

ORIGINAL ARTICLE

Enhancing genomic prediction with genome-wide association studies in multiparental maize populations

Y Bian¹ and JB Holland^{1,2}

Genome-wide association mapping using dense marker sets has identified some nucleotide variants affecting complex traits that have been validated with fine-mapping and functional analysis. However, many sequence variants associated with complex traits in maize have small effects and low repeatability. In contrast to genome-wide association study (GWAS), genomic prediction (GP) is typically based on models incorporating information from all available markers, rather than modeling effects of individual loci. We considered methods to integrate results of GWASs into GP models in the context of multiple interconnected families. We compared association tests based on a biallelic additive model constraining the effect of a single-nucleotide polymorphism (SNP) to be equal across all families in which it segregates to a model in which the effect of a SNP can vary across families. Association SNPs were then included as fixed effects into a GP model that also included the random effects of the whole genome background. Simulation studies revealed that the effectiveness of this joint approach depends on the extent of polygenicity of the traits. Congruent with this finding, cross-validation studies indicated that GP including the fixed effects of the most significantly associated SNPs along with the polygenic background was more accurate than the polygenic background model alone for moderately complex but not highly polygenic traits measured in the maize nested association mapping population. Individual SNPs with strong and robust association signals can effectively improve GP. Our approach provides a new integrative modeling approach for both reliable gene discovery and robust GP.

Heredity (2017) **118**, 585–593; doi:10.1038/hdy.2017.4; published online 15 February 2017

INTRODUCTION

Genomic prediction (GP) based on genome-wide markers has emerged as a powerful supplement to conventional plant and animal breeding (Hayes and Goddard, 2001; Meuwissen *et al.*, 2001; Bernardo and Yu, 2007; Heffner *et al.*, 2009; Crossa *et al.*, 2014). Exploiting recent dramatic decreases in genotyping costs, GP can accelerate selection cycles or increase effective population sizes under selection (Wang and Hill, 2000) and is likely to increase selection gains per unit of time (Heffner *et al.*, 2010; Butruille *et al.*, 2015). Most genomic selection models assume genetic architectures similar to Fisher's 'infinitesimal' model (Hill, 2014), in which each marker is assumed to be associated with very small genetic effects. In contrast, quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) attempt to find individual markers associated with larger amounts of genetic variation than expected for a polygenic background effect.

GP and GWAS attempt to model different aspects of genetic architecture and have complementary advantages. Whereas GWAS may serve as initial evidence for the role of a particular gene in the inheritance of a complex trait, GWAS models are generally poor at representing the combined simultaneous effects of many small effect variants. In particular, with dense marker data sets, researchers often have more markers than observations, resulting in strong dependencies among single-nucleotide polymorphism (SNP) association tests. In contrast, common GP models rely on assumptions about the

genetic architecture that permit simultaneous modeling of high-dimensional SNP data, often resulting in good prediction ability at the expense of low interpretability. Consequently, it seems inevitable to trade off model interpretability for model complexity in GP (Gianola and van Kaam, 2008; González-Recio *et al.*, 2008).

Genomic best linear unbiased prediction (GBLUP), or equivalently, Ridge regression best linear unbiased prediction (RR-BLUP; Habier *et al.*, 2013) assumes equal marker variances and no epistasis but works well for quantitative trait prediction in practice (Lorenzana and Bernardo, 2009; Guo *et al.*, 2011; Heffner *et al.*, 2011; Heslot *et al.*, 2012). GBLUP has become a widely used statistical method to predict genotypic values in breeding practice with less computational burden than more complex approaches (vanRaden, 2008; Hayes *et al.*, 2009; Piepho, 2009). GBLUP uses DNA markers, often SNPs, to calculate the realized genomic relationships between individuals to mimic the genetic relationships at QTLs. The prediction ability of GBLUP relies on linkage disequilibrium (LD) between SNP loci and QTL and additive genetic relationships at QTL (Habier *et al.*, 2013). GBLUP is expected to perform well when the quantitative trait is controlled by a large number of loci dispersed across the genome.

GP generally outperforms QTL-based marker-assisted selection in animal and plant breeding (Lorenzana and Bernardo, 2009; Moser *et al.*, 2009; Heffner *et al.*, 2011). The poor predictive accuracy of marker-assisted selection is due to inadequate power to detect small effect loci and treating markers that do not pass stringent thresholds as

¹Department of Crop Science, USDA-ARS, North Carolina State University, Raleigh, NC, USA and ²US Department of Agriculture-Agricultural Research Service, Plant Science Research Unit, Raleigh, NC, USA

Correspondence: Dr JB Holland, Department of Crop Science, USDA-ARS, North Carolina State University, Box 7620, Raleigh, NC 27695, USA.

E-mail: james_holland@ncsu.edu

Received 29 June 2016; revised 3 January 2017; accepted 12 January 2017; published online 15 February 2017

having zero effect, while conversely overestimating the effects of markers that are retained (Beavis, 1998; Xu, 2003; Schön *et al.*, 2004). GWAS has some limitations similar to QTL mapping due to the use of stringent thresholds designed to minimize false positive associations, resulting in limited repeatability of detection results across studies especially for small-effect associations (Bian *et al.*, 2014; Lazzaroni *et al.*, 2014). Differences in allele frequencies, genetic background, epistasis, population structure and linkage phases among populations are also major challenges facing identification of robust associations.

A variety of approaches have been proposed to account for variability in SNP-associated effects in GP. Bayesian models assume the genetic architecture that includes a mixture of major and small effect QTL and polygenic background and apply shrinkage and variable selection techniques, using assumptions about marker effect distributions and hyper parameters to tune the priors (Meuwissen *et al.*, 2001; Park and Casella, 2008; Habier *et al.*, 2011). Bayesian approaches to GP are more computationally intensive and do not scale well to large marker data sets. The parameter estimates from Bayesian or non-parametric methods (including some kernel methods and random forests) can be sensitive to priors and unstable across studies (Gianola *et al.*, 2009). More computationally efficient methods have been proposed to weigh the genomic relationship matrix to represent the relative size of variance explained by the corresponding loci (Zhang *et al.*, 2014; Ober *et al.*, 2015; Tiezzi *et al.*, 2015). These methods aim to improve prediction and are not intended as gene discovery methods.

Methods that specifically integrate GWAS and GP methods might provide a way to simultaneously model different aspects of the genetic architecture and provide more robust associations that could aid gene discovery while also improving GP. Spindel *et al.* (2016) proposed combining a small number of significant SNPs detected with GWAS as fixed effects in combination with an RR-BLUP model to capture polygenic effects in a panel of 332 rice breeding lines, resulting in superior prediction performance (Spindel *et al.*, 2016). The objectives of this study were to compare the prediction ability of models including only polygenic background effects to those integrating polygenic background effects with GWAS-based SNP 'discoveries' in the context of a large multiple family interconnected population in maize. A simulation study was conducted to compare the power and false discovery rate (FDR) of association tests and the prediction ability of different modeling approaches under two different genetic architectures, differing by the effect sizes of a subset of markers. Real data from the maize nested association mapping (NAM) were analyzed to compare predictive ability of different models for traits with different genetic architectures.

