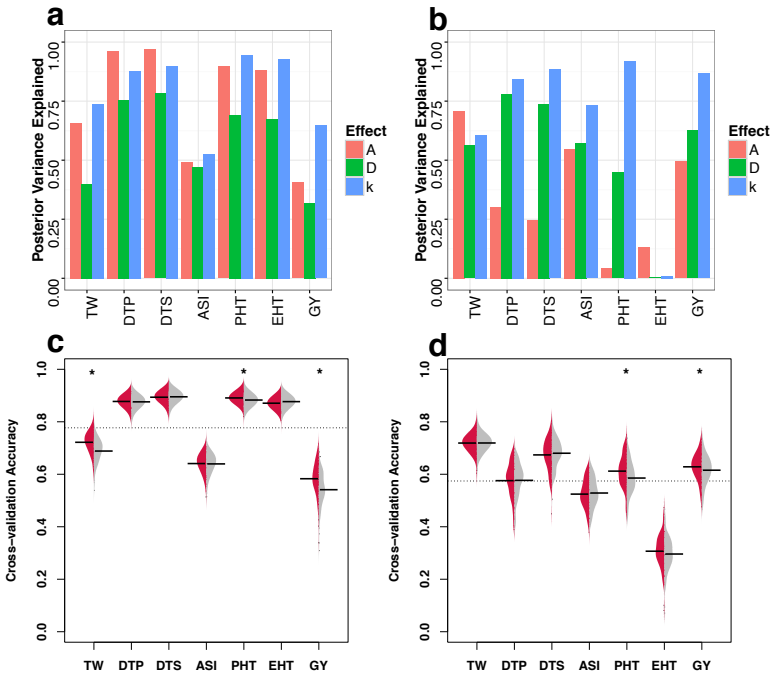


Figure 3. Genomic prediction models incorporating GERP. (a-b) Total posterior variance explained for traits *per se* **(a)** and heterosis (BPH) (**b**) under models of additivity (red), dominance (green), and incomplete dominance (blue). **(c-d)** Beanplots represent prediction accuracy estimated from cross-validation experiments for traits *per se* **(c)** and heterosis (BPH) **(d)** under a model of incomplete dominance. Prediction accuracy using estimated dominance values for each SNP is shown on the left (red) and permutation results 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 **Supplementary Figure S7**.

popularity [30]. 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.,* [31] for a recent example). We implemented a haplotype-based genomic prediction model that incorporates weights based on our *a priori* identification of deleterious alleles (see **Methods** and **Supplementary Figure S8**). We explored several different models and found that a model which incorporates both the GERP scores and the estimated levels of dominance (k) of deleterious alleles explained a greater amount of the posterior phenotypic variance for heterosis and most traits *per se* (**Figure 3a-b**). A simple additive model showed superior explanatory power for flowering time, however, consistent with previous association mapping results [26]. Cross-validation analyses (see **Methods**) showed that models incorporating observed GERP scores out-performed permutations (**Figure 3c-d**), even after controlling for differences between genic and nongenic regions (**Supplementary Figure S9**). Our model improved prediction accuracy of grain yield by more than 4%, and improvements were also seen for plant height (0.8%) and grain quality (3.3%). While our model showed no improvement for traits showing low levels of heterosis (**Figure 1a**), including GERP scores significantly improved prediction accuracy for heterosis of both plant

height and grain yield (by 2.6% and 1.3% respectively). 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 S3**) calculated directly from the pedigree of the hybrid population [32] (ANOVA FDR < 0.01 and difference in AIC < 0, **Table S4**).

Discussion

Taken together, our results provide support for an important role of deleterious variants in determining phenotypic variation for traits related to fitness. In maize and other organisms that have recently undergone substantial population growth, many new mutations are likely deleterious [4] and a large proportion of the phenotypic variance is expected to be explained by deleterious variants of small effect [6]. Though our population size is small, our partial diallel crossing design and GBLUP approach circumvent some of the problems with standard genome-wide association analyses, including genome-wide multiple testing thresholds and the inability to assess the effects of rare alleles. We show that *a priori* information on the fitness consequence of a variant is useful in predicting effect sizes and dominance for grain yield, a close surrogate of fitness in agronomic settings, as well as for correlated traits such as plant height and ear height. All three traits are well explained by a model allowing for incomplete dominance (**Figure 3a**), and incorporating GERP scores allows for improved prediction of both plant height and grain yield (**Figure 3c**). Though GERP scores do provide information for other traits (effect size for flowering time, prediction increase for grain quality), these impacts are lesser as expected for traits less correlated to plant fitness (**Supplementary Figure S10**). While the GERP scores used here reflect conservation of across relatively deep phylogenetic time, future efforts may be able to increase power by incorporating information from within-species polymorphism data [33] as well as other types of annotations that have been shown to contribute substantially to phenotypic variation (e.g. Wallace *et al.,* [34] and Rodgers-Melnick *et al.,* [35]).

Finally, our results also have implications for understanding the genetic basis of heterosis. Heterosis has been observed across many species, from yeast [36] to plants [37] and vertebrates [38], and a number of hypotheses have been put forth to explain the phenomenon [10, 39]. Of all of these explanations, complementation of recessive deleterious alleles [11, 39] remains the simplest genetic explanation, and one that is supported by considerable empirical evidence [40–42]. It remains controversial, however, because complementation of purely recessive mutations cannot fully explain a number of empirical observations, including patterns of heterosis and inbreeding depression in polyploid plants [10, 43, 44]. Our estimates, however, indicate that most deleterious SNPs show incomplete dominance (**Figure 2b**) for traits with high levels of heterosis, and our genomic prediction models find substantial 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 [45] and that GWAS hits exhibit incomplete dominance for phenotypes *per se* among hybrids [46]. We argue that allowing for incomplete dominance effectively unifies models of simple complementation with those of gene dosage [44]. Combined with observations that deleterious SNPs are enriched in low-recombination pericentromeric regions [24] (**Figure 1d**), such a model can satisfactorily explain changes in heterozygosity during breeding [15, 23], enrichment of yield QTL around centromeres [19], and even observed patterns of heterosis in polyploids (**Supplementary Figure S11**). 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.,* [47] and

Huang *et al.,* [46]), 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.

