



植物全基因组选择技术的研究进展 及其在玉米育种上的应用

孙 琦¹,李文兰¹,陈立涛²,赵 勠¹,李文才¹,于彦丽¹,孟昭东^{1*}

(1 山东省农业科学院玉米研究所,济南 250100;2 莱阳市种子公司,山东莱阳 265200)

摘要:全基因组选择技术通过全基因组中大量的单核苷酸多态性标记(SNP)和参照群体的表型数据建立BLUP模型估计出每一标记的育种值,称为估计育种值(GEBV),然后仅利用同样的分子标记估计出后代个体育种值并进行选择。该文就近年来国内外有关影响基因组选择效率的主要因素——参考群体的类型与大小、模型的建立方法、标记的类型及其数目、性状遗传力,以及对基因组选择效率的影响等方面的研究进展进行综述,并介绍了全基因组选择技术在玉米育种上应用概况以及对未来的展望。

关键词:全基因组选择;玉米;估计育种值

中图分类号:Q789 **文献标志码:**A

Research Progress on Plant Genomic Selection(GS) and Its Application in Maize Breeding

SUN Qi¹, LI Wenlan¹, CHEN Litao², ZHAO Meng¹, LI Wencai¹,
YU Yanli¹, MENG Zhaodong^{1*}

(1 Maize Institute, Shandong Academy of Agricultural Sciences, Jinan 250100, China; 2 Laiyang City Seed Corporation, Laiyang, Shandong 265200, China)

Abstract:Marker-assisted selection (MAS) technology could realize direct genetic selection, but it must base on QTL mapping. Genomic selection (GS), as the newest MAS method, has much advantage compared to traditional MAS technology, especially QTL mapping not necessary. In this paper, the factors affecting prediction accuracy of GS were reviewed, including training population type, prediction model, marker number, population size, population structure, hereditary of traits and so on. The application of GS in maize breeding was also introduced as well as hybrids performance prediction. We then predicated the future research and application of GS in maize breeding.

Key words:genomic selection (GS); maize; GEBV

With rapid development of the molecular biology and genomics, marker-assisted selection (MAS) emerged as the times require. MAS tech-

nology is as a kind of crop genetic improvement method combining the phenotypic and genetic value, which can realize genetic direct selection and effec-

tive polymerization^[1]. When complex traits controlled by multiple genes need to be improved, MAS has two aspects of flaws. First, selection of the progeny population is established on the quantity traits location (QTL) mapping. But the result of QTL mapping basing on the bi-parental populations has no universality and couldn't be applied accurately in breeding^[2]. Second, the important traits were controlled by lots of small effective genes, lack of appropriate statistic method and breeding technology which will apply quantity genes to complex traits improvement^[3]. New MAS technology-genomic selection (GS) emerged as the times require.

1 Origination and advantage of genomic selection (GS)

Meuwissen first put forward genomic selection (GS) breeding strategy. GS uses a “training population” of individuals that have been genotyped and phenotyped. Best linear unbiased prediction (BLUP) model is established on the basis of the genotyped result of an individual and its breeding value (Mean performance of crosses with same tester) for the training population. The breeding value of “Candidate population” is estimated by BLUP model and genotypic data, without cross to tester and phenotypes record^[4]. BLUP model takes genotypic data of untested individuals and produces genomic estimated breeding values (GEBVs). These GEBVs say nothing of the function of the underlying genes as the ideal selection criterion^[5]. Genomic selection basis of GEBVs is superior to traditional breeding for increasing gains per unit time even if both models show the same efficiency. In principle, phenotypes value of the candidate individuals is non-essential for the selection, hence shortening the length of the breeding cycle^[6].

Genomic selection have several merits compared to the traditional MAS. (1) QTL mapping is not necessary for GS. Genomic selection differs from previous strategies such as linkage and association mapping in that it abandons the objective to

map the effect of single gene and instead of focusing on the efficient estimation of breeding values on the basis of a large number of molecular markers, ideally covering the full genome^[5]. (2) Genomic selection is more precision especially for early selection. Genotyping uses high density molecular markers which can estimate all of the QTL effects and explain the genetic variance for most of the traits. But MAS only uses several markers in traits selection. So genomic selection is more accurate than MAS^[7]. (3) Genomic selection can shorten generation interval, accelerate genetic progress and reduce production cost. Genetic progress of GS is more than phenotypic selection 4%—25%. Cost of GS is less than traditional breeding 26%—56%^[8]. (4) Selection efficiency of low heritability traits is higher for GS than MAS. (5) The criterion of GS is breeding value, sum of all of the allele genetic effects for each individual. It is judged by the mean performance of its cross progeny, not the performance of itself. So GS is more accurate^[9].