MATERIALS AND METHODS

Plant phenotypes and genotypes

We used the maize NAM population comprising a set of ~5000 recombinant inbred lines (RILs) derived from crosses between a reference parent, inbred line B73, and 25 other diverse founder inbred lines of maize (McMullen *et al.*, 2009). We used predicted marginal values for RILs previously reported for three quantitative traits, resistance to southern leaf blight (SLB), resistance to gray leaf spot (GLS) and plant height (PHT), as phenotypes. The predicted marginal values were computed for all three traits as BLUPs adjusted for non-genetic effects from mixed-effect models in which lines were considered independent random effects (Bian *et al.*, 2014; Peiffer *et al.*, 2014; Benson *et al.*, 2015). A consensus linkage map consisting of 1476 markers with a uniform 1-cM inter-marker distance had been developed from stratified sampling of markers (Bian and Holland, 2015) from the 0.2-cM linkage map (Bian *et al.*, 2014; Swarts *et al.*, 2014) and was used for linkage analysis. An additive genomic

relationship matrix (*G* matrix) for the whole NAM panel computed from the high-density markers in maize HapMap V1 of 1.6 million SNPs (Peiffer *et al.*, 2014) was also adopted here to represent the polygenic additive genetic relationships (vanRaden, 2008), as previously used in (Bian and Holland, 2015). A total of 4354, 3225 and 4359 RILs with both genotypic and phenotypic data were available for analyzing SLB, GLS and PHT, respectively. A total of 1 328 174 single-nucleotide markers (most SNPs and small indels) in Maize HapMapV2 genotypes (Chia *et al.*, 2012) had complete and polymorphic loci for 26 NAM lines. Those markers were referred to as SNPs in this study, and their genotypes were projected into NAM RILs based on the linkage map (Bian *et al.*, 2014) and used for GWAS of the three agronomic traits. We then sampled 1 of every 10 of the GWAS HapMap V2 SNPs for each cM interval and ensured that at least two markers were kept for each interval, such that a total of 133 509 SNPs were used for the simulation analyses.

Simulation scheme

We simulated two quantitative traits with distinct genetic architectures using the real genotypes in NAM populations. Both architectures included a mixture of polygenic and oligogenic effects but differed in the number and effect sizes of the oligogenic effects. Heritability was 0.56 for both traits on a line mean-basis, with two sources of genetic variance: additive main effects of causal loci and additive polygenic backgrounds. We assumed that the *G* matrix represented the true pairwise additive genetic relationships and population structure among all RILs. For the first trait, 10 independent SNPs were randomly chosen from each of the 10 chromosomes and assigned identical additive effects as the true variants. For the polygenic effects, we assigned a variance component to genetic background effects, distributed according to the *G* matrix. When fit to the simulated phenotypes separately, the polygenic effects (*G* matrix) explain 32% of the phenotypic variance and the 10 causal loci collectively explained 27% of the phenotypic variance. As the causal and background effects have some correlation, the variance they explain when fit together is less than the sum of the individual component variations. For the second trait, we increased the number of causal loci from 10 to 100 true variants assigned with equal additive genotypic values but kept the variance associated with polygenic and all oligogenic effects the same. Every chromosome then had 10 nearly independent causal loci that were at least 1 cM apart. We hereafter refer to the first trait as 'oligogenic' and the second trait as 'polygenic,' although both traits include some contribution from both oligogenic and polygenic effects. The RIL samples were the same 4354 as used in the subsequent analysis of SLB. The simulation did not include epistatic interactions. A pseudocode was included in Supplementary Appendix S1.

Benchmark model: genomic BLUP models

Given the large amount of genotypic data available in this study, any model that became prohibitively computationally intensive was not employed because of the need to conduct many analyses for simulations and cross-validations. We therefore chose to use the GBLUP model to represent a conventional GP model. The GBLUP model was a mixed linear model expressed as follows:

$$y = 1\mu + Zu + e$$

where *y* is the $n \times 1$ vector for the trait values of *n* NAM RILs, μ is the intercept, *u* is an $n \times 1$ vector of genotype random effects and has covariance structure $var(u) = G\sigma_u^2$, where σ_u^2 is the additive genetic variance, *G* is the realized genomic relationship matrix, *Z* is a design matrix (identity matrix here) relating elements of *y* to elements of *u* and *e* is an $n \times 1$ vector of error effects with $e \sim N(0, I\sigma_e^2)$.

We tested an alternative mixed-effect model GBLUP by adding the NAM family indices (incidence matrix relating each individual RIL to its corresponding family) as fixed effects into the regular GBLUP model:

$$y = 1\mu + T\alpha + Zu + e,$$

where the new term *T* is the $n \times 24$ incidence matrix for the first 24 of the 25 NAM RIL families and α is the 24×1 vector for family mean effects relative to the reference family. This model was called the PGBLUP model, 'P' referring to 'population' effects.

GWAS discovery: main-effect and nested-effect model in GWAS

We used a mixed linear GWAS model to test for individual SNPs associated with significantly more genetic variation than expected for polygenic background variants. The main-effect GWAS model was:

$$y = 1\mu + T\alpha + x\beta + Zu + e,$$

where x is an $n \times 1$ marker genotype vector for n RILs at a marker locus and β is the marker main effect. The random background effects, u , were modeled with a covariance structure proportional to the G matrix, as was carried out for the GBLUP model.

For each NAM population, every chromosome pair of a RIL line is a mosaic of the two founder haplotypes (one is B73 and the other is an alternate founder). Marker effects at QTL could vary by populations due to epistatic interactions between the causal variants and other factors in the genetic background or due to functional allelic or haplotypic variation around the site being tested. To allow for this possibility, we developed a nested-effect model for GWAS that included separate coefficients representing the additive genetic effects of each founder allele (or the founder's haplotype that is in tight LD with the test locus). Using this nested-effect GWAS model, we can estimate the effects on the phenotype contributed by each of the alternate founder alleles. The marker genotype at a locus, as a continuous variable, is considered nested within population index, using the following mixed-effect model:

$$y = 1\mu + T\alpha + W\gamma + Zu + e,$$

similar to the previous models, but introducing W , an $n \times m$ design matrix, constructed by multiplying the marker genotype column into the population indices matrix and excluding the columns for populations in which the SNP is not segregating, and γ , a vector of marker nested effects, that is, a separate 'slope' for the marker loci within each population. In contrast to the main-effect model, this model estimates m allele effects per SNP, where m ranges from 1 to 25 and represents the number of populations in which the SNP is segregating.

