Incorporation of evolutionary constraint improves genomic prediction of hybrid phenotype

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

[*]Department of Plant Sciences, University of California, Davis, CA 95616, USA, [§]DuPont Pioneer, Johnston, IA 50131, USA, [†]US Department of Agriculture, Agricultural Research Service, Ithaca, NY 14853, USA, [‡]US Department of Agriculture, Agricultural Research Service, Columbia, MO 65211, USA, [**]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA, [§§]Department of Crop Sciences, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, [††]Center for Population Biology and Genome Center, University of California, Davis, CA 95616, USA

ABSTRACT 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 of the genome. We identifed haplotype blocks using an identity-by-decent analysis and scored these blocks on the basis of segregating putatively deleterious variants. 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.

KEYWORDS heterosis; deleterious; genomic selection; diallel; GERP; maize

The phenomenon of heterosis or hybrid vigor has been observed across many species, from yeast (Shapira *et al.* 2014) to plants (Shull 1908) and vertebrates (Gama *et al.* 2013). 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 (Birchler *et al.* 2003), overdominance (East 1936; Schwartz 1973; Krieger *et al.* 2010), and epistasis (Minvielle 1987; Schnell and Cockerham 1992). Complementation of recessive deleterious alleles, however, remains the simplest genetic explanation (Charlesworth and Willis 2009), and one that is supported by considerable empirical evidence (Xiao *et al.* 1995; Frascaroli *et al.* 2007). *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).*

Copyright © 2015 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Tuesday 7th July, 2015%
[1]Corresponding authors: Department of Plant Sciences, University of California, Davis, CA 95616, USA. Email: rossibarra@ucdavis.edu and jolyang@ucdavis.edu
[2]These authors contributed equally to this work
[3]Current address: KWS SAAT AG, Grimsehlstr. 31, 37555 Einbeck, Germany

Materials and Methods

Plant materials and phenotypic data

We selected 12 maize inbred lines, broadly representative of corn belt maize germplasm (Mikel and Dudley 2006), as parents of a partial diallel population. Each parent in a cross was used as both male and female and the resulting seed was bulked (Figure S1). We evaluated the 66 F1 hybrids, 12 inbred parents and two current commercial check hybrids in the field in Urbana, IL over three years (2009-2011) in an incomplete block design with three replicates each year. Plots consisted of four rows, 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), ear height (EHT, in cm), days to 50% silking (DTS), days to 50% pollen shed (DTP),

anthesis-silking interval (ASI, in days), grain yield adjusted to 15.5% moisture (adj GY, in bu/A), and test weight (TW, in pounds). Overall mean phenotypic values for each cross can be found at Table S1.

We estimated Best Linear Unbiased Estimates (BLUEs) of the genetic effects in ASReml-R (Gilmour *et al.* 2009) 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; ε, the model residuals.

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_{min} was used instead of BPH_{max} for ASI.

Sequencing and Genotyping

We extracted DNA from the 12 inbred lines following Doyle and Doyle (1987) and sheared the DNA on a Covaris (Woburn, Massachusetts) for library preparation. Libraries were prepared using an Illumina paired end libaray protocol with 180 bp fragments. Libraries were then sequenced at Cornell.

We trimmed raw sequence reads for adapter contamination with Scythe (https://github.com/vsbuffalo/scythe) and for quality *Sofiane: what qual score? do we need to say anything about overlapping reads?* 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 (Li and Durbin 2009), 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 the samtools utilities (Li *et al.* 2009). To deal with known issues with paralogy in maize (Chia *et al.* 2012), SNPs were filtered to be heterozygote in less than 3 inbred lines, have a mean minor allele depth of at least 4, have a mean depth over all individuals lower than 30 and have missing/heterozygote alleles in fewer than 6 inbred lines.

We used the fastIBD method implemented in BEAGLE (Browning and Browning 2009) to impute missing data and identify regions of identity by descent (IBD) between the 12 inbred lines. 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 S2). IBD blocks at least 1 Kb in size were kept for further analysis.