Acknowledgements

Financial support for this work came from NSF (grants IOS-0820619 and IOS-1238014), USDA (grant 2009-65300-05668 and Hatch projects CA-D-PLS-2066-H and ILLU-802-354), DuPont Pioneer, Kellogg Company, and Mars Incorporated. We would like to thank Tim Beissinger, Graham Coop, James Holland, Matthew Hufford, Emily Josephs, Peter Morrell, Michelle C. Stitzer, Kevin Thornton, and Stephen Wright for helpful discussion.

Author Contributions

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. generated the data. J.Y., S.M. and J.R.-I. analyzed data. A.B. provided statistical advice. J.Y. and J.R.-I. wrote the manuscript.

Competing Interests Statement

The authors declare no competing financial interests.

References

1. Adam Eyre-Walker and Peter D Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618, 2007.

2. Thomas Mitchell-Olds, John H Willis, and David B Goldstein. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, 8(11):845–856, 2007.

3. Yuval B Simons, Michael C Turchin, Jonathan K Pritchard, and Guy Sella. The deleterious mutation load is insensitive to recent population history. *Nature genetics*, 46(3):220, 2014.

4. Timothy Mathes Beissinger, Li Wang, Kate Crosby, Arun Durvasula, Matthew B Hufford, and Jeffrey Ross-Ibarra. Recent demography drives changes in linked selection across the maize genome. *Nature Plants*, 2(7):16084, 2016.

5. Stephan Peischl, Isabelle Dupanloup, Adrien Foucal, Michèle Jomphe, Vanessa Bruat, Jean-Cristophe Grenier, Alexandre Gouy, Elias Gbeha, Lars Bosshard, Elodie Hip-Ki, et al. Relaxed selection during a recent human expansion. *bioRxiv*, page 064691, 2016.

6. Kirk E Lohmueller. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet*, 10(5):e1004379, 2014.

7. Lawrence H Uricchio, Noah A Zaitlen, Chun Jimmie Ye, John S Witte, and Ryan D Hernandez. Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. *Genome research*, 26:863–873, 2016.

8. Jaleal Sanjak, Anthony D Long, and Kevin R Thornton. The genetic architecture of a complex trait is more sensitive to genetic model than population growth. *bioRxiv*, 2016.

9. Natalie J Nannas and R Kelly Dawe. Genetic and genomic toolbox of zea mays. *Genetics*, 199(3):655–669, 2015.

10. James A Birchler, Donald L Auger, and Nicole C Riddle. In search of the molecular basis of heterosis. *The Plant Cell*, 15(10):2236–2239, 2003.

11. James F Crow. 90 years ago: the beginning of hybrid maize. *Genetics*, 148(3):923–928, 1998.

12. Jer-Ming Chia, Chi Song, Peter J Bradbury, Denise Costich, Natalia de Leon, John Doebley, Robert J Elshire, Brandon Gaut, Laura Geller, Jeffrey C Glaubitz, Michael Gore, Kate E Guill, Jim Holland, Matthew B Hufford, Jinsheng Lai, Meng Li, Xin Liu, Yanli Lu, Richard McCombie, Rebecca Nelson, Jesse Poland, Boddupalli M Prasanna, Tanja Pyhäjärvi, Tingzhao Rong, Rajandeep S Sekhon, Qi Sun, Maud I Tenaillon, Feng Tian, Jun Wang, Xun Xu, Zhiwu Zhang, Shawn M Kaeppler, Jeffrey Ross-Ibarra, Michael D McMullen, Edward S Buckler, Gengyun Zhang, Yunbi Xu, and Doreen Ware. Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet*, 44(7):803–7, Jul 2012.

13. Michael D McMullen, Stephen Kresovich, Hector Sanchez Villeda, Peter Bradbury, Huihui Li, Qi Sun, Sherry Flint-Garcia, Jeffry Thornsberry, Charlotte Acharya, Christopher Bottoms, et al. Genetic properties of the maize nested association mapping population. *Science*, 325(5941):737–740, 2009.

14. Michael a Gore, Jer-Ming Chia, Robert J Elshire, Qi Sun, Elhan S Ersoz, Bonnie L Hurwitz, Jason a Peiffer, Michael D McMullen, George S Grills, Jeffrey Ross-Ibarra, Doreen H Ware, and Edward S Buckler. A first-generation haplotype map of maize. *Science (New York, N.Y.)*, 326(5956):1115–7, November 2009.

15. Justin P Gerke, Jode W Edwards, Katherine E Guill, Jeffrey Ross-Ibarra, and Michael D McMullen. The genomic impacts of drift and selection for hybrid performance in maize. *Genetics*, 201(3):1201–1211, 2015.

16. Sofiane Mezmouk and Jeffrey Ross-Ibarra. The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda, Md.)*, 4(January):163–71, 2014.

17. Mark A Mikel and John W Dudley. Evolution of north american dent corn from public to proprietary germplasm. *Crop science*, 46(3):1193–1205, 2006.

18. Joshua A Macke, Martin O Bohn, Kent D Rausch, and Rita H Mumm. Genetic factors underlying dry-milling efficiency and flaking-grit yield examined in us maize germplasm. *Crop Science*, 56(5):2516–2526, 2016.

19. a Larièpe, B Mangin, S Jasson, V Combes, F Dumas, P Jamin, C Lariagon, D Jolivot, D Madur, J Fiévet, A Gallais, P Dubreuil, A Charcosset, and L Moreau. 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*, 190(2):795–811, March 2012.

20. Eugene V Davydov, David L Goode, Marina Sirota, Gregory M Cooper, Arend Sidow, and Serafim Batzoglou. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS computational biology*, 6(12):e1001025, January 2010.

21. Robert Bukowski, Xiaosen Guo, Yanli Lu, Cheng Zou, Bing He, Zhengqin Rong, Bo Wang, Dawen Xu, Bicheng Yang, Chuanxiao Xie, et al. Construction of the third generation zea mays haplotype map. *bioRxiv*, page 026963, 2015.

