Theor Appl Genet (2012) 124:261–275
DOI 10.1007/s00122-011-1702-9

ORIGINAL PAPER

Evaluation of genome-wide selection efficiency in maize nested association mapping populations

Zhigang Guo · Dominic M. Tucker · Jianwei Lu · Venkata Kishore · Gilles Gay

Received: 1 March 2011 / Accepted: 7 September 2011 / Published online: 22 September 2011

Abstract In comparison to conventional marker-assisted selection (MAS), which utilizes only a subset of genetic markers associated with a trait to predict breeding values (BVs), genome-wide selection (GWS) improves prediction accuracies by incorporating all markers into a model simultaneously. This strategy avoids risks of missing quantitative trait loci (QTL) with small effects. Here, we evaluated the accuracy of prediction for three corn flowering traits days to silking, days to anthesis, and anthesis-silking interval with GWS based on cross-validation experiments using a large data set of 25 nested association mapping populations in maize (*Zea mays*). We found that GWS via ridge regression-best linear unbiased prediction (RR-BLUP) gave significantly higher predictions compared to MAS utilizing composite interval mapping (CIM). The CIM method may be selected over multiple linear regression to decrease over-estimations of the efficiency of GWS over a MAS strategy. The RR-BLUP method was the preferred method for estimating marker effects in GWS with prediction accuracies comparable to or greater than BayesA and BayesB. The accuracy with RR-BLUP increased with training sample proportion, marker density, and heritability until it reached a plateau. In general, gains in accuracy with RR-BLUP over CIM increased with decreases of these factors. Compared to training sample proportion, the accuracy of prediction with RR-BLUP was relatively insensitive to marker density.

Abbreviations

ASI	Anthesis silking interval
BV	Breeding value
CIM	Composition interval mapping
DA	Day till anthesis
DS	Days till silking
GWS	Genome-wide selection
ICE	Iterative conditional expectation
LOD	Logarithm of odds
MAS	Marker-assisted selection
MLR	Multiple linear regression
MCMC	Monte Carlo Markov chain
NAM	Nested association mapping
QTL	Quantitative trait loci
RIL	Recombinant inbred line
RR-BLUP	Ridge regression-best linear unbiased prediction
SNP	Single nucleotide polymorphisms

Introduction

Genome-wide selection (GWS) has been proposed as a method for increasing genetic gain for quantitative traits in animal and plant breeding (Meuwissen et al. 2001; Bernardo and Yu 2007; Habier et al. 2007; Hayes et al. 2009; Heffner et al. 2009; Jannink et al. 2010). Conventional marker-assisted selection (MAS) exploits only a subset of markers identified by least-squares procedures such as multiple linear regression (MLR) (Lande and Thompson 1990; Meuwissen et al. 2001; Lorenzana and Bernardo

Communicated by M. Sillanpaa.

Z. Guo (✉) · J. Lu · G. Gay
Syngenta Biotechnology, Inc., Cornwallis Road,
3054 E, Research Triangle Park, NC 27705-2257, USA
e-mail: zhigang.guo@syngenta.com

D. M. Tucker · V. Kishore
Syngenta Seeds, Inc., Thorps Road, 12101,
Clinton, IL 61727, USA

2009). In contrast, GWS incorporates all markers across the entire genome into a prediction model simultaneously, reducing the risk of missing or inaccurately considering quantitative trait loci (QTL) with lesser effects. Each marker is generally considered as a putative QTL and all the markers are combined to predict breeding values (BVs) of progeny with the GWS method.

In animal breeding, extensive simulations and empirical studies have been performed to verify advantages of GWS over MAS. The first simulation study was conducted by Meuwissen et al. (2001). In the study, three methods: ridge regression-best linear unbiased prediction (RR-BLUP), BayesA, and BayesB were proposed to estimate effects of each marker in GWS. Following, performances of each method were compared with MAS. The authors found that GWS significantly improved the accuracy of predicted BVs. Increases of 130% with RR-BLUP, 151% with BayesA, and 167% with BayesB, respectively, were reported over MAS. In a separate study, an empirical analysis in mice concluded that the prediction accuracies of GWS were 3–22% improvement over conventional methods, depending on the trait (Legarra et al. 2008). Recently, results in dairy cattle from Australia, New Zealand, United States, and the Netherlands showed that the average reliability of genetic prediction of GWS was 85% higher than that of conventional parent averages (Hayes et al. 2009).

Progresses in animal breeding have stimulated the development of GWS in plant breeding. Differing from outcrossing schemes in animal breeding, the focus of GWS in plant breeding has been on bi-parental breeding populations. Bernardo and Yu (2007) first performed simulations on maize (*Zea mays*) that compared the efficiency of GWS versus MAS. They concluded that the selection responses due to GWS via RR-BLUP were 18–43% greater than that due to MAS via MLR. The advantage of GWS increased with decreases in heritability. Similar results were reported when marker intervals or haplotypes, rather than individual markers, were used as predictor variables for GWS (Piyasatian et al. 2007). Thus far, only one empirical study was performed by Lorenzana and Bernardo (2009) that concluded predictions were more accurate with RR-BLUP in comparison to MLR. In addition, RR-BLUP was the method of choice for GWS strategies as it produced improved or comparable prediction accuracies compared to empirical Bayesian methods (Xu 2003, 2007).

The above empirical evaluation was based only on seven bi-parental plant populations: four maize, one Arabidopsis (*Arabidopsis thaliana*), and two barley (*Hordeum vulgare*) populations. Thus, it is necessary to verify genetic gains of GWS using a larger number of bi-parental populations. Secondly, in previous studies, GWS was compared to MLR methods (Meuwissen et al. 2001; Bernardo and Yu 2007; Lorenzana and Bernardo 2009). The MLR method serves as an adequate control method in association mapping, but may not be an appropriate control method for linkage mapping in bi-parental populations. In practice, the most widely utilized mapping method has been composite interval mapping (CIM, Zeng 1993, 1994) for this type of population, which has provided improved power and precision estimates of QTL positions and effects over MLR (Jansen 1993; Zeng 1993). Thus, the predictive advantages of GWS over MAS may be over-estimated if MLR is used as a control in bi-parental populations, rather than CIM. In addition, the influence of heritability on the genetic gain with GWS over MAS has not been determined using an empirical study in breeding populations. Finally, the combined effect of training sample sizes and marker densities on GWS needs to be carefully investigated in further studies.

Our objectives therefore were to: (1) evaluate the performance of GWS over MAS using CIM as a control in a large data set consisting of 25 bi-parental populations, (2) compare RR-BLUP with other Bayesian methods (BayesA and BayesB, Meuwissen et al. 2001) for estimating marker effects for GWS, and (3) determine effects of training sample proportion, marker density, and heritability on the accuracy of GWS prediction. The data used in the study was from a recent QTL mapping study of flowering time in maize (Buckler et al. 2009). A total of 4,699 recombinant inbred lines (RILs) were from 25 nested association mapping (NAM) populations containing 1,106 single nucleotide polymorphism (SNP) markers per RIL. Three flowering traits were chosen for analysis: days to silking (DS, female flowering), days to anthesis (DA, male flowering), and anthesis-silking interval (ASI). These traits are known to be controlled by numerous QTL with small effects (Buckler et al. 2009) that is particularly suitable for a GWS study (Bernardo and Yu 2007).

Materials and methods

NAM Populations

Phenotype and genotype data from 25 maize NAM populations were obtained from the Panzea website (http://www.panzea.org), a project aimed at investigating the genetic architecture of maize and teosinte (*Z. teosinte*). A total of 4,699 RILs were obtained from 25 bi-parental crosses between genetically diverse lines and B73 (Table 1). Genotypic data from 1,106 SNP markers covered a genetic map of 1,439 cM. On average, the size of an interval flanked by two adjacent markers in each NAM populations were $\sim$1.6 cM, indicating an extensive and dense coverage of the maize genome. For individual populations, numbers of polymorphic SNPs varied from 785 to 895 while genome size ranged from 1,371 to 1,397 cM. This indicated that the genomic coverage for individual populations was very close, even with different numbers of markers. Each

Table 1 Population information and genomic coverage regarding the 25 nested association mapping populations (NAM) used in the current study

Index	Crosses	Sample size	Marker number	Genomic coverage (cM)	Heritability (H^2)		
					DS	DA	ASI
1	B73 × B97	194	816	1,386	0.77	0.77	0.64
2	B73 × CML103	196	830	1,396	0.73	0.78	0.61
3	B73 × CML228	191	895	1,396	0.89	0.90	0.68
4	B73 × CML247	196	852	1,397	0.89	0.90	0.75
5	B73 × CML277	187	834	1,385	0.91	0.92	0.71
6	B73 × CML322	185	848	1,394	0.83	0.84	0.72
7	B73 × CML333	193	838	1,388	0.82	0.84	0.61
8	B73 × CML52	196	853	1,390	0.89	0.88	0.68
9	B73 × CML69	189	849	1,389	0.80	0.82	0.67
10	B73 × Hp301	192	806	1,387	0.85	0.83	0.67
11	B73 × Il14H	194	846	1,389	0.86	0.84	0.61
12	B73 × Ki11	193	836	1,386	0.91	0.90	0.69
13	B73 × Ki3	126	809	1,397	0.85	0.84	0.67
14	B73 × Ky21	196	820	1,395	0.81	0.75	0.66
15	B73 × M162 W	185	841	1,378	0.83	0.82	0.69
16	B73 × M37 W	194	804	1,389	0.83	0.83	0.65
17	B73 × Mo18 W	192	825	1,386	0.90	0.88	0.81
18	B73 × MS71	196	785	1,371	0.76	0.70	0.63
19	B73 × NC350	188	838	1,388	0.84	0.83	0.71
20	B73 × NC358	184	825	1,396	0.79	0.75	0.58
21	B73 × Oh43	193	824	1,389	0.81	0.71	0.71
22	B73 × Oh7B	181	796	1,396	0.80	0.81	0.71
23	B73 × P39	183	841	1,378	0.86	0.87	0.62
24	B73 × Tx303	188	815	1,394	0.88	0.85	0.80
25	B73 × Tzi8	187	860	1,381	0.89	0.87	0.81
Average		188	831	1,389	0.84	0.83	0.68

Heritability estimates for days to silking (DS), days to anthesis (DA) and anthesis-silking interval (ASI) were calculated by Buckler et al. (2009) and are also indicated in the table

RIL was evaluated for DS, DA, and ASI at four locations over 2 years, and phenotypic means of each line were used for our data analysis for simplicity (Buckler et al. 2009).