Genomic selection using IBD blocks incorporated with GERP scores

We used genome-wide estimates of evolutionary constraint (GERP Davydov *et al.* 2010) estimated by Rodgers-Melnick *et al.* (2015). Haplotype blocks were weighted by the summed GERP scores of all deleterious (GERP score $>$ 0) SNPs; blocks with no deleterious SNPs were excluded from further analysis. This estimation was calculated under both additive and dominant modes of inheritance using a custom python script available at (https://github.com/yangjl/zmSNPtools). For a particular SNP with a GERP score g, the non-reference homozygote was assigned a value of $2g$, the heterozygote a value of g, and the reference homozygote a value of 0. Under the dominant model, both the heterozygote and the non-reference homozygote were assigned a value of g, with the reference homozygote again assigned a value of 0. To conduct prediction, a 5-fold cross-validation method was used, dividing the diallel population randomly into training (80%) and validation sets (20%) 10 times. The BayesC option from GenSel4 (Habier *et al.* 2011) was used for model training, using 41,000 iterations and removing the first 1,000 as burn-in. 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 ($>$ 100Mb) windows which were circularly shuffled 10 times to estimate a null conservation score for each IBD blocks. Cross-validation experiments using the permuted data were conducted on the same training and validation sets.

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. And the other ten of proprietary inbreds (LH1, LH123HT, LH82, PH207, 4676A, PHG39, PHG47, PHG84, PHJ40, and PHZ51) that have expired from Plant Variety Protection (PVP) 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 (Table S1). 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 were not significant. As shown in Figure 1, BLUE values were normally distributed (normality test P values $>$ 0.05). Broad sense heritability for these traits ranged from 0.65 for ASI to 0.95 for PHT. Using the parental phenotypic data, we then estimated best-parent heterosis (BPH) for each trait. 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 (Falconer and Mackay 1996). GCA and SCA varied among traits (Table S2), 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 $\sim$10X, 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 (van Heerwaarden *et al.* 2012); for all individuals using 180,313 overlapping SNPs identified through genotyping

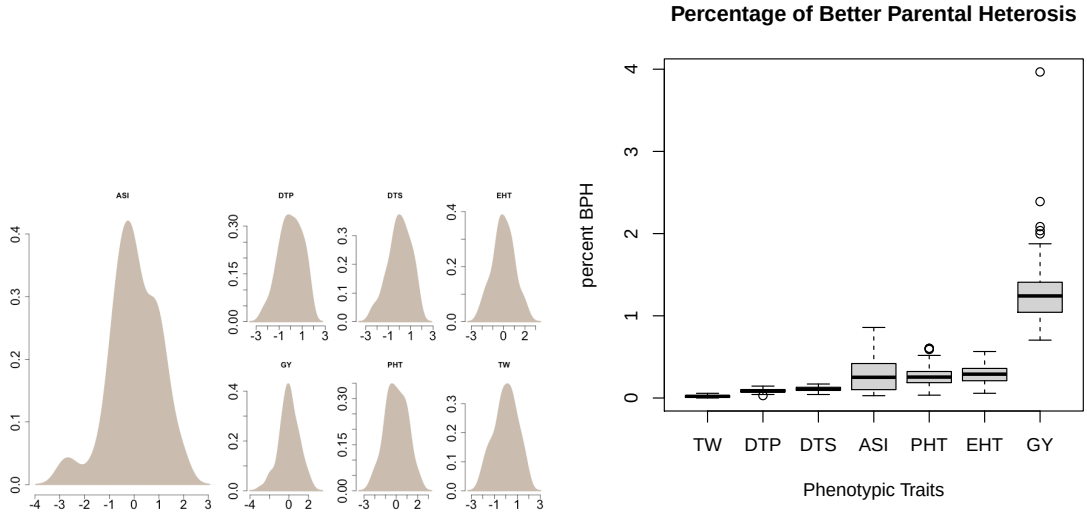


Figure 1 Phenotypic variability in an elite maize partial diallel. (Left) Density plots of the BLUE values for the seven phenotypic traits. On the x-axis, the phenotypic values were normalized. *does this include hybrids or just parents?* (Right) Boxplot of percent best parent heterosis (pBPH). In the plot, ASI was calculated using pBPHmin and the other six traits were calculated using pBPHmax.

by sequencing (GBS) (Romay *et al.* 2013); and for B73 and Mo17 using the 10,426,715 SNP from the HapMap2 project (Chia *et al.* 2012). Compared to corresponding SNPs identified by previous studies, a concordance rate of 99.1% was observed. *Sofiane: can we separate those numbers out by study? or just report for one study and mention that similar rates were seen in other studies? either way it would be nice to know what rate went with what data. also is concordance mean identical genotype? do we have minor allele rate (which is a bit more informative)? if not, skip it.*

More than 86 million bp of the genome were annotated as conserved, with GERP scores > 0. Nonetheless, 506,898 of these sites were found to segregate among the 12 inbred parents of our diallel (Figure 2A and S3). The minor allele frequency of SNPs at conserved sites was negatively correlated with GERP score (Figure 2B; 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.

