American Genetic Association

OXFORD

Original Article

Genetic Mapping with Background Control for Quantitative Trait Locus (QTL) in 8-Parental Pure-Line Populations

Jinhui Shi, Jiankang Wang, and Luyan Zhang

From the National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, No. 12 Zhongguancun South Street, Beijing 100081, China.

Address correspondence to L. Zhang at the address above, or e-mail: zhangluyan@caas.cn.

Abstract

Multiparental advanced generation intercross (MAGIC) populations provide abundant genetic variation for use in plant genetics and breeding. In this study, we developed a method for quantitative trait locus (QTL) detection in pure-line populations derived from 8-way crosses, based on the principles of inclusive composite interval mapping (ICIM). We considered 8 parents carrying different alleles with different effects. To estimate the 8 genotypic effects, 1-locus genetic model was first built. Then, an orthogonal linear model of phenotypes against marker variables was established to explain genetic effects of the locus. The linear model was estimated by stepwise regression and finally used for phenotype adjustment and background genetic variation control in QTL mapping. Simulation studies using 3 genetic models demonstrated that the proposed method had higher detection power, lower false discovery rate (FDR), and unbiased estimation of QTL locations compared with other methods. Marginal bias was observed in the estimation of QTL effects. An 8-parental recombinant inbred line (RIL) population previously reported in cowpea and analyzed by interval mapping (IM) was reanalyzed by ICIM and genome-wide association mapping implemented in software FarmCPU. The results indicated that ICIM identified more QTLs explaining more phenotypic variation than did IM; ICIM provided more information on the detected QTL than did FarmCPU; and most QTLs identified by IM and FarmCPU were also detected by ICIM.

Subject areas: Quantitative genetics and Mendelian inheritance
Keywords: 8-way cross, inclusive composite interval mapping, pure lines, quantitative trait locus

Since Mendel's experiments in plant hybridization were rediscovered, biparental segregating populations, such as double haploid line (DH), recombinant inbred lines (RIL), backcross (BC), and F_2 populations, have been widely used in genetic studies and also the main genetic materials to identify quantitative trait loci (QTLs) (Wang et al. 2014). Major objective of studies on QTL mapping is to identify chromosomal polymorphisms associated with phenotypic traits in parents. In biparental populations, genetic loci without variation between the 2 parents cannot be detected, and the number of recombination events is relatively limited, resulting in a lack of mapping precision (Huang et al. 2012). In addition, it is not clear whether the identified QTL has multiple alleles. Multiparental advanced generation intercross (MAGIC) populations are emerging resources within the field of genetics (Mackay and Powell 2007; Cavanagh et al. 2008). MAGIC populations are multifounder equivalents of the advanced intercross (Darvasi and Soller 1995), similar to the

heterogeneous stocks (HS; Mott et al. 2000) and the Collaborative Cross (CC; Churchill et al. 2004) used in mouse genetics.

MAGIC populations are normally created from 8-way crosses. Firstly, 4 single crosses are made from 8 homozygous parents. Secondly, two 4-way crosses are generated from the 4 single crosses. Thirdly, an 8-way cross is made from the two 4-way crosses. Finally, DHs are produced by embryo rescue and pollen culture technology, or RILs are produced by repeated selfing and single-seed descent from the 8-way cross. MAGIC populations possess a greater number of recombination events and higher genotypic diversity than other populations, and increase the number of QTLs, and precision and resolution in QTL detection (Cavanagh et al. 2008). In addition, similar to biparental populations, MAGIC populations usually have no population substructure, reducing the risk of false-positive QTLs that may be caused by population structure (Kover et al. 2009).

In plants, MAGIC and MAGIC-like populations have been developed in a wide range of species, such as *Arabidopsis thaliana* (Cavanagh et al. 2008; Kover et al. 2009; Huang et al. 2011), rice (Bandillo et al. 2013), maize (Dell'Acqua et al. 2015), wheat (Huang et al. 2012; Mackay et al. 2014), barley (Sannemann et al. 2015), and cowpea (Huynh et al. 2018). Furthermore, genome-wide association studies (GWAS) have been employed in some MAGIC populations. Based on collections of related individuals, GWAS take advantage of historical recombination events that have accumulated over thousands of generations (Korte et al. 2013). However, GWAS have some disadvantages. Random mating causes linkage disequilibrium (LD) decay, therefore LD in the population is low. The repeatability of mapping results by GWAS is poor between populations and between mapping methods. GWAS have less power to detect alleles with small genetic variation and low frequency (Ward and Kellis 2012). Moreover, the accumulated contribution of significant single nucleotide polymorphisms (SNPs) can explain only part of the genetic variation, which leads to the phenomenon of missing heritability (Eichler et al. 2010).

At present, there are several linkage mapping methods available for MAGIC populations from 8-way crosses. The most common method is interval mapping (IM), which tests each chromosomal position for association with the trait of interest (Lander and Botstein 1989). Mott et al. (2000) implemented IM, based on founder probabilities, in the R package *happy* to analyze the HS population of mice. R/happy was also applied to a multiparental population of *Drosophila* (King et al. 2012). Composite IM (CIM) is based on the idea of covariables (Jansen 1994; Zeng 1994), available in the package R/mpMap (Huang and George 2011) for multiparental populations. Verbyla et al. (2007) proposed a model called whole genome average IM (WGAIM), which considered population structure and nongenetic effects (such as spatial variation) in the mixed-model framework. Although WGAIM increased the correct detection of QTL, it also increased false discovery rate (FDR) (Verbyla et al. 2007). Wei and Xu (2016) developed a mixed-model method, which is available in R/MagicQTL, setting the parental effects to random in accordance with a normal distribution. The method was subdivided into Fixed-A, Fixed-B, Random-A, and Random-B, among which Fixed-B and Random-B had better performance. However, the computation complex of this method is high (Wei and Xu 2016). Broman et al. (2019) developed the R package *qtl2* for QTL mapping with high-dimensional data and multiparental populations (such as CC, HS, and so on). But QTL analysis in R/qtl2 is only conducted by genome scan with single-QTL model, instead of multiple-QTL model, and hard to explore the possibility of multiple causal SNPs in a QTL region (Broman et al. 2019).

