



Genomic Prediction for 25 Agronomic and Quality Traits in Alfalfa (*Medicago sativa*)

Congjun Jia^{1†}, Fuping Zhao^{1†}, Xuemin Wang¹, Jianlin Han^{2,3}, Haiming Zhao⁴, Guibo Liu⁴ and Zan Wang^{1*}

¹ Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, ² CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China, ³ International Livestock Research Institute (ILRI), Nairobi, Kenya, ⁴ Institute of Dryland Farming, Hebei Academy of Agriculture and Forestry Sciences, Hengshui, China

OPEN ACCESS

Edited by:

Yiwei Jiang,
Purdue University, United States

Reviewed by:

Quanzhen Wang,
Northwest A&F University, China
Lan Zhu,
Oklahoma State University,
United States

*Correspondence:

Zan Wang
wangzan@caas.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 14 March 2018

Accepted: 30 July 2018

Published: 20 August 2018

Citation:

Jia C, Zhao F, Wang X, Han J,
Zhao H, Liu G and Wang Z (2018)
Genomic Prediction for 25 Agronomic
and Quality Traits in Alfalfa (*Medicago*
sativa). *Front. Plant Sci.* 9:1220.
doi: 10.3389/fpls.2018.01220

Agronomic and quality traits in alfalfa are very important to forage industry. Genomic prediction (GP) based on genotyping-by-sequencing (GBS) data could shorten the breeding cycles and accelerate the genetic gains of these complex traits, if they display moderate to high prediction accuracies. The aim of this study was to investigate the predictive potentials of these traits in alfalfa. A total of 322 genotypes from 75 alfalfa accessions were used for GP of the agronomic and quality traits, which were related to yield and nutrition value, respectively, using BayesA, BayesB, and BayesC π methods. Ten-fold cross validation was used to evaluate the accuracy of GP represented by the correlation between genomic estimated breeding value (GEBV) and estimated breeding value (EBV). The accuracies ranged from 0.0021 to 0.6485 for different traits. For each trait, three GP methods displayed similar prediction accuracies. Among 15 quality traits, mineral element Ca had a moderate and the highest prediction accuracy (0.34). NDF digestibility after 48 h (NDFD 48 h) and 30 h (NDFD 30 h) and mineral element Mg had prediction accuracies varying from 0.20 to 0.25. Other traits, for example, fat and crude protein, showed low prediction accuracies (0.05 to 0.19). Among 10 agronomic traits, however, some displayed relatively high prediction accuracies. Plant height (PH) in fall (FH) had the highest prediction accuracy (0.65), followed by flowering date (FD) and plant regrowth (PR) with accuracies at 0.52 and 0.51, respectively. Leaf to stem ratio (LS), plant branch (PB), and biomass yield (BY) reached to moderate prediction accuracies ranging from 0.25 to 0.32. Our results revealed that a few agronomic traits, such as FH, FD, and PR, had relatively high prediction accuracies, therefore it is feasible to apply genomic selection (GS) for these traits in alfalfa breeding programs. Because of the limitations of population size and density of SNP markers, several traits displayed low accuracies which could be improved by a bigger reference population, higher density of SNP markers, and more powerful statistic tools.

Keywords: alfalfa, genomic prediction, agronomic trait, quality trait, estimated breeding value

INTRODUCTION

Alfalfa (*Medicago sativa* L) is the first most-important forage legume in the world, because of its high biomass yield (BY) and good nutritional quality. To meet the future demand of quantity and quality, the main objectives in alfalfa breeding programs are biomass related agronomic traits and nutrition value related quality traits. Though yield and quality of alfalfa have been improved by phenotypic selection, the genetic gain are relatively low compared to other crops, owing to many reasons, such as low heritability, complex genetic architecture, and high genotype-environment interaction (Annicchiarico et al., 2015a). Therefore, it is emergent that new breeding strategies should be introduced into alfalfa breeding programs to accelerate the genetic gain of targeted traits and thus to meet the increasing demands of forage production.