Genomic selection originated from animal breeding during last century. It has been widely used in dairy cattle breeding in America, Australia, New Zealand and so on^[10-11]. It was also applied in broiler chickens and pigs breeding^[12-13]. GS' application in plant breeding was developed in recent years, which focused on simulation studies. It is used in maize^[14], wheat^[15], tree^[16], sugar beet^[17], Barley^[18], triticale^[19] and so on.

Empirical study is performed in larger companies such as Monsanto and Pioneer-Dupond. Mark Sorrells and Jean-Luc Jannink are trying to use GS to increase the speed of variety improvement 3—4 times. The work is carried out with CYMMIT and performed four aspects to improve the yield of maize and wheat^[20].

Under the above context, the objective of this study is to review the essential factors affecting the GS in plant breeding. Maize is essential for global food security. More research of genomic selection on maize lauched in recent years^[21-23]. The paper will introduce the advance on the application of GS

in maize breeding. We then put forward the future research which should be carried out in maize breeding in China.

2 Affecting factors of genomic selection

Factors that affect GS prediction accuracy of include the number of markers used for estimating the GEBVs^[10], trait heritability^[7], calibration population size^[5], statistical models^[24], number and type of molecular markers^[25-26], linkage disequilibrium^[27], effective population size^[28], relationship between calibration and test set (TS)^[29-31] and population structure^[32-34].

2.1 Training population of genomic selection

In animal breeding, we only discussed GS in the context of population-wide linkage disequilibrium, where the population might be defined as an entire breed of cattle, pig, or chicken. The need for high marker densities in GS may be reduced if the candidate population consists of progeny of the training population. In that case, an evenly spaced low-density subset of the markers typed on the training population can be used on the candidates, and scores for the full complement of markers can be inferred by cosegregation^[35]. Because plants often produce very large full sibships (an F_2 population derived from a single F_1 by selfing is an example of such a sibship), however, there is also a tradition of QTL detection, MAS and GS within such sibships^[5]. Bernardo compared F_2 , BC_1 , and BC_2 populations from an adapted \times exotic maize cross as training population in the simulation experiment^[14]. The result indicates that genomewide selection should start at F_2 rather than backcross population, even when the number of favorable alleles is substantially larger in the adapted parent than in the exotic parent. Compared to natural populations, genetic basis of F_2 populations is simpler because F_2 populations derive from only two inbred lines. So the biparental population size might be smaller than that of natural populations. Simulation studies have previously indicated that for three cycles of genomewide selection in an a-

dapted \times exotic cross, a population size of $NC_0 = 144$ was generally sufficient^[21]. Low density markers are suitable to F_2 populations^[22]. But two disadvantages of F_2 populations exist. Biparental population requires separate model for training within each cross. The BLUP model is only suit for the progenies selection from the two parental lines. The progeny of F_2 population must be selected by the phenotypic value of F_3 testcrosses. Following progeny selection may be only according to BLUP model after F_3 .

F_2 as training population often be suit for cross-pollinated plant such as maize. Yusheng Zhao based on experimental data of six segregating populations from a half-diallel mating design with 788 testcross progenies from an elite maize breeding program^[23]. In the study of Vannesa *et al.*^[36], marker effects estimated in 255 diverse maize hybrids were used to predict grain yield, anthesis date, and anthesis-silking interval within the diversity panel and testcross progenies of 30 F_2 -derived lines from each of five populations.

Wegenast *et al.* suggested that genomic selection was applied in plant breeding, however, not only within a specific bi-parental cross or within a diverse panel of elite lines but also rather within and among crosses^[37]. Self-pollination plant often adopt natural population such as wheat or sugar. Würschum et al used 924 sugar beet lines as training population. The results suggest that a training population derived from intensively phenotyped and genotyped diverse lines from a breeding program does hold potential to build up robust calibration models for genomic selection^[17]. Hans *et al.* accessed the accuracy of GEBVs for rust resistance in 206 hexaploid wheat landraces^[15].