Cross-validation

Each NAM population, referred to as biparental populations in the current study, was randomly split into a training data set of four-fifths progenies and a validation data set of one fifth progenies (Legarra et al. 2008; Lorenzana and Bernardo 2009). Marker effects estimated from the training data set were used to predict the BVs of RILs in the validation set based on SNP genotype data only. The accuracy of prediction was defined as the square of correlation coefficient between predicted BVs and observed phenotypes for flowering traits in the validation sample.

Methods Used for QTL Identification in MAS

The MLR and CIM methods were used for the MAS strategy. Markers near QTL were first identified using stepwise regression based on a given significance level α, and the effects were estimated by multiple regression with MLR (Bernardo and Yu 2007; Lorenzana and Bernardo 2009). The CIM method was performed based on the additive genetic model

$$y_i = \mu + q_i a + \sum x_{ij} b_j + e_i$$

where y_i is the phenotype of individual i in a NAM population, μ the overall mean of the phenotype, q_i the genotype of the putative QTL, a the additive effect of the QTL, x_{ij} the genotype of the cofactor marker j of individual i, b_j the effect of marker j, and e_i residual error following a normal distribution $N(0, \sigma_e^2)$. QTL genotype q_i was not observed and was replaced in the model with its expectation, calculated from the probability distribution of QTL genotypes that was conditional on the closest flanking markers (Haley and Knott 1992). Missing the x_{ij} genotype data was similarly imputed. Cofactors in the model were selected by stepwise regression. The effects of QTL identified by CIM were estimated by multiple regression based on their conditional expectations (Utz et al. 2000).

Methods used for estimating marker effects in GWS

The genetic model for GWS was same as the one used by Bernardo and Yu (2007), written as

$$y_i = \mu + \sum z_{ij}u_j + e_i$$

where y_i was the phenotype of individual i in a NAM population, μ the overall mean of the phenotype, z_{ij} the genotype of the marker j of individual i with $j = 1, 2, \ldots, M$ where M was the total number of markers, u_j the additive effect of marker j, and e_i residual error following a normal distribution $N(0, \sigma_e^2)$. The key in the model was to simultaneously estimate effects of genome-wide markers. To do so, three methods: (1) RR-BLUP, (2) BayesA, and (3) BayesB were used to calculate u_j in the study, respectively. In RR-BLUP, each marker was assumed to have an equal genetic variance and the effects of markers were calculated by solving Henderson mixed model equation (Henderson 1984). In BayesA, it was assumed that $u_j \sim N(0, \sigma_{gj}^2)$ where σ_{gj}^2 was the genetic variance modeled by an inverted chi-square distribution $\chi^{-2}(v, S)$ with v the degrees of freedom and S the scale parameter. Markers effects were estimated by the Monte Carlo Markov Chain (MCMC) using Gibbs sampling as proposed by Meuwissen et al. (2001). BayesB was actually an extension of BayesA by sampling the marker effect u_j as $u_j = 0$ with probability r and $u_j \sim N(0, \sigma_{gj}^2)$ where $\sigma_{gj}^2 \sim \chi^{-2}(v, S)$ with probability $1 - r$ where r was the prior probability that a marker was not a QTL (Meuwissen et al. 2001). The difficulty in the BayesB method was that the posterior distribution of u_j cannot be expressed in the closed form of a known distribution. This problem was solved by using Metropolis-Hasting sampling. This significantly increased the computational time. To overcome computational time issues, we applied a more efficient algorithm of BayesB based on an iterative conditional expectation (ICE) algorithm (Meuwiseen et al. 2009). The BV of each line in a validation sample were predicted as

$$\hat{y}_i = \mu + \sum (z_{ij}u_j)$$

where $\hat{y_i}$ is the BV of individual i in the sample and μ and u_j were estimated from a training sample using methods described above. In GWS, all markers were needed to predict BVs of lines in the sample, while only a subset of markers and QTL were used for prediction in MLR and CIM in MAS. The genotype z_{ij} was replaced by the conditional expectation of a QTL for CIM.

Data analysis

A total of 100 replicates of cross-validations were performed to evaluate the performances of MLR, CIM, RR-BLUP, BayesA, and BayesB across the 25 NAM populations. At each replicate, a whole NAM population was randomly split into a training sample (four fifth population size) and a validation sample (one fifth population size). It must be noted that the proportion of a training sample varied at different levels when we investigated the effect of training sample proportion on GWS. Marker effects were first estimated by proposed methods based on phenotypic and genotypic data in the training sample.

The significance level was chosen to be $\alpha = 0.001$, 0.01, 0.05, 0.10, and 0.20 for stepwise selection in MLR. The first level $\alpha = 0.001$ was obtained by permutation tests (Buckler et al. 2009), and the last one $\alpha = 0.20$ was determined by our preliminary experiment. In this preliminary experiment, less stringent significance levels were applied, for example, $\alpha = 0.30$ and $\alpha = 0.40$ as suggested by Bernardo and Yu (2007). However, often numbers of markers significantly associated with flowering traits exceeded the number of RILs in individual NAM populations, causing the over-fitting of the general least-square model in this strategy. Therefore, it was determined $\alpha = 0.20$ was the lowest significance level of α to avoid this issue. With a given α, a marker was allowed to enter the model when its p value was less than the level. Following, the model was then re-evaluated. Markers with p values greater than α were dropped. This process was repeated until the next marker added had a p-value greater than α.

With CIM, cofactors were selected by stepwise selection as demonstrated above in MLR ($\alpha = 0.05$) with a cofactor window size of 10 cM. The whole genome was scanned by a fixed step size of 1 cM. A QTL was identified at the position where the logarithm of odds (LODs) score assumed its maximum in the region under consideration with a LOD threshold of 2.5 (Utz et al. 2000).

For RR-BLUP, the genetic variance V_g and environmental variance V_e were estimated based on heritabilities H^2 of traits (Table 1) and phenotypic variances V_t as $V_g = H^2 V_t$ and $V_e = V_t - V_g$. For each marker, the genetic variance was calculated as $\sigma_{gj}^2 = V_g/2\sum p_k(1 - p_k)$ with p_k was the allele frequency at marker locus k ($k = 1, 2, \ldots, M$) where M was the total number of markers (Habier et al. 2007). Given the equivalent genetic variance, effects of each marker were then calculated by solving the Henderson mixed model equation (Henderson 1984).

For BayesA, a total of 10,000 MCMC iterations were run with the first 2,000 iterations discarded for burn-in. The remaining 8,000 iterations were used to estimate marker effects. In addition, we performed a preliminary analysis for DS in NAM population 1 to verify the accuracy obtained from five independent BayesA analyses. This accuracy calculation result was actually negligible (data not shown). Therefore, the number of iterations used in the

BayesA method was sufficient to guarantee the stability of accuracy of prediction and was not highly impacted by stochastic error in the MCMC. As described in the preceding section, BayesB was implemented as the ICE algorithm (Meuwiseen et al. 2009) which needed the prior probability (r) that a marker was not associated with a QTL. The prior was calculated as $r = (M - N_q)/M$ with N_q the number of significant markers associated with traits of interest by stepwise selection conditional on $\alpha = 0.05$. The iterative BayesB calculation was stopped until changes of marker effect estimates between current and last iterations were small, less than 10^{-6}.

The BVs of RILs in the validation sample were then predicted by summing all marker effects. R^2 were then calculated with R the correlation coefficient between predicted BVs and observed phenotypes in the validation sample. Prediction accuracies for each method were actually calculated as $r^2_{BV:P}$, which was the average of R^2 across

Table 2 Prediction accuracy with MLR, CIM, RR-BLUP, BayesA and BayesB for DS at training sample proportion 0.8 and marker density 1.6 cM based on 25 NAM populations