Breeding value (BV), known as genetic merit of an individual which cannot be measured directly, is always the key issue in plant breeding programs. However, accurately estimated breeding value (EBV) is impossible to be achieved in complex traits by using phenotypic data alone. To improve the accuracy of prediction, incorporating information of genetic markers, known as marker-assisted selection (MAS), is an optional strategy. The superiority of MAS than phenotypic selection is determined by the percentage of the genetic variance accounted for by the QTLs associated with the markers (Meuwissen and Goddard, 1996). Unfortunately, the proportion of variation in complex traits explained by significant markers is usually very small (Hayes and Goddard, 2010). Therefore, many markers in linkage disequilibrium (LD) with QTLs contributed to targeted traits are needed to realize a relatively high prediction accuracy.

Due to the decreased cost of high-throughput genotyping methods, huge amount of genomic information of many non-model plants has been produced. Utilization of genotypic information in plant breeding has become a highly prioritized research area in recent years. Since dense genetic markers covering whole genome are available in many species, a new method for estimating breeding value, namely the genomic selection (GS) or genomic prediction (GP), showed a great potential for enhancing the accuracy of GP of BV (Meuwissen et al., 2001). It is assumed that all genes, with either large or small effects, affecting targeted traits are in LD with some markers that are distributed across the genome, paving the way to achieve a high accuracy of genomic estimated breeding value (GEBV) (Meuwissen, 2007). In a simulation study, the accuracy could be as high as 0.85 (Meuwissen et al., 2001). But this is not always the case in the real data. Several studies on GP have been done in wheat (Lado et al., 2013; Jiang et al., 2017; Sukumaran et al., 2017), maize (Riedelsheimer et al., 2012; Crossa et al., 2013; Pace et al., 2015), and other plants (Shu et al., 2013; Xu et al., 2014; Grenier et al., 2015), revealing a majority of the prediction accuracies between 0.05 and 0.8, depending on the traits, statistical methods, and experiment designs.

As mentioned above, GP can significantly improve the accuracy of estimation of breeding value. Therefore, it attracts

a great interest of plant breeders worldwide. Traits being targeted in plant breeding programs are either difficult or costly to be measured. Additionally, the targeted traits (e.g., yield, phenology, and adaptation to stress) in plant breeding are mostly quantitative traits, which are controlled by multiple genes and generally sensitive to environmental variables. Phenotypic selection, neglecting the underlying biological processes and the interactions between genes and environments, cannot make a significant genetic gain in a short time frame. Considering the genetic architecture of the quantitative traits, MAS is also not the best choice. GP, following its assumption, is thus an ideal tool to be used in the plant breeding programs. Many methods have been adopted for GP or GS. Bayesian methods and GBLUP, however, are those being frequently used. Bayesian methods exhibited more advances than GBLUP in terms of prediction accuracy following a simulation study (Meuwissen et al., 2001). No matter which method is used for GP, the density of markers across the whole genome is a determining factor. Typically, two types of high throughput genotyping methods of SNP array and whole-genome re-sequencing can be employed to generate high quality genotypes of markers. For important crop species, several SNP Bead chips at different marker densities have been developed (Ganal et al., 2012). Because of the lack of SNP array, genotyping by sequencing (GBS) is therefore an alternative to alfalfa genotyping. In the current study, we investigated the impact of three Bayes statistical methods on the prediction accuracies of alfalfa agronomic and quality traits with genotypic data obtained by GBS.

MATERIALS AND METHODS

Plant Materials and Experimental Designs

The alfalfa materials used in this study were consisted of 322 genotypes representing 75 tetraploid alfalfa accessions under the experimental designs as described in Wang et al. (2016).

