Incorporation of Evolutionary Constraint Improves Genomic Prediction of Hybrid Phenotypes

Jinliang Yang[1,2], Sofiane Mezmouk[1,2,3], Andy Baumgarten[4], Edward S. Buckler[5], Katherine E. Guill[6], Michael D. McMullen[6,7], Rita H. Mumm[8], and Jeffrey Ross-Ibarra[1,9,10]

Complementation of deleterious alleles has long been proposed as a major contributor to the hybrid vigor observed in 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. We resequenced the parents of the population, using genomic evolutionary rate profiling (GERP) to identify constrained sites across more than 86 Mb sites across the genome. We identifed haplotype blocks using an identity-by-decent (IBD) analysis and scored these blocks on the basis of segregating putatively deleterious variants. As a result, incorporating sequence conservation improves prediction accuracies in a five-fold cross-validation experiment for several traits *per se* as well as heterosis for those traits. These results provide strong empirical support for the simple complementation model of heterosis, and demonstrates the utility of incorporatin functional annotation and its potential usage in phenotypic prediction and plant breeding.

[1]Department of Plant Sciences, University of California, Davis, CA 95616, USA

[2]These authors contributed equally to this work

[3]Current address: KWS SAAT AG, Grimsehlstr. 31, 37555 Einbeck, Germany

[4]DuPont Pioneer, Johnston, IA 50131, USA

[5]US Department of Agriculture, Agricultural Research Service, Ithaca, NY 14853, USA

[6]US Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA

[7]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA

[8]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[9]Center for Population Biology and Genome Center, University of California, Davis, CA 95616, USA

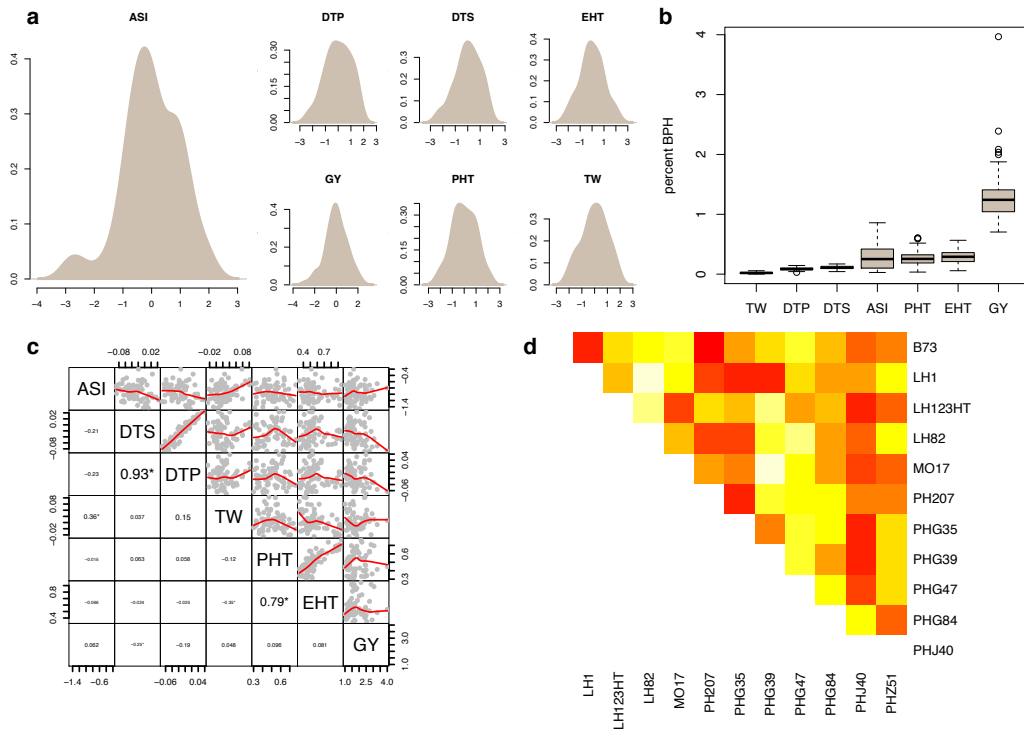[10]Correspondence should be addressed to J.R.-I. (rossibarra@ucdavis.edu).