The two GWAS models were implemented by a two-stage procedure (Supplementary Appendix S2). In the first stage, restricted maximum-likelihood (REML) estimates of variance components σ_u^2 and σ_e^2 were calculated from a mixed-effect model that omitted the $X\beta$ or $W\gamma$ term, using the EMMA algorithm (Kang *et al.*, 2008). The $1\mu + T\alpha$ terms were included as fixed effects in this baseline model, where GLS needed an additional flowering time fixed effect as a covariate (Benson *et al.*, 2015). In the second stage, the significance of each marker was measured by the additional variation explained, with one degree of freedom for the marker main-effect model or with m degrees of freedom for marker effect nested within population model, based on an F-test conditioning on previously estimated σ_u^2 and σ_e^2 (Kang *et al.*, 2010; Segura *et al.*, 2012). The variance components were estimated only once and used for all tests in the genome scan without re-estimation (Kang *et al.*, 2010; Zhang *et al.*, 2010). Although computationally efficient 'exact' linear mixed-effect model GWAS methods, such as FaSTLMM (Lippert *et al.*, 2011) and GEMMA (Zhou and Stephens, 2012), have been developed, we used our own implementation of EMMAX to incorporate test statistics extended to marker effects nested within families with more than one degree of freedom.

Resample model inclusion probability (Valdar *et al.*, 2009) was computed to measure repeatability of detection for GWAS. Resample model inclusion probability was calculated for each SNP as the proportion of cross-validation training data sets for which the marker passed a particular P -value threshold. The power was estimated by the proportion of the 1-cM intervals to which the causal loci had been assigned that had significant GWAS associations, and the FDR was estimated by the proportion of the 1-cM intervals that had significant GWAS associations but had not been assigned a causal locus. Similarly, to understand the influence of oligogenic versus polygenic architecture on GWAS performance, we examined power and FDR for the two simulated traits.

Evaluation of prediction accuracy and prediction models

We tested the hypothesis that incorporating association mapping discoveries from our new GWAS models can improve prediction abilities compared with GBLUP models. Cross-validation was used to estimate parameters of different models (including selection of significant markers for GWAS-based models) in training sets and to evaluate their prediction ability in independent validation

sets. For each trait, we created 50 pairs of disjoint training and test sets by randomly sampling 80% of RILs from each of the 25 NAM families for each training set and assigning the remaining lines to the corresponding test set. Prediction abilities were estimated separately for each of the 25 families and 50 replicated cross-validation samples as the average squared Pearson's correlation (R^2) between the observed (or simulated phenotypic) and predicted trait values. We gave a negative sign to R^2 values when the underlying correlation values were occasionally negative.

For the three real and two simulated traits, we fit the following prediction models in each of the 50 training sets for each trait:

1. GBLUP model using the estimated realized genomic relationship matrix (G) based on all markers.
2. REML-based linear mixed-effect models incorporating GWAS discoveries from main-effect models along with G .

In addition, for the three real traits, we fit five additional models:

3. PGBLUP model, equivalent to the GBLUP model but additionally incorporating a fixed population (family) main effect.
4. Bayesian Ridge regression-based linear mixed-effect models incorporating GWAS discoveries from main-effect models.
5. Bayesian Ridge regression-based linear mixed-effect models incorporating GWAS discoveries from nested-effect models.
6. Bayesian LASSO-based linear mixed-effect models incorporating GWAS discoveries from main-effect models.
7. Bayesian LASSO-based linear mixed-effect models incorporating GWAS discoveries from nested-effect models.

Model 2 (GBLUP+GWAS) included the most significant GWAS association within each cM interval that passed the P -value threshold in a training set, avoiding overrepresentation of the same genetic information. Selection of markers to include in a prediction model and estimation of their effects were both based solely on GWAS analysis in one training set. Because collinearity often occurred in the genotype matrix of the nested-effect GWAS discoveries, we did not report the prediction results from REML-based linear mixed-effect models incorporating the nested-effect GWAS discoveries.

Models 5–7 implemented the linear mixed-effect model in a Bayesian framework (Park and Casella, 2008; de los Campos *et al.*, 2009; Pérez *et al.*, 2010). The most significant GWAS association within each cM interval that passed the P -value threshold in a training set was included in the Bayesian models. We assigned Gaussian priors (Bayesian Ridge regression, models 4 and 5) or double-exponential priors (Bayesian LASSO, BL, models 6 and 7) to the significant SNPs from GWAS, flat priors to the fixed effects and also included the G matrix.

In Bayesian Ridge regression, marker regression coefficients were assigned with a common Gaussian prior, with mean zero and variance σ_β^2 . In the second level of the priors, the unknown variance parameter σ_β^2 was assigned a scaled-inverse Chi-squared density, with degrees of freedom df_β and scale S_β . The average prior variance of significant marker associations can be approximated as $MS_x \times \sigma_\beta^2$, where MS_x is the average sum of squares of the significant marker genotypes (de los Campos *et al.*, 2013). Equating $MS_x \times \sigma_\beta^2$ to the product of the assumed trait heritability explained by the GWAS causal loci (h^2) times the phenotypic variance (V_y), we obtained $\sigma_\beta^2 = h^2 \times V_y / MS_x$. The scaled-inverse chi-square density assigned to the variance parameter σ_β^2 was parameterized in such a way that the prior expected value $E(\sigma_\beta^2) = S_\beta / (df_\beta - 2)$, so that we computed the scale parameter S_β as $h^2 \times V_y \times (df_\beta - 2) / MS_x$.

In the BL models, the marginal distribution of marker effects is double exponential (Park and Casella, 2008), which was implemented as a mixture of scaled normal densities. In the first level of the hierarchy, marker effects were assigned independent Gaussian densities with mean zero and marker-specific variances (scale parameter times residual variance), and this prior induced marker-specific shrinkage on marker effect estimates, depending on the values of scale. In the second level, the marker-specific scale parameters were assigned IID exponential densities with a regularization parameter. In the last level of the hierarchy, the prior distribution of the regularization parameter was set to