Phenotypic Data Collection and Analysis

A total of 25 traits (Table 1), including 15 quality and 10 agronomic traits, were measured for all genotypes. All the plants were harvested at early flowering stage and prepared to measure the 15 quality traits using a FOSS 5000 scanning monochromator (FOSS, Denmark). The 15 quality traits included three fiber-related traits, four digestibility-related traits, and eight nutrition component traits being measured following the procedures described in our previous studies (Wang et al., 2016; Jia et al., 2017). Before harvesting, plant height (PH) of each plot was measured as nature height on every plant. Plant branch (PB) was measured as the number of primary branches arising from the main stem. The number of main stem node (SN) for each plot was directly counted since the first node on the main stem from every plant. The first inflorescence position (FP) was measured as the position of the first inflorescence on the stem. After harvesting, BY was measured as the fresh weight by clipping all six plants in each plot at a uniform height of 5 cm. The stems and leaves

TABLE 1 | Prediction accuracies of 25 traits.

Traits	Prediction accuracies		
	BayesA	BayesB	BayesC π
ADF	0.1847	0.1805	0.1824
aNDF	0.1843	0.1942	0.1958
dNDF30	0.2002	0.1933	0.1963
dNDF48	0.2538	0.2545	0.249
ADL	0.0783	0.0756	0.08
IVTDM30h	0.0907	0.0906	0.0889
IVTDM48h	0.0994	0.1029	0.1002
CP	0.0528	0.0524	0.0557
RUP	0.0828	0.0909	0.0746
Ash	0.087	0.0836	0.085
Ca	0.337	0.3422	0.3393
K	0.1663	0.1586	0.1572
Mg	0.2184	0.2183	0.2178
P	0.1178	0.1196	0.1203
Fat	0.1148	0.1203	0.1114
BY	0.2418	0.2511	0.2457
DM	0.1285	0.1257	0.1271
FD	0.5153	0.5139	0.5131
FH	0.6485	0.6466	0.6451
FP	0.0639	0.0596	0.0683
LS	0.3214	0.3215	0.3249
PB	0.2589	0.2598	0.2593
PH	0.1587	0.1601	0.1626
PR	0.5105	0.5111	0.5074
SN	0.0045	0.0047	0.0021

Abbreviations of traits are explained in Materials and Methods.

were separated and placed into a nylon net bag, naturally air-dried, and weighed separately to calculate the leaf to stem ratio (LS). Meanwhile, dry matter (DM) was defined as the sum of the weights of stems and leaves. Plant regrowth (PR) was measured as the PH two weeks after the first harvest. Flowering date (FD) was calculated by the date of opening of the first flower for the first two growth cycles. PH in fall (FH) was measured as the PH 21 days after the last harvest. The mean value of all six plants in each plot represents the trait value of a genotype grown in that plot. The measurements of all traits were performed on all genotypes under three consecutive years (2013, 2014, and 2015).

Linear mixed model was fitted to calculate the BLUP value and EBV for individual trait of each genotype as follows:

$$y_i = \mu + g_i + e_i + \varepsilon_i.$$

In this equation, y_i represents the phenotype of the i th genotype, μ is the grand mean value of the targeted trait in all environments, g_i is denoted as genetic effect, e_i is the environmental effect, and ε_i is the random error. The BLUP value was estimated for individual trait of each genotype based on the above-mentioned linear model using the lme4 model (Bates et al., 2011). The EBV of individual genotype was used as response value in GP equation to estimate marker effect.

DNA Isolation, GBS Library Construction, Sequencing, and Genotypic SNP Calling

Leaf tissues were collected from all genotypes and DNAs were extracted using the Qiagen DNeasy 96 Plant kit (Qiagen, CA, United States). DNA degradation and contamination were monitored on 1% agarose gels. DNA purity and concentration were checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, United States) and Qubit[®] DNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, United States), respectively. DNA was digested by *MseI* [New England Biolabs (NEB)] restriction enzyme. The reduced representation libraries were constructed for individual genotypes according to published GBS protocol (Elshire et al., 2011) and sequenced using Illumina HiSeq2000 platform. Raw data were submitted to the NCBI Sequence Read Archive with a reference number of SRP081825. The Tassel 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline (Lu et al., 2013) was used for *de novo* SNP discovery and genotype calling following Li et al. (2014).

SNP Imputation

After SNP calling, NPUTE was used to impute the GBS data (Roberts et al., 2007).