22. Motoo Kimura. *The neutral theory of molecular evolution*. Cambridge University Press, 1984.

23. Michael D McMullen, Stephen Kresovich, Hector Sanchez Villeda, Peter Bradbury, Huihui Li, Qi Sun, Sherry Flint-Garcia, Jeffry Thornsberry, Charlotte Acharya, Christopher Bottoms, Patrick Brown, Chris Browne, Magen Eller, Kate Guill, Carlos Harjes, Dallas Kroon, Nick Lepak, Sharon E Mitchell, Brooke Peterson, Gael Pressoir, Susan Romero, Marco Oropeza Rosas, Stella Salvo, Heather Yates, Mark Hanson, Elizabeth Jones, Stephen Smith, Jeffrey C Glaubitz, Major Goodman, Doreen Ware, James B Holland, and Edward S Buckler. Genetic properties of the maize nested association mapping population. *Science (New York, N.Y.)*, 325(5941):737–40, August 2009.

24. Eli Rodgers-Melnick, Peter J Bradbury, Robert J Elshire, Jeffrey C Glaubitz, Charlotte B Acharya, Sharon E Mitchell, Chunhui Li, Yongxiang Li, and Edward S Buckler. Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences*, 112(12):3823–3828, 2015.

25. Yang Da, Chunkao Wang, Shengwen Wang, and Guo Hu. Mixed model methods for genomic prediction and variance component estimation of additive and dominance effects using snp markers. *PloS one*, 9(1), 2014.

26. Edward S Buckler, James B Holland, Peter J Bradbury, Charlotte B Acharya, Patrick J Brown, Chris Browne, Elhan Ersoz, Sherry Flint-Garcia, Arturo Garcia, Jeffrey C Glaubitz, et al. The genetic architecture of maize flowering time. *Science*, 325(5941):714–718, 2009.

27. Jason A Peiffer, Maria C Romay, Michael A Gore, Sherry A Flint-Garcia, Zhiwu Zhang, Mark J Millard, Candice AC Gardner, Michael D McMullen, James B Holland, Peter J Bradbury, et al. The genetic architecture of maize height. *Genetics*, 196(4):1337–1356, 2014.

28. Henrik Kacser and James A Burns. The molecular basis of dominance. *Genetics*, 97(3-4):639–666, 1981.

29. Nitin Phadnis and James D Fry. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics*, 171(1):385–392, 2005.

30. Zeratsion Abera Desta and Rodomiro Ortiz. Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science*, 19(9):592–601, 2014.

31. Stefan M Edwards, Izel F Sørensen, Pernille Sarup, Trudy FC Mackay, and Peter Sørensen. Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in drosophila melanogaster. *Genetics*, pages 110–116, 2016.

32. Anthony J Greenberg, Sean R Hackett, Lawrence G Harshman, and Andrew G Clark. A hierarchical bayesian model for a novel sparse partial diallel crossing design. *Genetics*, 185(1):361–373, 2010.

33. Zoé Joly-Lopez, Jonathan M Flowers, and Michael D Purugganan. Developing maps of fitness consequences for plant genomes. *Current opinion in plant biology*, 30:101–107, 2016.

34. Jason G Wallace, Peter J Bradbury, Nengyi Zhang, Yves Gibon, Mark Stitt, and Edward S Buckler. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet*, 10(12):e1004845, 2014.

35. Eli Rodgers-Melnick, Daniel L. Vera, Hank W. Bass, and Edward S. Buckler. Open chromatin reveals the functional maize genome. *Proceedings of the National Academy of Sciences*, 113(22):E3177–E3184, 2016.

36. R Shapira, T Levy, S Shaked, E Fridman, and L David. Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. *Heredity*, 113(4):1–11, 2014.

37. George H Shull. The composition of a field of maize. *Journal of Heredity*, 1(1):296–301, 1908.

38. L. T. Gama, M. C. Bressan, E. C. Rodrigues, L. V. Rossato, O. C. Moreira, S. P. Alves, and R. J B Bessa. Heterosis for meat quality and fatty acid profiles in crosses among Bos indicus and Bos taurus finished on pasture or grain. *Meat Science*, 93(1):98–104, 2013.

39. Deborah Charlesworth and John H Willis. The genetics of inbreeding depression. *Nature reviews. Genetics*, 10(11):783–96, November 2009.

40. Antonio Augusto Franco Garcia, Shengchu Wang, Albrecht E Melchinger, and Zhao-Bang Zeng. Quantitative trait loci mapping and the genetic basis of heterosis in maize and rice. *Genetics*, 180(3):1707–1724, 2008.

41. Jinhua Xiao, Jiming Li, Longping Yuan, and Steven D Tanksley. Dominance is the major genetic basis of heterosis in rice as revealed by qtl analysis using molecular markers. *Genetics*, 140(2):745–754, 1995.

42. Li Wang, Ian K Greaves, Michael Groszmann, Li Min Wu, Elizabeth S Dennis, and W James Peacock. Hybrid mimics and hybrid vigor in arabidopsis. *Proceedings of the National Academy of Sciences*, 112(35):E4959–E4967, 2015.

43. Zhensheng Li, Bin Li, and Yiping Tong. The contribution of distant hybridization with decaploid agropyron elongatum to wheat improvement in china. *Journal of Genetics and Genomics*, 35(8):451–456, 2008.

44. Hong Yao, Anjali Dogra Gray, Donald L Auger, and James A Birchler. Genomic dosage effects on heterosis in triploid maize. *Proceedings of the National Academy of Sciences*, 110(7):2665–2669, 2013.

45. Daniel L Halligan and Peter D Keightley. Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics*, 40:151–172, 2009.

46. Xuehui Huang, Shihua Yang, Junyi Gong, Qiang Zhao, Qi Feng, Qilin Zhan, Yan Zhao, Wenjun Li, Benyi Cheng, Junhui Xia, et al. Genomic architecture of heterosis for yield traits in rice. *Nature*, 537(7622):629–633, 2016.

47. Mei Guo, Mary A Rupe, Jun Wei, Chris Winkler, Marymar Goncalves-Butruille, Ben P Weers, Sharon F Cerwick, Jo Ann Dieter, Keith E Duncan, Richard J Howard, et al. Maize argos1 (zar1) transgenic alleles increase hybrid maize yield. *Journal of experimental botany*, 65(1):249–260, 2014.