2.2 Prediction model of genomic selection

Genomic selection modeling takes advantage of the increasing abundance of molecular markers through modeling of many genetic loci with small effects^[26,35,38]. Over the last decade, simulation and empirical cross-validation studies in plants have shown GS is more effective than traditional MAS strategies that use only a subset of markers

with significant effects^[5-7,39].

Estimation methods of allelic effects include least squares regression^[40], ridge regression BLUP (RR-BLUP), principle component analysis^[41-42] and Bayes regression^[43]. In essence for least squares, chromosome fragments or markers are selected associated to the traits by genome-wide association studies (GWAS) at the same time and then the effect of the fragments is estimated^[44]. RR-BLUP method regards the fragment effects as random effects. The marker effect was estimated by linear mixed models. The sum of fragments effect is breeding value for an individual^[43]. Bayes methods combines the prior distribution of marker effect variance and data collection. Frequently used Bayes methods conclude Bayes A and Bayes B. Main difference between Bayes A and Bayes B is that Bayes A permits different variance for different markers and Bayes B permits that the variance of some markers is zero^[45].

Simulation studies show that the prediction accuracy of Bayes method is best and least squares is weakest. The accuracy rate of RR-BLUP is slightly smaller than Bayes A. Even so, RR-BLUP has four aspects superior to Bayesian method. First, Bayesian method is complex and need super computer. But computer requirement is lower and calculation speed is higher for RR-BLUP. Marker effects are estimated by RR-BLUP in SAS PROC IML^[46]. Second, prediction within families was more accurate in BLUP than Bayes B. Regression coefficient b of RR-BLUP is nearer to 1 than Bayes A^[47]. Habier *et al.* showed that RR-BLUP is more effective at capturing genetic relationships because it fits more markers into the prediction Model^[27]. In contrast, Bayes B is more effective at capturing LD between markers and QTL. Third, RR-BLUP is more accurate than other method when the number of QTLs increases or the heredity is higher^[18]. Fourth, BLUP led to lower inbreeding and a smaller reduction of genetic variance compared to Bayes and PLS^[48]. From above, we can conclud that BLUP methods is better than Bayesian regression for plant models.

In addition, machine-learning methods also can be used to predict the marker effect, including support vector machine (SVM), booting and random forest (RF). Ongutu *et al.* compared these methods for genomic selection. The result shows that the correlation between the predicted and true breeding values is 0.547 for boosting, 0.497 for SVMs, and 0.483 for RF, indicating better performance for boosting than for SVMs and RF^[49].

2.3 Other factors affecting prediction accuracy

In genome-wide selection methods, prediction accuracy is affected by population size (N), average hereditary of traits (h^2) and marker numbers (N_M)^[50]. Simulation studies showed that the population structure is also crucial for the prediction accuracy in genomic selection^[27].

Prediction accuracy increases with markers density. Markers number on a certain length genome also directly affects total information of genetic markers. If SSR markers density increases from 0.25 $Ne/morgan$ (Ne , effective population size) to 2 $Ne/morgan$, prediction accuracy will be improved from 0.63 to 0.83. If SNP markers density increases from 1 $Ne/morgan$ to 8 $Ne/morgan$, prediction accuracy will be improved from 0.69 to 0.86. Even at the highest tested densities of 2 Ne SSR markers per Morgan or 8 Ne SNP markers per Morgan, accuracy had not reached a plateau^[5]. Meanwhile, more markers number, more easy to get the Linkage disequilibrium (LD) markers. Emily found that in the biparental populations, there was no consistent gain in genome-wide prediction (rmp) from increasing marker density above one marker per 12.5 cM^[22]. Zhao *et al.* revealed that the accuracy was nearly reaching a plateau at 800 SNPs when the number of markers varied from 100 to 800^[23]. The reason is that genome is sufficiently saturated with markers when the prediction accuracy arrives at a plateau^[28,50]. The number of markers needed for accurate predictions of genotypic values depends on the extent of linkage disequilibrium (LD) between markers and QTL^[4] and also on the germplasm under consideration^[18].