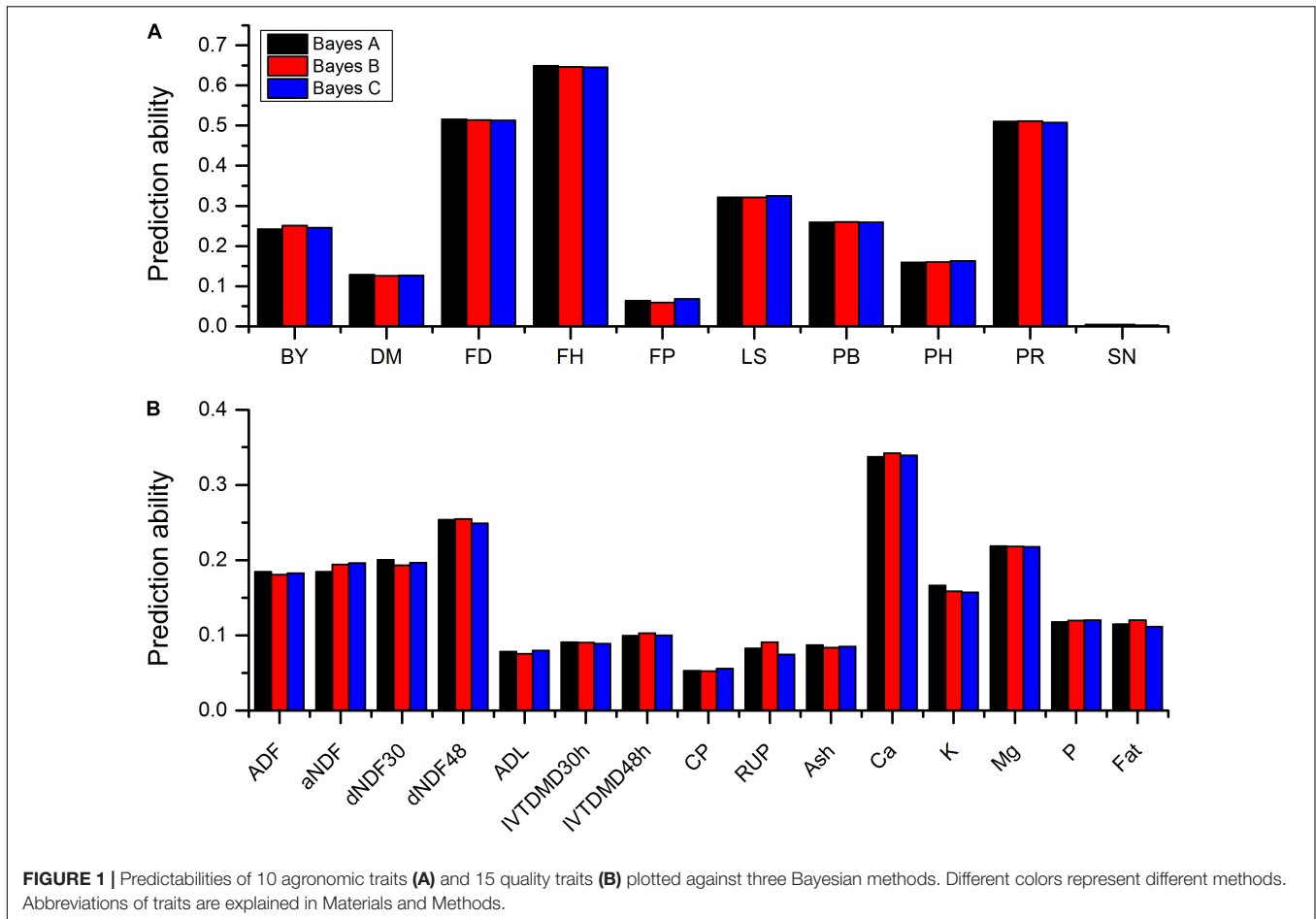
Statistical Methods for GP

Three regression methods with different prior assumptions of the distribution of marker effects were used to estimate SNP effects, namely the BayesA (Meuwissen et al., 2001), BayesB (Meuwissen et al., 2001), and BayesC π (Habier et al., 2011). A ten-fold cross validation was used to evaluate the accuracy of GP. The data were randomly split into 10 approximately equal-sized groups. For each cross validation, nine groups were used as the training population to estimate parameters and the remaining group (validation population) was used as the test sample. The linear model is denoted as follows:

$$y_i = \mu + \sum_{j=1}^m Z_{ij}\alpha_j + e_i$$

where, y_i is the EBV of one trait, μ is the overall mean, m is the number of markers, Z_{ij} is the j th SNP genotype of plant i , α_j is the average effect of allele substitution for SNP j , and e_i is the residual error with an assumed normal distribution $N(0, \sigma^2_e)$. SNP effects were estimated based on the training population using this equation. The GEBV for plant i in the validation population was predicted by summing up SNP effects over all loci. Predictive accuracy was measured as the correlation between the EBVs and GEBVs. Random sampling training and validation sets were repeated 10 times and the mean of correlations was calculated to measure the GP accuracy. All Bayes programs were run in BGLR package in R environment.¹ The number of Burn-in was 10000, thin was 20, and the total number of iteration was 30,000. Other priors of parameters were assigned following Perez and de los Campos (2014).

¹<http://www.r-project.org>



RESULTS

Phenotypic Variation

Since our previous works have described the phenotypic variations of some fiber-related traits (Wang et al., 2016) and crude protein and mineral elements (Jia et al., 2017), we will not describe them in this study. Instead, we want to represent the EBV variations of all traits incorporated in this study. The frequency distributions of EBVs for all 25 traits were symmetric as shown in **Supplementary Figure S1**.

GP Using Three Bayesian Methods

Sequencing of the GBS libraries yielded approximately 184.59 million raw reads and 178.2 million clean reads in all 322 alfalfa genotypes. After imputation, 44,757 high quality SNPs were obtained and used for GP. The results of prediction accuracies of three Bayesian methods are shown in **Table 1**. The predictabilities drawn from the ten-fold cross validation varied across different traits. SN had the lowest predictability (0.0021) but FH had the highest predictability (0.6485). Some quality traits such as crude protein (CP), RUP, and ADL exhibited relatively low prediction accuracies (< 0.1) while the remaining quality traits such as fat, K, and Ca showed low to moderate predictabilities (0.11–0.34). Agronomic traits hold

similar patterns except three traits that had relatively high predictabilities with FH to be the highest (0.65), followed by FD (0.52), and PR (0.51). Other traits, such as LS, PB, and BY displayed moderate predictabilities (0.24–0.32). Similar to BayesA method, BayesB and BayesC π methods did not reveal any significant difference from each other in terms of the predictabilities of all quality and agronomic traits (**Table 1** and **Figure 1**). The predictabilities among the three Bayesian methods are shown in **Figure 1**. From the bar-plotting, only minor differences were observed among the three methods for all 25 traits, it was therefore hard to determine which method was the best.

DISCUSSION

Since GS was proposed by Meuwissen et al. (2001), many studies have been conducted in major crop species (Heffner et al., 2011a,b; Zhao et al., 2013; Iwata et al., 2015; Spindel et al., 2015) and farm animals (Fang et al., 2017; Hay and Roberts, 2017; Tan et al., 2017). The application of GPs to alfalfa BY and forage quality breeding were also initiated recently (Annicchiarico et al., 2015b; Li et al., 2015; Biazzi et al., 2017). In alfalfa industry, BY and forage quality are the key traits for genetic improvement.

Other than the direct traits such as PH, BY, and DM that can inflect the BY of alfalfa, some phenology-related agronomic traits such as FH can also affect the BY. In this study, we therefore investigated the possibility of GP applied to alfalfa germplasm resources and GS applied to 10 important agronomic traits and 15 forage quality traits of alfalfa production.

