



OPEN ACCESS

EDITED BY

Xueqiang Wang,
Zhejiang University, China

REVIEWED BY

Zitong Li,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia
Li Li,
University of New England, Australia

*CORRESPONDENCE

Hailan Liu
✉ hlzju@hotmail.com

SPECIALTY SECTION

This article was submitted to
Plant Bioinformatics,
a section of the journal
Frontiers in Plant Science

RECEIVED 04 November 2022

ACCEPTED 06 December 2022

PUBLISHED 21 December 2022

CITATION

Liu H, Xia C and Lan H (2022) An
efficient genomic prediction method
without the direct inverse of the
genomic relationship matrix.
Front. Plant Sci. 13:1089937.
doi: 10.3389/fpls.2022.1089937

COPYRIGHT

© 2022 Liu, Xia and Lan. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply
with these terms.

An efficient genomic prediction method without the direct inverse of the genomic relationship matrix

Hailan Liu *, Chao Xia and Hai Lan

Maize Research Institute, Sichuan Agricultural University, Chengdu, Sichuan, China

GBLUP, the most widely used genomic prediction (GP) method, consumes large and increasing amounts of computational resources as the training population size increases due to the inverse of the genomic relationship matrix (GRM). Therefore, in this study, we developed a new genomic prediction method (RHEPCG) that avoids the direct inverse of the GRM by combining randomized Haseman–Elston (HE) regression (RHE-reg) and a preconditioned conjugate gradient (PCG). The simulation results demonstrate that RHEPCG, in most cases, not only achieves similar predictive accuracy with GBLUP but also significantly reduces computational time. As for the real data, RHEPCG shows similar or better predictive accuracy for seven traits of the *Arabidopsis thaliana* F2 population and four traits of the *Sorghum bicolor* RIL population compared with GBLUP. This indicates that RHEPCG is a practical alternative to GBLUP and has better computational efficiency.

KEYWORDS

genomic prediction, GBLUP, genomic relationship matrix, randomized Haseman–Elston regression, preconditioned conjugate gradient

Introduction

Currently, genomic prediction (GP) has been widely applied to many species, such as dairy cattle, dairy sheep, maize, and wheat (Pszczola et al., 2011; Duchemin et al., 2012; Crossa et al., 2014). For example, significant achievements have been made in the genetic improvement of dairy cattle via GP in many countries, such as the United States, Australia, Canada, New Zealand, and France (Hayes et al., 2009; Winkelman et al., 2015; García-Ruiz et al., 2016; Weller et al., 2017). Moreover, GP helps to optimize the breeding procedure when used with many other breeding technologies. For example, it can accelerate the selection of superior pure lines from the large numbers of those generated by doubled haploid (DH) technology, which is otherwise a significant problem in terms of the consumption of time and money (Wang et al., 2020). Additionally, GP can rapidly

increase the frequencies of favorable alleles when combined with genome editing (GE) (Jenko et al., 2015; Bastiaansen et al., 2018).

Many computational methods of GP have been proposed and GBLUP is the most widely used (Meuwissen et al., 2001; Daetwyler et al., 2013; Mouresan et al., 2019). For conventional GBLUP, restricted maximum likelihood (REML) is often used to estimate heritability, and its computational complexity is cubic of the training population size (Xu et al., 2014). The fact that the inverse of the genomic relationship matrix (GRM) is essential when estimating the heritability in REML and calculating the best linear unbiased prediction (BLUP) contributes to the decrease of the computational efficiency of GBLUP when the size of the training population increases. To improve computational efficiency, methods such as IBS-based HE regression, algorithm for proven and young (APY), updating the inverse, recursive algorithm, spectral decomposition, and the preconditioned conjugate gradient (PCG) algorithm are employed (Kang et al., 2008; Legarra and Misztal, 2008; Misztal et al., 2009; Endelman, 2011; Faux et al., 2012; Meyer et al., 2013; Chen, 2014; Misztal, 2016; Liu and Chen, 2017; Masuda et al., 2017; Vandenplas et al., 2018; Vandenplas et al., 2020). In particular, PCG solves mixed model equations (MMEs) via iteration instead of by directly inverting the GRM. Recently, we proposed a fast GP method (SHEAPY) combining randomized Haseman–Elston regression (RHE-reg) and a modified APY (Liu and Chen, 2022). In the SHEAPY, RHE-reg is used to estimate heritability because of its high computational speed. In this study, we continue to combine it with PCG, calculating marker values to develop a new GP method (RHEPCG), which can significantly improve computational efficiency without the direct inverse of GRM.