Index	Crosses	MLR $r^{2a}_{BV:P}$	N_{QTL}	CIM $r^{2b}_{BV:P}$	N_{QTL}	RR-BLUP $r^{2c}_{BV:P}$	N_{QTL}	BayesA $r^{2d}_{BV:P}$	N_{QTL}	BayesB $r^{2e}_{BV:P}$	N_{QTL}
1	B73 × B97	0.30 (0.05)	49	0.35 (0.18)	27	0.50 (0.68, 0.43)	816	0.41 (−0.18)	816	0.31 (−0.38)	816
2	B73 × CML103	0.10 (0.05)	40	0.09 (**−0.04**)	20	0.27 (1.86, 1.98)	830	0.25 (−0.07)	830	0.11 (−0.60)	830
3	B73 × CML228	0.35 (0.001)	4	0.34 (**−0.01**)	29	0.45 (0.31, 0.32)	895	0.50 (0.10)	895	0.45 (**−0.01**)	895
4	B73 × CML247	0.38 (0.01)	17	0.41 (0.08)	29	0.55 (0.45, 0.34)	852	0.49 (−0.10)	852	0.47 (−0.14)	852
5	B73 × CML277	0.47 (0.01)	10	0.44 (−0.07)	24	0.50 (0.08, 0.15)	834	0.52 (**0.03**)	834	0.56 (0.11)	834
6	B73 × CML322	0.30 (0.01)	14	0.34 (0.11)	33	0.54 (0.77, 0.60)	848	0.48 (−0.11)	848	0.36 (−0.33)	848
7	B73 × CML333	0.21 (0.01)	12	0.24 (0.13)	21	0.40 (0.85, 0.64)	838	0.39 (**−0.01**)	838	0.34 (−0.14)	838
8	B73 × CML52	0.22 (0.05)	65	0.25 (0.11)	30	0.38 (0.72, 0.54)	853	0.38 (**−0.02**)	853	0.32 (−0.17)	853
9	B73 × CML69	0.24 (0.01)	11	0.25 (**0.04**)	21	0.35 (0.45, 0.39)	849	0.37 (**0.08**)	849	0.29 (−0.16)	849
10	B73 × Hp301	0.38 (0.01)	16	0.46 (0.21)	27	0.57 (0.52, 0.26)	806	0.56 (**−0.03**)	806	0.46 (−0.20)	806
11	B73 × Il14H	0.36 (0.01)	16	0.42 (0.18)	26	0.53 (0.48, 0.26)	846	0.54 (**0.01**)	846	0.45 (−0.15)	846
12	B73 × Ki11	0.46 (0.001)	4	0.44 (**−0.04**)	23	0.50 (0.09, 0.14)	836	0.57 (0.12)	836	0.55 (0.09)	836
13	B73 × Ki3	0.31 (0.01)	10	0.29 (**−0.08**)	46	0.41 (0.32, 0.43)	809	0.45 (0.11)	809	0.41 (**0.01**)	809
14	B73 × Ky21	0.34 (0.01)	15	0.38 (0.12)	27	0.54 (0.60, 0.43)	820	0.48 (−0.11)	820	0.36 (−0.33)	820
15	B73 × M162 W	0.33 (0.001)	4	0.31 (−0.08)	25	0.42 (0.27, 0.37)	841	0.41 (**−0.05**)	841	0.41 (**−0.03**)	841
16	B73 × M37 W	0.33 (0.01)	10	0.33 (**0.00**)	23	0.43 (0.30, 0.30)	804	0.43 (**−0.01**)	804	0.36 (−0.17)	804
17	B73 × Mo18 W	0.39 (0.01)	15	0.44 (0.13)	34	0.56 (0.43, 0.26)	825	0.45 (−0.20)	825	0.47 (−0.16)	825
18	B73 × MS71	0.19 (0.01)	8	0.22 (0.16)	19	0.32 (0.71, 0.47)	785	0.35 (0.07)	785	0.29 (−0.11)	785
19	B73 × NC350	0.31 (0.01)	11	0.31 (**−0.03**)	22	0.37 (0.17, 0.21)	838	0.39 (0.04)	838	0.32 (−0.15)	838
20	B73 × NC358	0.37 (0.01)	16	0.40 (**0.07**)	32	0.55 (0.50, 0.40)	825	0.49 (−0.11)	825	0.36 (−0.35)	825
21	B73 × Oh43	0.24 (0.01)	11	0.28 (0.14)	24	0.32 (0.31, 0.15)	824	0.27 (−0.15)	824	0.21 (−0.34)	824
22	B73 × Oh7B	0.18 (0.05)	46	0.23 (0.27)	24	0.38 (1.14, 0.68)	796	0.34 (−0.13)	796	0.28 (−0.27)	796
23	B73 × P39	0.22 (0.05)	48	0.25 (0.12)	25	0.44 (1.01, 0.80)	841	0.35 (−0.20)	841	0.33 (−0.26)	841
24	B73 × Tx303	0.15 (0.01)	9	0.17 (0.14)	25	0.26 (0.74, 0.53)	815	0.26 (**0.02**)	815	0.19 (−0.26)	815
25	B73 × Tzi8	0.27 (0.01)	11	0.28 (**0.04**)	24	0.29 (0.09, **0.05**)	860	0.40 (0.37)	860	0.33 (0.12)	860
Min		0.10	4	0.09	19	0.26	785	0.25	785	0.11	785
Max		0.47	65	0.46	46	0.57	895	0.57	895	0.56	895
Aver		0.30	19	0.32 (0.07)	26	0.43 (0.43, 0.34)	831	0.42 (−0.03)	831	0.36 (−0.16)	831

Bold in parentheses indicates the gain with one method over another one is not significant at $\alpha = 0.05$

NAM nested association mapping, *MLR* multiple linear regression, *CIM* composite interval mapping, *RR-BLUP* ridge regression-best linear unbiased prediction, *DS* days to silking, *Min* minimum accuracy of prediction across 25 NAM populations, *Max* maximum accuracy of prediction across 25 NAM populations, *Aver* average of accuracy of prediction across 25 NAM populations, N_{QTL} number of QTL

[a] In parentheses is the significance level which gave the highest $r^2_{BV:P}$ in MLR

[b] In parentheses is the gain in accuracy with CIM over with MLR

[c] The first value in parentheses is the gain with RR-BLUP over MLR, and the second one the gain with RR-BLUP over CIM

[d] In parentheses is the gain in accuracy with BayesA over RR-BLUP

[e] In parentheses is the gain in accuracy with BayesB over RR-BLUP

all replicates. Though five models with different α were used for prediction with MLR in MAS, only the model which produced the highest level of accuracy was considered as the true prediction of MLR. A pairwise t test ($\alpha = 0.05$) was applied to compare overall accuracies between various methods.

To determine effects of training sample sizes, random subsets with proportions of 0.20, 0.40, 0.60, and 0.80 of each population were selected from each NAM family. Effects of marker density on GWS, within each NAM family, were tested by setting a conditional genetic distance criteria c (for example, $c = 5$ cM) between two flanking markers to ensure that each chromosome was evenly covered by a set of SNPs. Out of 1106 SNPs, separate sets of polymorphic markers were first extracted for each of the NAM populations (Table 1). Following, based on the identified polymorphic markers in each of the NAM populations, varying numbers of makers were obtained corresponding to differing values of c without compromising the overall genomic coverage. In detail, the fist marker on each of the ten chromosomes was always selected, and its position was recorded as p_1. Following, the

Table 3 Prediction accuracy with MLR, CIM, RR-BLUP, BayesA and BayesB for DA at training sample proportion 0.8 and marker density 1.6 cM based on 25 NAM populations

Index	Crosses	MLR $r^{2a}_{BV:P}$	N_{QTL}	CIM $r^{2b}_{BV:P}$	N_{QTL}	RR-BLUP $r^{2c}_{BV:P}$	N_{QTL}	BayesA $r^{2d}_{BV:P}$	N_{QTL}	BayesB $r^{2e}_{BV:P}$	N_{QTL}
1	B73 × B97	0.27 (0.01)	13	0.30 (0.10)	25	0.41 (0.54, 0.39)	816	0.41 (**0.00**)	816	0.32 (−0.22)	816
2	B73 × CML103	0.08 (0.05)	29	0.10 (0.20)	15	0.20 (1.48, 1.06)	830	0.21 (**0.06**)	830	0.11 (−0.44)	830
3	B73 × CML228	0.34 (0.01)	15	0.39 (0.16)	31	0.52 (0.54, 0.32)	895	0.46 (−0.11)	895	0.46 (−0.11)	895
4	B73 × CML247	0.41 (0.01)	16	0.46 (0.11)	27	0.57 (0.37, 0.23)	852	0.54 (−0.05)	852	0.56 (**−0.01**)	852
5	B73 × CML277	0.46 (0.01)	9	0.45 (**−0.03**)	22	0.45 (**−0.02**, **0.01**)	834	0.46 (**0.02**)	834	0.50 (0.10)	834
6	B73 × CML322	0.31 (0.05)	67	0.37 (0.17)	31	0.53 (0.71, 0.45)	848	0.50 (−0.06)	848	0.39 (−0.27)	848
7	B73 × CML333	0.30 (0.01)	13	0.36 (0.18)	23	0.49 (0.61, 0.35)	838	0.50 (**0.03**)	838	0.44 (−0.09)	838
8	B73 × CML52	0.25 (0.01)	14	0.30 (0.19)	29	0.44 (0.77, 0.48)	853	0.41 (−0.06)	853	0.36 (−0.19)	853
9	B73 × CML69	0.29 (0.001)	11	0.30 (**0.03**)	25	0.39 (0.38, 0.30)	849	0.42 (0.06)	849	0.39 (**−0.01**)	849
10	B73 × Hp301	0.34 (0.05)	75	0.41 (0.20)	34	0.54 (0.57, 0.31)	806	0.54 (**0.00**)	806	0.42 (−0.22)	806
11	B73 × Il14H	0.38 (0.01)	14	0.39 (**0.03**)	26	0.48 (0.28, 0.25)	846	0.50 (0.03)	846	0.44 (−0.09)	846
12	B73 × Ki11	0.44 (0.001)	4	0.44 (**0.00**)	20	0.46 (**0.04**, **0.04**)	836	0.54 (0.16)	836	0.52 (0.13)	836
13	B73 × Ki3	0.20 (0.01)	11	0.25 (0.23)	43	0.42 (1.12, 0.72)	809	0.38 (−0.11)	809	0.29 (−0.31)	809
14	B73 × Ky21	0.29 (0.01)	15	0.34 (0.19)	32	0.50 (0.73, 0.46)	820	0.45 (−0.09)	820	0.28 (−0.44)	820
15	B73 × M162 W	0.38 (0.001)	4	0.32 (−0.16)	23	0.38 (**0.00**, 0.20)	841	0.42 (0.08)	841	0.40 (0.04)	841
16	B73 × M37 W	0.36 (0.01)	15	0.40 (0.12)	29	0.55 (0.54, 0.37)	804	0.52 (−0.06)	804	0.47 (−0.14)	804
17	B73 × Mo18 W	0.40 (0.01)	14	0.42 (**0.06**)	26	0.53 (0.35, 0.28)	825	0.49 (−0.08)	825	0.53 (**−0.01**)	825
18	B73 × MS71	0.19 (0.01)	9	0.21 (0.13)	14	0.25 (0.32, 0.17)	785	0.30 (0.20)	785	0.30 (0.21)	785
19	B73 × NC350	0.25 (0.01)	11	0.27 (0.09)	19	0.32 (0.31, 0.20)	838	0.33 (**0.01**)	838	0.29 (−0.11)	838
20	B73 × NC358	0.34 (0.01)	15	0.38 (0.09)	28	0.51 (0.48, 0.35)	825	0.51 (**0.00**)	825	0.41 (−0.19)	825
21	B73 × Oh43	0.22 (0.01)	10	0.25 (0.14)	20	0.29 (0.31, 0.15)	824	0.29 (**0.00**)	824	0.22 (−0.24)	824
22	B73 × Oh7B	0.38 (0.01)	12	0.43 (0.12)	23	0.47 (0.22, 0.09)	796	0.43 (−0.07)	796	0.40 (−0.14)	796
23	B73 × P39	0.26 (0.01)	12	0.29 (0.11)	25	0.49 (0.87, 0.69)	841	0.40 (−0.18)	841	0.33 (−0.32)	841
24	B73 × Tx303	0.15 (0.01)	8	0.15 (**0.00**)	18	0.17 (0.12, 0.12)	815	0.20 (0.16)	815	0.16 (**−0.07**)	815
25	B73 × Tzi8	0.30 (0.01)	12	0.29 (**−0.03**)	30	0.38 (0.25, 0.29)	860	0.47 (0.24)	860	0.38 (**0.01**)	860
Min		0.08	4	0.10	14	0.17	785	0.20	785	0.11	785
Max		0.46	75	0.46	43	0.57	895	0.54	895	0.56	895
Aver		0.30	17	0.33 (0.10)	26	0.43 (0.43, 0.30)	831	0.43 (0.00)	831	0.37 (−0.14)	831