Several methods, such as random regression BLUP, Bayesian methods and GBLUP, were employed to estimate GP and GS. Some simulation studies on different species suggested Bayesian methods to be superior than GBLUP in terms of the prediction accuracy (Meuwissen et al., 2001; Fernando et al., 2007; Clark et al., 2010; Zhang et al., 2010; Calus and Veerkamp, 2011). Compared with other methods, Bayesian methods also possessed other advantages (Gonzalez-Recio et al., 2010). In this study, we used the empirical data of 25 traits of 322 genotypes of 75 alfalfa accessions to compare the performance of GP following three statistical approaches of BayesA, BayesB, and BayesC π . The BayesA method is based on the assumption that the prior distribution of variances of SNPs followed the scaled inverted chi-square distribution, implicating many SNPs with small effects and a small proportion of SNPs with moderate effects. BayesB assumes that many of the SNPs have no effect and the prior distribution of the variances of SNPs is a mixture of a distribution with zero variance and an inverse chi-squared distribution (Meuwissen et al., 2001). BayesC π , however, treats the prior probability π that a SNP has zero effect as unknown (Habier et al., 2011). **Figure 1** shows that these three Bayesian methods demonstrated very similar prediction accuracies across all 25 traits, irrespective of their different assumptions. BayesA, BayesB, and BayesC π identified six, five, and four quality traits as well as three, four, and three agronomic traits having the highest accuracies, respectively.

Besides the methods of GP discussed above, there are other factors affecting the prediction accuracies. One of them is the population composition and structure. Therefore, EBVs were directly used as the response variable to GP rather than phenotypes in the study. Since EBVs were corrected for non-genetic effects, it can be readily captured by SNPs using the Bayes methods. Methods of imputation for SNP genotypes are also important (Moghaddar et al., 2015).

Compared to previous studies, there were some differences in the accuracies of prediction for both agronomic and quality traits. For example, Biazzi et al. (2017) reported a very low accuracy (~ 0.1) for LS which had nonetheless a moderate value at 0.32 in our study. DM showed a low accuracy (0.13) in our study, but Annicchiarico et al. (2015b) identified a moderate value of 0.35 in two genetically distinguished alfalfa populations. For BY, previous study showed moderate to high accuracies (0.21–0.66, Li et al., 2015) while it had an accuracy at 0.25 in the present study.

REFERENCES

Annicchiarico, P., Barrett, B., Brummer, E. C., Julier, B., and Marshall, A. H. (2015a). Achievements and challenges in improving temperate perennial forage legumes. *Crit. Rev. Plant Sci.* 34, 327–380. doi: 10.1080/07352689.2014.898462

All the 15 quality traits had relatively low prediction accuracies due probably to their low heritabilities (Wang et al., 2016; Jia et al., 2017) determined by the genetic complexity of these traits. Biazzi et al. (2017) detected moderate prediction accuracy values for stem dNDF and leaf protein content (0.3–0.4) followed by leaf ADL and dNDF while the remaining traits showed low to very low accuracies. In our study, the accuracy of dNDF was almost moderate, similar to that of leaf dNDF but slightly lower than stem dNDF. These differences may be attributed to different sizes of reference populations, training populations, and number of markers. Different statistical models may lead to such discrepancies. The methods of imputation of SNP genotypes can also affect the accuracy of prediction (Moghaddar et al., 2015).

The present study was an attempt to predict alfalfa GEBVs of 25 important traits associated with BY and forage quality using three Bayesian statistical methods. Overall, they all exhibited similar predictabilities. Some traits possessed relatively high prediction accuracies (e.g., FH, FD, and PR with accuracies of 0.65, 0.52, and 0.51, respectively). Therefore, it is feasible to apply GS on these traits in alfalfa breeding programs. While GS/GP may be poorly effective for other traits such as ADL, crude protein, and RUP with low prediction accuracies.

AUTHOR CONTRIBUTIONS

ZW designed the experiments. HZ, XW, and GL phenotyped the traits. CJ and FZ analyzed the data and drafted the manuscript. ZW and JH revised the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was supported by the earmarked fund for China Agriculture Research System (CARS34), National Natural Science Foundation of China (No. 31761143013), and Agricultural Science and Technology Innovation Program (No. ASTIP-IAS-10) of China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01220/full#supplementary-material>

FIGURE S1 | Distribution of EBVs for 25 traits. Abbreviations of traits are explained in Materials and Methods.

