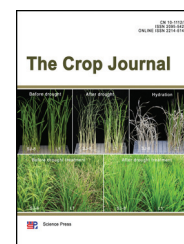
Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect



## Factors affecting genomic selection revealed by empirical evidence in maize

Xiaogang Liu<sup>a</sup>, Hongwu Wang<sup>a</sup>, Hui Wang<sup>a</sup>, Zifeng Guo<sup>a</sup>, Xiaojie Xu<sup>a</sup>, Jiacheng Liu<sup>a</sup>, Shanhong Wang<sup>a</sup>, Wen-Xue Li<sup>a</sup>, Cheng Zou<sup>a</sup>, Boddupalli M. Prasanna<sup>c</sup>, Michael S. Olsen<sup>c</sup>, Changling Huang<sup>a,\*</sup>, Yunbi Xu<sup>a,b,\*\*</sup>

<sup>a</sup>Institute of Crop Science, National Key Facility of Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>b</sup>International Maize and Wheat Improvement Center (CIMMYT), El Batán 56130, Texcoco, Mexico

<sup>c</sup>International Maize and Wheat Improvement Center (CIMMYT), ICRAF campus, United Nations Avenue, Nairobi, Kenya

### ARTICLE INFO

#### Article history:

Received 28 December 2017

Received in revised form 5 March 2018

Accepted 17 March 2018

Available online xxxx

#### Keywords:

Marker density

Population size

Population relationship

*Zea mays*

### ABSTRACT

Genomic selection (GS) as a promising molecular breeding strategy has been widely implemented and evaluated for plant breeding, because it has remarkable superiority in enhancing genetic gain, reducing breeding time and expenditure, and accelerating the breeding process. In this study the factors affecting prediction accuracy ( $r_{MG}$ ) in GS were evaluated systematically, using six agronomic traits (plant height, ear height, ear length, ear diameter, grain yield per plant and hundred-kernel weight) evaluated in one natural and two biparental populations. The factors examined included marker density, population size, heritability, statistical model, population relationships and the ratio of population size between the training and testing sets, the last being revealed by resampling individuals in different proportions from a population. Prediction accuracy continuously increased as marker density and population size increased and was positively correlated with heritability;  $r_{MG}$  showed a slight gain when the training set increased to three times as large as the testing set. Low predictive performance between unrelated populations could be attributed to different allele frequencies, and predictive ability and prediction accuracy could be improved by including more related lines in the training population. Among the seven statistical models examined, including ridge regression best linear unbiased prediction (RR-BLUP), genomic BLUP (GBLUP), BayesA, BayesB, BayesC, Bayesian least absolute shrinkage and selection operator (Bayesian LASSO), and reproducing kernel Hilbert space (RKHS), the RKHS and additive-dominance model (Add + Dom model) showed credible ability for capturing non-additive effects, particularly for complex traits with low heritability. Empirical evidence generated in this study for GS-relevant factors will help plant breeders to develop GS-assisted breeding strategies for more efficient development of varieties.

© 2018 Crop Science Society of China and Institute of Crop Science, CAAS. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Corresponding author.

\*\* Correspondence to: Y. Xu, Institute of Crop Science, National Key Facility of Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

E-mail addresses: [huangchangling@caas.cn](mailto:huangchangling@caas.cn) (C. Huang), [y.xu@cgiar.org](mailto:y.xu@cgiar.org) (Y. Xu).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

<https://doi.org/10.1016/j.cj.2018.03.005>

Please cite this article as: X. Liu, et al., Factors affecting genomic selection revealed by empirical evidence in maize, *The Crop Journal* (2018), <https://doi.org/10.1016/j.cj.2018.03.005>

## 1. Introduction

Genomic selection (GS), or genome-wide selection, has become increasingly important in the field of molecular plant breeding with development of high-throughput, cost-effective genotyping technology [1–3]. The GS strategy based on genome-wide polymorphic markers was initially proposed in 2001 with several publications of statistical models [4]. GS can be applied in improving preselection accuracies for complex agronomic traits based on genomic information. The hypothesis of GS primarily depends on the fact that chromosome segments contributing to phenotypic variation are in high linkage disequilibrium (LD) with a minimum of one marker locus within the genome [4]. The GS procedure, it usually utilizes both phenotypic and genotypic data from a training population (TP) to train a statistical model, which can be used to estimate genomic-estimated breeding values (GEBVs) for precise selection of each individual from a candidate (breeding) population that is genotyped without phenotyping [5]. The GS strategy obviously has advantages in comparison with previous molecular breeding technologies such as marker-assisted selection (MAS) and marker-assisted recurrent selection (MARS), which depend on identified significant markers, tagged genes or mapped quantitative trait loci (QTL). Moreover, MAS or MARS has some shortcomings because the search for significant marker-QTL associations has low power in capturing genes with minor effects [6,7]. GS removes the requirement to unearth QTL, and can directly estimate all marker effects in whole genome and capture genetic loci with minor effects for complex traits [8,9]. Beyond that, GS can substantially enhance the rate of annual genetic gain by accelerating breeding cycles and by reducing time and cost because selection of the candidate population depends only on genotypes of individuals without need for phenotypic records [10,11].