Inclusive CIM (ICIM) was originally proposed for bi-parental populations (Li et al. 2007; Zhang et al. 2008; Wang 2009) and then extended to 4-way-cross F_1 populations (Zhang et al. 2015) and pure-line populations from 4-way crosses (Zhang et al. 2017). Background control was used to increase QTL detection power and reduce FDR. Based on the ICIM principle, we developed a QTL mapping methodology for DH and RIL populations from 8-way crosses. Our objectives in this study were 1) to present an orthogonal linear model of phenotypes against marker variables to explain the genetic effects; 2) to propose an algorithm of 1-dimensional scanning with background control to estimate QTL locations and genotypic effects of the 8 parents; and 3) to investigate the efficiency of the approach by simulation studies and an actual population in cowpea.

Materials and Methods

Classification of Markers in Pure-Line Populations from 8-Way Crosses

Based on the number of identifiable alleles in the 8 parents, a total of 4139 marker categories can be differentiated in the pure-line populations from 8-way crosses (Zhang et al. 2019). Markers belonging to category ABCDEFGH represented the ideal situation, in which the parents had 8 identifiable alleles, denoted by *A*, *B*, *C*, *D*, *E*, *F*, *G*, and *H*. Their corresponding genotypes were denoted by *AA*, *BB*, *CC*, *DD*, *EE*, *FF*, *GG*, and *HH*. Markers belonging to the remaining categories were called incomplete loci. Linkage analysis and map construction were described in Zhang et al. (2019). Based on the constructed linkage map, incomplete and missing markers can be imputed. After imputation, all markers belonged to category ABCDEFGH, and no missing marker types remained. Therefore, in the following QTL mapping study, all markers were assumed to have 8 identifiable alleles, and there were no missing marker types as well.

One-Locus Model for Pure-Line Populations from 8-Way Crosses

The 8 alleles at 1 QTL were designated by A_q, B_q, C_q, D_q, E_q, F_q, and G_q. The genotypic value of an individual with a known QTL genotype was defined as

$$\mu_k = \mu + a_k w_k, k = 1, 2, \ldots, 8 \tag{1}$$

where μ_k was the kth genotypic value of the QTL; μ was the overall mean of the 8 QTL genotypes; a_k was the kth genotypic effect; and w_k was the indicator of QTL genotype, valued at 1 for the kth parental allele and 0 for other parental alleles. Mean and genotypic effects were calculated as

$$\mu = \frac{1}{8} \sum_{i=1}^{8} \mu_i, a_k = \frac{1}{8} \left(7\mu_k - \sum_{i=1, i \neq k}^{8} \mu_i \right), k = 1, 2, \ldots, 8 \tag{2}$$

When there was no segregation distortion, the genetic variation contributed by the QTL was defined as

$$V_Q = \frac{1}{8} \sum_{i=1}^{8} \mu_i^2 - \left(\frac{1}{8} \sum_{i=1}^{8} \mu_i \right)^2 = \frac{1}{8} \sum_{i=1}^{8} a_i^2 \tag{3}$$

It is worth noting that there was 1 restriction on the 9 genetic parameters (i.e., μ and a_k, k = 1, 2, …, 8) to be estimated in Equation 2, that is, sum of the 8 genotypic effects must be equal to 0. One orthogonal model equivalent to Equation 1 but with no restriction was built in Equation 4 to avoid the complexity caused by the restriction.

$$G = \mu + b_1 u + b_2 v + b_3 s + b_4 uv + b_5 us + b_6 vs + b_7 uvs \tag{4}$$

where G was the genotypic value of an individual with known QTL genotype, and the definition of b_i was as given below.

$$b_1 = \frac{a_1 + a_2 + a_3 + a_4}{4}, b_2 = \frac{a_1 + a_2 + a_5 + a_6}{4}$$

$$b_3 = \frac{a_1 + a_3 + a_5 + a_7}{4}, b_4 = \frac{a_1 + a_2 + a_7 + a_8}{4}$$

$$b_5 = \frac{a_1 + a_3 + a_6 + a_8}{4}, b_6 = \frac{a_1 + a_4 + a_5 + a_8}{4}$$

$$b_7 = \frac{a_1 + a_4 + a_6 + a_7}{4} \tag{5}$$

where u, v, and s were the 3 basic orthogonal variables for different QTL genotypes in Equation 4, and their values were given in Table 1. The other orthogonal variables were derived from u, v, and s. Let **X** represent the 8 × 8 design matrix in Equation 4; it can be easily seen that $\mathbf{X}^T\mathbf{X}$ was a diagonal matrix, indicating the orthogonality of the model of Equation 4.