Different marker type has different polymorphism information content (PIC). Comparing SSR and SNP markers, they found that for similar accuracies, the SNP markers required a density of 2 to 3 times that of the SSR^[5].

Simulation studies showed that the population size is crucial for the prediction accuracy in genomic selection^[27]. The result of Emily *et al.* indicated that prediction accuracy rmp increased as population size N increased. In the biparental maize population and with the highest markers number N_M , (1 213 markers) and hereditary $h^2 = 0.30$, the prediction accuracy for grain yield was $rmp = 0.19$ with $N = 48$, $rmp = 0.26$ with $N = 96$, and $rmp = 0.33$ with $N = 192$ ^[22]. Zhao Yusheng observed a monotonic increase in the prediction accuracy for grain yield with increasing population size without any substantial decrease in the slope^[23]. The study of Bernardo also indicated that larger population size would get higher prediction precision^[14]. But F_2 population size of $NC_0 = 144$ was generally sufficient^[21].

Training population structure is also an important factor affecting prediction accuracy of genomic selection for multi-parental populations. Training population structure set methods conclude random sampling, unidirectional sampling (selecting individuals with highest genotypic values), bi-directional sampling (selecting individuals with highest or lowest genotypic values)^[50-51]. This bi-directional selection showed to be much more powerful than random sampling^[52]. Yusheng Zhao observed a substantial loss in the accuracy to predict genomic breeding values in unidirectional selected populations. Bidirectional selection is a valuable approach to efficiently implement genomic selection in applied plant breeding programs^[53].

For the same trait within the same population, prediction accuracy (rmp) will remain unchanged for different combinations of population size (N) and trait hereditary (h^2). Decrease on h^2 can be compensated by a proportional increase in N (and vice versa) so that rmp is maintained. On the other hand, traits with initially low h^2 can be evalua-

ted with larger N or the h^2 for a subset of traits can be increased by the use of additional testing resources. Different traits, however, vary in their prediction accuracy even when N , h^2 , and N_M (markers number) are constant. Yield traits had lower prediction accuracy than other traits despite the constant N , h^2 , and N_M . Simulation results indicated that rmp is also lowest for yield traits even when its h^2 is as high as other traits. Plant height and lodging are always predicted most accurately followed by flowering time^[22]. Empirical evidence and experience on the predictability of different traits are necessary in designing training populations.

3 Genomic selection in maize breeding

3.1 Origination of GS in maize

The key technology of GS is the maize hybrid prediction by BLUP model with markers effects or coefficient of parentage. It was used to predict the single-cross performance in maize hybrid breeding at first. The BLUP model is established based on the tested hybrids data and the markers information of their parents. The performance of untested hybrids is predicted by the BLUP model and the markers data of the parents^[54].

Bernardo devoted himself to hybrids prediction by BLUP model in maize^[55-58]. The coefficient of relative between theory and actual observation was $0.688 \sim 0.800$ by RFLP markers^[54]. BLUP is suitable for hybrid performance prediction since the trait only has moderate heritability. Prediction accuracy of molecular marker effects is higher than phylogenetic relationship^[58]. With the development of molecular markers, new molecular marker type emerged. Simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) were widely used. Manje Gowda *et al.* found that prediction accuracy of flower time and plant height was above 0.8 with SSR markers in maize^[19]. Research of Massman *et al.* indicated that prediction accuracy of grain yield was 0.8, and root lodging ratio was 0.87 using SSR markers^[59]. But the prediction

effect of grain yield was only 0.50~0.66, and root lodging ratio was only 0.31~0.45 with coefficient of parentage^[55]. Then it indicated that molecular markers was more suitable for hybrid performance prediction than coefficient of parentage.

Then scientists found that BLUP was not only used to hybrid performance prediction, but also the breeding value of individuals among the maize population. So BLUP was used to individuals selection of F_2 population in selection and breeding of inbred lines. Hybrid performance prediction lay the foundation for the genome-wide selection in maize.