GS has long been practiced in livestock and animal breeding [12–14]. For plant breeding, simulation analysis and empirical evaluation of prediction accuracy of GS had been accomplished by using cross-validation method in the experimental populations of *Arabidopsis* [15], rice [16,17], wheat [18,19], barley [20,21], maize [3,22–24], and forest trees [25–27]. Prediction accuracy ( $r_{MG}$ ) is regarded as a vital parameter to evaluate the performance of GS in breeding programs. It is usually defined as Pearson's correlation ( $r$ ) between the true breeding value and the GEBVs of candidate individuals. The factors affecting the estimate accuracy of GEBV will have an influence on prediction ability of GS. These key factors are more or less interrelated in a comprehensive manner. They generally include model performances, relationship between training and breeding populations (BP), heritability of target trait, population size of both TP and BP, population structure, and marker density.  $r_{MG}$  varies with particular GS statistical models that depend on prior assumptions and treatment of marker effects [28–31]. Several statistical models have been applied in genomic prediction, including ridge regression best linear unbiased prediction (RR-BLUP) [32,33], genomic best linear unbiased prediction (GBLUP) [3,31,34,35], Bayesian models [4,28,30,36,37], and machine learning models [38–42]. Moreover, developing optimum models with consideration of

genotype  $\times$  environment interaction can significantly improve the predictive ability in multi-environment trials [43–46]. Designing the composition of the TP with reference to BP is an important factor for maintaining a high degree of prediction accuracy in GS breeding programs [47–49]. Agronomic traits with high heritability are regulated by major-effect genes and are rarely affected by environment, and thus they will be positively correlated with higher  $r_{MG}$  with good selection response [50,51]. High marker density can improve the proportion of genetic variation explained by molecular markers such as single nucleotide polymorphisms (SNPs), and thus result in high prediction accuracy, the latter being used as a selection criterion to assist plant breeders to select target traits with precision [51,52]. Population structure, as a specific factor affecting GEBV prediction, can give rise to biased estimates in GS [53–55]. Hence, taking all GS-relevant factors into account can be deemed a rational strategy that will be more helpful for plant breeders in making selections based on the superiority of breeding individuals rather than their phenotypic data alone.

In this research article, we focus on better understanding GS-relevant factors to seek a few available measures to instruct plant breeders in reasonably designing GS breeding programs for enhancing genetic gain per unit time while reducing cost. The major challenge in GS is how to integrate empirical results into plant breeding practices to improve crop yield and create more economic values. The datasets used in the study contain genotypic data from the 55 K SNP array and phenotypic data for six complex traits generated for one natural and two biparental populations. The objectives of the case study were to (1) evaluate the prediction accuracy of different traits in three maize populations; (2) assess the effects of six GS-relevant factors on prediction accuracy, including population size, marker density, heritability, model performance, relationship between TP and BP, and the relative population size between the training and testing sets; and (3) utilize the results to make recommendations for plant breeders in implementing GS in commercial breeding programs.

## 2. Materials and methods

### 2.1. Plant materials

The experiment started with one natural population (NAT) and three biparental populations, including recombinant inbred line (RIL), doubled haploid (DH) and  $F_{2:3}$  populations. The natural population comprised 435 maize elite inbred lines derived from temperate and tropical regions around the world, so that it has rich genetic diversity. The three biparental populations were derived from a widely grown single-cross maize hybrid, Zhongdan 909, with elite Chinese inbred lines Zheng 58 and HD568 as parents, and included 212 RILs, 79 DH lines, and 304  $F_{2:3}$  families. As the RILs were generated with many generations of selfing, and DH population had only a limited number of individuals, we combined the RILs and DH lines as a single population, and named it the RIL&DH population.

Download English Version:

<https://daneshyari.com/en/article/8408659>

Download Persian Version:

<https://daneshyari.com/article/8408659>

[Daneshyari.com](https://daneshyari.com)