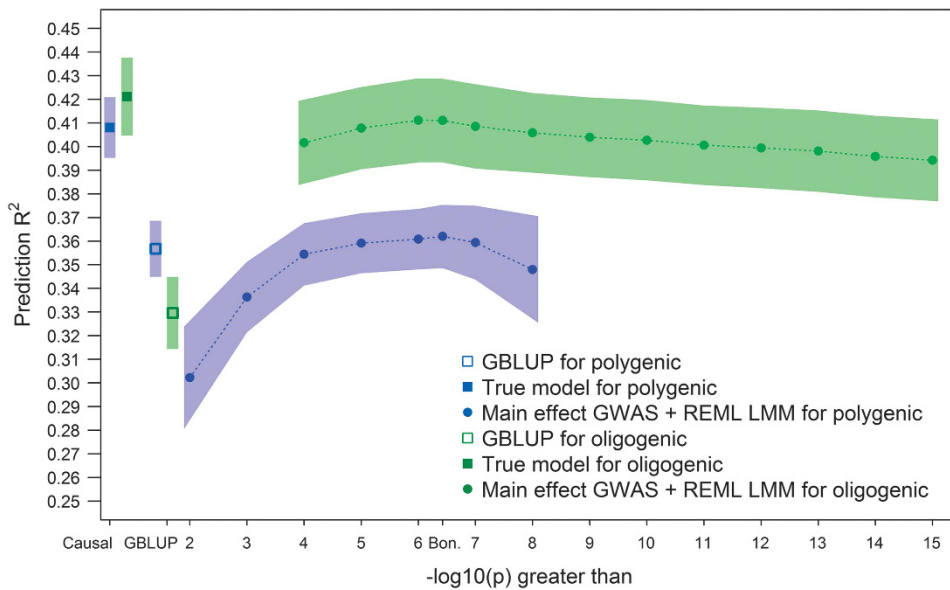


Figure 1 Prediction ability (R^2) of the simulated oligogenic and polygenic traits. 'Causal' represents the model including the causal variants as fixed effects (with effects estimated in the training set), 'GBLUP' represents the 'background only model' with no fixed effects of SNPs and 'Main-effect GWAS+REML LMM' denotes linear mixed models combining significant GWAS associations and polygenic background effects. The horizontal axis indicates the P -value thresholds for declaring a significant locus, 'Bon.', represents the Bonferroni correction P -value. The margins represent 95% confidence intervals of the mean prediction R^2 across 50 cross-validation runs.

follow a Gamma density with user defined rate (r) and scale (s) parameters. We set $s=1.01$, which should allow for a relatively uninformative prior. We solved for the rate (r) parameter to match the expected fraction of variance accounted for by the GWAS causal loci (h^2). A flat prior was used for all other fixed effects (covariates), specifically with a Gaussian prior with mean zero and variance equal to 1×10^{10} . Degrees of freedom were all set to 5. The residual variance was assigned a scaled-inverse chi-square prior density with degrees of freedom 5 and scale parameter, $(1-H^2) V_y \times (df_p - 2)$, where H^2 was the estimated line mean-basis heritability. The additive genetic variance of background genome (represented by the G matrix) was assigned with a similar prior to the residual variance. The line mean-basis heritability was assumed to be 0.85 based on previous analyses of NAM populations, and 80% of the genetic variance was assigned to the G matrix term (0.68), and the remainder of genetic variance was due to the additional causal loci (h^2 assigned 0.17) *a priori*. R package 'BLR' (de los Campos *et al.*, 2009) was used to conduct the Bayesian modeling, which utilized a Gibbs sampler to draw samples from the posterior distributions. A total of 50 000 Markov chain Monte Carlo samples were drawn from the resulting posterior distribution, the initial 10 000 iterations were discarded as burn-in and the thinning interval was 5. Trace plots were examined for mixing status.

To test the hypothesis that prediction ability within a population was related to the genetic similarity of the population's parents, we estimated the partial correlation between the genetic similarity of the parents and the prediction ability across the seven different models. For each trait, the vector of prediction R^2 values (averaged across 50 replicates) for each of 25 families \times 7 models was adjusted for mean differences among the 7 models. Then the correlation between the residual prediction R^2 values and the Jaccard genetic similarity based on all SNPs for each pair of parents was estimated to obtain the partial correlation.

RESULTS

Simulation results: power, FDR and prediction R^2

The main-effect GWAS model was implemented for the two simulated traits. Power and FDR were computed across a wide range of P -value significance thresholds. Power to detect SNP-trait associations in the oligogenic trait was substantially higher than that in the polygenic trait (Supplementary Figures S1 and S2). As the P -value threshold

decreased from 10^{-4} to 10^{-15} , power decreased from 0.89 to 0.36 and FDR dropped from 0.88 to 0.30 in the oligogenic trait, and for the same amount of change in FDR, the power decreased from 0.53 to 0.01 in the polygenic trait. Power and FDR were highly related, so that there was no clear optimal P -value threshold for GWAS. At the Bonferroni-adjusted P -value of 3.7×10^{-7} , an average of 27 SNP loci were discovered, of which about 7 were true (co-residing with the true causal variant within 1-cM intervals), resulting in an FDR of 0.72 and power of 0.70 for the oligogenic trait. For the polygenic trait, FDR was 0.26 and power was 0.01 at this same threshold.

Next, for the two simulated traits and different P -values, we incorporated significantly associated SNPs (selecting only the most significant SNP from each 1-cM linkage map interval that passed the P -value threshold) into the REML-based linear mixed-effect model to obtain predicted values of the two traits. The prediction abilities of the REML-based linear mixed-effect model were compared with other benchmark models (Figure 1). The prediction R^2 in the GBLUP model was 0.33 for the oligogenic trait and 0.36 for the polygenic trait. Both R^2 were greater than the simulated proportion of variance truly due to polygenic background (0.32) because the G matrix is also associated with some of the variation due to the 'major' genes.

We also examined the prediction R^2 in a mixed-effect model that fits the exact causal SNPs and G matrix as we had simulated. Because this 'causal model' has the correct predictors specified, it can be considered a theoretically optimal model. The 'causal model' had prediction ability R^2 from 14 to 15 percentage points lower than the heritability (0.56). Although the causal loci were known, the marker effects and background variance components required re-estimation for each training data set. Estimation errors in the causal model can be due to genetic sampling variance (affecting causal allele frequencies and introducing correlations among causal markers) as well as experimental error. Bernardo (2001) observed that, for polygenic traits, even knowing which genes are causal does not solve problems in estimating their effects.

The integrated GWAS-GP model performed well for the oligogenic trait: the prediction R^2 increased from 0.33 in the basic GBLUP model to a maximum of 0.41 at the optimal P -value threshold ($P = 10^{-6}$; Figure 1). Within the commonly used P -value range, from 10^{-4} to 10^{-7} , the GWAS-GP model had stable prediction ability, although it began to decrease as stringency increased further (Figure 1). In contrast, the GWAS-GP model at its optimal P -value threshold provided very little improvement over GBLUP for the polygenic trait and provided substantially worse prediction ability outside of the P -value range between 10^{-4} and 10^{-7} . For both traits, prediction ability was near optimal at the Bonferroni corrected P -value, and therefore we used this threshold in the analyses of the three agronomic traits.