Bold in parentheses indicates the gain with one method over another one is not significant at $\alpha = 0.05$

NAM nested association mapping, *MLR* multiple linear regression, *CIM* composite interval mapping, *RR-BLUP* ridge regression-best linear unbiased prediction, *DA* days to anthesis, *Min* minimum accuracy of prediction across 25 NAM populations, *Max* maximum accuracy of prediction across 25 NAM populations, *Aver* average of accuracy of prediction across 25 NAM populations, N_{QTL} number of QTL

[a] In parentheses is the significance level which gave the highest $r^2_{BV:P}$ in MLR

[b] In parentheses is the gain in accuracy with CIM over with MLR

[c] The first value in parentheses is the gain with RR-BLUP over MLR, and the second one the gain with RR-BLUP over CIM

[d] In parentheses is the gain in accuracy with BayesA over RR-BLUP

[e] In parentheses is the gain in accuracy with BayesB over RR-BLUP

markers surrounding the position $p_2 = p_1 + c$ were examined and the second marker was selected which was closest to p_2. This p_2 was replaced with the position of the second marker. This process was repeated until the last maker on the chromosome was selected and was included if its genetic distance from the last selected marker was greater than c. A cross-validation was performed based on different marker densities of $c = 1.6$, 5, 10, 20, 30, 40, and 50 cM for each method, NAM population, and trait. All the identified polymorphic markers were included in the model at $c = 1.6$ cM as the mean marker distance of the entire marker set was ~ 1.6 cM. Note that marker density 1.6 cM was always used when evaluating the effect of training sample proportion on accuracy of prediction, and training sample proportion 0.8 was used when evaluating the effect of marker density in each of the NAM populations.

The heritability influence on GWS accuracies was tested by a cross-validation procedure as described below. In each NAM population, the heritabilities H^2 for DS, DA, and ASI (Table 1) were calculated based on phenotypic data from four locations over 2 years (Buckler et al. 2009). Overall, H^2 ranged from 0.73 to 0.91 for DS, from 0.70 to 0.92 for

Table 4 Prediction accuracy with MLR, CIM, RR-BLUP, BayesA and BayesB for ASI at training sample proportion 0.8 and marker density 1.6 cM based on 25 NAM populations

Index	Crosses	MLR $r^{2a}_{BV:P}$	N_{QTL}	CIM $r^{2b}_{BV:P}$	N_{QTL}	RR-BLUP $r^{2c}_{BV:P}$	N_{QTL}	BayesA $r^{2d}_{BV:P}$	N_{QTL}	BayesB $r^{2e}_{BV:P}$	N_{QTL}
1	B73 × B97	0.11 (0.01)	8	0.14 (0.26)	16	0.19 (0.72, 0.36)	816	0.19 (**0.01**)	816	0.15 (−0.20)	816
2	B73 × CML103	0.16 (0.01)	8	0.16 (**0.00**)	19	0.30 (0.86, 0.85)	830	0.31 (**0.02**)	830	0.26 (−0.14)	830
3	B73 × CML228	0.16 (0.01)	12	0.19 (0.20)	26	0.30 (0.90, 0.58)	895	0.30 (**0.01**)	895	0.20 (−0.33)	895
4	B73 × CML247	0.11 (0.01)	6	0.11 (**0.00**)	13	0.13 (0.21, 0.22)	852	0.18 (0.35)	852	0.17 (0.27)	852
5	B73 × CML277	0.19 (0.01)	9	0.16 (−0.11)	19	0.22 (0.20, 0.35)	834	0.27 (0.21)	834	0.20 (−0.10)	834
6	B73 × CML322	0.12 (0.05)	54	0.15 (0.29)	25	0.31 (1.68, 1.07)	848	0.33 (0.04)	848	0.23 (−0.27)	848
7	B73 × CML333	0.14 (0.05)	45	0.16 (**0.09**)	22	0.24 (0.65, 0.52)	838	0.24 (**0.00**)	838	0.16 (−0.33)	838
8	B73 × CML52	0.13 (0.01)	9	0.13 (**−0.00**)	18	0.22 (0.68, 0.76)	853	0.25 (0.14)	853	0.19 (−0.14)	853
9	B73 × CML69	0.17 (0.05)	40	0.18 (**0.02**)	19	0.21 (0.23, 0.21)	849	0.22 (**0.04**)	849	0.19 (−0.10)	849
10	B73 × Hp301	0.29 (0.01)	8	0.27 (−0.10)	14	0.38 (0.29, 0.43)	806	0.38 (**0.00**)	806	0.38 (**0.00**)	806
11	B73 × Il14H	0.21 (0.01)	8	0.20 (**−0.05**)	15	0.23 (0.07, 0.12)	846	0.21 (−0.08)	846	0.24 (**0.05**)	846
12	B73 × Ki11	0.10 (0.05)	44	0.11 (**0.10**)	21	0.22 (1.11, 0.92)	836	0.21 (**−0.03**)	836	0.15 (−0.31)	836
13	B73 × Ki3	0.11 (0.001)	2	0.12 (**0.07**)	30	0.19 (0.75, 0.64)	809	0.18 (−0.04)	809	0.15 (−0.21)	809
14	B73 × Ky21	0.21 (0.01)	9	0.24 (0.17)	16	0.31 (0.49, 0.28)	820	0.31 (**0.00**)	820	0.27 (−0.13)	820
15	B73 × M162 W	0.16 (0.001)	4	0.16 (**−0.00**)	16	0.22 (0.25, 0.37)	841	0.23 (0.08)	841	0.16 (−0.26)	841
16	B73 × M37 W	0.11 (0.001)	2	0.08 (−0.29)	13	0.15 (0.32, 0.85)	804	0.15 (**−0.01**)	804	0.15 (−0.01)	804
17	B73 × Mo18 W	0.17 (0.01)	12	0.22 (0.33)	23	0.27 (0.64, 0.23)	825	0.32 (0.17)	825	0.19 (−0.30)	825
18	B73 × MS71	0.24 (0.01)	11	0.26 (**0.08**)	18	0.40 (0.66, 0.53)	785	0.41 (**0.01**)	785	0.34 (−0.16)	785
19	B73 × NC350	0.07 (0.01)	6	0.07 (**0.00**)	12	0.12 (0.71, 0.71)	838	0.12 (**0.00**)	838	0.15 (0.30)	838
20	B73 × NC358	0.12 (0.01)	8	0.13 (**0.14**)	25	0.23 (0.96, 0.72)	825	0.26 (0.12)	825	0.14 (−0.39)	825
21	B73 × Oh43	0.29 (0.05)	56	0.35 (0.22)	29	0.42 (0.48, 0.21)	824	0.43 (0.02)	824	0.26 (−0.39)	824
22	B73 × Oh7B	0.14 (0.01)	8	0.16 (0.15)	17	0.23 (0.71, 0.49)	796	0.28 (0.21)	796	0.24 (**0.03**)	796
23	B73 × P39	0.13 (0.01)	8	0.15 (**0.12**)	27	0.27 (0.99, 0.78)	841	0.30 (0.11)	841	0.20 (−0.25)	841
24	B73 × Tx303	0.13 (0.05)	58	0.18 (0.34)	27	0.32 (1.44, 0.81)	815	0.31 (−0.02)	815	0.18 (−0.44)	815
25	B73 × Tzi8	0.08 (0.01)	7	0.10 (0.20)	19	0.10 (0.20, **0.00**)	860	0.10 (0.00)	860	0.10 (**0.00**)	860
Min		0.07	2	0.07	12	0.10	785	0.10	785	0.10	785
Max		0.29	58	0.35	30	0.42	895	0.43	895	0.38	895
Aver		0.15	18	0.17 (0.13)	20	0.25 (0.67, 0.47)	831	0.26 (0.04)	831	0.20 (−0.20)	831

Bold in parentheses indicates the gain with one method over another one is not significant at $\alpha = 0.05$

NAM nested association mapping, *MLR* multiple linear regression, *CIM* composite interval mapping, *RR-BLUP* ridge regression-best linear unbiased prediction, *ASI* anthesis-silking interval, *Min* minimum accuracy of prediction across 25 NAM populations, *Max* maximum accuracy of prediction across 25 NAM populations, *Aver* average of accuracy of prediction across 25 NAM populations, N_{QTL} number of QTL

[a] In parentheses is the significance level which gave the highest $r^2_{BV:P}$ in MLR

[b] In parentheses is the gain in accuracy with CIM over with MLR

[c] The first value in parentheses is the gain with RR-BLUP over MLR, and the second one the gain with RR-BLUP over CIM

[d] In parentheses is the gain in accuracy with BayesA over RR-BLUP

[e] In parentheses is the gain in accuracy with BayesB over RR-BLUP

DA, and 0.58 to 0.81 for ASI. On average, across 25 NAM populations, DS and DA showed high heritabilities of 0.84 and 0.83, while ASI showed a moderate heritability of 0.68. In order to investigate relationship between H^2 and prediction accuracy, accuracies of predictions were obtained with training sample proportion 0.80 and marker density 1.6 cM from each NAM population regardless of traits of interest.