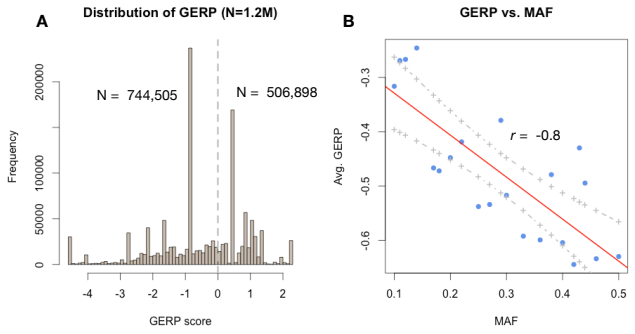


Figure 2 Distribution of GERP scores and relationship between GERP scores and MAFs. **(A)** Histogram of GERP scores at ∼1.3 million SNPs. **(B)** Plot of average GERP scores in bins (bin size = 0.01) of minor allele frequency (MAF). Red and grey lines define the regression and its 95% confidence interval.

Phenotypic prediction

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 (Rodgers-Melnick *et al.* 2015) to sum the individual effects of deleterious alleles within IBD blocks under both an additive and dominant model (see Methods and Figure S4).

A Bayesian-based statistical method (BayesC) (Habier *et al.* 2011) 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 S3). 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 3 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 3 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 S5 and Table S4).

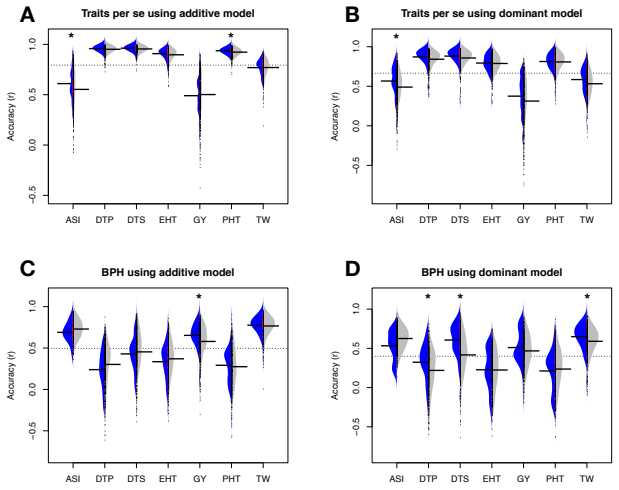


Figure 3 Beanplots of cross-validation accuracies using SNPs with positive GERP scores. Cross-validation experiments were conducted using selected SNPs and permuted data for traits *per se* (**A, B**) and BPH (**C, D**) under additive (**A and C**) and dominant (**B and D**) models. Accuraries from the 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*

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 4, 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 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 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:

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

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

$$Y_{ij} = \mu + GCA_i + GCA_j + SCA_{ij} + \varepsilon \quad (3)$$

$$Y_{ij} = \mu + GCA_i + GCA_j + SCA_{ij} + G_{ij} + \varepsilon \quad (4)$$

where Y_{ij} is the BLUE value of the hybrid crossed between the i^{th} inbred and j^{th} inbred; μ, the overall mean; GCA_i, the general combining ability of the i^{th} inbred; GCA_j, the general combining ability of the j^{th} inbred; SCA_{ij}, the specific combining ability of between the i^{th} and j^{th} inbreds; G_{ij}, breeding values estimated by our GS model for hybrid crossed between the i^{th} 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 S5). *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?*

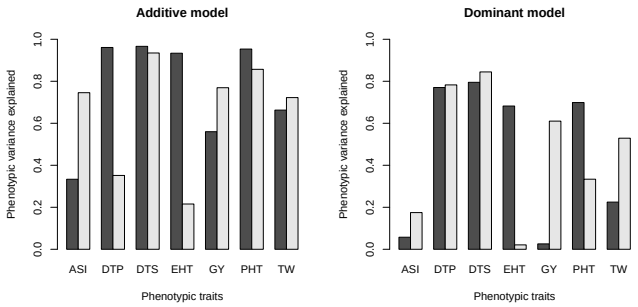


Figure 4 Posterior phenotypic variance explained by deleterious genic SNPs in IBD blocks using additive and dominant models. Dark color indicates trait *per se* and grey color indicates BPH.