3.2 Application of genomic selection in maize

Bernardo's laboratory began to study applying GS to maize breeding in Minnesota University of America^[21]. They did plenty of simulation and empirical experiments. Piepho in German and Robert in Brazil also tried to study using GS in maize breeding^[60-61]. GS utility in maize breeding consist of two sides, hybrids performance prediction and improvement of inbred lines. He devoted to inbred lines improvement using GS. The BLUP model of biparental populations from two inbred lines is only suit for the progeny of the parents. Genomewide selection as proposed in maize involves two steps^[21]. First, a segregating maize population is genotyped and evaluated for testcross performance of F_3 family. Based on the genotypic and phenotypic data, breeding values associated with a large set of markers (e. g. , 256 to 512 markers) are calculated for the traits of interest. Significance tests for markers are not used, and the effects of all markers are fitted as random effects in a linear model by best linear unbiased prediction (BLUP). Second, two or three generations of selection based on all markers are conducted in a year-round nursery (e. g. , Hawaii or Puerto Rico) or greenhouse. Trait values are predicted as the sum of an individual plant's marker values across all markers, and selection is subsequently based on these genome-wide prediction. According to the steps, Emily (2013b) introgressed semidwarf germplasm to U. S. Corn belt inbred and found that genomewide selection from Cycle 1 until Cycle 5 either maintained

or improved on the gains from phenotypic selection achieved in Cycle 1^[62].

The results of Bernardo indicated that a useful strategy for the rapid improvement of an adapted×exotic cross involves 7 to 8 cycles of genomewide selection starting in the F_2 ^[14]. Benjamin *et al.* demonstrated that progressive selfing had a significant and positive impact on genomic selection gains. In particular, selfing to the F_8 produced a 72% increase over F_2 gains^[63]. However, most of the gains are realized by the F_5 generation (95% of the F_8 gains). Also note that the F_8 and DH performed similarly, consistent with previous observations^[64].

In the research of Bernardo, the training population is the specific bi-parental populations from the two parental lines, so the BLUP model is suit for the progeny of the two inbred lines. Other experiments of GS in maize are about multi-parental populations as training population. Study of Yusheng Zhao was based on experimental data of six segregating populations from a half-diallel mating design. As for maize up to three generations are feasible per year, selection gain per unit time is high and, consequently, genomic selection holds great promise for maize breeding programs^[23]. These result of the study might be as genomic prediction model for further breeding elite maize lines between the six populations. In the study of Vanessa *et al.* , marker effects estimated in 255 diverse maize hybrids were used to predict grain yield, anthesis date, and anthesis-silking interval within the diversity panel and testcross progenies of 30 F_2 -derived lines from each of five populations^[36]. Potential uses for genomic prediction in maize hybrid breeding are discussed emphasizing the need of (1) a clear definition of the breeding scenario in which genomic prediction should be applied (i. e. , prediction among or within populations), (2) a detailed analysis of the population structure before performing cross validation, and (3) larger training sets with strong genetic relationship to the validation set.

4 Future research in maize breeding

GS is just beginning to be implemented, but it will take long time to be used in maize breeding. In previous study, training population was only from several inbred lines, even if two inbred lines. It couldn't be implemented by other breeding program. Future research should focus on two sides of work. First, we should commit to build a generalized prediction model for some kinds of inbred lines such as yield, quality and so on. But these traits were complex composed of a great deal of genes. Traditional MAS technology couldn't realize the traits selection in maize breeding. 973 Plan "Basic study on breeding of genome-wide selection of yield

and quality traits in maize" has been carried out in 2014. The plan will systematically analyze the genetic basis of maize yield and quality, and then build genome-wide selection breeding model. It will afford new technology for maize breeding. Second, in China, abiotic stress tolerance also reduces the yield seriously in maize especially drought tolerance. Drought is the foremost factor restricting maize production, often resulting in 20–50% maize yield reduction every year in China^[65]. If we establish prediction model of drought tolerance, it will afford the theory and technology support of maize breeding. Consequently, our research team will carry out study on the genomic selection program of drought tolerance.