Materials and methods

Genetic model and the linear system of MMEs

Herein, we only focus on additive effects, and the basic model is described as:

$$y = X\theta + Zu + e, \quad (1)$$

in which y is the $n \times 1$ vector of the standardized phenotypic values; θ is a fixed effect; X is the $n \times 1$ vector of the incidence; Z is the $n \times m$ matrix of the standardized genotypic values; u is the $m \times 1$ vector of SNP marker effects; and e is the $n \times 1$ vector of the residual error.

On the basis of the above genetic model of additive effect, the linear system of MMEs was as follows:

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + I \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{\theta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}, \quad (2)$$

in which I is the identity matrix, σ_g^2 is the variance of SNP marker effects, and σ_e^2 is the residual variance.

To solve MMEs, randomized HE-reg based on IBS was used to estimate heritability, which was then introduced into Equation 2 with residual variance. Then, PCG was used to solve Equation 2 to obtain marker values.

Estimating heritability via randomized HE-reg based on IBS

IBS-based RHE regression is a method of moment that can reduce computational time and memory to $O(\frac{nmk}{\max(\log_3^{(n)}, \log_3^{(m)})} + nm)$ and $O(nm)$, respectively (n , m , and k represent the number of samples, the number of markers, and the length of random vector, respectively) (Wu and Sankararaman, 2018; Liu and Chen, 2022). Here, it was used to estimate the heritability:

$$y_i y_j = b_0 + b_1 \omega_{ij} + e \quad (3)$$

in which y_i and y_j represent the phenotypic values of individuals i and j from the training population; b_0 is the intercept; b_1 is the regression coefficient; $\omega_{ij} = \frac{Z_i Z_j^T}{m}$ is the genetic relatedness between a pair of individuals i and j ; z_i and z_j are the genotype vector of individuals i and j ; and e is residual error. For a trait, its phenotypic variance is $\hat{\sigma}_y^2 = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}$, its additive genetic variance is $\hat{\sigma}_g^2 = \hat{b}_1$, and its error variance is $\hat{\sigma}_e^2 = \hat{\sigma}_y^2 - \hat{\sigma}_g^2$. The computational equation of $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$ is described as:

$$\begin{bmatrix} tr[\omega^2] & tr[\omega] \\ tr[\omega] & n \end{bmatrix} \begin{bmatrix} \hat{\sigma}_g^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} y^T \omega y \\ y^T y \end{bmatrix}, \quad (4)$$

in which $\omega = \frac{ZZ^T}{m}$ corresponds to the genomic related matrix between individuals and $tr[\omega] = n$. To accelerate computational efficiency, $tr[\hat{\omega}^2]$ was calculated via a randomized estimation. The equation is as below:

$$tr[\hat{\omega}^2] = \frac{1}{s} \frac{1}{m^2} \sum_{s=1}^S (w_s' ZZ^T) (ZZ^T w_s). \quad (5)$$

In the equation, S represents the rounds of randomization implemented, and was set as 5 throughout the study; each entry of w_s comes from a standard normal distribution $N(0,1)$.