The phenomenon of heterosis or hybrid vigor has been observed across many species, from yeast[1] to plants[2] and vertebrates[3]. Hybrid vigor is particularly important in agriculture, where hybrid breeding is fundamental to the production of a number of crops including both rice **(CITE)** and maize **(CITE)**. A number of hypotheses have been put forth to explain the phenomenon, including gene dosage[4], overdominance[5–7,12], and epistasis[8,9]. Complementation of recessive deleterious alleles, however, remains the simplest genetic explanation[10], and one that is supported by considerable empirical evidence **(CITE)** [11]. *i'm not sure these are all appropriate. huang2015genomic for example i think doesn't actually measure heterosis, they just do GWAS on hybrids. and frascaroli if I read it right says that heterosis for grain yield is mostly overdominance (dominance > 1).*

RESULTS

Genetic values, heritability and heterosis

A partial diallel population was created using 12 maize inbred lines. Two of them are important public inbreds, B73 and Mo17, the former of which was selected as the maize reference genome[13]. And the other ten are proprietary inbreds (LH1, LH123HT, LH82, PH207, 4676A, PHG39, PHG47, PHG84, PHJ40, and PHZ51) that have expired from Plant Variety Protection Act (PVPA)[14] and represent much of the lineage of key heterotic germplasm pools used in present-day commercial corn hybrids. From this population, phenotypic data were collected for seven traits of interest during 2009-2011: anthesis-silking interval (ASI, in days), days to 50% pollen shed (DTP), days to 50% silking (DTS), ear height (EHT, in cm), grain yield adjusted to 15.5% moisture (GY, in bu/A), plant height (PHT, in cm), and test weight (TW, in pounds).

Best linear unbiased estimators (BLUEs) for genotypes of the seven traits were derived from mixed linear models (**Supplementary Table 1**). In the models, all fixed effects were significant (Wald test P value < 0.05) for all traits ex-

Figure 1 (**a-b**) Phenotypic variability in an elite maize partial diallel. (Left) Density plots of the BLUE values for the seven phenotypic traits. On the x-axis, the phenotypic values were normalized. *does this include hybrids or just parents?* (Right) Boxplot of percent best parent heterosis (pBPH). In the plot, ASI was calculated using pBPHmin and the other six traits were calculated using pBPHmax. *i sort of messed up the formatting of this figure. maybe include as two separate figures? do as a panel? the boxplot should be in the main text though.*

cept ASI, for which the effect of replicates within environments were not significant. As shown in **Figure 1a**, BLUE values were normally distributed (Shapiro-Wilk normality test *P* values > 0.05). Using the parental phenotypic data, we then estimated best-parent heterosis (BPH) for each trait (**Figure 1b**). Because the selected inbred lines are commercially relevant and fairly elite in performance, hybrids in this population exhibit relatively low hybrid vigor (overall mean percent BPH = $0.3\% \pm 0.4\%$) for most traits except GY (mean percent BPH = $95\% \pm 16\%$, Figure 1). Finally, general and specific combining ability (GCA and SCA) were estimated following [15]. GCA and SCA varied among traits (Table **??**), but B73, PHG47 and PHG39 showed the greatest GCA for grain yield.

Sequence variation and evolutionary constraint

All twelve inbreds were sequenced to an average depth of $\sim10X$, resulting in a filtered set of 13.8 million SNPs. We estimated the allelic error rate using three independent data sets: for all individuals using 41,292 overlapping SNPs on the maize SNP50 bead chip[16]; for all individuals using 180,313 overlapping SNPs identified through genotyping by sequencing (GBS)[17]; and for B73 and Mo17 using the 10,426,715 SNP from the HapMap2 project[18]. Compared to correspond-

ing SNPs identified by previous studies, a concordance rate of 99.1% was observed.

The minor allele frequency of SNPs at conserved sites was negatively correlated with GERP score (**Figure 2a**; *P* value $<$ 0.05, $r = -0.8$), consistent with the idea that variants at sites with more positive GERP scores are more deleterious and more strongly impacted by purifying selection.

More than 86 million bp of the maize reference genome were annotated as conserved[19], with GERP scores > 0. Nonetheless, 506,898 (or 0.6%) of these sites were found to segregate among the 12 inbred parents of our diallel (**Figure 2A** and **Supplementary Figure 1**). On average, each inbred parent carried 179,144 putative deleterious variants, ranged from 156,386 (PHG35) to 195,959 (PHG84). Mo17 carried 189,241 putative deleterious variants, ranked as the third place.