GWAS of the three agronomic traits

We implemented two GWAS models for each of the two quantitative disease resistance traits (SLB and GLS) and PHT. The main-effect model assumed a constant difference between the SNP allele effects across families, whereas the nested-effect model allowed the allele effects to vary across families. The discoveries from the two models were obviously clustered in a relatively few genomic regions (Figure 2 and Supplementary Figures S3–S8). However, main-effect SNP GWAS models identified strong and repeatable associations on chromosome 6 for SLB and chromosome 1 for PHT alone, while nested-effect GWAS models appeared more powerful for associations on chromosome 4 for GLS and chromosome 7 for PHT.

The average number of SNP association discoveries was 21 (range 18–25) for SLB, 14 (range 7–24) for GLS and 0.9 (range 0–5) for PHT, when fitting the main-effect GWAS model in the 50 training sets. The average number of significant loci was 21 (18–26) for SLB, 23 (13–32) for GLS and 1 (0–4) for PHT, when fitting the nested-effect GWAS model in the 50 training sets.

Prediction of SLB, GLS and PHT

We investigated the impact of fitting the significant association SNPs into the GBLUP and PGBLUP models and evaluated the performance relative to these two baseline GBLUP models (Figure 3, Supplementary Tables S1–S6). The Bayesian models showed good convergence after the burn-in period (Supplementary Figure S10). For SLB, discoveries from both GWAS models significantly improved prediction R^2 over the GBLUP/PGBLUP models. For GLS, the nested-effect GWAS associations increased prediction R^2 significantly over both baseline GBLUP models, while the main-effect model increased prediction ability significantly only over the GBLUP model but not the PGBLUP model. For PHT, there appeared no improvement over the baseline models, although the new approach did not introduce overfitting or decrease prediction accuracy.

We investigated how well the significant main-effect GWAS associations alone could predict by computing accuracy of prediction based on reduced models, which used the same model covariates and the same significant GWAS associations, omitting the G matrix derived from the genomic background. A reduced model of the ‘main-effect GWAS+REML LMM’ model (in Figure 3), which left out the effect of G matrix of HapMap genotypes, generated substantially diminished prediction R^2 , decreased from 0.53 to 0.19 for SLB, from 0.24 to 0.13 for GLS, and from 0.49 to 0.01 for PHT. The prediction R^2 stayed the same when fitting those significant GWAS associations as fixed and random effects (Supplementary Table S7).

DISCUSSION

In this study, we proposed new GWAS models for the maize multiparental NAM population and tested whether their results could be used to enhance GP. Previous GWAS analyses in the maize NAM have used population main effects and joint linkage QTL effects to adjust for polygenic background effects (Kump *et al.*, 2011; Tian *et al.*, 2011). In contrast, the model used here adjusted for background effects by fitting a polygenic background effect with covariance

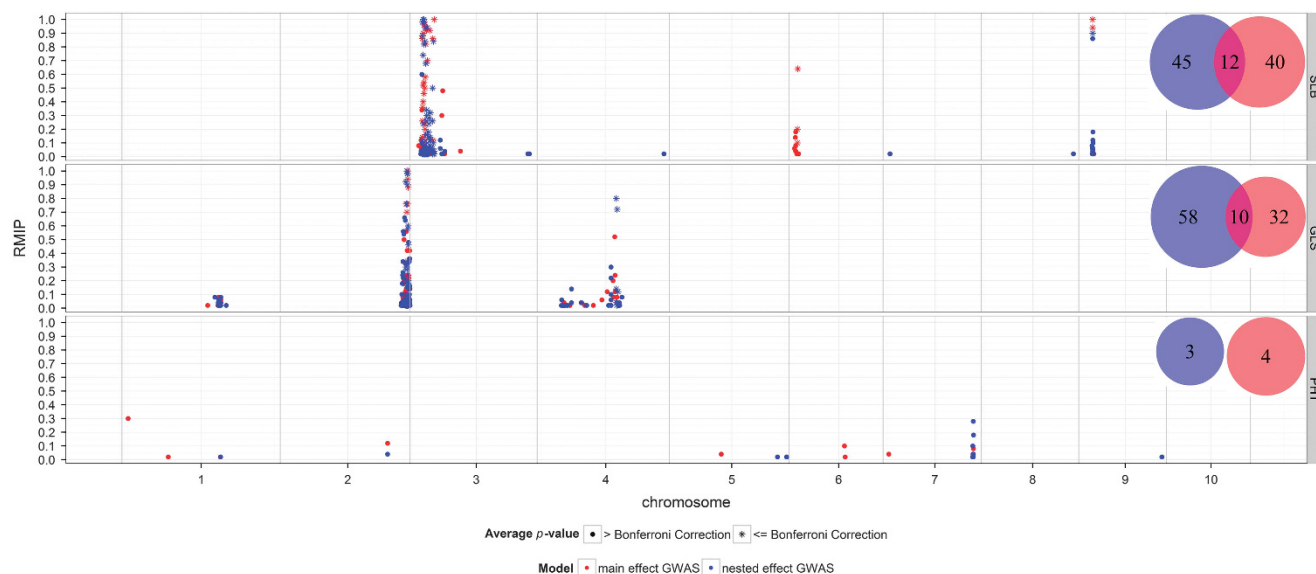


Figure 2 GWAS repeatability plot for the three agronomic traits. Each data point represents the resample model inclusion probability (RMIP) of an SNP with a significant association in one or more of 50 GWAS analyses of training data sets containing 80% of lines. Two filters were applied to the significant SNPs: first, only one associated locus per cM interval was considered, certifying the chosen SNPs to be the most competitive in the neighboring region, and second, the chosen SNPs’ P -values needed to pass the Bonferroni correction (P -value $< 3.8 \times 10^{-8}$). Asterisks denote those loci that have average P -values significant at the Bonferroni corrected P -value across 50 GWAS analyses. Venn diagrams show the numbers of highly repeatable loci (RMIP ≥ 0.05) identified in the two models.