Discussion

In this study we have identified more than 500,000 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 (Mezmouk and Ross-Ibarra 2014; Rodgers-Melnick *et al.* 2015) and the role of selection preventing such alleles from reaching high frequency. Nonetheless, each inbred line carries a large number of deleterious variants, averaging $\sim 100,000$ 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.*

The large number of linked deleterious alleles present means that there is likely insufficient recombination in standard breeding 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 (Mezmouk and Ross-Ibarra 2014).

Our models did not increase the prediction accuracies equally well for traits *per se* and their heterosis transformations. This is not surprising, however, given the variation in genetic architecture of different phenotypic traits — flowering time, for example, appears primarily determined by many loci of small additive effect (Buckler *et al.* 2009), and prediction accuracy is highest and the vast majority of phenotypic variance can be explained under simple additive models. Traits while **X** *better to use another maize example if we can rather than rice...* . Consistent with this, we observed that the additive model increased prediction accuracies for ASI as a trait *per se* and *ideally above use a phenotype we can compare in this way... dominant model increased the prediction accuracy for trait* **X**.

As genotyping costs continue to decline, genomic prediction models are increasing in popularity (Desta and Ortiz 2014). 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.

Acknowledgements

Financial support for this work came from NSF (grants IOS-0820619 and IOS-1238014), USDA (grants 2009-65300-05668 and **X**), DuPont Pioneer, and N2 Genetics. We'd like to thank Graham Coop, James Holland, **X**, and **X** reviewers for helpful discussion.

Literature Cited

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

Browning, B. L. and S. R. Browning, 2009 A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet **84**: 210–23.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J. C. Glaubitz, *et al.*, 2009 The genetic architecture of maize flowering time. Science **325**: 714–718.

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

Chia, J.-M., C. Song, P. J. Bradbury, D. Costich, N. de Leon, J. Doebley, R. J. Elshire, B. Gaut, L. Geller, J. C. Glaubitz, M. Gore, K. E. Guill, J. Holland, M. B. Hufford, J. Lai, M. Li, X. Liu, Y. Lu, R. McCombie, R. Nelson, J. Poland, B. M. Prasanna, T. Pyhäjärvi, T. Rong, R. S. Sekhon, Q. Sun, M. I. Tenaillon, F. Tian, J. Wang, X. Xu, Z. Zhang, S. M. Kaeppler, J. Ross-Ibarra, M. D. McMullen, E. S. Buckler, G. Zhang, Y. Xu, and D. Ware, 2012 Maize hapmap2 identifies extant variation from a genome in flux. Nat Genet **44**: 803–7.

Davydov, E. V., D. L. Goode, M. Sirota, G. M. Cooper, A. Sidow, and S. Batzoglou, 2010 Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS computational biology **6**: e1001025.

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

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

East, E. M., 1936 Heterosis. Genetics **21**: 375.

Falconer, D. and T. Mackay, 1996 *Introduction to Quantitative Genetics*. Longman.

Frascaroli, E., M. A. Cane, P. Landi, G. Pea, L. Gianfranceschi, M. Villa, M. Morgante, and M. E. Pè, 2007 Classical genetic and quantitative trait loci analyses of heterosis in a maize hybrid between two elite inbred lines. Genetics **176**: 625–644.

Gama, L. T., M. C. Bressan, E. C. Rodrigues, L. V. Rossato, O. C. Moreira, S. P. Alves, and R. J. B. Bessa, 2013 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.

Gilmour, A. R., B. Gogel, B. Cullis, and R. Thompson, 2009 Asreml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK .

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

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

Li, H. and R. Durbin, 2009 Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics **25**: 1754–60.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup, 2009 The sequence alignment/map format and samtools. Bioinformatics **25**: 2078–9.

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

Mikel, M. A. and J. W. Dudley, 2006 Evolution of north american dent corn from public to proprietary germplasm. Crop science **46**: 1193–1205.

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

Rodgers-Melnick, E., P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler, 2015 Recombination in diverse maize is stable, predictable, and associated with genetic load. Proceedings of the National Academy of Sciences **112**: 3823–3828.

Romay, M. C., M. J. Millard, J. C. Glaubitz, J. A. Peiffer, K. L. Swarts, T. M. Casstevens, R. J. Elshire, C. B. Acharya, S. E. Mitchell, S. A. Flint-Garcia, M. D. McMullen, J. B. Holland, E. S. Buckler, and C. A. Gardner, 2013 Comprehensive genotyping of the usa national maize inbred seed bank. Genome Biol **14**: R55.