Assume 1 QTL is located between 2 markers; A_1, B_1, C_1, D_1, E_1, F_1, G_1, and H_1 are the 8 alleles of the left flanking marker; and A_2, B_2, C_2, D_2, E_2, F_2, G_2, and H_2 are the 8 alleles of the right flanking marker. Supplementary Tables S1 and S2 showed the frequency of the QTL genotype for each marker class in DH and RIL populations from 8-way crosses, respectively, where r represented the 1-meiosis recombination frequency, and R was the recombination frequency accumulated during the repeated generations of selfing. The relationship between these 2 variables was $R = 2r/1 + 2r$. A quantity analogous to the 3-point coincidence for RILs derived from 8-way crosses can be referred in Teuscher and Broman (2007), and information for 2-locus genotype probabilities in RILs derived from 8-way crosses was provided in Broman (2012). Similar to variables u, v, and s in Equation 4, 3 orthogonal variables, denoted by x, y, and z, were defined for each marker locus. The values of x, y, and z were the same as those of u, v, and s, as given in Table 1.

Similar to the effects that occur in biparental F_2 populations, 4-way-cross F_1 populations and pure-line populations from 4-way crosses, if there is 1 QTL between 2 flanking markers, the QTL effects will cause both main effects and interactions between markers (Zhang et al. 2008, 2015, 2017). However, the coefficients of marker interactions are much smaller than those of the marker main effects, and most of the QTL variation can be absorbed by the main effects of neighboring markers. Therefore, in this study, marker interactions were ignored.

The Inclusive Linear Model for Multiple QTL

For simplicity, we assumed that there were m QTLs located at m intervals within $m + 1$ markers. For intervals with no QTL, the QTL effects can be set at 0. Similar to 1-locus model, the genotypic value of an individual in 1 DH or RIL population from an 8-way cross was shown in Equation 6, indicating that genotypic vales were the summation of the overall mean and QTL effects.

$$G = \mu + \sum_{j=1}^{m} \left(b_{1j} u_j + b_{2j} v_j + b_{3j} s_j + b_{4j} u_j v_j \right.$$
$$\left. + b_{5j} u_j s_j + b_{6j} v_j s_j + b_{7j} u_j v_j s_j \right) \tag{6}$$

where u_j, v_j, and s_j were indicators for genotypes at the jth QTL. The inclusive linear model containing all markers was given by Equation 7.

$$P = E(G) + \varepsilon = \mu + \sum_{j=1}^{m+1} \left(c_{1j} x_j + c_{2j} y_j + c_{3j} z_j + c_{4j} x_j y_j \right.$$
$$\left. + c_{5j} x_j z_j + c_{6j} y_j z_j + c_{7j} x_j y_j z_j \right) + \varepsilon \tag{7}$$

where P was the phenotypic value of the trait of interest; G was the corresponding genotypic value; ε was the random error, following a normal distribution with a mean of 0; and c_{1j} to c_{7j} were the effects of the jth marker. Phenotypic value was explained by marker effects in Equation 7. For large populations, the coefficients of an individual marker in Equation 7 were affected only by the QTL located within the left and right intervals of the marker. That is, the 14 variables of the 2 closest markers could absorb most effects of the QTL. Therefore, the linear model of Equation 7 can be used to control background genetic variation in QTL mapping.

Background-Controlled 1-Dimensional Scanning

A 2-stage mapping strategy was considered in QTL scanning. First, marker variables with significant coefficients in Equation 7 were identified by stepwise regression (Efroymson 1960). In each step of the regression, 1 variable which has not yet been in the model

Table 1. Values of the orthogonal variables for different QTL genotypes

Variable	$A_q A_q$	$B_q B_q$	$C_q C_q$	$D_q D_q$	$E_q E_q$	$F_q F_q$	$G_q G_q$	$H_q H_q$
U	1	1	1	1	−1	−1	−1	−1
V	1	1	−1	−1	1	1	−1	−1
S	1	−1	1	−1	1	−1	1	−1
$u \times v$	1	1	−1	−1	−1	−1	1	1
$u \times s$	1	−1	1	−1	−1	1	−1	1
$v \times s$	1	−1	−1	1	1	−1	−1	1
$u \times v \times s$	1	−1	−1	1	−1	1	1	−1

These indicators were designed to reveal the relationship between genotypic values and QTL genotypes (Equation 4).

but explains the largest variation is added to the model, based on a preassigned probability. Once a new variable has entered to the model, the existing variables will be double checked to determine whether some of them need to be removed from the model, which is based on another preassigned probability. The process stops until no new variable can enter into the model. Coefficients of nonsignificant variables (i.e., those not in the stepwise regression model) were set to 0. Second, during 1-dimensional scanning, the phenotypic values were adjusted by Equation 8 and then used in QTL detection.

$$\Delta P_i = P_i - \sum_{j \neq t, t+1} \left(\hat{c}_{1j} x_{ij} + \hat{c}_{2j} y_{ij} + \hat{c}_{3j} z_{ij} + \hat{c}_{4j} x_{ij} y_{ij} \right.$$
$$\left. + \hat{c}_{5j} x_{ij} z_{ij} + \hat{c}_{6j} y_{ij} z_{ij} + \hat{c}_{7j} x_{ij} y_{ij} z_{ij} \right)$$
(8)

where P_i represented the phenotypic value of the ith pure line; ΔP_i denoted the adjusted phenotypic value; the hat symbol meant "estimated"; and t and $t + 1$ were the 2 flanking markers of the present scanning position. Please be noted that all coefficients of marker variables did not change once they were estimated in the first step. At a testing position in marker interval $[t, t + 1]$, if coefficients of variables for the 2 flanking markers were all 0, ΔP_i only contained the model residual. Otherwise, the nonzero effects of the flanking markers were kept in ΔP_i, which was caused by the QTL at the current scanning interval. QTL effects in the other intervals were all excluded due to the background control. In other words, ΔP_i as defined in Equation 8 contained QTL information in the current interval and did not change until the testing position moved to the next interval.