48. D.S. Falconer and T.F.C. Mackay. *Introduction to Quantitative Genetics*. Longman, 1996.

49. J Doyle and JL Doyle. Genomic plant dna preparation from fresh tissue-ctab method. *Phytochem Bull*, 19(11):11–15, 1987.

50. Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–60, Jul 2009.

51. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–9, Aug 2009.

52. Chunkao Wang, Dzianis Prakapenka, Shengwen Wang, Sujata Pulugurta, Hakizumwami Birali Runesha, and Yang Da. GVCBLUP: a computer package for genomic prediction and variance component estimation of additive and dominance effects. *BMC Bioinformatics*, 15(1):1–9, 2014.

53. Michael Lynch, Bruce Walsh, et al. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA, 1998.

54. Brian L Browning and Sharon R Browning. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*, 84(2):210–23, Feb 2009.

55. David Habier, Rohan L Fernando, Kadir Kizilkaya, and Dorian J Garrick. Extension of the bayesian alphabet for genomic selection. *BMC bioinformatics*, 12(1):186, 2011.

56. Naomi R Wray, Jian Yang, Ben J Hayes, Alkes L Price, Michael E Goddard, and Peter M Visscher. Pitfalls of predicting complex traits from snps. *Nature Reviews Genetics*, 14(7):507–515, 2013.

57. Joost van Heerwaarden, Matthew B Hufford, and Jeffrey Ross-Ibarra. Historical genomics of north american maize. *Proc Natl Acad Sci U S A*, 109(31):12420–5, Jul 2012.

58. Maria C Romay, Mark J Millard, Jeffrey C Glaubitz, Jason A Peiffer, Kelly L Swarts, Terry M Casstevens, Robert J Elshire, Charlotte B Acharya, Sharon E Mitchell, Sherry A Flint-Garcia, Michael D McMullen, James B Holland, Edward S Buckler, and Candice A Gardner. Comprehensive genotyping of the usa national maize inbred seed bank. *Genome Biol*, 14(6):R55, Jun 2013.

59. Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1):1, 2015.

Methods

Plant materials and phenotypic data. We selected 12 maize inbred lines as parents of a partial diallel population. Each parent in a cross was used as either male or female and the resulting seed was bulked (**Figure S1**). Field performance of the 66 F1 hybrids and 12 inbred parents was 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. 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 model:

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

where Y_{ijkl} is the phenotypic value of the l^{th} genotype evaluated in the k^{th} block of the j^{th} replicate within the i^{th} year; μ, the overall mean; ς_i, the fixed effect of the i^{th} year; δ_{ij}, the fixed effect of the j^{th} replicate nested in the i^{th} year; β_{jk}, the random effect of the k^{th} block nested in the j^{th} replicate; α_l, the the fixed genetic effect of the l^{th} individual; $\varsigma_i \cdot \alpha_l$, the interaction effect of the l^{th} individual with the i^{th} year; and ε, the model residuals. We calculated the broad sense heritability (H^2) of traits based on the analysis of all individuals (inbred prarents, hybrid progeny, and checks) assuming 6 degrees of freedom (3 years of evaluation with 3 replicates each year) following the equation of $H^2 = V_G/(V_G + V_E/6)$. In the calculation of V_G, we considered genetic effects as random.

In the models, all fixed effects were significant (Wald test P value < 0.05) for all traits except ASI, for which the effect of replicates within environments was not significant. 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, **Supplementary Figure S1**), though some traits were highly correlated (e.g. Spearman correlation $(r) = 0.98$ for DTS and DTP, **Supplementary Figure S10**).

We estimated best-parent heterosis (BPH) as:

$$BPH_{min,ij} = \hat{G_{ij}} - min(\hat{G_i}, \hat{G_j})$$

$$BPH_{max,ij} = \hat{G_{ij}} - max(\hat{G_i}, \hat{G_j})$$

where $\hat{G_{ij}}$, $\hat{G_i}$ and $\hat{G_j}$ are the genetic values of the hybrid and its two parents i and j. BPH_{max} was used for all traits except ASI, for which we calculated BPH_{min}. General combining ability (GCA) was estimated following Falconer and Mackay [48] and the estimated values can be found in (**Table S3**).

Sequencing and Genotyping. We extracted DNA from the 12 inbred lines following [49] 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.

We trimmed raw sequence reads for adapter contamination with Scythe (`https://github.com/vsbuffalo/scythe`) and for quality and sequence length ($\geq$ 20

nucleotides) with Sickle (`https://github.com/najoshi/sickle`). We mapped filtered reads to the maize B73 reference genome (AGPv2) with bwa-mem [50], keeping reads with mapping quality (MAPQ) 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 [51]. To deal with known issues with paralogy in maize [12], 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. Alignments and genotypes for each of the 12 inbreds are available at iPlant (`/iplant/home/yangjl/pvp_diallel_data/bam_BWA-mem`).

Estimating effect sizes and dominance of GERP-SNPs. We used genomic evolutionary rate profiling (GERP) [20] estimated from a multi-species whole-genome alignment of 13 plant genomes [24]; the alignment and estimated GERP scores are available at iPlant (`/iplant/home/yangjl/pvp_diallel_data/GERPv2`). At each SNP, we assigned the minor allele (as defined by the multi-species alignment) as putatively deleterious.

We estimated the additive and dominant effects of individual GERP-SNPs using a GBLUP model [25] implemented in GVCBLUP [52]:

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

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. 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 ($k = d/a$) [53] for the 107,346 SNPs that each explained greater than the mean genome-wide variance (total variance explained divided by total number of GERP-SNPs).

To compare the effect of random SNPs with GERP-SNPs, we created 10 sets of random SNPs, sampled to match the minor allele frequency distribution (in bins of 10%) and recombination rate (in quartiles) of deleterious SNPs. For each sampled set we fit the above model and obtained the SNP effects and variance explained.

To test the relationship between GERP score and dominance under a simple model of mutation-selection equilibrium, we estimated the selection coefficient s from the relationship between effect size on yield (here treated as fitness) and GERP score. We then assigned the beneficial allele at each SNP a random dominance value in the range of $0 \geq k \geq 1$ and calculated the equilibrium allele frequency p for each SNP under mutation-selection balance using $p = \sqrt{\frac{\mu}{s}}$ for values of $k > 0.98$ and $p = \frac{2\mu}{k+1}$ for $k \leq 0.98$. We then simulated data using binomial sampling to choose SNPs in a sample of size $n = 12$.