Preconditioned conjugate gradient (PCG) algorithm

When the MMEs are described as $Ax = b$, in which A is the coefficient matrix, x is the vector of solutions, and b is the right-hand side, the PCG is used to solve the linear system of MMEs and compute marker effects. As it does not need to invert GRM like conventional methods, a much higher efficiency can be achieved (Vandenplas et al., 2019). Its code is as follows (Tsuruta et al., 2001; Vandenplas et al., 2018):

When $n = 0$,

$x_0=1$; $e_0=0$; $\alpha_0=1$ (1 is a vector of containing 1.);

$$r_0 = \mathbf{b} - \mathbf{A}x_0 \quad ;$$

$$p_0 = \mathbf{M}^{-1}r_0 \quad ;$$

When $n=1, 2, \dots$,

$$w_n = \mathbf{M}^{-1}p_{n-1} \quad ;$$

$$\alpha_n = p'_{n-1}w_n \quad ;$$

$$\beta_n = \alpha_n / \alpha_{n-1} \quad ;$$

$$\alpha_{n-1} = \alpha_n \quad ;$$

$$e_n = w_n + e_{n-1}\beta_n \quad ;$$

$$q_n = \mathbf{A}e_n \quad ;$$

$$\epsilon_n = p'_{n-1}w_n / e'_n q_n \quad ;$$

$$x_n = x_{n-1} + e_n \epsilon_n \quad ;$$

$$r_n = r_{n-1} - \epsilon_n q_n$$

Until convergence.

End.

Here, $\mathbf{A} = \begin{bmatrix} x^T x & x^T z \\ z^T x & z^T z + 1 \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix}$, \mathbf{M} is the preconditioner matrix, and $\mathbf{M} = \text{diag}(\mathbf{A})$; r , p , and w are vectors, $x = \begin{bmatrix} 1 \\ \alpha \\ \beta \end{bmatrix}$, and $b = \begin{bmatrix} x^T y \\ z^T y \end{bmatrix}$. To solve MMEs, $\hat{\sigma}_g^2$ via RHE regression based on IBS and $\hat{\sigma}_e^2$ are introduced into \mathbf{A} matrix.

Simulated data

The F2 population was simulated to evaluate the performance and cost time of GBLUP and RHEPCG. We simulated a chromosome with a length of 2,000 cM (the recombination rate was c between the i^{th} and $(i+1)^{\text{th}}$ markers), and all markers in this chromosome were defined as QTL, the effects of which followed a standard normal distribution. A series of different training population sizes (1,000, 1,200, 2,000, 6,000, 10,000, 15,000, and 20,000), candidate population sizes (100, 200, 300, and 400), and heritability (0.2, 0.4, 0.6, 0.65, and 0.8) were simulated. Each simulation scenario included 10 replications.

Real data

Two sets of data (*Arabidopsis thaliana* and *Sorghum bicolor*) were used to evaluate the predictive accuracy of GBLUP and RHEPCG. (1) An *A. thaliana* F2 population (P15) with 434

individuals derived from a cross between Br-0 and C24 was obtained from the study by Salomé et al. It consisted of a total of 233 SNP markers and seven traits, including DTF1 (days until visible flower buds in the center of the rosette), DTF2 (days until inflorescence stem reached 1 cm in height), DTF3 (days until first open flower), RLN (rosette leaf number), CLN (cauline leaf number), TLN (total leaf number: sum of RLN and CLN), and LIR1 (leaf initiation rate [RLN/DTF1]) (Salomé et al., 2011). (2) A *S. bicolor* RIL population with 399 individuals derived from a cross between *S. bicolor* BTx623 and *S. bicolor* IS3620C was obtained from the study by Kong et al. It consisted of a total of 381 bins and five traits, including PH (plant height), BTF (base to flag length), FTR (flag to rachis length), ND (number of nodes), and FL (days to flowering). The phenotype data were obtained from the University of Georgia Plant Science Farm, Watkinsville, GA, USA on May 10, 2011 (Kong et al., 2018).

Implementation and computations

The GBLUP and RHEPCG were written in R language (R Core Team, 2017) and run on a server of the CentOS Linux operating system (Intel (R) Xeon (R) CPU E7-4870 @2.40GHz) with 80 CPUs and 755G memory. The RHEPCG program is available from the authors. The squared correlation coefficient (r^2) between the phenotypes and the predicted genotypic values were defined as the prediction accuracy.