We calculated the number of complementation of these putative deleterious variants among F1 hybrids, except ones that have B73 as one of their parents. The best combination (Mo17 and PHG35) complemented each other at 412,042 sites, however, it still carried 94,856 homozygous putative deleterious alleles. The correlation of the complmentation significantly (Pearson correlation test, P-value = 0.02) correlated with the SCA of the GY trait and negative correlated with the cumula-
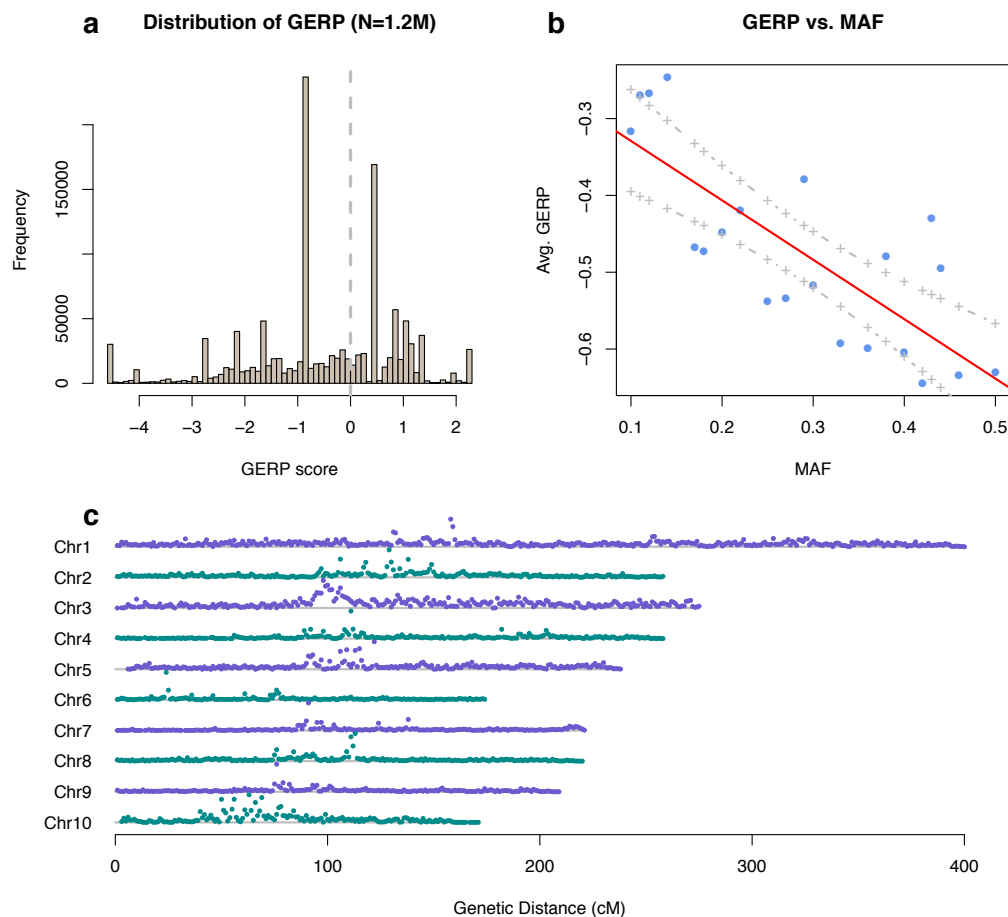
a   **Distribution of GERP (N=1.2M)**   **b**   **GERP vs. MAF**

c

Figure 2 (**a-b**) Distribution of GERP scores and relationship between GERP scores and MAFs. **(A)** Histogram of GERP scores at ∼1.3 million SNPs. **(B)** Plot of average GERP scores in bins (bin size = 0.01) of minor allele frequency (MAF). Red and grey lines define the regression and its 95% confidence interval.

tive deleterious scores (**Figure 2a**), which indicate hybrid combining ability might be determined by the complementation of deleterious alleles. **correlation of the number of deleterious complementation and SCA, MPH and trait per se? Or just because of the genetic distance?**

mean and variation of the deleterious variants per cM We also calculated mean deleterious variants per cM. As shown in (**Figure 2d**), large amount of deleterious variants were accumulated around the centromeric regions.

this should be mentioned here but also in the results. is there any big variation? some lines more than others? is mo17 worse than the pvps? given the comment below i added about linkage, can we calculate mean deleterious variants per cM? that would be sweet to know and to point out that there are tons around the centromere. i think a bit more detail on distribution worthwhile since even Eli's paper is only using GBS. then add here a few sentences comparing to Eli's result and McMullen 2009.