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 [54]. We then defined haplotype blocks as contiguous regions within which there were no IBD break points across all pairwise comparisons of the parental lines (**Supplementary Figure S8**). IBD blocks at least 1 Kb in size were kept for further analysis.

Haplotype blocks were weighted by the summed GERP scores of all putatively deleterious (GERP score > 0) SNPs using a custom python script (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 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. The BayesC option from GenSel4 [55] was used for genomic selection model training, using 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, dominant or incomplete dominant models for the ith hybrid in the jth haplotype block.

To conduct prediction, a 5-fold cross-validation method was used, dividing the diallel population randomly into training (80%) and validation sets (20%) 100 times. After model training, prediction accuracies were obtained by comparing the predicted breeding values with the observed phenotypes in the corresponding validation sets. For comparison, GERP scores were permuted using 50k SNP ($\approx$ 100Mb or larger) windows which were circularly shuffled 10 times to estimate a null conservation score for each IBD block. Permutations were conducted 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 and nongenic regions. Cross-validation experiments using the permuted data were conducted on the same training and validation sets.

Then, using 100% of the data, we derived the correlations between breeding values estimated from the above function 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, but the two statistics will converge to the same value when there is no error in SNP/haplotype effect estimation [56].

Finally, to evaluate the utility of our genomic prediction model over a classical model of general combining ability, we compared the following equations:

$$Y_{ij} = \mu + GCA_i + GCA_j + \varepsilon \quad (1)$$

$$Y_{ij} = \mu + GCA_i + GCA_j + G_{ij} + \varepsilon \quad (2)$$

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

1 Supplementary Information

1.1 Genotype Comparisons.

We estimated the allelic error rate using three independent data sets: for all individuals using 41,292 overlapping SNPs from the maize SNP50 bead chip [57]; for all individuals using 180,313 overlapping SNPs identified through genotyping-by-sequencing (GBS) [58]; and for B73 and Mo17 using 10,426,715 SNP from the HapMap2 project [12]. Compared to corresponding SNPs identified by previous studies, a mean genotypic concordance rate of 99.1% was observed.

1.2 Recombination Rates.

We estimated the deleteriousness with respect to the recombination rates (cM/Mb) using GERP > 0 SNPs. We obtained the NAM genetic map from Panzea website (`http://www.panzea.org/`) and divided the recombination rates using the three quantiles (cM/Mb $= 0.15$ at 25%, cM/Mb $= 0.55$ at 50%, and cM/Mb $= 1.74$ at 75%). As a result, the mean GERP score in low recombination regions ($< 25\%$) is significantly higher than in high recombination regions (P value < 0.05 as compare to 25-50%, P value < 0.01 to 50-75%, and P value < 0.01 to $> 75\%$, **Figure 1d**).

1.3 Deleterious Genetic Load and Allele Frequency.

We obtained maize HapMap3.2 [21] and computed allele frequencies across $>$ 1,200 maize lines using PLINK 1.9 [59]. To compare the deleterious genetic load, we extracted the genotypic data of the 23 landrace samples (i.e. the BKN lines) and our elite maize parental lines from the HapMap3.2 data. We computed the number of non-major GERP-SNPs divided by total number of non-missing sites as a measure of deleterious genetic load carried by each line. We computed the fixed and segregating loads for landrace and maize samples seperately.

2 Supporting Tables

Table S1. BLUE values of the seven phenotypic traits. (`https://github.com/yangjl/pvpDiallel/blob/master/doc/STable_trait_matrix.csv`)

Table S2. Levels of heterosis of the seven phenotypic traits. (`https://github.com/yangjl/pvpDiallel/blob/master/doc/STable_heterosis.csv`)

Table S3. General combining ability and specific combining ability of the seven phenotypic traits. (`https://github.com/yangjl/pvpDiallel/blob/master/doc/STable_CA.csv`)

Table S4. Model comparisons P values and AICs. (`https://github.com/yangjl/pvpDiallel/blob/master/doc/STable_model_comp.csv`)

3 Supporting Figures

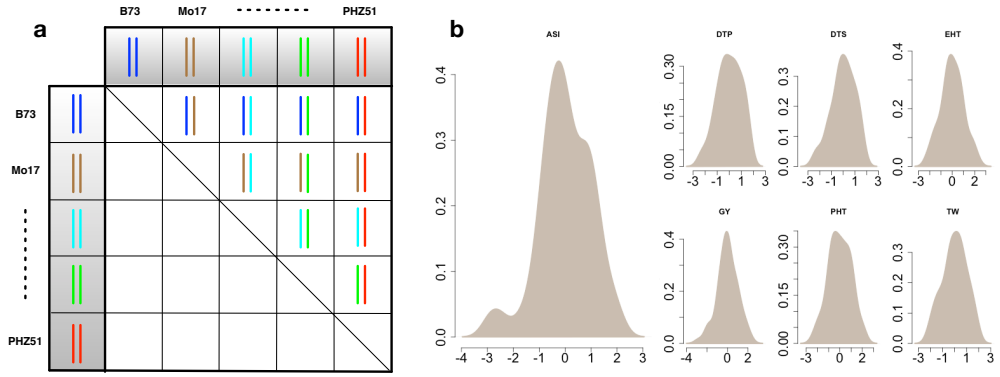


Figure S1. A half-diallel population and distributions of phenotypes. (**a**) Twelve maize inbred lines were selected and crossed in a half-diallel fashion. Each inbred lines was used as both male and female and the resulting F1 seed was bulked. (**b**) Density plots of normalized BLUE values for the 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).