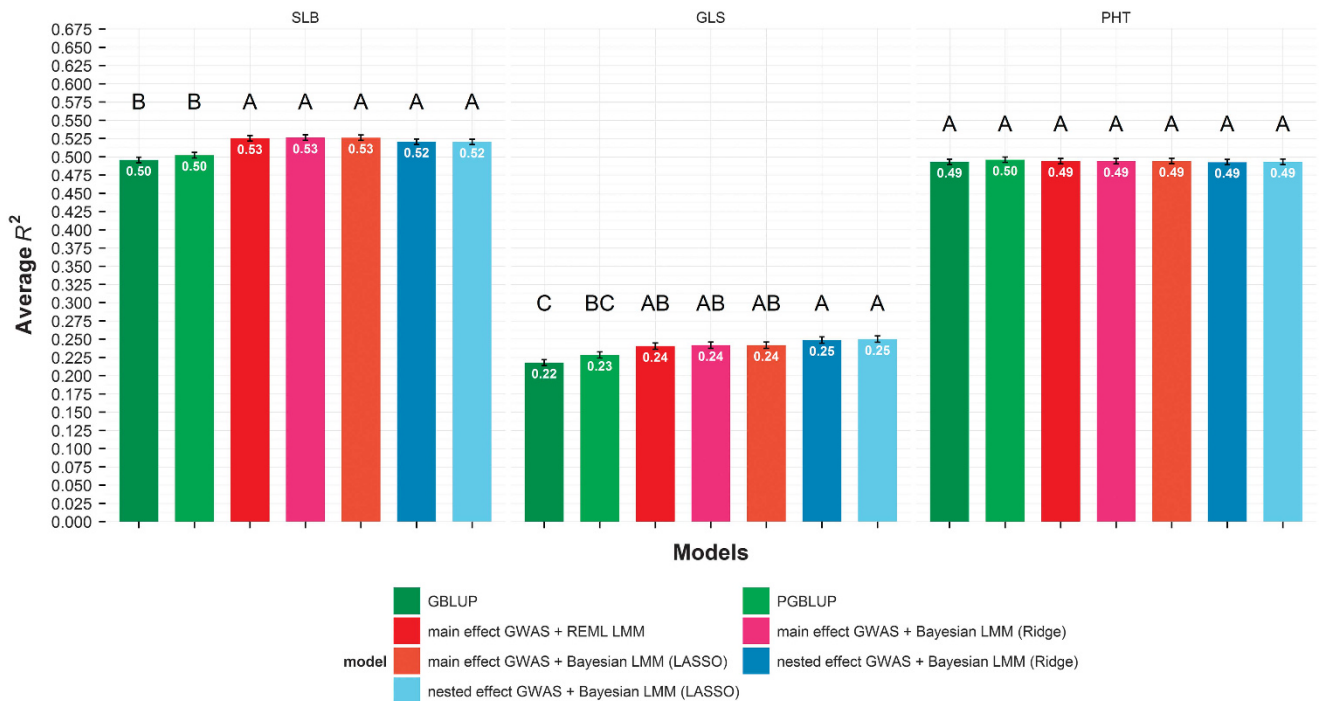


Figure 3 Within-family prediction ability (R^2) for three agronomic traits using GBLUP and new models. S.e. bars are imposed on each mean prediction ability value. Models with different letters above the bar indicate the average prediction R^2 values are significantly different ($\alpha < 0.05$; Tukey's honestly significant difference). Prediction R^2 values are averages across 25 NAM populations and 50 runs of cross-validations. The main-effect and nested-effect GWAS models failed to detect significant SNPs in 21 and 29 runs, respectively. For those cases, the prediction R^2 for the respective models was equal to the corresponding GBLUP models.

relationships among all NAM lines estimated by genome-wide SNPs (and summarized in the **G** matrix). This model along with Bonferroni adjustment to declare the significance threshold for individual SNP-trait associations resulted in substantially more stringent testing, with only strong and most competitive associations being declared as discoveries (Figure 2).

Simulation studies showed that the main-effect model had good power to detect individual SNPs explaining ~3% of phenotypic variation. In contrast, polygenes (100 causal loci, each causing ~0.3% of phenotypic variation) were very difficult to distinguish from background effects without a very high FDR. One way to bridge the disconnected interest between mapping gene and GP is to have an integrative modeling system to diagnose oligogenicity and exploit reliable association loci in prediction performance. Our GWAS associations for the two disease-resistance traits improved prediction accuracy significantly.

Power in main-effect versus nested-effect GWAS models

Previous GWAS analyses of the maize NAM population, and other multiparental populations for other species alike, assumed that individual SNP effects are equal across populations in which they segregate (Kump *et al.*, 2011; Tian *et al.*, 2011). By nesting SNP genotype effects in founders, we allow the flexibility of allelic effects across the populations. SNP effects may vary across families because they may tag different clusters (haplotypes) of functional variants in different founders or they may interact epistatically with the genetic background. Statistically, the new nested-effect GWAS model is an extension of the NAM joint-family linkage QTL model (Bian *et al.*, 2014), except that identity in state between parents at the tested SNP results in zero SNP effect in the new model, and the genetic background effects are controlled with a random polygenic effect

rather than by model selection. Density of GWAS associations in putative regions of interest appeared higher when using the nested-effect GWAS model. This suggests that, in general, the reduction in power to detect QTL due to increasing degrees of freedom used to model marker effects nested within families is counterbalanced by more accurately estimating distinct allelic effects across families at a common locus.

There may be specific cases, however, where the additive SNP effect model may be more realistic in some cases because LD among loci is weak in diverse maize, the current marker coverage is extremely high across the genome and the importance of epistasis in maize is uncertain. The choice of best model for specific populations and traits would be best guided by empirical studies on the effect of modeling QTL as main effects or nested effects on prediction accuracies.

Genetic architecture and implication to breeding practice

Although most agronomic traits are quantitatively inherited, relatively large effect genes may still be involved in their inheritance and their presence has a large impact on which GP model is optimal. A previous simulation study showed that when a few (1–3) major genes are present for a quantitative trait and each major gene accounts for 10% of genetic variance, fitting these major genes as fixed effects are beneficial to the genome-wide prediction model (Bernardo, 2014). In contrast, with polygenicity, many genetic effects must be estimated, and therefore estimates of genetic effects tend to be poor even with large sample sizes. The number of marker associations identified in maize NAM GWAS varies widely among traits and generally matches the assumptions about genetic architecture, with presumably more oligogenic characters such as metabolic traits having fewer associations identified than more polygenic traits, such as disease and plant

architecture traits (Wallace *et al.*, 2014). The GWAS-GP model is a flexible approach to capturing signal across the genetic architecture spectrum suggested by the results of Wallace *et al.* (2014).

Our GWAS approach distinguished the purely polygenic trait (PHT) from the ones that may have major genes, and by explicitly modeling the effects of highly confident loci, if any, in addition to the infinitesimal model, the new GP model had better prediction ability for traits, such as SLB and GLS. In contrast, the small number of PHT associations detected and their limited repeatability suggest that the infinitesimal genetic model should be sufficient for PHT (Supplementary Figures S7 and S8). The true effects of underlying QTL should be small for PHT, and so the accuracy with which these effects were estimated was expected to be low although a large sample was used. Being conservative about the GWAS discoveries (<1 on average) guarded against overfitting. The prediction R^2 varied among NAM populations for all the traits we studied (Supplementary Figure S9 and Supplementary Tables S8–S10). This phenomenon may occur because the segregating QTL are population dependent (Ogut *et al.*, 2015) and the founders' genetic relatedness to the common parental line B73 varies substantially. The correlations between the genetic similarity between parents and the prediction ability within their families were $r = 0.29$ ($P < 0.001$), 0.10 ($P < 0.176$) and -0.15 ($P < 0.043$) for SLB, GLS and PHT, respectively (Supplementary Table S11). This result suggests that there is no simple relationship between genetic diversity of parents and within-family prediction accuracy that is consistent across traits; the true relationship is likely a function of similarity at shared QTL regions across families.