Partition of the phenotypic variance

A mixed linear model approach that partitions genetic values into additive (or allele subsitition effects) and dominant deviations were employed (**X** Guo Hu et al.) to estimate the effects of GERP SNPs. According to Linch (**X**), a degree of dominance (k) was defined as estimated dominant effect divided by estimated additive effect for a given GERP SNP. The value of $k > 1$ indicates the GERP SNP shows an overdominant effect; $0 < k < 1$ indicates a partial dominant effect; $-1 < k < 0$ indicates a partial recessive effect; and $k < -1$ indicates a underdominant effect. Note reference alleles at GERP sites were considered as beneficial alleles and the alternative alleles were deemed as deleterious alleles. Importantly, phenotypic traits with high level of heterosis have an excess of alleles showing dominant and overdominant effects. A correlation test showed that the number of constrast between dominance and overdominance loci and recessive and underdominance loci were signficated correlated with their levels of heterosis (P value < 0.01,
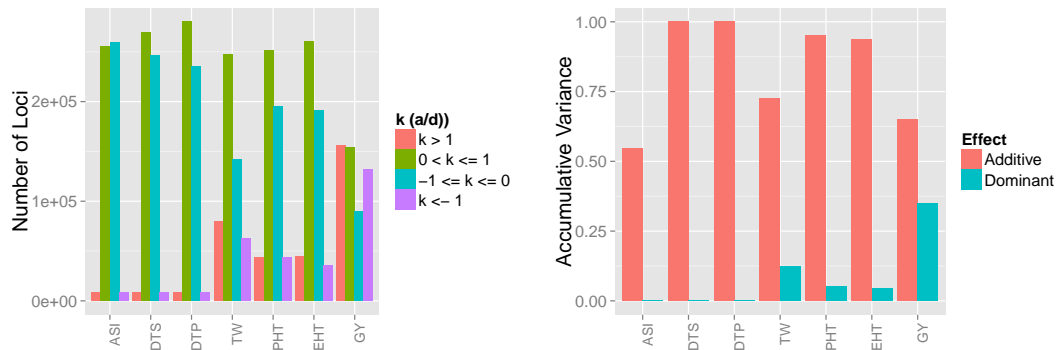
Figure 3 (**a-b**) Modes of inheritance of GERP SNPs and genetic variance partitioned into additive and dominant variance.

Figure 3a). Consistent with this, the accumulative variance explained by dominant effects were positively correlated with the levels of heterosis (P value < 0.01, **Figure 3b**)

Note the reasons that a large number of recessive loci were detected at GERP sites were: 1) the GERP may no the causal SNPs, the effects they accounted for might be the variants in LD with them; and 2) we did not apply any cutoff on effect detection, by chance a small negative dominant effects will be assigned;

SNPs with high GERP scores tend to have larger additive effects, especially for traits showing high levels of heterosis (Figure 4a, supplementary Table **X**). While, dominant effects stay constant across the range of GERP score (Figure 4b). The average dominant effects are larger than for traits (GY, EHT, PHT) showing high level of heterosis, but the average dominant effects are almost =0 for traits showing mediate or no heterosis (ASI, DTS, DTP, TW). The degree of dominance significantly (P value ¡ 0.01) increased as the GERP score increase (Figure 4c).

Phenotypic prediction

the relevant text: ” pnas201221966 2665..2669

However, if complementation of recessive mutations were the sole basis of heterosis, one would predict that the two types of triploid hybrids would be equivalent for their heterotic response because the recessive mutations would be equally complemented. Moreover, the complementation hypothesis would also lead one to predict that heterosis should be basically equivalent between dip- loid and triploid hybrids. Neither of these predictions was met. ” we can't test heterosis of ABB and BBA because we didn't make the AAA and BBB

The small sample size of our diallel and the general low frequency of deleterious SNPs precludes association-based approaches to evaluate the impact of variants on phenotypic variation. To alleviate this limitation, we conceived a haplotype-based genomic selection approach in which we use estimates of evolutionary constraint across the genome to sum the individual effects of deleterious alleles within IBD blocks.

under additive, dominant and incomplete dominant models (see **Methods and Supplementary Figure 2**).