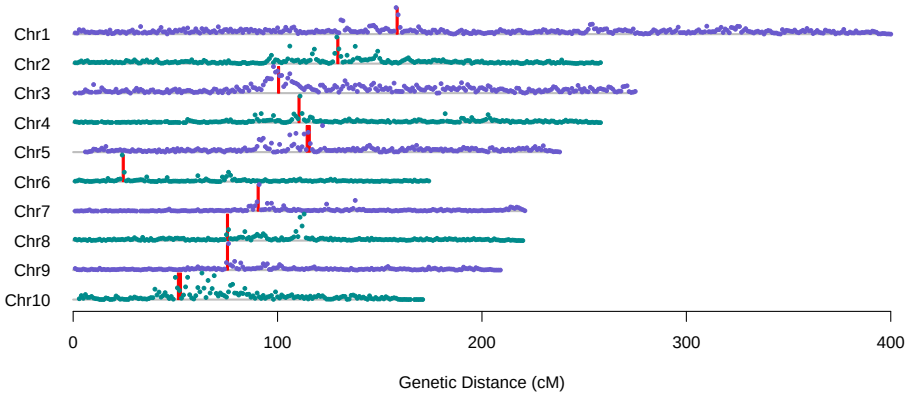


Figure S2. Segregating genetic load across ten maize chromosomes. Dots indicate mean GERP scores of putatively deleterious SNPs (GERP scores > 0) carried by our 12 elite maize lines (bin size = 1 cM). Vertical red lines indicate centromeres.

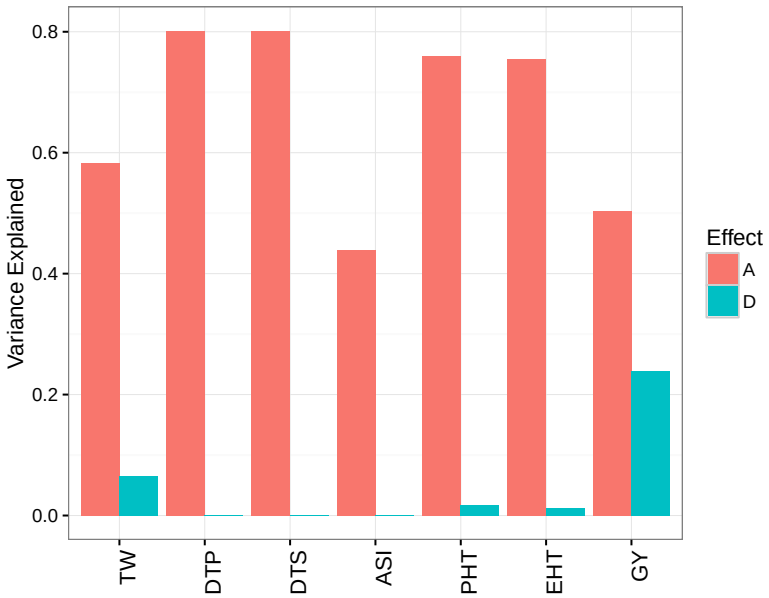


Figure S3. Cumulative variance explained by GERP-SNPs. Additive and dominance effects are indicated by red and blue colors respectively.

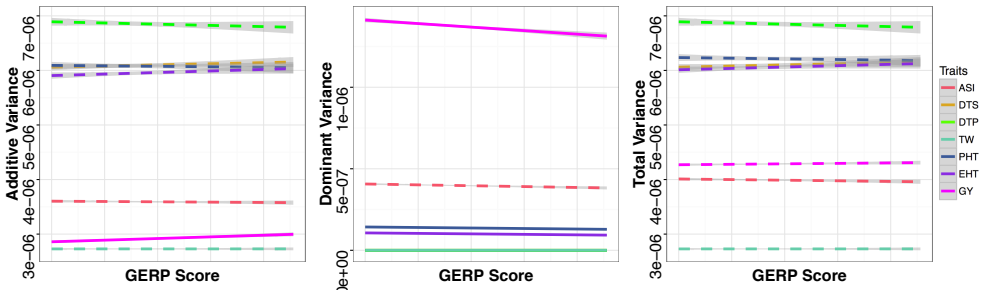


Figure S4. Linear regressions of GERP-SNPs' additive variance, dominance variance and total variance of seven traits *per se* agaginst their GERP scores. Solid and dashed lines represent significant and non-significant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for the 107,346 SNPs with $> 1\times$ of the mean genome-wide variance explained.

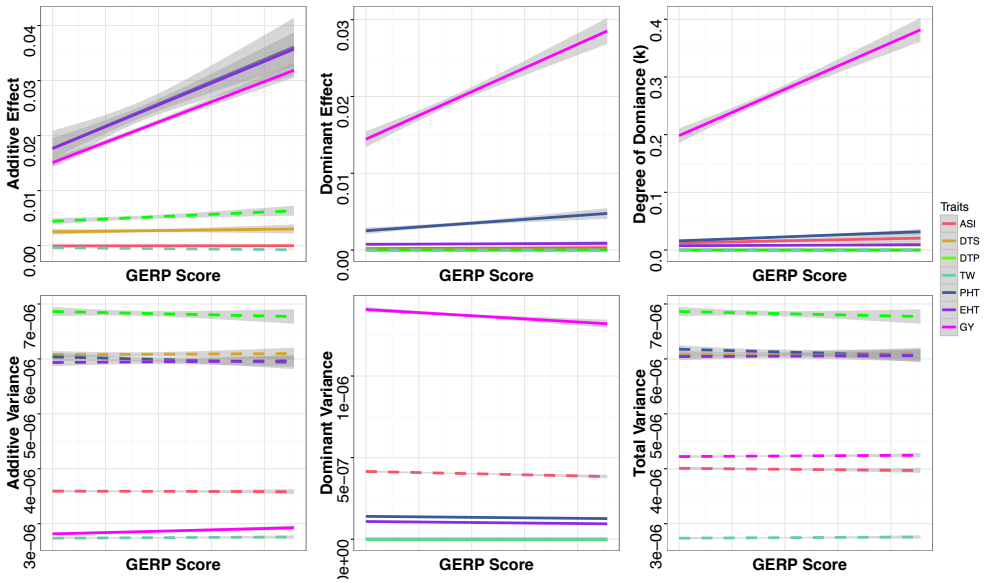


Figure S5. Linear regressions after filtering out GERP-SNPs located in regions in the lowest quartiles of recombination. Solid and dashed lines represent significant and non-significant linear regressions, with grey bands representing 95% confidence intervals. Data are only shown for GERP-SNPs with $> 1\times$ of the mean genome-wide variance explained and with $>$ 1st quantile of the recombination rate (cM/Mb).