Annicchiarico, P., Nazzicari, N., Li, X., Wei, Y., Pecetti, L., and Brummer, E. C. (2015b). Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics* 16:1020. doi: 10.1186/s12864-015-2212-y

Bates, D., Mächler, M., and Dai, B. (2011). *lme4: Linear Mixed-Effects Models Using Eigen and S4 Classes*. Available at: <http://lme4.r-forge.r-project.org/>

- Biazzi, E., Nazzicari, N., Pecetti, L., Brummer, E. C., Palmonari, A., Tava, A., et al. (2017). Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS One* 12:e0169234. doi: 10.1371/journal.pone.0169234
- Calus, M. P. L., and Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genet. Sel. Evol.* 43:26. doi: 10.1186/1297-9686-43-26
- Clark, S. A., Hickey, J. M., and Van Der Werf, J. H. J. (2010). How robust are genomic selection methods? *Anim. Prod. Sci.* 50:VIII.
- Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with genotyping-by-sequencing. *G3* 3, 1903–1926. doi: 10.1534/g3.113.008227
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Fang, L., Sahana, G., Ma, P., Su, G., Yu, Y., Zhang, S., et al. (2017). Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics* 18:604. doi: 10.1186/s12864-017-4004-z
- Fernando, R. L., Habier, D., Stricker, C., Dekkers, J. C. M., and Totir, L. R. (2007). Genomic selection. *Acta Agr. Scand. A Anim. Sci.* 57, 192–195. doi: 10.1080/09064700801959395
- Ganal, M. W., Polley, A., Graner, E. M., Plieske, J., Wieseke, R., Luerssen, H., et al. (2012). Large SNP arrays for genotyping in crop plants. *J. Biosci.* 37, 821–828. doi: 10.1007/s12038-012-9225-3
- Gonzalez-Recio, O., Weigel, K. A., Gianola, D., Naya, H., and Rosa, G. J. M. (2010). L-2-Boosting algorithm applied to high-dimensional problems in genomic selection. *Genet. Res.* 92, 227–237. doi: 10.1017/S001667231000261
- Grenier, C., Cao, T. V., Ospina, Y., Quintero, C., Chatel, M. H., Tohme, J., et al. (2015). Accuracy of genomic selection in a rice synthetic population developed for recurrent selection breeding. *PLoS One* 10:e0136594. doi: 10.1371/journal.pone.0136594
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12:186. doi: 10.1186/1471-2105-12-186
- Hay, E. H., and Roberts, A. (2017). Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed. *J. Anim. Sci.* 95, 1467–1471. doi: 10.2527/jas.2016.1355
- Hayes, B., and Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome* 53, 876–883. doi: 10.1139/G10-076
- Heffner, E. L., Jannink, J. L., Iwata, H., Souza, E., and Sorrells, M. E. (2011a). Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* 51, 2597–2606. doi: 10.2135/cropsci2011.05.0253
- Heffner, E. L., Jannink, J. L., and Sorrells, M. E. (2011b). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75. doi: 10.3835/plantgenome2010.12.0029
- Iwata, H., Ebana, K., Uga, Y., and Hayashi, T. (2015). Genomic prediction of biological shape: elliptic fourier analysis and kernel partial least squares (PLS) regression applied to grain shape prediction in rice (*Oryza sativa* L.). *PLoS One* 10:e0120610. doi: 10.1371/journal.pone.0120610
- Jia, C., Wu, X., Chen, M., Wang, Y., Liu, X., Gong, P., et al. (2017). Identification of genetic loci associated with crude protein and mineral concentrations in alfalfa (*Medicago sativa*) using association mapping. *BMC Plant Biol.* 17:97. doi: 10.1186/s12870-017-1047-x
- Jiang, Y., Schulthess, A. W., Rodemann, B., Ling, J., Plieske, J., Kollers, S., et al. (2017). Validating the prediction accuracies of marker-assisted and genomic selection of *Fusarium* head blight resistance in wheat using an independent sample. *Theor. Appl. Genet.* 130, 471–482. doi: 10.1007/s00122-016-2827-7