At a testing position in marker interval $[t, t + 1]$, the phenotypes of the 8 QTL genotypes followed the normal distribution $N\left(\mu_k, \sigma^2\right)$, where $k = 1, 2, \ldots, 8$. To test for the existence of a QTL at the current scanning position, the null and alternative hypotheses were set as follows.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_8$$

H_A: at least 2 of $\mu_1, \mu_2, \ldots, \mu_8$ were not equal.

Under the null hypothesis, the 8 QTL genotypes followed the same normal distribution $N\left(\mu_0, \sigma_0^2\right)$. The mean value and variance were calculated as

$$\hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^{n} \Delta P_i, \text{ and } \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\Delta P_i - \hat{\mu}_0\right)^2$$
(9)

where n was the mapping population size.

Under the alternative hypothesis H_A, all pure lines could be classified into 64 marker classes. The sample size of each marker class was represented by n_i ($i = 1, 2, \ldots, 64$). The log-likelihood function was

$$\ln L_A = \sum_{j=1}^{64} \sum_{i \in S_j} \ln \left[\sum_{k=1}^{8} \pi_{jk} f\left(\Delta P_i; \mu_k, \sigma^2\right) \right]$$
(10)

where S_j denoted the set of pure lines belonging to the jth marker class ($j = 1, 2, \ldots, 64$); π_{jk} was the proportion of the kth ($k = 1, 2, \ldots, 8$) QTL genotype in each marker class (Supplementary Tables S1 and S2); and $f\left(\bullet; \mu_k, \sigma^2\right)$ represented the density function of the normal distribution $N\left(\mu_k, \sigma^2\right)$.

QTL genotypes were unknown before genetic mapping, so maximum likelihood estimates of the parameters in Equation 10 were calculated via the expectation–maximization (EM) algorithm (Dempster et al. 1977). The EM algorithm has been widely applied in many QTL detection algorithms, where the QTL genotype were treated as unknown variables (Kao 1999). Most pure lines in marker classes 1, 10, 19, 28, 37, 46, 55, and 64 had genotypes $A_q A_q$, $B_q B_q$, $C_q C_q$, $D_q D_q$, $E_q E_q$, $F_q F_q$, $G_q G_q$, and $H_q H_q$, respectively. Initial values of the parameters used in the EM algorithm were defined as

$$\mu_1^{(0)} = \frac{1}{n_1} \sum_{i \in S_1} \Delta P_i, \mu_2^{(0)} = \frac{1}{n_{10}} \sum_{i \in S_2} \Delta P_i, \mu_3^{(0)} = \frac{1}{n_{19}} \sum_{i \in S_3} \Delta P_i$$

$$\mu_4^{(0)} = \frac{1}{n_{28}} \sum_{i \in S_4} \Delta P_i, \mu_5^{(0)} = \frac{1}{n_{37}} \sum_{i \in S_5} \Delta P_i, \mu_6^{(0)} = \frac{1}{n_{46}} \sum_{i \in S_6} \Delta P_i,$$

$$\mu_7^{(0)} = \frac{1}{n_{55}} \sum_{i \in S_7} \Delta P_i, \mu_8^{(0)} = \frac{1}{n_{64}} \sum_{i \in S_8} \Delta P_i, \text{ and}$$

$$\sigma^{2(0)} = \frac{1}{n_1 + n_{10} + n_{19} + n_{28} + n_{37} + n_{46} + n_{55} + n_{64}}$$
$$\left[\sum_{i \in S_1} \left(\Delta P_i - \mu_1^{(0)}\right)^2 + \sum_{i \in S_2} \left(\Delta P_i - \mu_2^{(0)}\right)^2 \right.$$
$$+ \sum_{i \in S_3} \left(\Delta P_i - \mu_3^{(0)}\right)^2 + \sum_{i \in S_4} \left(\Delta P_i - \mu_4^{(0)}\right)^2$$
$$+ \sum_{i \in S_5} \left(\Delta P_i - \mu_5^{(0)}\right)^2 + \sum_{i \in S_6} \left(\Delta P_i - \mu_6^{(0)}\right)^2$$
$$\left. + \sum_{i \in S_7} \left(\Delta P_i - \mu_7^{(0)}\right)^2 + \sum_{i \in S_8} \left(\Delta P_i - \mu_8^{(0)}\right)^2 \right]$$
(11)

In the E-step, the posterior probability of the ith pure line belonging to the kth QTL genotype was calculated as

$$w_{ik}^{(t)} = \frac{\pi_{jk} f\left(\Delta P_i; \mu_k^{(t)}, \sigma^{2(t)}\right)}{\sum_{l=1}^{8} \pi_{jl} f\left(\Delta P_i; \mu_l^{(t)}, \sigma^{2(t)}\right)}$$
(12)

where $i \in S_j$, and t was the step number in EM iterations starting from 0. In the M-step, means and variance in the log-likelihood function were updated by

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{n} w_{ik}^{(t)} \Delta P_i}{\sum_{i=1}^{n} w_{ik}^{(t)}} (k = 1, 2, \ldots, 8) \text{ and}$$

$$\sigma^{2(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{8} w_{ik}^{(t)} \left(\Delta P_i - \mu_k^{(t+1)}\right)^2$$
(13)

The EM iteration continued until the difference in likelihood function between 2 consecutive iterations reached a preassigned

precision. From the estimation under the 2 hypotheses, the logarithm of odds (LOD) score or likelihood ratio test (LRT) between H_A and H_0 was calculated by Equation 14,

$$LOD = \log_{10} \frac{\max L_A}{\max L_0} \text{ or } LRT = -2 \ln \frac{\max L_0}{\max L_A} \sim \chi^2(df = 7) \tag{14}$$

where $\max L_A$ was the maximum value of the likelihood function under H_A, and $\max L_0$ was the maximum value of the likelihood function under H_0. LRT approximates a χ^2 distribution with the degree freedom equal to the parental number minus 1.