A Bayesian-based statistical method (BayesC)[20] using a 5-fold cross-validation approach was was employed for model training. In general, average prediction accuracies were higher using the additive model (mean $r = 0.81$ and 0.49 for traits *per se* and BPH) than the dominant model (mean $r = 0.70$ and 0.42), and accuracies for heterosis traits were lower than for traits *per se* (Table **??**). A student-t test was employed to compare the mean prediction accuracies between data with real GERP information and data with permutated GERP. To account for multiple traits and multiple transformations, the FDR approach was used to correct the obtained P-values. As a result, incorporating evolutionary constraint information improved prediction accuracy for ASI and PHT *per se* under an additive model and for ASI under a dominant model (FDR < 0.05, Figure 6 A and B). GERP scores also improved prediction accuracies of heterosis (BPH) for GY under the additive model and DTP, DTS and TW under the dominant model (FDR ¡ 0.05, Figure C and D). To rule out the possible confounding of high GERP scores and genic annotations, we re-permuted the data using only deleterious (GERP > 0) genic SNPs. Though this resulted in fewer SNPs (316, 983), the model prediction accuracies remained significantly improved for GY *per se* under the additive model and for BPH of GY and PHT under the additive model (Figure and *Supplementary Table*).

Posterior phenotypic variance explained and model comparisons

To learn why the prediction performace varied among traits *per se* and heterosis, we obtained the posterior variance explained by our models using the complete set of data. As shown in Figure 7, additive models explained more phenotypic variance for traits *per se* of DTP, DTS, EHT and PHT; but explained less phenotypic variance for heterosis (BPH) of ASI, GY and TW. In contrast, a larger proportion of the phenotypic variance could be explained by the dominant models for heterosis (BPH) of ASI, GY and TW. *is this correct? the figure looks to me to*
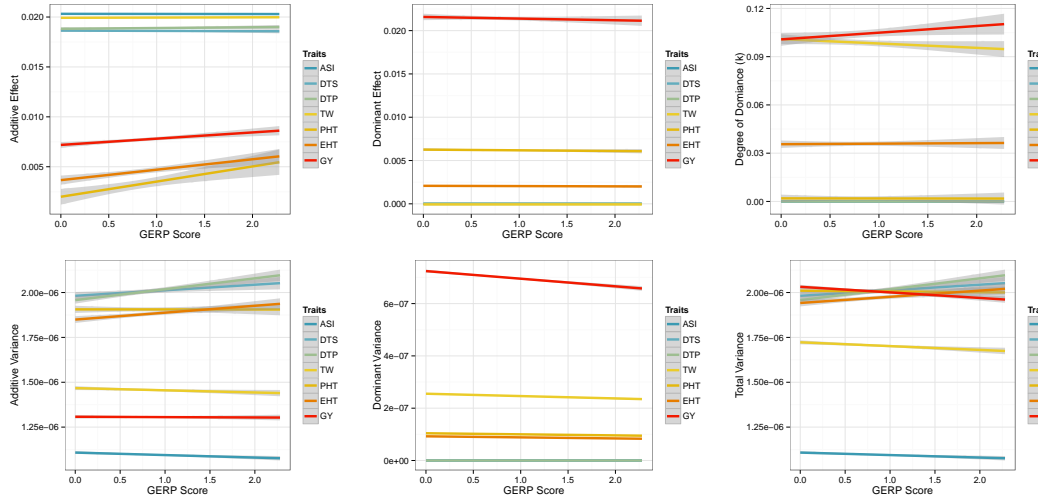
Figure 4 (**a-f**) Relationship of GERP score with different modes of inheritance and variance. GERP score of the putative deleterious SNPs were fitted using a linear model with their additive effects (a), dominant effects (b), degree of dominace (c), additive variance (d), dominant variance (e) and total variance (f) for seven phenotypic traits, respectively. Grey color indicates the 95% confidence interval.

disagree with this statement. This difference was particularly striking for grain yield under the dominant model, where only 3% of the variance in trait *per se* could be explained but 61% of the variance in BPH was explained. *I think a sentence or two here connecting variance explained with prediction accuracy would be helpful*