Schnell, F. and C. Cockerham, 1992 Multiplicative vs. arbitrary gene action in heterosis. Genetics **131**: 461–469.

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

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

Shull, G. H., 1908 The composition of a field of maize. Journal of Heredity pp. 296–301.

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

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

Table S1 BLUE values of the seven phenotypic traits. (https://github.com/RILAB/pvpDiallel/blob/master/manuscript/Figure_Table/Table_S1.trait_matrix.csv)

Table S2 General combining ability and specific combining ability of the seven phenotypic traits. (https://github.com/RILAB/pvpDiallel/blob/master/manuscript/Figure_Table/Table_S2.CA.csv)

Supporting Information

Table S3 Cross-validation results using all genome-wide deleterious SNPs. (https://github.com/RILAB/pvpDiallel/blob/master/manuscript/Figure_Table/Table_S3_allsnps_FDR.csv)

Table S4 Cross-validation results using deleterious genic SNPs. (https://github.com/RILAB/pvpDiallel/blob/master/manuscript/Figure_Table/Table_S4_genicsnps_FDR.csv)

Table S5 ANOVA P values of model comparisons. (https://github.com/RILAB/pvpDiallel/blob/master/manuscript/Figure_Table/Table_S5_model_comp.csv)

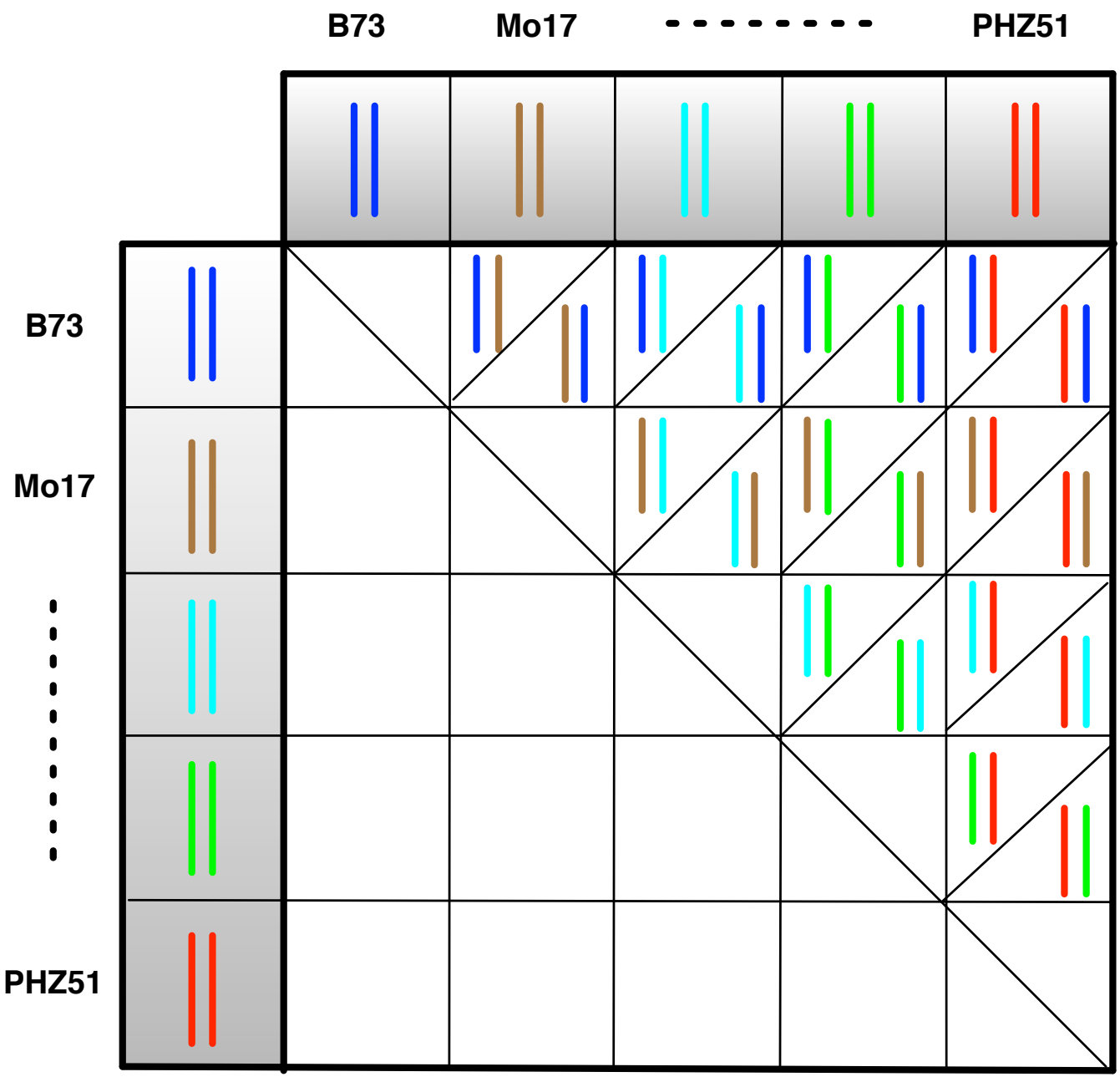
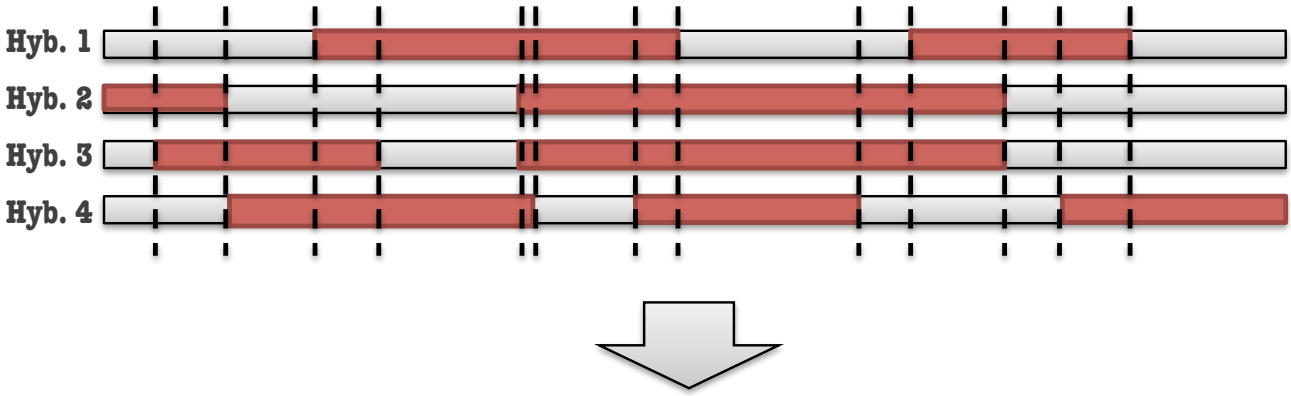