QTL Models in the Simulation Study

In this study, we considered 3 QTL models to verify the efficiency of ICIM in 8-parental pure-line populations. In model I, 4 chromosomes were considered, and each chromosome was 150 cM in length. Seventy-six markers were evenly distributed on each chromosome, and the distance between any 2 adjacent markers was set to 2 cM. Six QTLs, represented by QTL1 to QTL6, were located on 4 chromosomes (Table 2). QTL1 and QTL6 were 2 independent QTLs with genetic variances of 6 and 4; QTL2 and QTL3 were linked in coupling with genetic variances of 6 and 9; and QTL4 and QTL5 were linked in repulsion with genetic variances of 6 and 9. Total genetic variance was equal to 40.01, and the random error variance was set to 20, such that the broad-sense heritability was equal to 0.67. The population size was set to 200. In model II, 8 chromosomes were considered, and each chromosome was also 150 cM in length. Marker density was the same as that in model I. Eight independent QTLs, denoted by QTL1 to QTL8 and with different percentages of variance explained (PVEs), were distributed on 8 chromosomes. Genetic variance of QTL1 to QTL8 was set from 0.5 to 7.5 (Table 2). Total genetic variance was equal to 32, and the random error

variance was set to 18. Therefore, the broad-sense heritability was equal to 0.64. Three population sizes were considered, that is, 200, 400, and 600.

Model III was the same as that in Wei et al. (2016) and was used to compare ICIM with Fixed-B, Random-B, and R/qtl2. The first 5 chromosomes in the linkage map of the MAGIC mouse population with 458 individuals (Churchill et al. 2004) were used in the simulation, including 2250 markers in total. The marker density was approximately 0.19 cM. Seven QTLs were considered, denoted by QTL1 to QTL7, among which QTL1 and QTL7 had smaller genetic variances than the others (Table 2). The random error variance was set to 1, and the broad-sense heritability was equal to 0.67.

For each QTL model, one thousand RIL populations from 8-way crosses were generated without missing data by the genetics and breeding simulation tools QuGene and QuLine (available from: http://sites.google.com/view/qu-gene; Wang et al. 2003, 2004). ICIM was implemented in software Genetic analysis of multiparental pure-line populations (GAPL) v1.2 (Zhang et al. 2019), which is freely available from http://www.isbreeding.net. The scanning step was set to 1 cM in models I and II, and 0.1 cM in model III. Probabilities of adding and removing variables in stepwise regression were set to 0.001 and 0.002, respectively. For model III, Fixed-B and Random-B implemented in the R/MagicQTL package, and R/qtl2 package were used and compared with ICIM. The scanning step was set to 0.1 cM. The other parameters were set to their default values.

For each model, additional one thousand populations were simulated for the null-QTL model to evaluate the empirical distribution of the test statistics and obtain the LOD threshold. Population size was the same as that in the respective simulated QTL model. The LOD score was estimated for ICIM and R/qtl2; the $\log_{10}P$ value of the Wald statistic (denoted Wald.LOGP) was estimated for Fixed-B and Random-B. The largest LOD score (or Wald.LOGP) from each simulated population was recorded, and the 95th percentile was

Table 2. Predefined locations and genotypic effects for 3 QTL models used in the simulation study