Heterosis transformations are largely determined by the accuracies of the parental phenotypes. To control for uncertainty of parental phenotypes, we estimated combining ability *do we need to explain what this is and the difference between SCA and GCA?* directly from the hybrid population itself. We extracted the breeding values estimated under both additive and dominant models using our haplotype blocks and incorporating GERP scores. We then applied the following models:

$$Y_{ij} = \mu + GCA_i + GCA_j + \varepsilon \tag{1}$$

$$Y_{ij} = \mu + GCA_i + GCA_j + G_{ij} + \varepsilon \tag{2}$$

$$Y_{ij} = \mu + GCA_i + GCA_j + SCA_{ij} + \varepsilon \tag{3}$$

$$Y_{ij} = \mu + GCA_i + GCA_j + SCA_{ij} + G_{ij} + \varepsilon \tag{4}$$

where Y_{ij} is the BLUE value of the hybrid crossed between the ith inbred and j^{th} inbred; μ, the overall mean; GCA_i, the general combining ability of the ith inbred; GCA_j, the general combining ability of the jth inbred; SCA_{ij}, the specific combining ability of between the ith and j^{th} inbreds; G_{ij}, breeding values estimated by our GS model for hybrid crossed between the ith inbred and j^{th} inbred; ε, the model residuals.

Consistent with the previous analysis, haplotype blocks coded with the dominant mode of inheritance significantly improved the fit of models for heterosis for ASI and GY (equation 1 vs. equation 2, ANOVA P value < 0.05, Table **??**). *i'm having trouble following here. GY gives a p of 0.04 which is*

not significant after bonferonni (maybe after FDR?) under the dominant model, but DOES give a significant p¡4E-16 under the additive model. why not mention/discuss that? Comparison of models 3 and 4, however, show no real difference (ANOVA P value = **X**), indicating that specific combining ability captures most of the parental interactions and the our haplotype blocks are unable to detect higher order interactions. *there are also at least three tests that are significant after multiple correction for 4 vs 3. is that not meaningful?*

DISCUSSION

In this study we have identified more than 500,000 evolutionary conserved (GERP > 0) sites in the genome segregating for putatively deleterious alleles in a panel of elite maize lines. The non-reference alleles at these SNPs are found at low frequency, consistent with previous observations[19,21] and the role of selection preventing such alleles from reaching high frequency. Nonetheless, each inbred line carries a large number of deleterious variants, averaging $\sim 200,000$ per line. *this should be mentioned here but also in the results. is there any big variation? some lines more than others? is mo17 worse than the pvps? given the comment below i added about linkage, can we calculate mean deleterious variants per cM? that would be sweet to know and to point out that there are tons around the centromere. i think a bit more detail on distribution worthwhile since even Eli's paper is only using GBS. then add here a few sentences comparing to Eli's result and McMullen 2009.*

Across lines, however, the majority of these deleterious mutations were maintained at low frequency, consistent with previous observations[19].

The large number of linked deleterious alleles present means that there is likely insufficient recombination in standard breed-
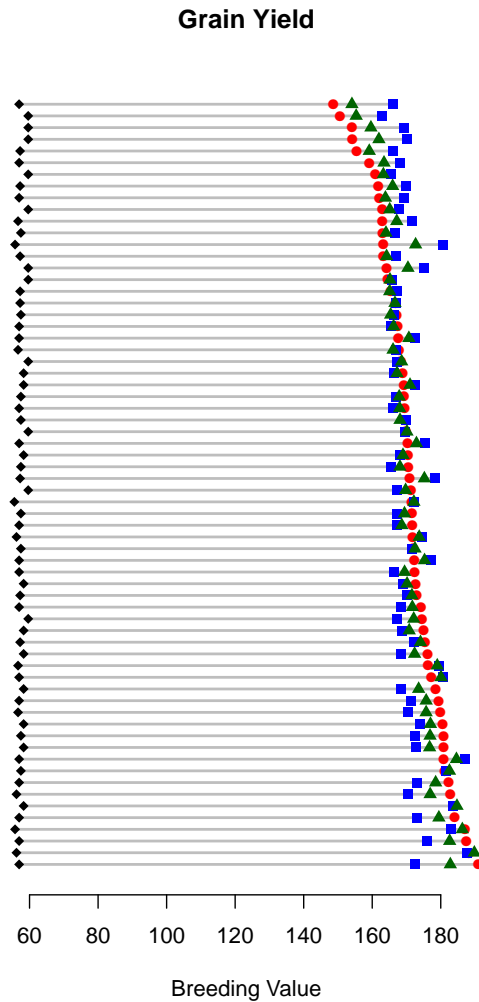
Grain Yield