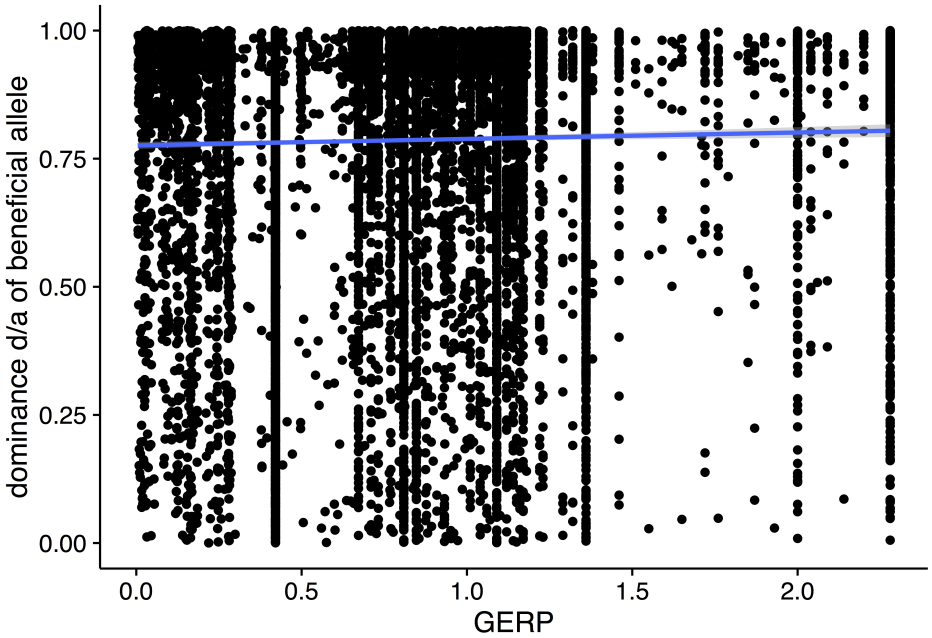


Figure S6. Regression of degree of dominance (k) on GERP scores for simulated data. Solid blue line indicates the regression line fitted to the simulated data (see **Methods** for details).

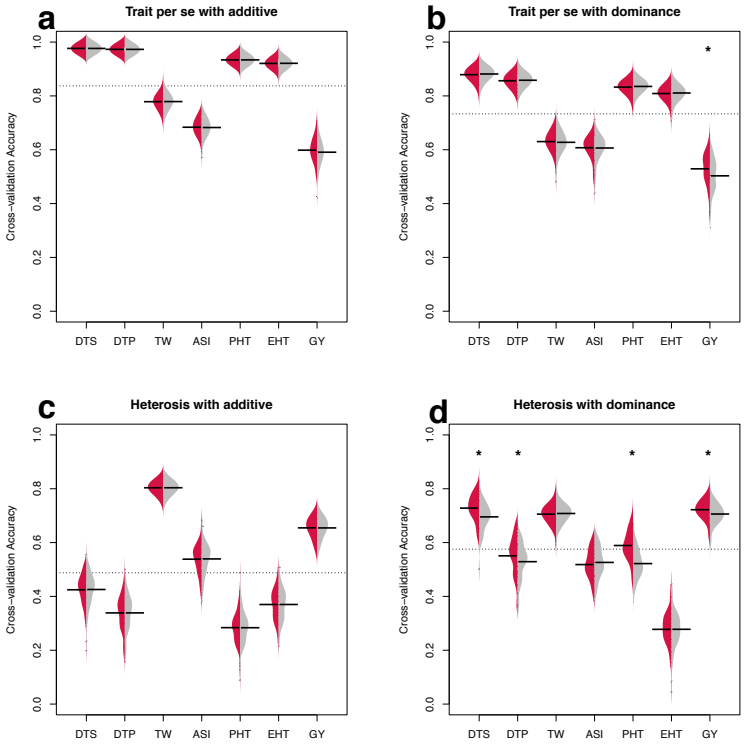
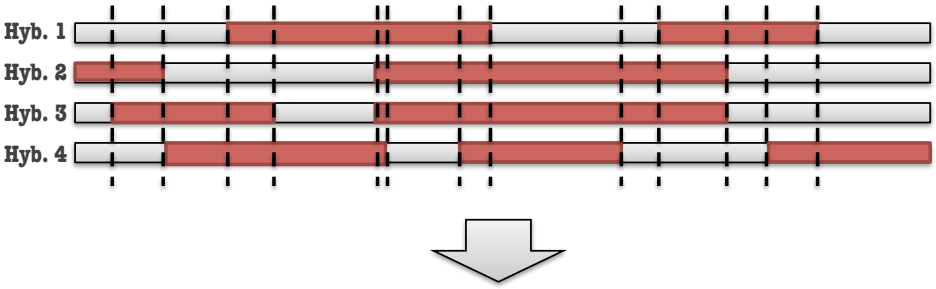


Figure S7. Cross-validation prediction accuracy for trait *per se* and heterosis. Beanplots represent prediction accuracy estimated from cross-validation experiments for traits *per se* (**a, b**) and heterosis (BPH) (**c, d**) under additive (**a, c**) and dominance (**b, d**) models. Prediction accuracy using real data is shown on the left (red) and permutation results on the right (grey). Horizontal bars indicate mean accuracy and the grey dashed lines indicate the overall mean accuracy. Stars indicate significantly (permutation FDR < 0.05) higher than cross-validation accuracy.



	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

Figure S8. 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).