| Model | QTL | Chr. | Pos. (cM) | Genotypic effect | | | | | | | | V_Q[a] | PVE (%)[b] |
				a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8		
Model I	QTL1	1	35	1.79	2.57	1.94	−2.41	−1.49	1.65	0.76	−4.81	6	10.00
	QTL2	2	25	2.79	1.87	2.54	−3.54	−1.91	−2.11	2.37	−2.01	6	10.00
	QTL3	2	55	3.09	2.34	2.94	−2.57	−3.51	−2.88	3.53	−2.94	9	15.00
	QTL4	3	25	2.79	1.87	2.54	−3.54	−1.91	−2.11	2.37	−2.01	6	10.00
	QTL5	3	55	−2.57	−3.51	−2.88	3.09	2.34	2.94	−2.94	3.53	9	15.00
	QTL6	4	35	1.29	1.07	1.74	−1.61	−2.09	1.65	1.66	−3.71	4	6.67
Model II	QTL1	1	55	0.39	0.27	0.79	−0.85	−0.38	0.61	0.47	−1.3	0.5	1.00
	QTL2	2	55	0.89	1.07	0.79	−1.85	−1.38	1.05	0.91	−1.48	1.5	3.00
	QTL3	3	55	1.19	1.47	1.04	−1.45	−1.98	1.05	1.26	−2.58	2.5	5.00
	QTL4	4	55	1.79	1.47	1.04	−2.41	−2.98	1.65	1.13	−1.69	3.5	7.00
	QTL5	5	55	1.79	1.97	1.04	−2.41	−2.98	1.65	1.69	−2.75	4.5	9.00
	QTL6	6	55	2.79	1.97	1.04	−2.41	−2.98	1.65	1.39	−3.45	5.5	11.00
	QTL7	7	55	2.79	1.97	2.04	−3.41	−2.99	1.65	1.3	−3.35	6.5	13.00
	QTL8	8	55	2.79	0.97	2.04	−3.41	−2.99	1.65	2.89	−3.94	7.5	15.00
Model III	QTL1	1	41.35	−0.14	0.17	0.49	−0.61	0.30	−0.36	0.09	0.06	0.11	3.58
	QTL2	2	21.16	−0.99	−0.23	−0.17	0.66	0.06	0.28	−0.19	0.57	0.24	7.82
	QTL3	3	58.79	0.69	−0.09	−0.53	−0.56	1.05	−0.57	0.09	−0.08	0.32	10.42
	QTL4	3	65.18	−0.68	0.14	1.07	0.05	−0.31	0.26	−0.97	0.45	0.37	12.05
	QTL5	4	27.42	0.88	0.01	0.98	0.32	−0.47	0.11	−1.07	−0.74	0.47	15.31
	QTL6	4	41.19	−0.71	0.32	−0.19	0.07	−0.38	−0.59	1.48	−0.01	0.43	14.01
	QTL7	5	28.65	−0.50	−0.08	0.09	−0.16	0.06	0.70	−0.40	0.27	0.13	4.23

[a]Genetic variance of individual QTLs.

[b]Percentage of phenotypic variance explained by individual QTLs.

used as the threshold, which controlled the genome-wide type I error below 0.05 for the null-QTL model.

To compare ICIM, Fixed-B, Random-B, and R/qtl2, the detection power and FDR were taken into consideration. The length of support interval was set to 10 cM. If a peak higher than the threshold was detected within ±5 cM around the true position of the predefined QTL, the peak was considered as a true positive. All QTLs detected outside of the support interval were treated as false positives. If multiple peaks occurred within the support interval, only the highest peak was counted. FDR was defined as the percentage of false positives out of the total number of true and false positives (Benjamini and Hochberg 1995; Li et al. 2010). Locations and effects of QTLs were estimated from significant peaks.

One Real MAGIC Population in Cowpea

One population consisting of 305 RILs from an 8-way cross in cowpea (*Vigna unguiculata* L. Walp.) was used in this study, which was developed by Huynh et al. (2018). In total, 32 114 SNPs were distributed on 11 chromosomes. The genetic map constructed from the population was 979.48 cM in length. Flowering time under long-day conditions (FTL) and flowering time under short-day conditions

(FTS) were used for QTL mapping by ICIM in GAPL (Zhang et al. 2019). The scanning step was set to 0.1 cM. Probabilities of entering and removing variables in stepwise regression were set to 0.001 and 0.002, respectively. The LOD threshold was determined by permutation tests with 1000 runs, and the type I error was set to 0.05. GWAS was conducted using FarmCPU software (Liu et al. 2016) for comparison. Physical positions of SNPs were achieved from the cowpea genome V1.0 (http://www.phytozome.net/). Threshold of *P* value was determined by Bonferroni correction at the 0.05 significance level.

Results

Power Analysis and Mapping Results for Model I

Using the null-QTL model, the LOD threshold for ICIM was calculated to be 6.21. Detection power for QTL1 to QTL6 was shown in Figure 1A. QTL3 had the highest detection power of 90.2%; QTL4 had the lowest detection power of 27.6%; and FDR was 29.41%. QTL1 and QTL6 were located on 2 different chromosomes independently, but the genetic variance of QTL1 was larger than that of QTL6. Therefore, detection power for QTL1 was higher than that
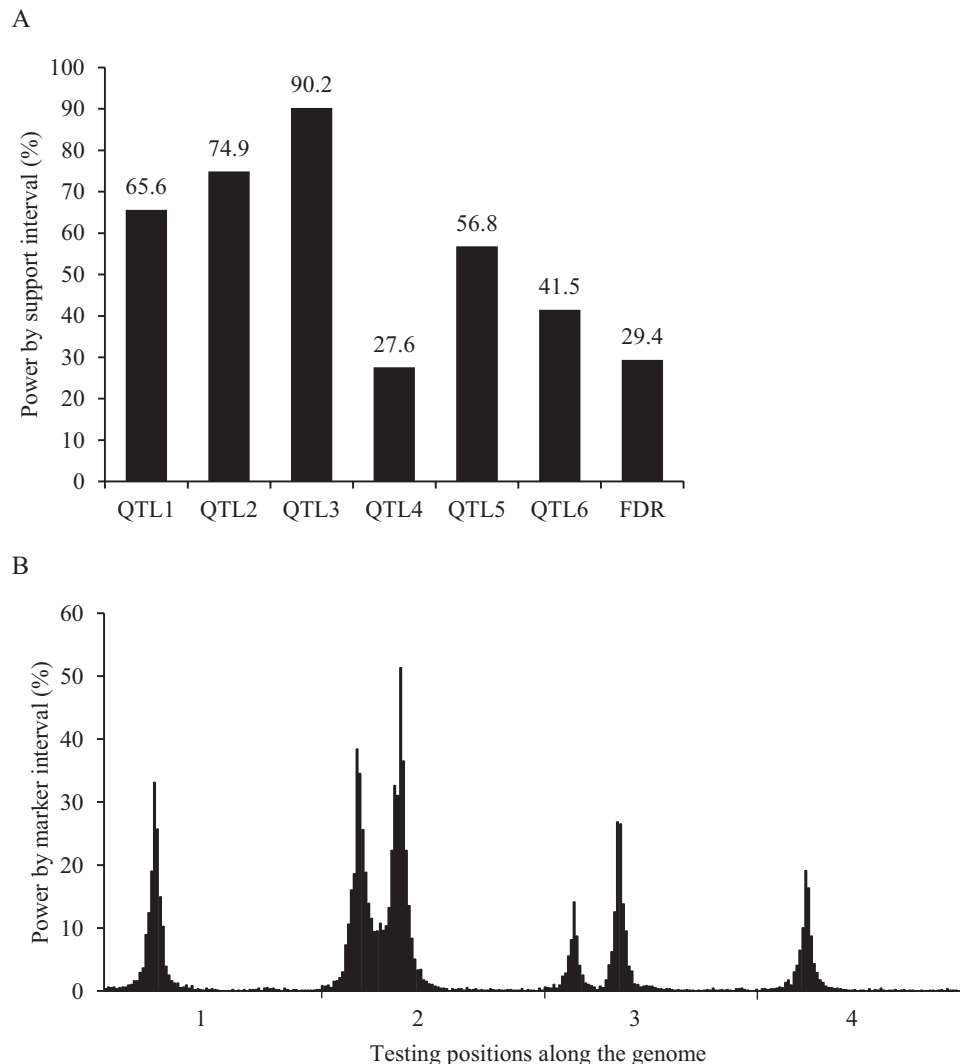