Figure 5 Beanplots of cross-validation accuracies using SNPs with positive GERP scores. Cross-validation experiments were conducted using selected SNPs and permuted data for traits *per se* (**A, B**) and BPH (**C, D**) under additive (**A and C**) and dominant (**B and D**) models. Accuraries from the real data were plotted on the left (blue) and permutation results on the right (grey). Horizotal bars indicate mean accuracies for each trait and the grey dashed lines indicate the overall mean accuracy. Stars indicate significantly (FDR ¡ 0.05) higher cross-validation accuracies. *need to explain red and black points*

Figure 6 Beanplots of cross-validation accuracies using SNPs with positive GERP scores. Cross-validation experiments were conducted using selected SNPs and permuted data for traits *per se* (**A, B**) and BPH (**C, D**) under additive (**A and C**) and dominant (**B and D**) models. Accuraries from the real data were plotted on the left (blue) and permutation results on the right (grey). Horizotal bars indicate mean accuracies for each trait and the grey dashed lines indicate the overall mean accuracy. Stars indicate significantly (FDR ¡ 0.05) higher cross-validation accuracies. *need to explain red and black points*

Figure 7 Posterior phenotypic variance explained by deleterious genic SNPs in IBD blocks using additive and dominant models. Dark color indicates trait *per se* and grey color indicates BPH.

ing programs to completely purge all such alleles. Instead, breeders have devised a strategy — hybrid breeding — that circumvents much of this problem via complementation. Consistent with this idea, our results show that prediction accuracies for both traits *per se* as well as heterosis increased when SNPs were weighted by their likelihood of being deleterious. Because there are likely thousands of deleterious alleles involved in complementation, many with relatively small effects,
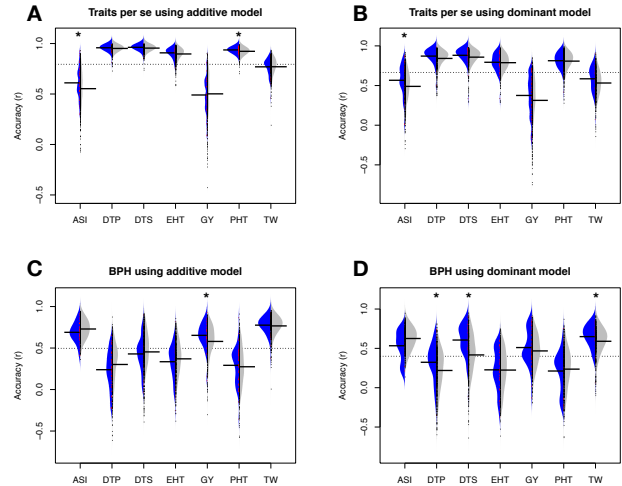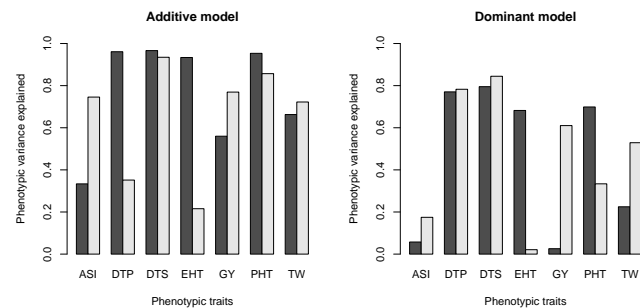
traditional GWAS approaches with genome-wide thresholds of multiple testing may not have the power to detect such effects. Using a liberal significance threshold, however, even GWAS methods have identified an enrichment for deleterious genic SNPs among associated markers for a number of traits[21].

Our models did not increase the prediction accuracies equally well for traits *per se* and their heterosis transformations. This is not surprising, given the variation in genetic architecture

of different phenotypic traits — flowering time, for example, appears primarily determined by many loci of small additive effect[22], and prediction accuracy is highest and the vast majority of phenotypic variance can be explained under simple additive models. *add other example from our data and maize lit. add sentence or two of correlation between prediction accuracies and BPH seen in boxplots – traits with higher BPH have lower heritability and thus harder to predict normally. also add back text on how difference in prediction accuracies can be explained by heritability. why use broad sense?*

Here a paragraph about how additive often beat recessive and that this is consistent with a role for partial dominance. we cite Birchler arguing how complete dominance is unlikely and doesn't jive with much data. we cite Huang showing partial dominance for phenotypes per se among hybrids. can cite[23] reviewing several studies finding evidence of partial dominance. while we don't have enough power to estimate dominance, the fact that an additive model works well suggests high-ish values of h. this would be consistent with, for example, alleles affecting plant height being predominantly additive (cite[24]), but plant height is correlated with yield, so alleles affecting height will have a weak, but additive, effect on yield.