Results

In the current study, we first compared the accuracy of prediction with MLR, CIM, RR-BLUP, BayesA, and BayesB methods for traits DS (Table 2), DA (Table 3), and ASI (Table 4) at training sample proportion 0.8 and marker density 1.6 cM. First, we compared two control methods MLR and CIM in order to select an appropriate control in MAS. Overall, CIM gave higher prediction accuracies than MLR in 68, 80, and 64% of NAM populations for DS, DA, and ASI. On average, across all populations, the prediction accuracy with CIM was 0.32 for DS, 0.33 for DA, and 0.17 for ASI, showing 7, 10, and 13% increases in precision over MLR for these three traits, respectively. As a result of this, the increases in accuracy with GWS over MAS (more detail discussed in the next section) were over-estimated when using MLR, compared to the results from CIM. In order to compensate for this issue, the CIM strategy was used as a control method in MAS in the current study.

Next, we evaluated the performance of GWS via RR-BLUP over MAS with the CIM method for DS (Table 2), DA (Table 3), and ASI (Table 4). We found that the accuracy of prediction with RR-BLUP was greater than that of traditional CIM. The accuracy with RR-BLUP ranged from 0.26 to 0.57 for DS, 0.17 to 0.57 for DA, and 0.10 to 0.42 for ASI, while with CIM ranged from 0.09 to 0.46 for DS, 0.10 to 0.46 for DA, and 0.07 to 0.35 for ASI. The gain in accuracy with RR-BLUP over CIM ranged from 0.05 to 1.98 for DS, 0.01 to 1.06 for DA, and 0.00 to 1.07 for ASI. We also found that these gains were statistically significant at $\alpha = 0.05$ in most cases. On average, across all the populations, the accuracy with RR-BLUP was 0.43, 0.43, and 0.25 for DS, DA, and ASI, 34, 30, and 47% higher than that with CIM.

Finally, RR-BLUP was compared with two Bayesian methods BayesA and BayesB to determine the favorable method for estimating marker effects in GWS (Tables 2, 3, 4). It was concluded that BayesA provided

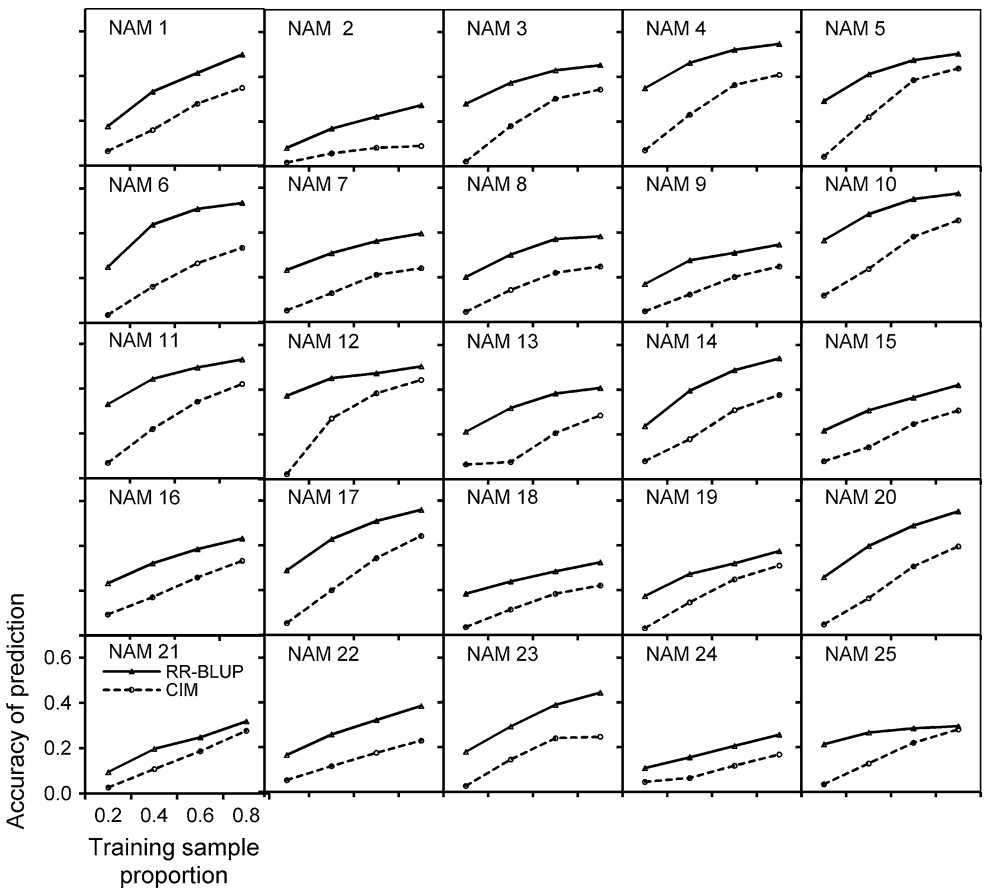

Fig. 1 Prediction accuracy using RR-BLUP and CIM with different proportions of training samples for DS at marker density 1.6 cM based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *DS* days to silking, *NAM* nested association mapping

better accuracies over BayesB and both gave estimates lower than or comparable to RR-BLUP. For example, for the DS trait, out of the 25 NAM populations, 10 populations showed significant decreases in prediction with BayesA over RR-BLUP. The decrease ranged from −0.07 to −0.20. Nine populations showed no differences while only 6 populations showing increases ranging from 0.04 to 0.37. Similar trends were seen for BayesB but to an even lower degree of gain over RR-BLUP. On average, across 25 NAM populations, gains with BayesA over RR-BLUP were -0.03 for DS, 0.00 for DA, and 0.05 for ASI, while with BayesB over RR-BLUP was -0.16, -0.14, and -0.20 for three traits DS, DA, and ASI, respectively.

Figures 1, 2, 3 shows the prediction accuracy with RR-BLUP and CIM with four levels of training sample proportions 0.2, 0.4, 0.6, and 0.8 at marker density $c = 1.6$ cM for DS, DA, and ASI, respectively. We found that the precision increased with the proportion of training sample for both RR-BLUP and CIM. This was due to improved estimations of marker effects and power of QTL detection. These accuracies gradually reached a plateau. Training sample proportion corresponding to a peak of 0.4 for RR-BLUP compared to 0.6 for CIM. This suggested that RR-BLUP reached the plateau faster than CIM. In general, the RR-BLUP method was relatively insensitive to training sample proportion compared to CIM. When training sample proportion decreased, accuracies with CIM declined faster than that with RR-BLUP. As a result, gains in prediction with RR-BLUP over CIM improved at the low level of training sample proportion.

Prediction estimates with RR-BLUP and CIM at different marker density levels were also examined for DS (Fig. 4), DA (Fig. 5), and ASI (Fig. 6) based on training sample proportion of 0.8. Marker densities of 1.6, 5, 10, 20, 30, 40, and 50 cM were used corresponding to different numbers of polymorphic markers selected in each NAM population via methods described in the previous section.

As expected, prediction accuracies improved with marker density due to enlarged linkage disequilibrium (LD) between QTL and markers. However, we found that these improvements tended to be zero when marker density exceeded a threshold of 10 cM. The precision remained unchanged with increases of marker density for both RR-BLUP and CIM at this point. This suggested that the density selected based on an interval size of 10 cM was sufficient in the NAM populations to capture LD between QTL and markers.

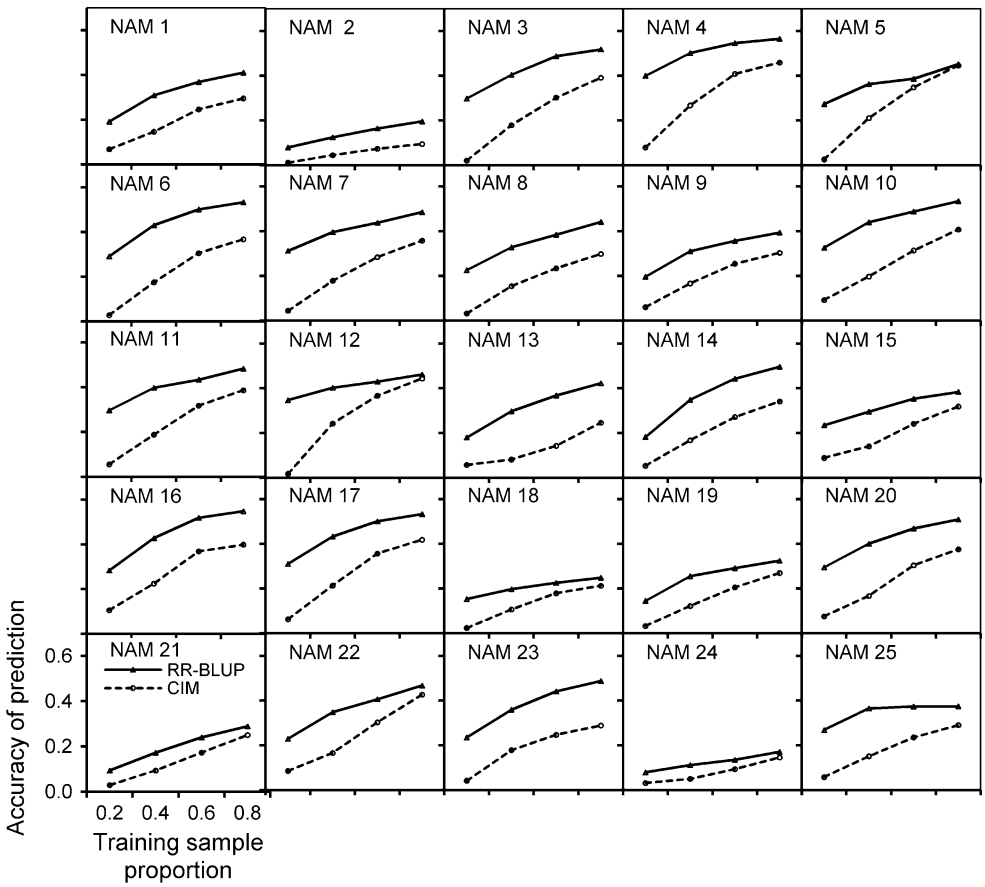

Fig. 2 Prediction accuracy using RR-BLUP and CIM with different proportions of training samples for DA at marker density 1.6 cM based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *DA* days to anthesis, *NAM* nested association mapping