Results and discussion

Comparison of GBLUP and RHEPCG in simulated F2 population studies

A series of simulations of the F2 population at different levels of parameters, including training population size, candidate population size, and heritability were used to assess the estimated heritability, predictive accuracy, and consumption time of GBLUP and RHEPCG.

Table 1 shows the predictive accuracy and computational time of the GBLUP and RHEPCG at different training population sizes (1,000, 2,000, 6,000, 10,000, 15,000, and 20,000). As the training population size increased, both methods demonstrated an obvious uptrend of predictive accuracy. When the training population size was 1,000, RHEPCG was slightly better than GBLUP in predictive accuracy ($r_{RHEPCG}^2 = 0.597 \pm 0.007$ vs $r_{GBLUP}^2 = 0.541 \pm 0.027$), but when the training population size was 10,000, an opposite result was achieved ($r_{GBLUP}^2 = 0.640 \pm 0.014$ vs $r_{RHEPCG}^2 = 0.607 \pm 0.023$). In other conditions, both methods performed similarly (for example, when a training population size was 20,000 $r_{GBLUP}^2 = 0.653 \pm 0.017$ vs. $r_{RHEPCG}^2 = 0.648 \pm 0.021$). With the enlargement of the training population, the predictive accuracy approximated the true heritability, which as some studies have demonstrated, is

TABLE 1 Comparison of the estimated heritability, predictive accuracy, and computational time of GBLUP and RHEPCG at the different training population sizes based on 10 simulations in the *Arabidopsis thaliana* F2 population.

Training size	GBLUP			RHEPCG		
	\hat{h}_{GBLUP}^2	r_{GBLUP}^2	Average time of each simulation (s)	\hat{h}_{RHEPCG}^2	r_{RHEPCG}^2	Average time of each simulation (s)
1,000	0.718 ±0.055	0.541 ±0.027	135	0.637 ±0.042	0.597 ±0.007	31
2,000	0.648 ±0.029	0.576 ±0.011	465	0.625 ±0.051	0.570 ±0.010	58
6,000	0.668 ±0.052	0.615 ±0.011	1,580	0.621 ±0.055	0.594 ±0.013	179
10,000	0.680 ±0.042	0.640 ±0.014	7,932	0.600 ±0.043	0.607 ±0.023	526
15,000	0.650 ±0.038	0.651 ±0.013	22,576	0.724 ±0.059	0.643 ±0.024	820
20,000	0.698 ±0.042	0.653 ±0.017	53,666	0.728 ±0.043	0.648 ±0.021	1,237

The training population sizes were 1,000, 2,000, 6,000, 10,000, 15,000, and 20,000; the candidate population size was 100; the length of the chromosome was 2,000; heritability (h^2) was set as 0.65; the recombination rate c was set as 0.01. \hat{h}_{GBLUP}^2 and \hat{h}_{RHEPCG}^2 represent the estimated heritability via GBLUP and RHEPCG, respectively; r_{GBLUP}^2 and r_{RHEPCG}^2 represent the squared correlation coefficient between the phenotypes and the predicted genotypic values and are defined as the prediction accuracy; the values after \pm represent the corresponding standard error.

the upper bound of predictive accuracy (de los Campos et al., 2013; Liu and Chen, 2018). Meanwhile, RHEPCG was significantly faster than GBLUP (for example, when a training population size was 20,000 $T_{GBLUP}=53666s$ vs $T_{RHEPCG}=1237s$). When the training population size was 1,000, 2,000, 6,000, 10,000, 15,000, and 20,000, the computational time of GBLUP was 4, 8, 9, 15, 28, and 43 times that of RHEPCG, respectively. In other words, the larger the training population size, the more obvious the advantage of the computational efficiency of RHEPCG becomes.

In Table 2, the predictive accuracy of both methods was similar at different candidate population sizes (100, 200, 300, and 400), which means the latter has no significant impact on the former. Table 3 shows that the predictive accuracy of GBLUP and RHEPCG increased when heritability varied from 0.2 to 0.4, 0.6, and 0.8. According to further analysis, the correlation between the estimated heritability and the predictive accuracy was 0.999 for both GBLUP and RHEPCG ($P_{Two-tailed}=0.001$),

and our results are consistent with Daetwyler et al. (2008) in that heritability can significantly influence predictive accuracy.