Figure 1. Power analysis from 1000 simulated populations for each predefined QTL (**A**) and each marker interval on the genome (**B**) for model I. The simulated population size was 200. Support interval for each predefined QTL in panel A was set to 10 cM. The last group of bars in panel A represented FDR.

for QTL6. Detection power of unlinked QTLs depends on genetic variance caused by individual QTLs. Detection power can also be affected by linkage. QTL1, QTL2, and QTL4 had the same genetic variance, but QTL2 was linked with QTL3 in coupling; QTL4 was linked with QTL5 in repulsion (Table 2). Compared with QTL1, QTL2 had higher detection power due to the coupling linkage phase; the repulsion linkage phase reduced the detection power for QTL4.

Figure 1B shows detection power in every marker interval in the whole genome across 1000 simulation runs. False positives were low in the marker interval where no QTL was located, and detection powers were significantly high around the predefined QTL. Higher detection powers were observed in marker intervals near QTLs with higher genetic variances. In other words, a QTL was less likely to be detected in chromosomal regions far from the predefined QTL. When QTLs were linked in the coupling phase, detection powers around the 2 linked QTLs were increased (Figure 1B). A ghost QTL was detected between the 2 linked QTLs. When QTLs were linked in the repulsion phase, detection power around the 2 linked QTLs decreased (Figure 1B). Both linkage phases complicated the QTL detection.

Table 3 shows LOD scores, locations, and genetic effects of QTLs estimated by ICIM, averaged from 1000 simulations. Unbiased estimations of QTL positions and effects were approximately achieved. Taking QTL1 as an example, the estimated position was 34.99 cM (Table 3), corresponding to the true position 35.00 cM (Table 2). The standard error was 2.11. The estimated effects were 1.05, 2.42, 2.00, −2.30, −1.54, 1.71, 1.00, and −4.33, which were close to the true effects as given in Table 2.

Power Analysis and Mapping Results for Model II

Under the null-QTL model, the LOD threshold averaged across the 3 population sizes was 6.53. Detection powers for QTL1 to QTL8 were shown in Figure 2. Detection power was clearly increased with the genetic variance of QTL and population size. QTL1 to QTL8 were arranged in the order of increasing genetic variance. When population size was 200, detection power ranged from 2.8% for QTL1 to 81.4% for QTL8. For a population size of 400, power for the 8 QTLs ranged from 5.1% to 100%. For a population size of 600, power ranged from 13.4% to 100%. To achieve the power higher than 90%, PVE needs to be larger than 9% for a population size of 400, and larger than 5% for a population size of 600. FDR for the 3 population sizes was 35.22%, 27.06%, and 25.62%, respectively. A larger population size corresponded to a reduced FDR.

Supplementary Tables S3, S4, and S5 provide the estimated LOD scores, locations, and effects of the 8 QTLs for population sizes of 200, 400, and 600, respectively. Generally speaking, QTLs with greater genetic variance resulted in a greater LOD score. For instance, for a population size of 200, the LOD score was 8.45 for QTL1, with the lowest genetic variance, and 12.23 for QTL8, with the greatest genetic variance (Supplementary Table S3). The estimated QTL locations were unbiased. Moreover, with an increase in population size, the bias in position estimation decreased. The same was true for standard errors. Taking QTL1 as an example, the estimated positions were 54.61, 55.29, and 54.98 cM for population sizes 200, 400, and 600 (Supplementary Tables S3–S5), respectively, whereas the true position was 55 cM. Standard errors were 3.05, 2.87, and 2.36 for the 3 population sizes, respectively.

The estimated effects in Supplementary Tables S3–S5 and their true values were shown in Figure 3. Some QTL effects were overestimated, but some were underestimated. For example, the a_8 effect of QTL1 was overestimated with percentages of bias equal to 82.37, 31.98, and 22.78 for population sizes of 200, 400, and 600, respectively; the a_8 effect of QTL8 was underestimated with percentages of bias equal to 13.09, 8.93, and 6.40 for population sizes of 200, 400, and 600, respectively. With an increase in population size and the genetic variance of QTLs, the bias in estimation of QTL effects decreased. In general, the estimated effects asymptotically approached unbiasedness (Figure 3).