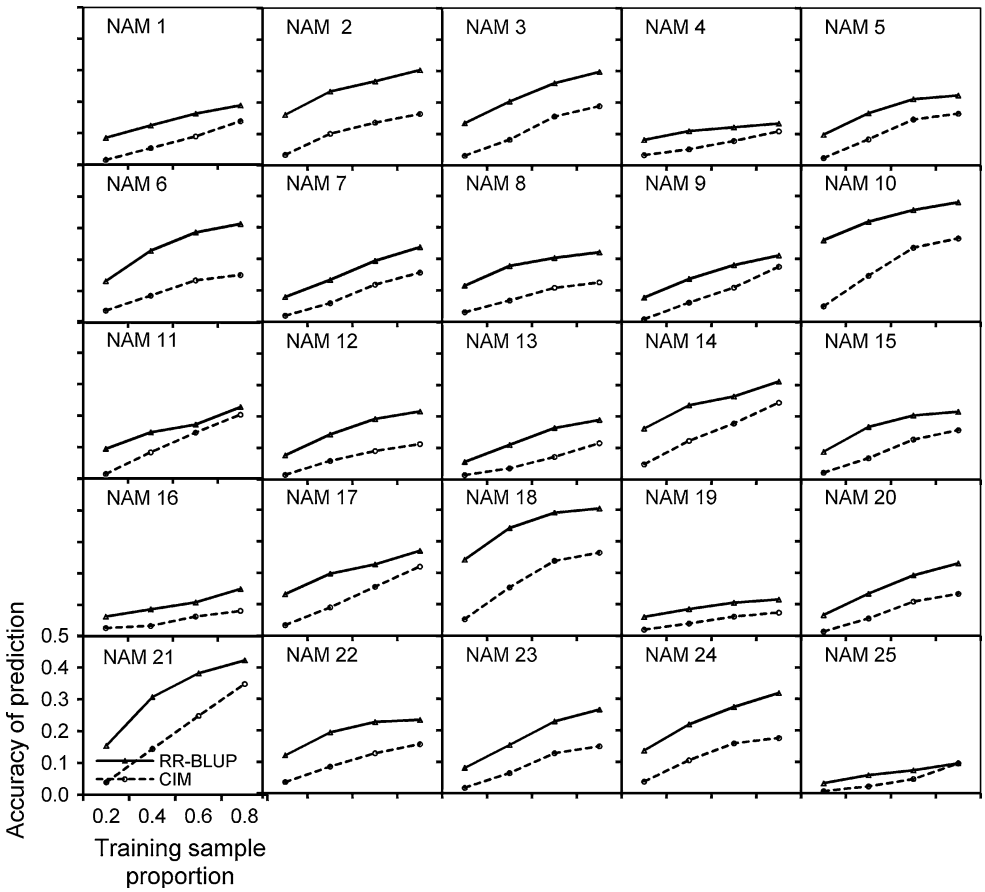

Fig. 3 Prediction accuracy using RR-BLUP and CIM with different proportions of training samples for ASI at marker density 1.6 cM based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *ASI* anthesis-silking interval, *NAM* nested association mapping

Both RR-BLUP and CIM methods decreased in accuracy with decreasing marker density with CIM decreasing more rapidly than RR-BLUP when maker density levels were lower than 10 cM. Consequently, gains in prediction with RR-BLUP over CIM improved with the reduction of marker density. In some cases, such as population 1, 2, 6, 19, 21, 24 in Fig. 4, population 2, 19, and 24 in Fig. 5, and population 3, 8, 9, 13, 15, 20, 24, and 25 in Fig. 6, the prediction with CIM was close to zero with low marker density due to difficulties in QTL detection caused by low LD between QTL and markers. In contrast, RR-BLUP still gave 0.03 to 0.17 accuracies in these situations, though the accuracies were not as large as that obtained with high marker density.

Although prediction accuracies decreased with marker densities at the range of 1.6 cM to 20 cM, some unexpected fluctuations occurred at the range of 20 cM to 50 cM (Figs. 4, 5, 6). In some situation prediction accuracy at lower marker density was greater than that of higher marker density. For example, in the NAM 1 population (Fig. 4), the prediction accuracy with RR-BLUP was 0.29 at $c = 40$ cM compared to 0.26 at $c = 30$ cM. The numbers of markers in the GWS model at these two levels were 52 at 30 cM and 38 at 40 cM, overall very minor differences in the whole genome marker coverage. In addition, when marker densities are at these low levels, lower densities may actually capture more QTL with great effects than high marker density levels by chance, as genome-wide coverage is very minimal at the levels of $c = 30$ and $c = 40$.

Combined effects of marker density and training sample proportion based on NAM Population 1 (B73 × B97) for the trait DS were calculated (Fig. 7). The prediction ability increased with both marker density and training sample proportion. This gain with RR-BLUP over CIM increased with reductions of these two factors. These observations were actually consistent to results obtained from a separate analysis as shown in Figs. 1, 2, 3, 4, 5, 6. Influences of training sample proportion on accuracy with RR-BLUP and CIM were greater than that of marker density. Therefore, predictions were relatively insensitive to marker density compared to training sample proportion. Similar trends were seen for other populations and traits (data not shown).

Relationships between heritability H^2 and prediction precision for RR-BLUP and CIM at different levels of training sample proportions (0.20, 0.40, 0.60, and 0.80) were tested at marker density $c = 1.6$ cM (Fig. 8). The prediction accuracy increased with H^2 for both RR-BLUP and CIM. One exception was CIM with a training sample proportion of 0.20. Precisions with CIM remained low and

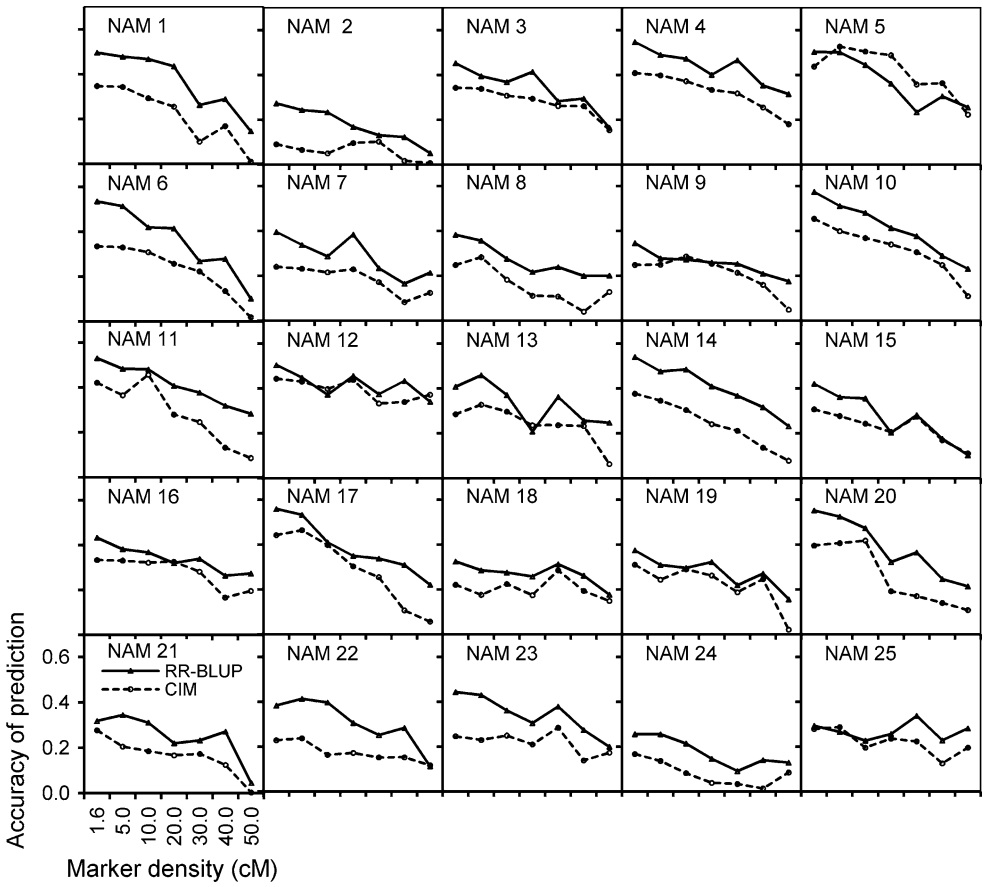

Fig. 4 Prediction accuracy using RR-BLUP and CIM with different marker density levels for DS at training sample proportion 0.8 based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *DS* days to silking, *NAM* nested association mapping

unchanged due to a poor power of QTL detection at this sample size. This suggested that H^2 had little effect on QTL detection with CIM when sample sizes were extremely low. More importantly, gains in prediction with RR-BLUP over CIM improved with the decrease of H^2 at different levels of training sample proportions and gains were statistically significant in most cases as shown in Tables 2, 3, 4. On average, for the case with training sample proportion 0.80, the value of RR-BLUP over CIM across the predictions with moderate H^2 (0.60–0.70) was 12% higher than that obtained from predictions with high H^2 (0.80–0.90). Similar trends were also noted at other levels of training sample proportion when fluctuations caused by training sample proportion were disregarded.