Currently, IBS-based RHE regression is used to estimate gene-environmental heritability and multi-trait genetic correlation (Kerin and Marchini, 2020; Wu et al., 2022). Therefore, RHEPCG can also be applied to such data via the incorporation of these effects into the model in the future.

Comparison of GBLUP and RHEPCG in studies of the *A. thaliana* F2 and *S. bicolor* RIL populations

A comparison of GBLUP and RHEPCG based on seven traits of the *A. thaliana* F2 population was performed in this study. Table 4 shows a significant difference between the estimated heritability via GBLUP and that via RHEPCG in

TABLE 2 Comparison of the predictive accuracy of GBLUP and RHEPCG at the different candidate population sizes based on 10 simulations in the *Arabidopsis thaliana* F2 population.

Candidate population	r_{GBLUP}^2	r_{RHEPCG}^2
100	0.581 ± 0.022	0.560 ± 0.016
200	0.561 ± 0.015	0.568 ± 0.016
300	0.561 ± 0.014	0.578 ± 0.008
400	0.580 ± 0.008	0.571 ± 0.011

The training population sizes was 1,200; the candidate population sizes were 100, 200, 300, and 400; the length of the chromosome was 2,000; heritability (h^2) was set as 0.65; the recombination rate (c) was set as 0.01. r_{GBLUP}^2 and r_{RHEPCG}^2 represent the squared correlation coefficient between the phenotypes and the predicted genotypic values and are defined as the prediction accuracy; the values after \pm represent the corresponding standard error.

TABLE 3 Comparison of the estimated heritability and predictive accuracy of GBLUP and RHEPCG at different levels of heritability based on 10 simulations in the *Arabidopsis thaliana* F2 population.

h^2	GBLUP		RHEPCG	
	\hat{h}_{GBLUP}^2	r_{GBLUP}^2	\hat{h}_{RHEPCG}^2	\hat{h}_{GBLUP}^2
0.2	0.220 ± 0.014	0.165 ± 0.021	0.203 ± 0.009	0.127 ± 0.017
0.4	0.460 ± 0.021	0.346 ± 0.027	0.406 ± 0.014	0.304 ± 0.018
0.6	0.708 ± 0.032	0.535 ± 0.025	0.611 ± 0.019	0.506 ± 0.016
0.8	0.910 ± 0.034	0.667 ± 0.032	0.817 ± 0.024	0.726 ± 0.012

The training population sizes was 1,200; the candidate population size was 100; the length of the chromosome was 2,000; heritability (h^2) was set as 0.2, 0.4, 0.6, and 0.8; the recombination rate (c) was set as 0.01. \hat{h}_{GBLUP}^2 and \hat{h}_{RHEPCG}^2 represent the estimated heritability via GBLUP and RHEPCG, respectively; r_{GBLUP}^2 and r_{RHEPCG}^2 represent the squared correlation coefficient between the phenotypes and the predicted genotypic values and are defined as the prediction accuracy; the values after ± represent the corresponding standard error.

TABLE 4 Comparison of the predictive accuracy between GBLUP and RHEPCG in seven traits from the *Arabidopsis thaliana* F2 (P15) population based on 10 simulations.

Trait	Training	Candidate	GBLUP		RHEPCG	
			\hat{h}_{GBLUP}^2	r_{GBLUP}^2	\hat{h}_{RHEPCG}^2	r_{RHEPCG}^2
DTF1	300	133	0.731 ± 0.025	0.383 ± 0.017	0.323 ± 0.034	0.368 ± 0.024
DTF2	300	134	0.604 ± 0.035	0.406 ± 0.024	0.389 ± 0.029	0.401 ± 0.023
DTF3	300	134	0.645 ± 0.033	0.351 ± 0.024	0.311 ± 0.030	0.321 ± 0.025
RLN	300	131	0.923 ± 0.025	0.555 ± 0.046	0.524 ± 0.030	0.640 ± 0.015
CLN	300	130	0.570 ± 0.017	0.350 ± 0.017	0.282 ± 0.019	0.342 ± 0.018
TLN	300	130	0.910 ± 0.022	0.572 ± 0.029	0.421 ± 0.028	0.619 ± 0.011
LIR1	300	131	0.449 ± 0.033	0.208 ± 0.012	0.162 ± 0.011	0.197 ± 0.017