We were not able to model simultaneously nested-effect GWAS discoveries with the **G** matrix in the frequentist model because of high and common collinearity among nested marker effects. In contrast, the Bayesian models imposed shrinkage on correlated significant GWAS discoveries, allowing SNP effects nested within families to be used in prediction models. Fitting the millions of markers available in Bayesian analyses was not computationally feasible for this study. Therefore, we preselected significant GWAS associations to use in both frequentist and Bayesian predictions models, greatly reducing computation time and facilitating convergence for the Bayesian methods.

The genetic architectures provided by this study imply that diverse breeding strategies are needed. Targeted introgression of favorable alleles or elimination of deleterious alleles can be employed in breeding if repeatable marker–trait associations are found. If no major marker–trait association is identified, however, phenotypic selection or GP should be given priority over targeting specific regions or loci. Highly repeatable and robust associations are those that are most likely to improve selection response when incorporated with GP models. These associations can also be considered higher priority targets for high-resolution genetic analysis to aid in the discovery of genes and genetic pathways underlying complex traits.

Using an extremely dense genome-wide marker set to represent the polygenic contribution may have reduced power to detect associations between traits and individual SNPs. If causal loci are in strong LD with other markers, the relationships modeled by the **G** matrix will account for some shared proportion of genetic variations at major effect loci as well as the polygenic loci, resulting in a reduction in the power of GWAS. Nevertheless, loci with moderate-to-strong effects can still stand out of the background and be captured as fixed effects. Alternative approaches that make locally appropriate corrections for genetic background and population structure effects during association scans could be more powerful. This can be accomplished by scanning with a **G** matrix calculated on all chromosomes but the current one

(Dell'Acqua *et al.*, 2015) and by testing within-interval SNP–trait associations with and without the boundary linkage markers included as covariates (a three-dimensional GWAS scan in the mixed-effect model context). Alternatively, an iterative process could be used in which significant polymorphisms identified in an initial scan could be included as covariates in additional scans, while updating the **G** matrix by recomputing it after excluding markers exceeding some correlation threshold with each fitted covariate marker.

Although GWAS had low power of detection for PHT due to its highly polygenic genetic architectures, our strategies incorporating the association mapping discoveries into prediction did not cause overfitting. It was also clear from the lack of repeatable GWAS associations that there were no strong effects to improve the polygenic model prediction. In this way, the repeatability of associations across data resamples is a useful diagnostic to predict the utility of the associations in GP. Accounting for the different population mean effects using the PGBLUP model did not benefit prediction in NAM populations (Figure 3) beyond the conventional GBLUP model, as the **G** matrix alone appears to mostly account for the effects of variation among family means.

This study investigated within-family prediction performed jointly on multiple inter-related families. The training set included samples of all families used to evaluate the accuracy of prediction, and this lends itself to reasonable accuracy in prediction. Accurate GP from training data in one family to individuals in other families is a more difficult problem (Windhausen *et al.*, 2012; Lehermeier *et al.*, 2014; Peiffer *et al.*, 2014). More research is needed to identify best practices for modeling in cross-population prediction, as it is crucial to commercial breeding. How this strategy would perform in training on multiple related or unrelated small biparental families to predict progeny from other crosses is an important question. To address the question, studies need to be carried out in optimizing the allocation of resources in GP using the current method.

Data availability

Seeds of NAM lines have been deposited at USDA Maize Genetics Cooperation Stock Center (<http://maizecoop.cropsci.uiuc.edu/nam-rils.php>). Files S1 to S21 are available for download at: <http://dx.doi.org/10.5061/dryad.cd3hv> (Bian and Holland, 2017).

File S1 (NAMSLB.RData) contains mean values for SLB disease scores and linkage map marker scores for NAM RILs formatted as an R data object. File S2 (NAMPHL.RData) contains mean values for PHL and linkage map marker scores for NAM RILs formatted as an R data object. File S3 (NAMGLS.RData) contains mean values for GLS disease scores and linkage map marker scores for NAM RILs formatted as an R data object. File S4 (relMat.RData) contains the realized additive genomic relationship matrix for NAM lines based on 1.6M HapMap I markers (courtesy of Dr Jason Peiffer). File S5 (Map.RData) contains the information of 0.2-cM density (7386 markers) linkage map including chromosome, cM position and AGP v2 coordinate. File S6 (emma.r) contains the EMMA-related functions in R (Kang *et al.* 2008). File S7 (Parents26_chr1.RData) contains 213 899 HapMap II marker genotypes for NAM parents on Chromosome 1. File S8 (Parents26_chr2.RData) contains 150 890 HapMap II marker genotypes for NAM parents on Chromosome 2. File S9 (Parents26_chr3.RData) contains 152 260 HapMap II marker genotypes for NAM parents on Chromosome 3. File S10 (Parents26_chr4.RData) contains 150 951 HapMap II marker genotypes for NAM parents on Chromosome 4. File S11 (Parents26_chr5.RData) contains 150 874 HapMap II marker genotypes for NAM parents on Chromosome 5. File S12 (Parents26_chr6.RData) contains 98 027 HapMap II

marker genotypes for NAM parents on Chromosome 6. File S13 (Parents26_chr7.RData) contains 114 988 HapMap II marker genotypes for NAM parents on Chromosome 7. File S14 (Parents26_chr8.RData) contains 107 405 HapMap II marker genotypes for NAM parents on Chromosome 8. File S15 (Parents26_chr9.RData) contains 97042 HapMap II marker genotypes for NAM parents on Chromosome 9. File S16 (Parents26_chr10.RData) contains 91 838 HapMap II marker genotypes for NAM parents on Chromosome 10. File S17 (NAM_parent_hung.txt) contains 'Family ID', 'Parent' name, 'Origin' and 'Subpop' for NAM parents. File S18 (GWAS script.r) contains the scripts for projection of parental marker genotypes into NAM RILs and GWAS additive and nested-effect models. File S19 (Simulation script.r) contains the R script to simulate the oligogenic and polygenic trait phenotypic values and compute the variance explained by each genetic component. File S20 (Bayesian linear mixed models.r) contains the R script to obtain the prediction from the two Bayesian LMM models based on GWAS results. File S21 (LMM demo.r) contains a numeric example to demonstrate the equivalence between the LMM realization in this paper and one-step linear mixed model.