Discussion

While MLR has been widely used as a control method in simulated and empirical studies, it tends to bias gains in accuracy with GWS over MAS in bi-parental plant populations. The advantage of MLR lies in its simplicity and fast computation. Both stepwise regression for identifying QTL markers and multiple regression for estimating QTL effects are easily calculated. However, MLR fails to localize a QTL that lies between two flanking markers. In contrast, CIM gives more accurate estimations of QTL positions and effects by localizing a QTL at any position between two flanking markers (Jansen 1993; Zeng 1993, 1994). The CIM method results in higher prediction accuracies than MLR, which was demonstrated in the current study. As a result, advantages of GWS over MAS may be overestimated when MLR, rather than CIM, is used as a control for comparison purposes. The CIM has also been widely utilized in bi-parental QTL mapping and marker-assisted breeding projects (Utz et al. 2000). Thus, we suggest that CIM may be used as an appropriate control method in order to assess the performance of GWS in plant bi-parental populations.

Gains in precision with GWS over MAS may be affected by three key factors: training sample proportion, marker density and heritability. It has been shown that in order to maximize effectiveness of MAS, QTL positions and their respective magnitude must be estimated with a high precision (Utz et al. 2000). Often it is difficult to satisfy these conditions due to limited sample size, low heritability and low marker density in true breeding populations. This phenomenon causes low detection power,

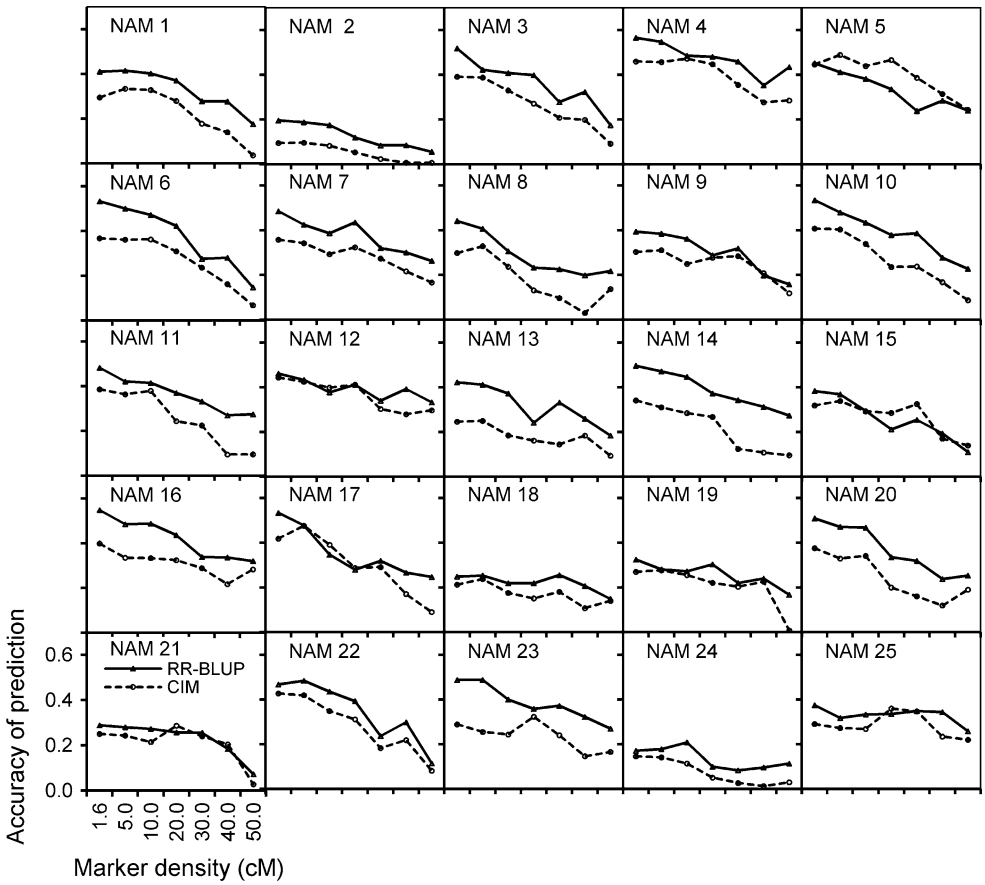

Fig. 5 Prediction accuracy using RR-BLUP and CIM with different marker density levels for DA at training sample proportion 0.8 based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *DA* days to anthesis, *NAM* nested association mapping

biased estimates of positions and magnitudes of QTL (Beavis 1994; Utz et al. 2000). In contrast, GWS is relatively insensitive to these factors. Gains in precision with GWS over MAS improve with decreases in these factors mentioned above. The GWS method tends to capture the genetic variation of minor QTL that are missed by MAS using several markers surrounding these QTL. Also, GWS shrinks overestimated QTL effects towards its true value by using RR-BLUP or other Bayesian approaches (Xu 2003).

The influence of marker density is lower than that of training sample proportions for GWS compared to MAS. Similar to Habier et al. (2007) simulation study, the RR-BLUP strategy tends to utilize genetic relationships between RILs in a population for prediction. Thus, reducing or increasing marker density does not necessarily influence genetic relationships between RILs in a NAM population. Another possible explanation is the GWS method allows multiple markers to detect a single QTL, while a single or flanking marker represents the QTL effect with MAS. Therefore, decreases of prediction accuracy with marker density in MAS are greater than that of GWS. When marker number reaches an area of high density, the prediction accuracy will not further improve. Marker densities corresponding to a mean genetic distance of 10 cM between markers is sufficient to capture LD between QTL and markers for GWS in bi-parental plant populations based on results from the current study.

The results in the paper show that RR-BLUP is a preferred method for estimating marker effects in GWS, providing accuracy of prediction higher than or comparable to BayesA and BayesB. This finding was consistent to results reported by Lorenzana and Bernardo (2009), although they compared RR-BLUP with an empirical Bayesian method, which was actually an extension of BayesA (Xu 2003, 2007). However, it has recently been reported that BayesB provided at least 60% higher accuracy than RR-BLUP based on a human simulation study (Meuwissen and Goddard 2010). Since our results were based on traits controlled by many QTL with small genetic effects, gains in accuracy with BayesB over RR-BLUP would not be expected in our study due to the polygenic nature for these traits (Buckler et al. 2009). It is unexpected that BayesB was worse than BayesA in our study. One explanation is that BayesB was performed using an ICE-based algorithm, rather than the original MCMC-based BayesB for the ease of intensive computations. The ICE-based BayesB has been found to produce slightly lower accuracies than MCMC-based BayesB (Meuwiseen et al. 2009). Another cause may be due to the difference between

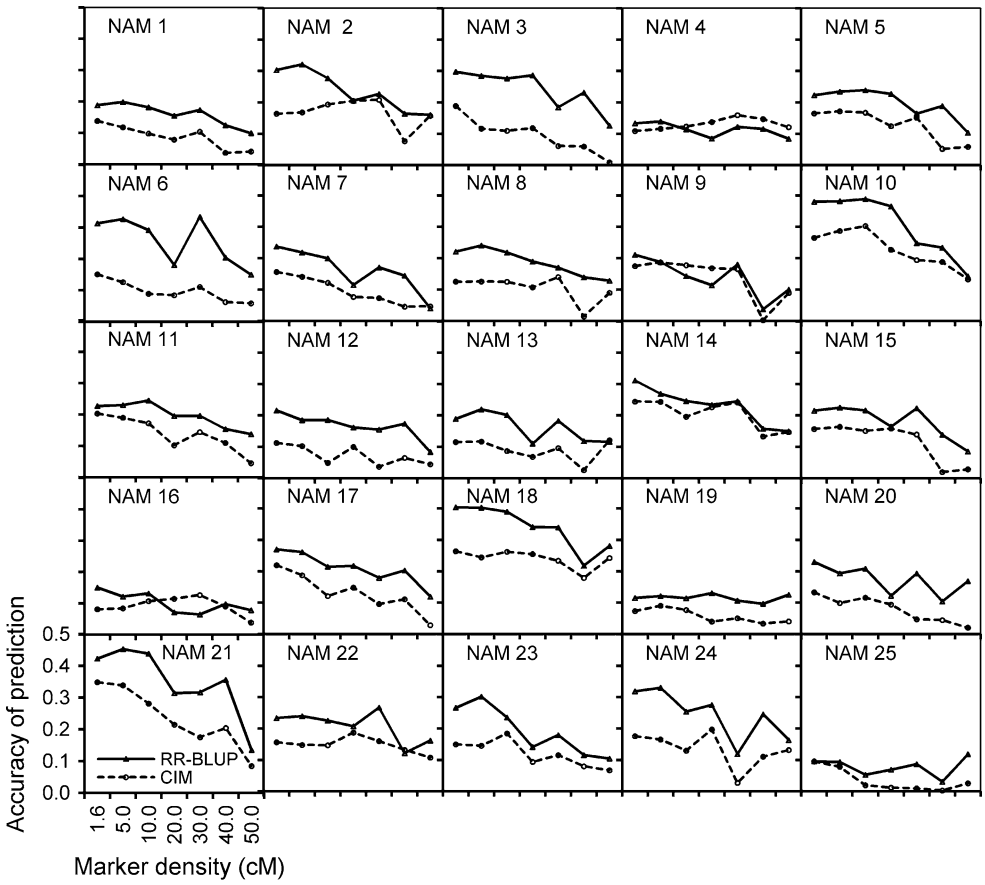

Fig. 6 Prediction accuracy using RR-BLUP and CIM with different marker density levels for ASI at training sample proportion 0.8 based on 25 NAM populations. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *ASI* anthesis-silking interval, *NAM* nested association mapping

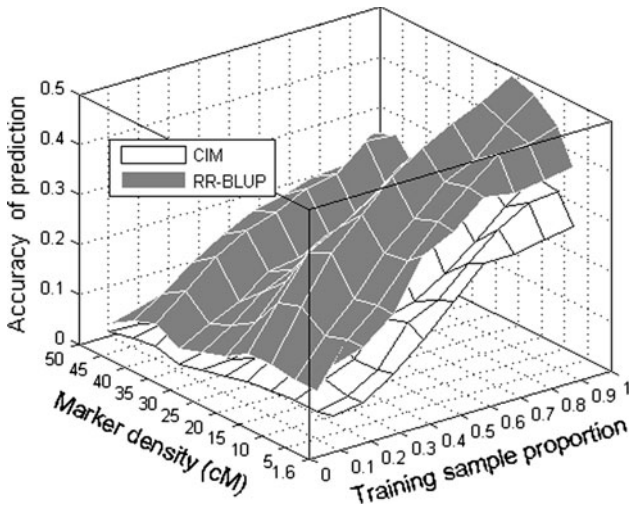

Fig. 7 A 3D plot of prediction accuracy with different proportions of training samples and markers density levels for DS based on the NAM population derived from B73 × B97. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *DS* days to silking, *NAM* nested association mapping

simulated data and real data used in this study. The conclusion that BayesB outperforms RR-BLUP is based only on simulation studies (Meuwissen et al. 2001; Meuwissen and Goddard 2010; Habier et al. 2007) with no empirical data supporting their respective conclusions. Thus, it is reasonable to infer that BayesB may benefit from the modeled data compared to RR-BLUP while RR-BLUP is superior when analyzing various types of real data from complex traits.