\hat{h}_{GBLUP}^2 and \hat{h}_{RHEPCG}^2 represent the estimated heritability via GBLUP and RHEPCG, respectively; r_{GBLUP}^2 and r_{RHEPCG}^2 represent the squared correlation coefficient between the phenotypes and the predicted genotypic values and are defined as the prediction accuracy; the values after ± represent the corresponding standard error.

TABLE 5 Comparison of the predictive accuracy between GBLUP and RHEPCG in five traits from the *Sorghum bicolor* RIL population based on 10 simulations.

Trait	Training	Candidate	GBLUP		RHEPCG	
			\hat{h}_{GBLUP}^2	r_{GBLUP}^2	\hat{h}_{RHEPCG}^2	r_{RHEPCG}^2
PH	300	88	0.653 ± 0.013	0.289 ± 0.011	0.331 ± 0.038	0.257 ± 0.016
BTF	300	88	0.619 ± 0.011	0.335 ± 0.019	0.436 ± 0.047	0.302 ± 0.017
FTR	300	88	0.337 ± 0.008	0.182 ± 0.018	0.493 ± 0.056	0.155 ± 0.026
ND	300	88	0.490 ± 0.011	0.306 ± 0.015	0.747 ± 0.065	0.197 ± 0.039
FL	300	93	0.629 ± 0.014	0.371 ± 0.014	0.622 ± 0.043	0.393 ± 0.031

\hat{h}_{GBLUP}^2 and \hat{h}_{RHEPCG}^2 represent the estimated heritability via GBLUP and RHEPCG, respectively; r_{GBLUP}^2 and r_{RHEPCG}^2 represent the squared correlation coefficient between the phenotypes and the predicted genotypic values and are defined as the prediction accuracy; the values after ± represent the corresponding standard error.

seven traits of *A. thaliana* F2 (P15). Meanwhile, the seven traits were used to evaluate the predictive accuracy of GBLUP and RHEPCG (Table 4). The two methods showed similar predictive accuracy in six traits: DTF1, DTF2, DTF3, CLN, TLN and LIR1 (for example, the predictive accuracies of DTF1 were $r_{GBLUP}^2 = 0.383 \pm 0.017$ and $r_{RHEPCG}^2 = 0.368 \pm 0.024$). RHEPCG was

significantly better than GBLUP for RLN (the predictive accuracies of RLN were $r_{GBLUP}^2 = 0.555 \pm 0.046$ and $r_{RHEPCG}^2 = 0.640 \pm 0.015$).

In addition, the predictive accuracy of GBLUP and RHEPCG was evaluated based on five traits of the *S. bicolor* RIL population (Table 5). The estimated heritability of PH, BTF, FTR, and ND

via GBLUP differed significantly from that via RHEPCG, and the two methods had similar predictive accuracy for PH, BTF, FTR, and FL (for example, the predictive accuracies of PH were $r_{GBLUP}^2 = 0.289 \pm 0.011$ and $r_{RHEPCG}^2 = 0.257 \pm 0.016$), and the predictive accuracy of GBLUP was significantly superior to that of RHEPCG for ND (the predictive accuracies of ND were $r_{GBLUP}^2 = 0.306 \pm 0.015$ and $r_{RHEPCG}^2 = 0.197 \pm 0.039$).

These results show that GBLUP and RHEPCG have different estimated heritability in some traits of *A. thaliana* and *S. bicolor*. According to Chen (2016), strong selection can lead to differences in the estimated heritability via LMM and HE, and therefore, these traits are very likely to have undergone strong selection. In the future, we will investigate the influence of strong selection on predictive accuracy.