References:

- [1] STUBER C W, POLACCO M, SENIOR M L. Synergy of empirical breeding, marker-assisted selection, and genomics to increase crop yield potential [J]. *Crop Science*, 1999, 39: 1 571-1 583.
- [2] MOOSE S P, MUMM R H. Molecular plant breeding as the foundation for 21st century crop improvement [J]. *Plant Physiology*, 2008, 147: 969-977.
- [3] BERNARDO R. Molecular markers and selection for complex traits in plants: learning from the last 20 years [J]. *Crop Science*, 2008, 48:1 649-1 664.
- [4] MEUWISSENT H, HAYES B J, GODDARD M E. Prediction of total genetic value using genome-wide dense marker maps [J]. *Genetics*, 2001, 157: 1 819-1 829.
- [5] JANNINK J L, LORENZ A J, IWATA H. Genomic selection in plant breeding: from theory to practice [J]. *Briefings in Functional Genomics*, 2010, 9(2):166-177.
- [6] HEFFNER E L, JANNINK J L, IWATA H, et al. Genomic selection accuracy for grain quality traits in biparental wheat populations [J]. *Crop Science*, 2011, 51: 2 597-2 606.
- [7] HEFFNER E L, SORRELLS M E, JANNINK J L. Genomic selection for crop improvement [J]. *Crop Science*, 2009, 49: 1-12.
- [8] MAYOR P J, BERNARDO R. Genomewide selection and marker-assisted recurrent selection in doubled haploid versus F₂ populations [J]. *Crop Science*, 2009, 49:1 719-1 725.
- [9] MASSMAN J M, JUNG H J G, BERNARDO R. Genome-wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol
- [10] SCHAEFFER L R. Strategy for applying genome-wide selection in dairy cattle [J]. *Journal of Animal Breeding Genetic*, 2006, 123: 218-223.
- [11] GODDARD M E, HAYES B J. Genomic selection [J]. *Journal of Animal Breeding Genetics*, 2007, 124: 323-330.
- [12] DAETWYLER H D, VILLANUEVA B, BIJMA P. Inbreeding in genome-wide selection [J]. *Journal of Animal Breeding Genetics*, 2007, 124: 369-376.
- [13] TU L, WOOLLIAMS J A, SIGBJORN L. The accuracy of genomic selection in norwegian red cattle assessed by cross validation [J]. *Genetics*, 2009, 183: 1 119-1 126.
- [14] BERNARDO R. Genomewide selection for rapid introgression of exotic germplasm in maize [J]. *Crop Science*, 2009, 49: 419-425.
- [15] HANS D D, BANSAL U K, BARIANA H S, et al. Genomic prediction for rust resistance in diverse wheat landraces [J]. *Theory and Applied Genetics*, 2014, 127: 1 795-1 803.
- [16] MARIE D, BOUVET J M. Genomic selection in tree breeding: testing accuracy of prediction models including dominance effect [J]. *BMC Proceedings*, 2011, 5(Suppl 7): 1-2.
- [17] WÜRSCHUM T, REIF J C, KRAFT T, et al. Genomic selection in sugar beet breeding populations [J]. *BMC Genetics*, 2013, 14: 85-92.
- [18] ZHONG S Q, DEKKERS J C M, FERNANDO R L, et al. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study [J]. *Genetics*, 2009, 182(1): 355-364.
- [19] GOWDA M, ZHAO Y S, MAURER H P, et al. Best linear

unbiased prediction of triticale hybrid performance[J]. *Eu-phytica*, 2013, 191: 223-230.

[20] 吴永升, 邵俊明, 周瑞阳, 等. 植物数量性状全基因组选择研究进展[J]. 西南农业学报, 2012, 25(4): 1 510-1 514.

WU Y S, SHAO J M, ZHOU R Y, et al. Reviews of genome-wide selection for quantitative traits in plants [J]. *Southwest China Journal of Agricultural Sciences*, 2012, 25(4): 1 510-1 514.

[21] BERNARDO R, YU J. Prospects for genome-wide selection for quantitative traits in maize[J]. *Crop Science*, 2007, 47: 1 082-1 090.

[22] EMILY C, BERNARDO R. Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers[J]. *Plant Genome*, 2013a, 6(1): 1-7.

[23] ZHAO Y S, GOWDA M, LIU W X, et al. Accuracy of genomic selection in European maize elite breeding populations [J]. *Theoretical and Applied Genetics*, 2012a, 124: 769-776.