Additive genetic models only have been considered in this study and epistasis has been omitted for simplicity. Theoretically, including epistasis in GWS may increase prediction accuracy if epistasis is important and it can be modeled precisely (Lorenzana and Bernardo 2009). However, in practice, both simulations and empirical studies showed no advantage or even poor prediction ability by incorporating QTL pairwise epistatic effects into a GWS model (Lee et al. 2008; Lorenzana and Bernardo 2009; Piyasatian et al. 2007). This may be caused by too much noise being integrated into the model when combining main and epistatic effects of QTL. Still another explanation is the need for large sample sizes to utilize the genetic variation from epistasis as indicated by Lorenzana and Bernardo (2009). Additional studies are required to validate or disprove these theories. In the current study, since few genetic interactions were detected by the original QTL mapping study by Buckler et al. (2009), overall impacts of epistasis may be limited or even negligible on the conclusions obtained from this study.

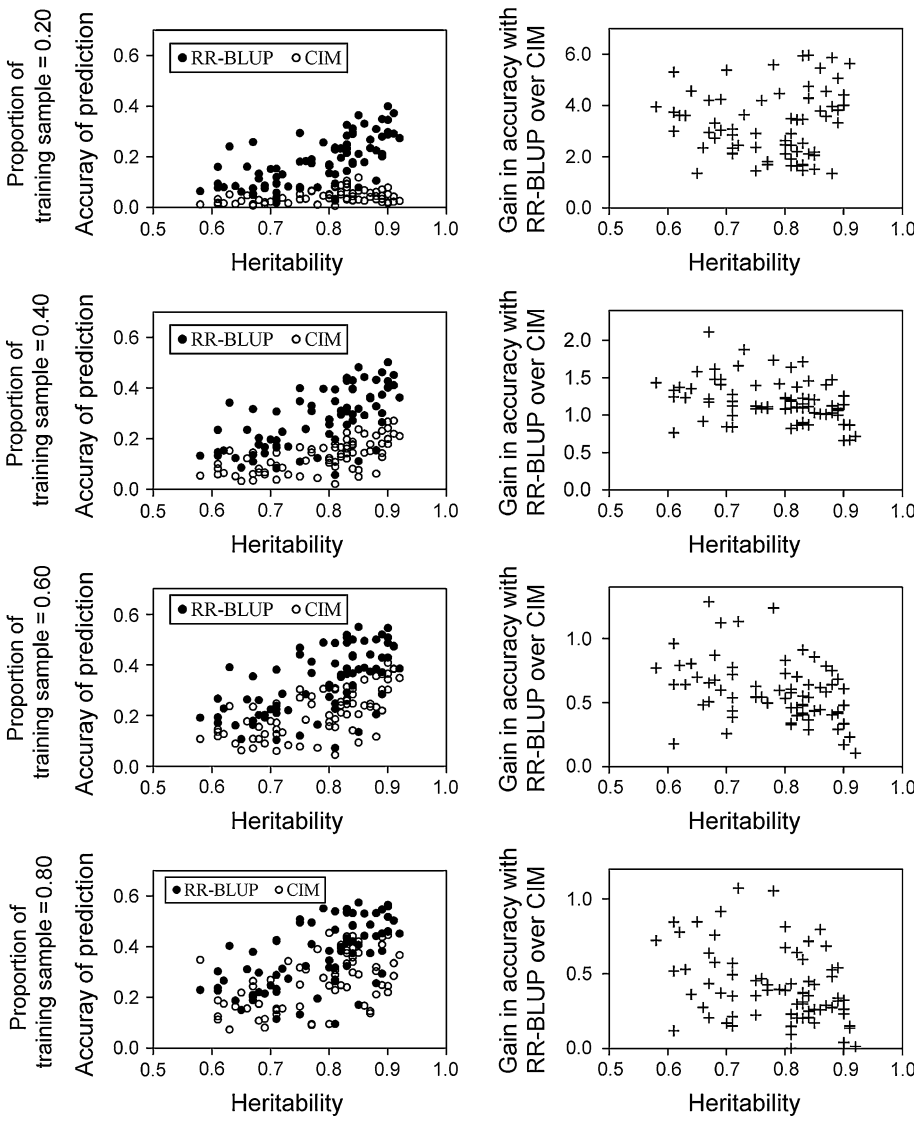

Fig. 8 Prediction accuracy using RR-BLUP and CIM with 25 NAM populations with different heritabilities for DS, DA and ASI based on four levels (0.20, 0.40, 0.60, 0.80) of proportions of training sample at marker density 1.6 cM. *RR-BLUP* ridge regression-best linear unbiased prediction, *CIM* composite interval mapping, *NAM* nested association mapping, *DS* days to silking, *DA* days to anthesis, *ASI* anthesis-silking interval

In practice, application of GWS in breeding may be determined by the objectives, resources of breeding programs, and the genetic architecture of traits. If the objective is focused on manipulating large-effect QTL such as plant diseases, MAS may be sufficient to pyramid favorable alleles by a successive backcrossing strategy (Jannink et al. 2010). In comparison to GWS, MAS for a few major QTL markers would be cheaper and a more efficient selection strategy. However, if the trait is controlled by many small-effect QTL, such as yield, it is difficult to identify the major effect QTL. The GWS strategy would bypass this problem by incorporating all major and minor QTL in a single model but genotyping costs increase. In addition, GWS can gradually improve a trait by incorporating multiple generations of recombination (Bernardo and Yu 2007).

Finally, it is important to note that NAM populations were developed by crossing 25 diverse exotic lines with a common elite line B73 in order to dissect the genetic architecture of complex traits in maize. More genetic variation and QTL exist in these populations compared to "traditional" corn breeding populations that are elite by elite parent inbreds, raising concerns of the application of the current study to plant breeding. However, both NAM and breeding populations are bi-parental-based, following a similar genetic recombination process by continued selfing to generate a population. In addition, similar prediction accuracies of GWS have been reported by Lorenzana and Bernardo (2009) using four, non-exotic maize populations. Therefore, although the number of QTL segregating and exotic type alleles in these NAM populations significantly differs from a commercial plant breeding program, our conclusions should not significantly differ.

Our conclusions were mainly based on maize flowering time traits in 25 NAM populations. The study showed that

the GWS strategy outperformed MAS via the CIM-method based on bi-parental segregation populations. However, the advantages GWS identified in this study should not be expected for any trait in any crop. In order to further validate the efficiency of GWS and get a broader understanding, further investigations are needed with additional traits and crops. Also, due to the complex genetic architecture of traits in breeding programs such as yield, one cannot expect one method will always be superior. Therefore, MAS and GWS may be used in complementarily in breeding programs or applied at different stages of cultivar development.

Acknowledgments The authors of the current manuscript would like to thank researchers and institutions who contributed to the Panzea database (http://www.panzea.org/). In addition, the authors would like express gratitude to the two anonymous reviewers for their detailed input in assessment of the manuscript.

References

Beavis WD (1994) QTL analysis: power, precision and accuracy. In: Paterson AH (ed) Molecular dissection of complex traits. CRC Press, Boca Raton, pp 145–162

Bernardo R, Yu J (2007) Prospects for genomewide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC (2009) The genetic architecture of maize flowering time. Science 325:714–718

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397

Haley C, Knott S (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324

Hayes B, Bowman P, Chamberlain A, Goddard M (2009) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, ON

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Jansen RC (1993) Interval mapping of multiple quantitative trait loci. Genetics 135:205–211

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756

Lee SH, van der Werf JHJ, Hayes BJ, Goddard ME, Visscher PM (2008) Predicting unobserved phenotypes for complex traits from whole-geome SNP data. PloS Genet 4:e1000231. doi: 10.1371/journal.pgen.1000231

Legarra A, Robert-Granie C, Manfredi E, Elsen JM (2008) Performance of genomic selection in mice. Genetics 180:611–618

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor Appl Genet 120:151–161

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Meuwissen TH, Goddard ME (2010) Accurate prediction of genetic values for complex traits by whole genome resequencing. Genetics 185:623–631

Meuwiseen TH, Solberg TR, Shepherd R, Woolliams JA (2009) A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. Genet Sel Evol 41(1):2

Piyasatian N, Fernando R, Dekkers JCM (2007) Genomic selection for marker-assisted improvement in line crosses. Theor Appl Genet 115:665–674

Utz HF, Melchinger AE, Schön CC (2000) Bias and sampling error of the estimated proportion of genotypic variance explained by quantitative trait loci determined from experimental data in maize using cross validation and validation with independent samples. Genetics 154:1839–1849

Xu S (2003) Estimating polygenic effects using markers of the entire genome. Genetics 163:789–801

Xu S (2007) An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics 63:513–521

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zeng ZB (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc Natl Acad Sci 90:10972–10976