ACKNOWLEDGEMENTS

We thank the High Performance Computing Services at NC State University. Research was supported by United States National Science Foundation grant IOS-1238014 and the US Department of Agriculture, Agricultural Research Service. YB was partly supported by a Monsanto Co. fellowship.

- Beavis WD (1998). QTL analyses: power, precision, and accuracy. In: Paterson AJ (ed). *Molecular Dissection of Complex Traits*. CRC Press: Boca Raton, FL, USA, pp 145–162.
- Benson JM, Poland JA, Benson BM, Stromberg EL, Nelson RJ (2015). Resistance to gray leaf spot of maize: genetic architecture and mechanisms elucidated through nested association mapping and near-isogenic line analysis. *PLoS Genet* **11**: e1005045.
- Bernardo R (2001). What if we knew all the genes for a quantitative trait in hybrid crops?. *Crop Sci* **41**: 1–4.
- Bernardo R (2014). Genomewide selection when major genes are known. *Crop Sci* **54**: 68–75.
- Bernardo R, Yu J (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci* **47**: 1082–1090.
- Bian Y, Holland JB (2015). Ensemble learning of QTL models improves prediction of complex traits. *G3* **5**: 2073–2084.
- Bian Y, Holland JB (2017). Data from: Enhancing genomic prediction with genome-wide association studies in multiparental maize populations. *Dryad Digital Repository*. doi:10.5061/dryad.cd3hv.
- Bian Y, Yang Q, Balint-Kurti PJ, Wissner RJ, Holland JB (2014). Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. *BMC Genomics* **15**: 1068.
- Butruille DV, Birru FH, Boerboom ML, Cargill EJ, Davis DA, Dhungana P *et al.* (2015). Maize breeding in the United States: views from within Monsanto. In: Janick J (ed). *Plant Breeding Reviews*: vol. 39, John Wiley & Sons, Inc.: New York, NY, USA, pp 199–282.
- Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J *et al.* (2012). Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* **44**: 803–807.
- Crossa J, Perez P, Hickey J, Burgueno J, Ornela L, Ceron-Rojas J *et al.* (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* **112**: 48–60.
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* **193**: 327–345.
- de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E *et al.* (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **182**: 375–385.
- Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G *et al.* (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol* **16**: 1–23.
- Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* **183**: 347–363.
- Gianola D, van Kaam JB (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* **178**: 2289–2303.
- González-Reco G, Gianola D, Long N, Weigel KA, Rosa GJ, Avendaño S (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics* **178**: 2305–2313.
- Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2011). Evaluation of genome-wide selection efficiency in maize nested association mapping populations. *Theor Appl Genet* **124**: 261–275.
- Habier D, Fernando RL, Garrick DJ (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics* **194**: 597–607.
- Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinform* **12**: 1–12.
- Hayes B, Goddard M (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.
- Hayes BJ, Visscher PM, Goddard ME (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* **91**: 47–60.
- Heffner EL, Jannink J-L, Sorrells ME (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* **4**: 65–75.
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci* **50**: 1681–1690.
- Heffner EL, Sorrells ME, Jannink J-L (2009). Genomic selection for crop improvement. *Crop Sci* **49**: 1–12.
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci* **52**: 146–160.
- Hill WG (2014). Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics* **196**: 1–16.
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB *et al.* (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* **42**: 348–354.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ *et al.* (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.
- Kump KL, Bradbury PJ, Wissner RJ, Buckler ES, Belcher AR, Oropeza-Rosales MA *et al.* (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat Genet* **43**: 163–168.
- Lazzeroni L, Lu Y, Belitskaya-Levy I (2014). P-values in genomics: apparent precision masks high uncertainty. *Mol Psychiatry* **19**: 1336–1340.
- Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L (2014). Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* **198**: 3–16.
- Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011). FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**: 833–835.
- Lorenzana RE, Bernardo R (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet* **120**: 151–161.
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q *et al.* (2009). Genetic properties of the maize nested association mapping population. *Science* **325**: 737–740.
- Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.
- Moser G, Tier B, Crump RE, Khatkar MS, Raadsma HW (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Genet Sel Evol* **41**: 56.
- Ober U, Huang W, Magwire M, Schlather M, Simianer H, Mackay TFC (2015). Accounting for genetic architecture improves sequence based genomic prediction for a *Drosophila* fitness trait. *PLoS One* **10**: e0126880.
- Ogut F, Bian Y, Bradbury PJ, Holland JB (2015). Joint-multiple family linkage analysis predicts within-family variation better than single-family analysis of the maize nested association mapping population. *Heredity* **114**: 552–563.
- Park T, Casella G (2008). The Bayesian LASSO. *J Am Stat Assoc* **103**: 681–686.
- Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ *et al.* (2014). The genetic architecture of maize height. *Genetics* **196**: 1337–1356.
- Pérez P, de los Campos G, Crossa J, Gianola D (2010). Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian linear regression package in R. *Plant Genome* **3**: 106–116.
- Piepho H-P (2009). Ridge regression and extensions for genomewide selection in maize. *Crop Sci* **49**: 1165–1176.
- Schön CC, Utz HF, Groh S, Truberg B, Openshaw S, Melchinger AE (2004). Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* **167**: 485–498.
- Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U, Long Q *et al.* (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* **44**: 825–830.
- Spindel JE, Begum H, Akdemir D, Collard B, Redona E, Jannink JL *et al.* (2016). Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity* **116**: 395–408.
- Swarts K, Li H, Navarro J, An D, Romay M, Hearne S *et al.* (2014). Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome* **7**: 3.
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S *et al.* (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* **43**: 159–162.
- Tiezzi F, Parker-Gaddis KL, Cole JB, Clay JS, Maltecca C (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS One* **10**: e0114919.
- Valdar W, Holmes CC, Mott R, Flint J (2009). Mapping in structured populations by resample model averaging. *Genetics* **182**: 1263–1277.
- vanRaden PM (2008). Efficient methods to compute genomic predictions. *J Dairy Sci* **91**: 4414–4423.

- Wallace JG, Bradbury PJ, Zhang N, Gibon Y, Stitt M, Buckler ES (2014). Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet* **10**: e1004845.
- Wang J, Hill WG (2000). Marker-assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics* **154**: 475–489.
- Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink J-L, Sorrells ME *et al.* (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3* **2**: 1427–1436.
- Xu S (2003). Theoretical basis of the Beavis effect. *Genetics* **165**: 2259–2268.
- Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA *et al.* (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* **42**: 355–360.
- Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J *et al.* (2014). Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. *PLoS One* **9**: e93017.
- Zhou X, Stephens M (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**: 821–824.

Supplementary Information accompanies this paper on Heredity website (<http://www.nature.com/hdy>)