[24] HESLOT N, YANG H P, SORRELLS M E, et al. Genomic selection in plant breeding: a comparison of models[J]. *Crop Science*, 2012, 52: 146-160.

[25] CHEN X, SULLIVAN P F. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput [J]. *Pharmacogenetics*, 2003, 3: 77-96.

[26] POLAND J, RIFE T W. Genotyping-by-sequencing for plant breeding and genetics[J]. *Plant Genetics*, 2012, 5: 92-102.

[27] HABIER D, FERNANDO R L, DEKKERS J C M. The impact of genetic relationship information on genome-assisted breeding values[J]. *Genetics*, 2007, 177: 2 389-2 397.

[28] DAETWYLER H D, VILLANUEVA B, WOOLLIAMS J A. Accuracy of predicting the genetic risk of disease using a genome-wide approach[J]. *PLoS One*, 2008, 3: 3 395.

[29] ALBRECHT T, WIMMER V, AUINGER H J, et al. Genome-based prediction of testcross values in maize[J]. *Theoretical and Applied Genetics*, 2011, 123: 339-350.

[30] CLARK S, HICKEY J, WERF J. Different models of genetic variation and their effect on genomic evaluation[J]. *Genetic Selection Evolution*, 2011, 43: 18.

[31] PSZCZOLA M, STRABEL T, MULDER H A, et al. Reliability of direct genomic values for animals with different relationships within and to the reference population[J]. *Journal of Dairy Science*, 2012, 95z: 389-400.

[32] SAATCHI M, MCCLURE M C, MCKAY S D, et al. Accuracies of genomic breeding values in American Angus beef cattle using k-means clustering for cross-validation[J]. *Genetic Selection Evolution*, 2011, 43: 40.

[33] WINDHAUSEN V S, ATLIN G N, CROSSA J, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments[J]. *Genes Genomes Genetic*, 2012, 2:1 427-1 436.

[34] GUO Z, TUCKER D M, BASTEN C J, et al. The impact of population structure on genomic prediction in stratified populations[J]. *Theoretical and Applied Genetics*, 2014, 127: 749-762

[35] HABIER D, FERNANDO R L, DEKKERS J C M. Genomic selection using low-density marker panels [J]. *Genetics*, 2009, 182: 343-353.

[36] VANESSA S W, ATLIN G N, HICKEY J M, et al. Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments[J]. *Genomic Selection*, 2012, 2(14):1 427-1 436.

[37] WEGENAST T, LONGIN C F H, UTZ H F, et al. Hybrid maize breeding with doubled haploids IV. Number versus size of crosses and importance of parental selection in two-stage selection for testcross performance[J]. *Theoretical and Applied Genetics*, 2008, 117: 251-260.

[38] SOLBERG T R, SONESSON A K, WOOLLIAMS J A, et al. Genomic selection using different marker types and densities[J]. *Journal of Animal Breeding Genetics*, 2008, 86(10): 2 447-2 454.

[39] LORENZANA R E, BERNARDO R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations [J]. *Theoretical and Applied Genetics*, 2009, 120:151-161.

[40] WOLD H, JOHNSON NL, KOTZ S. Partial least squares [C]. *Encyclopedia of Statistical Science*. New York: Wiley, 1985:581-91.

[41] SOLBERG T R, SONESSON A K, WOOLLIAMS J A, et al. Reducing dimensionality for prediction of genome-wide breeding values[J]. *Genetic Selection Evolution*, 2009, 41(1): 29.

[42] CROSSA J, CAMPOS G, PÉREZ P, et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers[J]. *Genetics*, 2010, 186(2): 713-724.

[43] MEUWISSEN T H E, SOLBERG T R, SHEPHERD R, et al. A fast algorithm for Bayes B type of prediction of genome-wide estimates of genetic value[J]. *Genetic Selection Evolution*, 2009, 41:2.

[44] WANG W Y S, BARRATT B J, CLAYTON D G, et al. Genome wide association studies: theoretical and practical concerns[J]. *Nature Review Genetics*, 2005, 6(2): 109-118.

[45] 李恒德, 包振民, 孙效文. 基因组选择及其应用[J]. 遗传, 2011, 33(12): 1 308-1 316.