- Lado, B., Matus, I., Rodríguez, A., Inostroza, L., Poland, J., Belzile, F., et al. (2013). Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3* 3, 2105–2114. doi: 10.1534/g3.113.007807
- Li, X., Wei, Y., Acharya, A., Hansen, J. L., Crawford, J. L., Viands, D. R., et al. (2015). Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. *Plant Genome* 8:90. doi: 10.3835/plantgenome2014.12.0090
- Li, X., Wei, Y., Acharya, A., Jiang, Q., Kang, J., and Brummer, E. C. (2014). A saturated genetic linkage map of autotetraploid alfalfa (*Medicago sativa* L.) developed using genotyping-by-sequencing is highly syntenous with the *Medicago truncatula* genome. *G3* 3, 1971–1979. doi: 10.1534/g3.114.012245
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi: 10.1371/journal.pgen.1003215
- Meuwissen, T. (2007). Genomic selection : marker assisted selection on a genome wide scale. *J. Anim. Breed. Genet.* 124, 321–322. doi: 10.1111/j.1439-0388.2007.00708.x
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Meuwissen, T. H. E., and Goddard, M. E. (1996). The use of marker haplotypes in animal breeding schemes. *Genet. Sel. Evol.* 28, 161–176. doi: 10.1186/1297-9686-28-2-161
- Moghaddar, N., Gore, K. P., Daetwyler, H. D., Hayes, B. J., and Van Der Werf, J. H. J. (2015). Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. *Genet. Sel. Evol.* 47:97. doi: 10.1186/s12711-015-0175-8
- Pace, J., Yu, X. Q., and Lubberstedt, T. (2015). Genomic prediction of seedling root length in maize (*Zea mays* L.). *Plant J.* 83, 903–912. doi: 10.1111/tpj.12937
- Perez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., et al. (2012). Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44, 217–220. doi: 10.1038/ng.1033
- Roberts, A., McMillan, L., Wang, W., Parker, J., Rusyn, I., and Threadgill, D. (2007). Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. *Bioinformatics* 23, i401–i407. doi: 10.1093/bioinformatics/btm220
- Shu, Y. J., Yu, D. S., Wang, D., Bai, X., Zhu, Y. M., and Guo, C. H. (2013). Genomic selection of seed weight based on low-density SCAR markers in soybean. *Genet. Mol. Res.* 12, 2178–2188. doi: 10.4238/2013.July.3.2
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11:e1005350. doi: 10.1371/journal.pgen.1005350
- Sukumaran, S., Crossa, J., Jarquin, D., Lopes, M., and Reynolds, M. P. (2017). Genomic prediction with pedigree and genotype x environment interaction in spring wheat grown in south and west Asia, North Africa, and Mexico. *G3* 3, 481–495. doi: 10.1534/g3.116.036251
- Tan, C., Wu, Z., Ren, J., Huang, Z., Liu, D., He, X., et al. (2017). Genome-wide association study and accuracy of genomic prediction for teat number in duroc pigs using genotyping-by-sequencing. *Genet. Sel. Evol.* 49:35. doi: 10.1186/s12711-017-0311-8
- Wang, Z., Qiang, H., Zhao, H., Xu, R., Zhang, Z., Gao, H., et al. (2016). Association mapping for fiber-related traits and digestibility in alfalfa (*Medicago sativa*). *Front. Plant Sci.* 7:331. doi: 10.3389/fpls.2016.00331

- Xu, S. H., Zhu, D., and Zhang, Q. F. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111
- Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning DJ, and Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One* 5:e12648. doi: 10.1371/journal.pone.0012648
- Zhao, Y. S., Gowda, M., Liu, W. X., Wurschum, T., Maurer, H. P., Longin, F. H., et al. (2013). Choice of shrinkage parameter and prediction of genomic breeding values in elite maize breeding populations. *Plant Breed.* 132, 99–106. doi: 10.1111/pbr.12008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Jia, Zhao, Wang, Han, Zhao, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.