Figure S1 Twelve maize inbred lines were selected and crossed in a partial diallel fashion. Each inbred lines was used as both male and female and the resulting F1s were bulked.



	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 S2 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 genotypes in each block are coded as heterozygotes (0) or homozygotes (1).

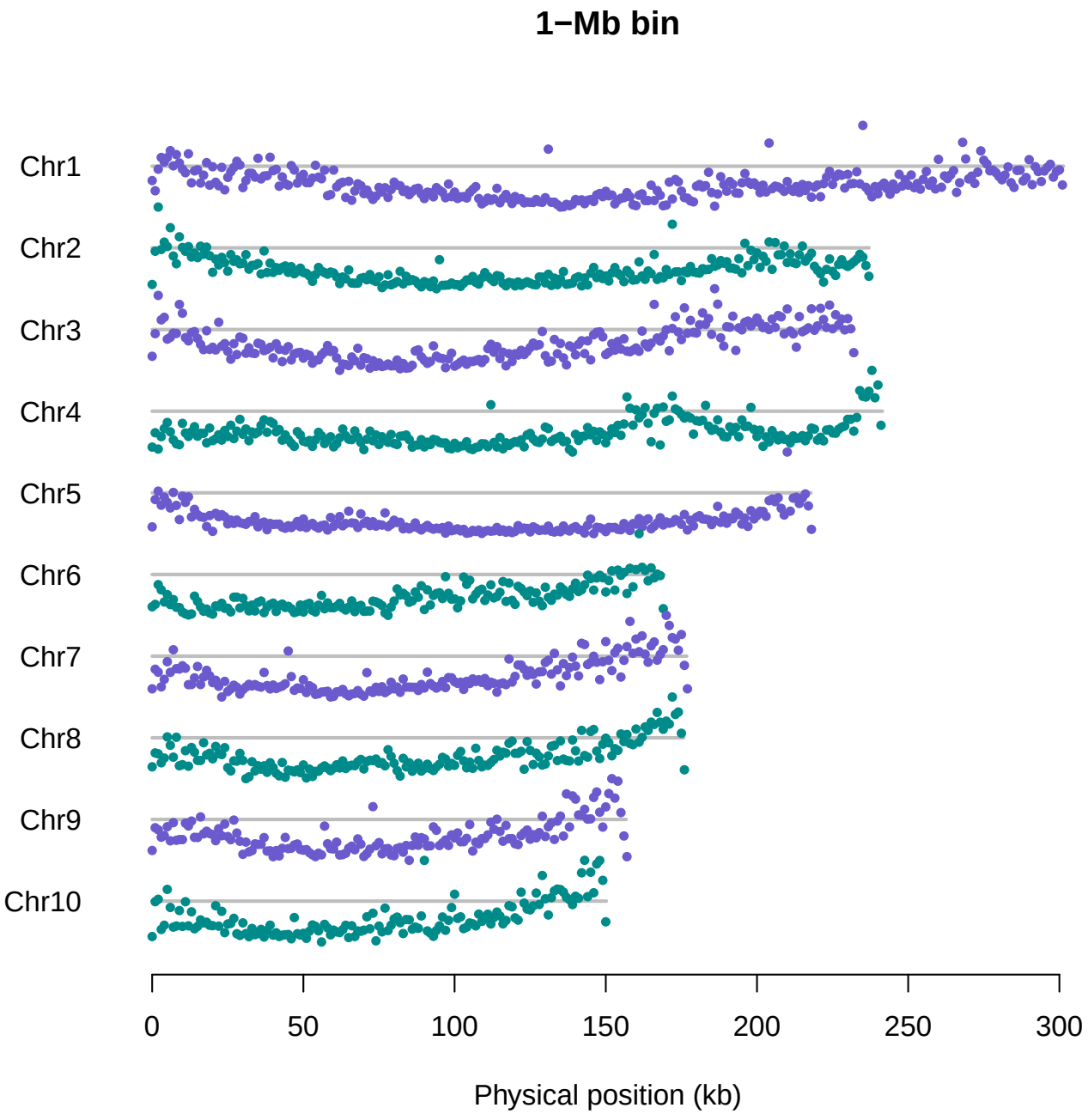


Figure S3 GERP score distribution across the genome. Shown are mean GERP scores in a 1-Mb bin region.

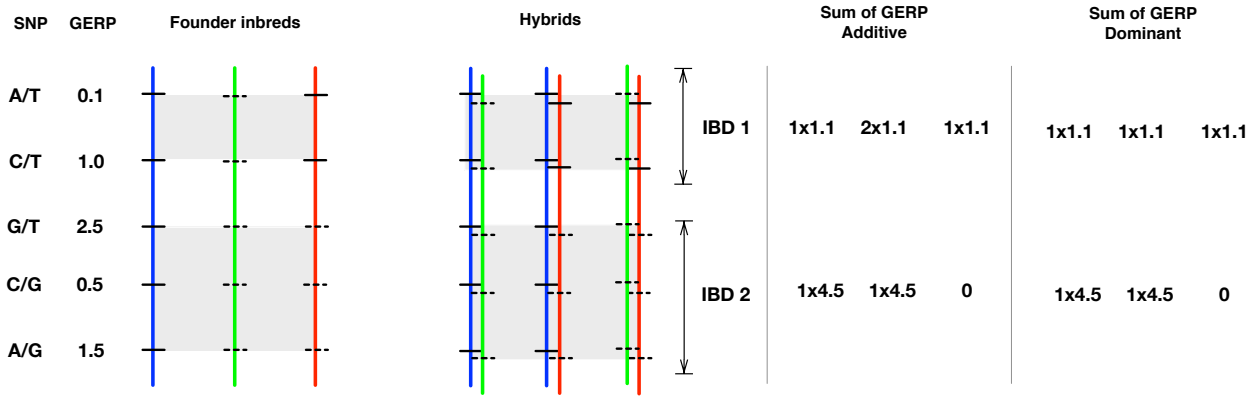


Figure S4 Incoporation of conservation information into IBD blocks. Regions of the genome that are identical by descent (IBD) among the 12 inbreds were identified using Beagle (Browning and Browning 2009). The GERP scores of SNPs in an IBD block were summed under both additive and dominant models. For a particular SNP with GERP score g, the homozygous non-reference genotype was assigned a value of $2g$, the heterozygote assigned a value of g, and the reference homozygote a value of 0. Under the dominant model, both the heterozygote and the non-reference homozygote were assigned a value of g, with the reference homozygote again assigned a value of 0.