LI H D, BAO Z M, SUN X W. Genomic selection and its application[J]. *Hereditas*, 2011, 33(12): 1 308-1 316.

[46] SAS Institute. The SAS system for Windows. Release 9.2.

SAS Inst. , Cary, NC, 2009.

[47] LUND M S, SAHANA G, KONING D J, *et al.* Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection[J]. *BMC Proceedings*, 2009, 3(Suppl. 1):S1.

[48] BASTIAANSEN J W, COSTER A, CALUS M P, *et al.* Long-term response to genomic selection: effects of estimation method and reference population structure for different genetic architectures[J]. *Genetics Selection Evolution*, 2012, 4: 3-16.

[49] OGUTU J O , PIEPHO H P, TORBEN S S. A comparison of random forests, boosting and support vector machines for genomic selection[J]. *BMC Proceedings*, 2011, 5(Suppl 3): S11.

[50] DAETWYLER H D, WONG R P, VILLANUEVA B, *et al.* The impact of genetic architecture on genome-wide evaluation methods[J]. *Genetics*, 2010, 185: 1 021-1 031.

[51] JULIO I, JANNINK J L, AKDEMIR D, *et al.* Training set optimization under population structure in genomic selection [J]. *Theoretical and Applied Genetics*, 2014, 128 (1): 145-158.

[52] NAVABI A, MATHER D E, BERNIER J, *et al.* QTL detection with bidirectional and unidirectional selective genotyping: marker-based and trait-based analyses[J]. *Theoretical and Applied Genetics*, 2009, 118: 347-358.

[53] ZHAO Y S, GOWDA M, LONGIN F H, *et al.* Impact of selective genotyping in the training population on accuracy and bias of genomic selection[J]. *Theoretical and Applied Genetics*, 2012b, 125: 707-713.

[54] 翟虎渠, 王建康. 应用数理遗传学[M]. 北京:中国农业科学出版社,2007:185-204.

ZHAI H Q, WANG J K. Applied Quantitative Genetics [M]. Beijing: China Agricultural Science and Technology Publishing House, 2007: 185-204.

[55] BERNARDO R. Prediction of maize single-cross performance using RFLPs and information from related hybrids[J]. *Crop Science*, 1994, 34: 20-25.

[56] BERNARDO R. Genetic models for predicting maize single-cross performance in unbalanced yield trial data[J]. *Crop Science*, 1995, 35: 141-147.

[57] BERNARDO R. Best linear unbiased prediction of maize single cross performance[J]. *Crop Science*, 1996, 36: 50-56.

[58] BERNARDO R. Marker-assisted best linear unbiased prediction of single-cross performance[J]. *Crop Science*, 1999, 39: 1 277-1 282.

[59] MASSMAN J M, GORDILLO A, LORENZANA R E, *et al.* Genomewide predictions from maize single-cross data[J]. *Theoretical and Applied Genetics*, 2013, 126:13-22.

[60] PIEPHO H P. Ridge regression and extensions for genome-wide selection in maize[J]. *Crop Science*, 2009, 49:1 165-1 176.

[61] ROBERTO F N, JULIO C D, ÉDER C M L, *et al.* Genome wide selection in for tropical maize root traits under conditions of nitrogen and phosphorus stress[J]. *Acta Scientiarum*, 2012, 34(4):389-395.

[62] EMILY C, BERNARDO R. Genomewide selection to introgress semidwarf maize germplasm into U. S. Corn Belt inbreds[J]. *Crop Science*, 2013b, 53: 1 427-1 436.

[63] BENJAMIN M, COMBE J L, TANKSLEY S D. Selfing for the design of genomic selection experiments in biparental plant populations [J]. *Theoretical and Applied Genetics*, 2013, 126: 2 907-2 920.

[64] BORDES J, CHARMET G, VAULX R D, *et al.* Doubled-haploid versus single-seed descent and S1-family variation for testcross performance in a maize population[J]. *Euphytica*, 2007, 154: 41-51.

[65] HU, R F, MENG E C, ZHANG S H, *et al.* Prioritization for maize research and development in China[J]. *Scientia Agricultura Sinica*, 2004, 37: 781-787.

(编辑:宋亚珍)