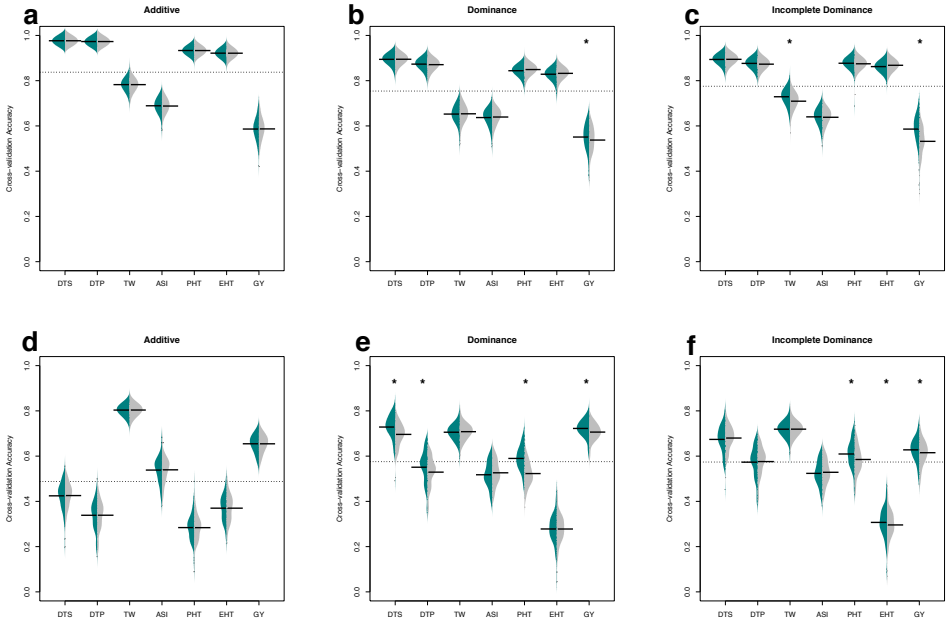


Figure S9. Cross-validation accuracy using GERP-SNPs in genic regions. Beanplots represent prediction accuracy estimated from cross-validation experiments for traits *per se* (**a, b, c**) and heterosis (BPH) (**d, e, f**) under additive (**a, d**), dominance (**b, e**), and incomplete dominance (**c, 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.

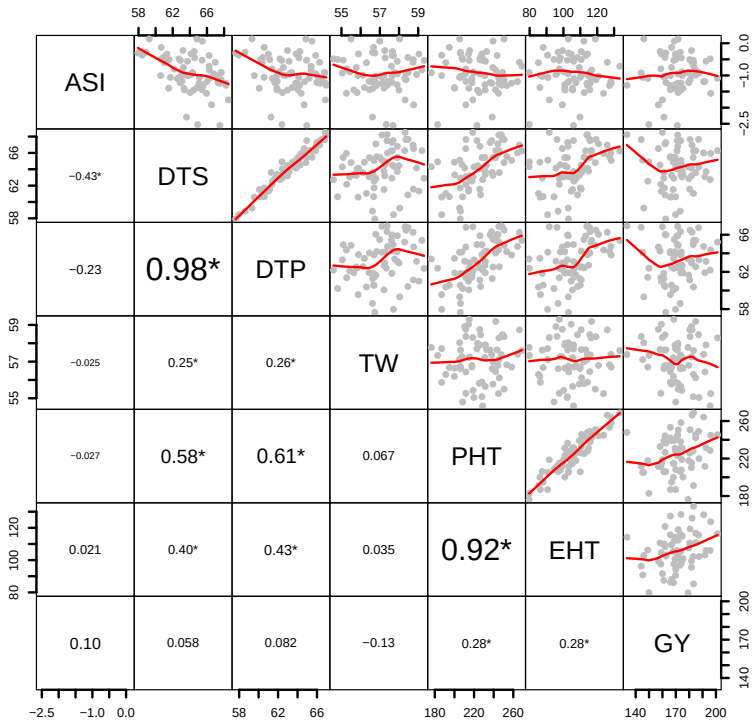


Figure S10. Pairwise correlation plots of seven phenotypic traits. The upper right panels show the scatter plots of all possible pairwise comparisons of two traits. Red line is a fitted smooth curve using "loess" method. In the lower left panels, the numbers are the Spearman correlation coefficients (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).

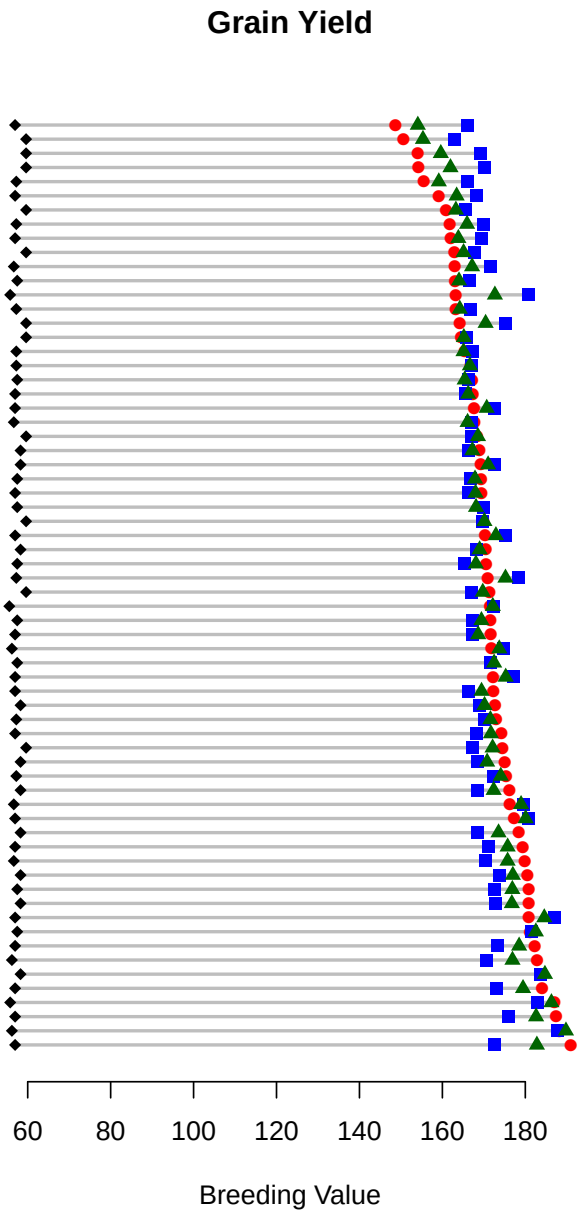


Figure S11. Breeding values of diploid and simulated triploid hybrids. Each line represents the posterior breeding values of a diploid hybrid (red circle), its best parent (black diamond), and predicted breeding values of AAB triploid (blue square) and ABB triploid (green triangle).