Conclusion

We present a new computing method of genomic prediction (RHEPCG) that does not require direct inversion of the GRM. Compared with GBLUP, it can significantly reduce computational time while maintaining predictive accuracy.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

References

- Bastiaansen, J. W. M., Bovenhuis, H., Groenen, M. A. M., Megens, H. J., and Mulder, H. A. (2018). The impact of genome editing on the introduction of monogenic traits in livestock. *Genet. Selection Evol.* 50, 18. doi: 10.1186/s12711-018-0389-7
- Chen, G. B. (2014). Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Front. Genet.* 5. doi: 10.3389/fgene.2014.00107
- Chen, G. B. (2016). On the reconciliation of missing heritability for genome-wide association studies. *Eur. J. Hum. Genet.* 24, 1810–1816. doi: 10.1038/ejhg.2016.89
- Crossa, J., Pérez, P., Hickey, J., Burguño, J., Ornella, L., Cerón-Rojas, J., et al. (2014). Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* 112, 48–60. doi: 10.1038/hdy.2013.16
- Daetwyler, H. D., Calus, M. P. L., Pong-Wong, R., de los Campos, G., and Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193, 347–365. doi: 10.1534/genetics.112.147983
- Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3, e3395. doi: 10.1371/journal.pone.0003395
- de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C., and Sorensen, D. (2013). Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* 9, e1003608. doi: 10.1371/journal.pgen.1003608
- Duchemin, S. I., Colombani, C., Legarra, A., Baloché, G., Larroque, H., Astruc, J. M., et al. (2012). Genomic selection in the French lacauine dairy sheep breed. *J. Dairy Sci.* 95, 2723–2733. doi: 10.3168/jds.2011-4980
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with r package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Faux, P., Gengler, N., and Misztal, I. (2012). A recursive algorithm for decomposition and creation of the inverse of the genomic relationship matrix. *J. Dairy Sci.* 95, 6093–6102. doi: 10.3168/jds.2011-5249
- García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggins, G. R., Ruiz-López, F. J., and van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A.* 113, E3995–E4004. doi: 10.1073/pnas.1519061113
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 433–443. doi: 10.3168/jds.2008-1646
- Jenko, J., Gorjanc, G., Cleveland, M. A., Varshney, R. K., Whitelaw, C. B. A., Woolliams, J. A., et al. (2015). Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genet. Selection Evol.* 47, 55. doi: 10.1186/s12711-015-0135-3
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101
- Kerin, M., and Marchini, J. (2020). A non-linear regression method for estimation of gene–environment heritability. *Bioinformatics* 36, 5632–5639. doi: 10.1093/bioinformatics/btaa1079
- Kong, W., Kim, C., Zhang, D., Guo, H., Tan, X., Jin, H., et al. (2018). Genotyping by sequencing of 393 *Sorghum bicolor* BTx6233xIS3620C recombinant inbred lines

Author contributions

HLL conceived and performed the study, interpreted the results, and wrote the manuscript. HL and CX interpreted the results and wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (32271984).

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

improves sensitivity and resolution of QTL detection. *G3 Genes|Genomes|Genetics* 8, 2563–2572. doi: 10.1534/g3.118.200173

Legarra, A., and Misztal, I. (2008). Technical note: Computing strategies in genome-wide selection. *J. Dairy Sci.* 91, 360–366. doi: 10.3168/jds.2007-0403

Liu, H., and Chen, G. B. (2017). A fast genomic selection approach for large genomic data. *Theor. Appl. Genet.* 130, 1277–1284. doi: 10.1007/s00122-017-2887-3

Liu, H., and Chen, G. B. (2018). A new genomic prediction method with additive-dominance effects in the least-squares framework. *Heredity* 121, 196–204. doi: 10.1038/s41437-018-0099-5

Liu, H., and Chen, G. B. (2022). A novel genomic prediction method combining randomized haseman-elston regression with a modified algorithm for proven and young for large genomic data. *Crop J.* 10, 550–554. doi: 10.1016/j.cj.2021.09.001