Discuss limitations of size of population, missing GERP data b/c of alignment with Tripsacum, missing information on non-SNP variants, etc.

As genotyping costs continue to decline, genomic prediction models are increasing in popularity[25]. Most previous work on genomic prediction, however, focuses exclusively on statistical properties of the models, ignoring potentially useful biological information (but see **(CITE)** for a recent example). In addition to providing evidence in favor of the simple complementation model for heterosis, our work here shows the utility of incorporating functional information about the variants used in genomic prediction models. As our functional annotations of genomes improve, we predict that including such information in genomic prediction will be vital to the development of more powerful predictive models for plant breeding.

METHODS

Methods and any associated references are available in a separate pdf file.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

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

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

1. 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* **113**, 1–11 (2014).

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

3. Gama, L. T. *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* **93**, 98–104 (2013).

4. Birchler, J. A., Auger, D. L. & Riddle, N. C. In search of the molecular basis of heterosis. *The Plant Cell* **15**, 2236–2239 (2003).

5. East, E. M. Heterosis. *Genetics* **21**, 375 (1936).

6. Schwartz, D. Single gene heterosis for alcohol dehydrogenase in maize: the nature of the subunit interaction. *Theoretical and Applied Genetics* **43**, 117–120 (1973).

7. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene single flower truss drives heterosis for yield in tomato. *Nature genetics* **42**, 459–463 (2010).

8. Minvielle, F. Dominance is not necessary for heterosis: a two-locus model. *Genetical research* **49**, 245–247 (1987).

9. Schnell, F. & Cockerham, C. Multiplicative vs. arbitrary gene action in heterosis. *Genetics* **131**, 461–469 (1992).

10. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nature reviews. Genetics* **10**, 783–96 (2009).

11. Xiao, J., Li, J., Yuan, L. & Tanksley, S. D. Dominance is the major genetic basis of heterosis in rice as revealed by qtl analysis using molecular markers. *Genetics* **140**, 745–754 (1995).

12. Frascaroli, E. *et al.* Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. *Genetics* **176**, 625–644 (2007).

13. Schnable, P. S. *et al.* The b73 maize genome: complexity, diversity, and dynamics. *science* **326**, 1112–1115 (2009).

14. Nelson, P. T. *et al.* Molecular characterization of maize inbreds with expired us plant variety protection. *Crop science* **48**, 1673–1685 (2008).

15. Falconer, D. & Mackay, T. *Introduction to Quantitative Genetics*. (, 1996).

16. van Heerwaarden, J., Hufford, M. B. & Ross-Ibarra, J. Historical genomics of north american maize. *Proc Natl Acad Sci U S A* **109**, 12420–5 (2012).

17. Romay, M. C. *et al.* Comprehensive genotyping of the usa national maize inbred seed bank. *Genome Biol* **14**, R55 (2013).

18. Chia, J.-M. *et al.* Maize hapmap2 identifies extant variation from a genome in flux. *Nat Genet* **44**, 803–7 (2012).

19. Rodgers-Melnick, E. *et al.* Recombination in diverse maize is stable, predictable, and associated with genetic load. *Proceedings of the National Academy of Sciences* **112**, 3823–3828 (2015).

20. Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC bioinformatics* **12**, 186 (2011).

21. Mezmouk, S. & Ross-Ibarra, J. The pattern and distribution of deleterious mutations in maize. *G3 (Bethesda, Md.)* **4**, 163–71 (2014).

22. Buckler, E. S. *et al.* The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).

23. Halligan, D. L. & Keightley, P. D. Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution, and Systematics* **40**, 151–172 (2009).

24. Peiffer, J. A. *et al.* The genetic architecture of maize height. *Genetics* **196**, 1337–1356 (2014).

25. Desta, Z. A. & Ortiz, R. Genomic selection: genome-wide prediction in plant improvement. *Trends in plant science* **19**, 592–601 (2014).