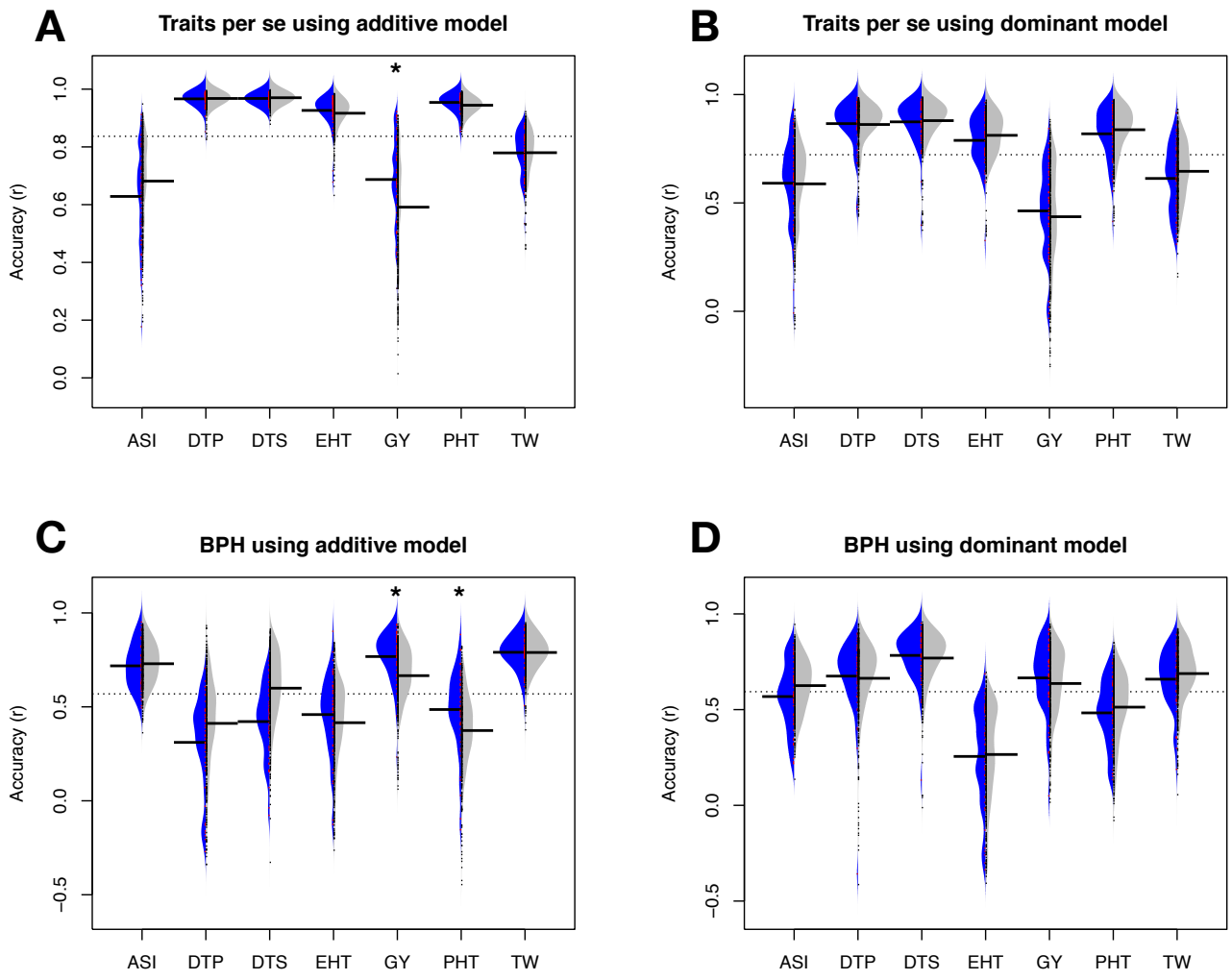


Figure S5 Cross-validation accuracies using genic SNPs. Cross-validation experiments were conducted using genic SNPs and compared to circular-shuffled data for traits *per se* (**A, B**) and pBPH (**C, D**) under additive (**A, C**) and dominant (**B, D**) models. Distirbutions show accuracty of prediction from real data (blue) and permutations (grey), with horizontal bars to indicate mean accuracy. Stars indicate significantly higher cross-validation accuracy for the real data. The average accuracy across all traits is shown with the grey dotted line. *need to explain red points here too.*