Masuda, Y., Misztal, I., Legarra, A., Tsuruta, S., Lourenco, D. A. L., Fragomeni, B. O., et al. (2017). Technical note: Avoiding the direct inversion of the numerator relationship matrix for genotyped animals in single-step genomic best linear unbiased prediction solved with the preconditioned conjugate gradient. *J. Anim. Sci.* 95, 49–52. doi: 10.2527/jas.2016.0699

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Meyer, K., Tier, B., and Graser, H. U. (2013). Technical note: updating the inverse of the genomic relationship matrix. *J. Anim. Sci.* 91, 2583–2586. doi: 10.2527/jas.2012-6056

Misztal, I. (2016). Inexpensive computation of the inverse of the genomic relationship matrix in populations with small effective population size. *Genetics* 202, 401–409. doi: 10.1534/genetics.115.182089

Misztal, I., Legarra, A., and Aguilar, I. (2009). Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92, 4648–4655. doi: 10.3168/jds.2009-2064

Mouresan, E. F., Selle, M., and Rönnegård, L. (2019). Genomic prediction including SNP-specific variance predictors. *G3 Genes|Genomes|Genetics* 9, 3333–3343. doi: 10.1534/g3.119.400381

Pszczola, M., Mulder, H. A., and Calus, M. P. L. (2011). Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. *J. Dairy Sci.* 94, 431–441. doi: 10.3168/jds.2009-2840

R Core Team (2017). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).

Salomé, P. A., Bomblies, K., Laitinen, R. A. E., Yant, L., Mott, R., and Weigel, D. (2011). Genetic architecture of flowering-time variation in *Arabidopsis thaliana*. *Genetics* 188, 421–433. doi: 10.1534/genetics.111.126607

Tsuruta, S., Misztal, I., and Strandén, I. (2001). Use of preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.* 79, 1166–1172. doi: 10.2527/2001.7951166x

Vandenplas, J., Calus, M. P. L., Eding, H., and Vuik, C. (2019). A second-level diagonal preconditioner for single-step SNPBLUP. *Genet. Selection Evol.* 51, 30. doi: 10.1186/s12711-019-0472-8

Vandenplas, J., Eding, H., Bosmans, M., and Calus, M. P. L. (2020). Computational strategies for the preconditioned conjugate gradient method applied to ssSNPBLUP, with an application to a multivariate maternal model. *Genet. Selection Evol.* 52, 24. doi: 10.1186/s12711-020-00543-9

Vandenplas, J., Eding, H., Calus, M. P. L., and Vuik, C. (2018). Deflated preconditioned conjugate gradient method for solving single-step BLUP models efficiently. *Genet. Selection Evol.* 50, 51. doi: 10.1186/s12711-018-0429-3

Wang, N., Wang, H., Zhang, A., Liu, Y., Yu, D., Hao, Z., et al. (2020). Genomic prediction across years in a maize doubled haploid breeding program to accelerate early-stage testcross testing. *Theor. Appl. Genet.* 133, 2869–2879. doi: 10.1007/s00122-020-03638-5

Weller, J. I., Ezra, E., and Ron, M. (2017). Invited review: A perspective on the future of genomic selection in dairy cattle. *J. Dairy Sci.* 100, 8633–8644. doi: 10.3168/jds.2017-12879

Winkelman, A. M., Johnson, D. L., and Harris, B. L. (2015). Application of genomic evaluation to dairy cattle in New Zealand. *J. Dairy Sci.* 98, 659–675. doi: 10.3168/jds.2014-8560

Wu, Y., Burch, K. S., Ganna, A., Pajukanta, P., Pasaniuc, B., and Sankaraman, S. (2022). Fast estimation of genetic correlation for biobank-scale data. *Am. J. Hum. Genet.* 109, 24–32. doi: 10.1016/j.ajhg.2021.11.015

Wu, Y., and Sankaraman, S. (2018). A scalable estimator of SNP heritability for biobank-scale data. *Bioinformatics* 34, i187–i194. doi: 10.1093/bioinformatics/bty253

Xu, S., Zhu, D., and Zhang, Q. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111