Power Analysis and Mapping Results for Model III

Using the null-QTL model, the estimated LOD threshold was 5.76 and 6.08 for ICIM and R/qtl2, and the estimated Wald.LOGP threshold was 3.56 and 2.34 for Fixed-B and Random-B, respectively. Detection power and FDR obtained by the 4 mapping methods were shown in Figure 4. ICIM achieved higher power for each QTL than the other 3 methods, especially for QTLs with smaller genetic variances. For instance, power for QTL7 was 87.6%, which was 24.2% higher than that for R/qtl2, and almost twice that for Fixed-B and Random-B (Figure 4). FDR for ICIM was 25.29%, which was 4.61% and 3.61% lower than that for Fixed-B and Random-B, respectively, but 19.47% higher than R/qtl2. Although R/qtl2 achieved the lowest FDR, detection power by R/qtl2 was much lower than the other 3 methods for all QTLs.

The estimated positions of the 7 QTLs from ICIM, Fixed-B, Random-B, and R/qtl2 were shown in Table 4. The 4 methods achieved approximately unbiased estimation for the 7 QTL positions. ICIM had the lowest bias for 3 QTLs; Fixed-B and Random-B,

Table 3. Estimated LOD scores, locations, and genotypic effects by ICIM for model I

Variable	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6
LOD	10.0 (3.07)[a]	11.66 (4.03)	13.79 (5.02)	9.05 (2.35)	10.28 (3.43)	8.85 (2.34)
Pos. (cM)	34.99 (2.11)	25.60 (2.21)	54.69 (2.09)	24.50 (2.23)	55.43 (1.88)	35.04 (2.21)
a_1	1.05 (1.43)	2.69 (1.75)	2.90 (1.76)	2.03 (1.57)	−1.55 (1.51)	0.54 (1.53)
a_2	2.42 (1.23)	1.58 (1.82)	2.20 (1.79)	0.51 (1.60)	−2.38 (1.43)	1.28 (1.34)
a_3	2.00 (1.27)	2.38 (1.75)	2.74 (1.82)	2.16 (1.52)	−2.31 (1.40)	1.86 (1.22)
a_4	−2.30 (1.29)	−3.36 (1.58)	−2.60 (1.74)	−2.97 (1.24)	2.17 (1.44)	−1.78 (1.35)
a_5	−1.54 (1.35)	−1.88 (1.83)	−3.07 (1.77)	−0.92 (1.75)	1.49 (1.52)	−2.24 (1.29)
a_6	1.71 (1.39)	−2.17 (1.69)	−2.51 (1.80)	−−1.36 (1.59)	1.94 (1.53)	1.92 (1.28)
a_7	1.00 (1.40)	2.71 (1.54)	3.02 (1.74)	2.17 (1.22)	−2.25 (1.51)	2.08 (1.17)
a_8	−4.33 (1.12)	−1.95 (1.87)	−2.68 (1.79)	−1.63 (1.39)	2.91 (1.47)	−3.67 (1.01)

Each value was the average from 1000 simulations.

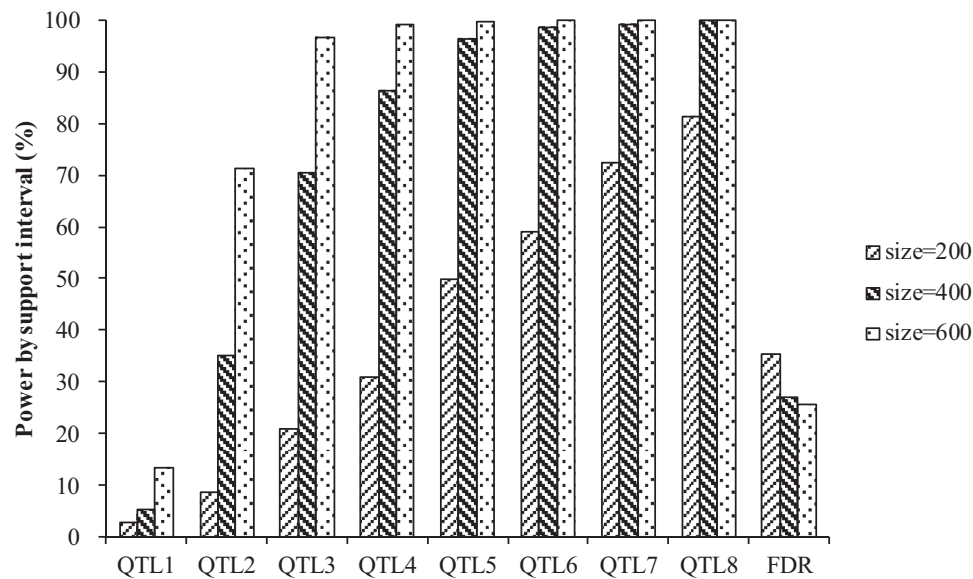[a]The number in parentheses is the standard error.

Figure 2. Power analysis from 1000 simulated populations for model II and population sizes 200, 400, and 600. Support interval for each predefined QTL was set to 10 cM. The last group of bars represented FDR.

for 1; and R/qtl2, for 3. Supplementary Figure S1 showed the estimated effects. Some QTL effects were overestimated, and some were underestimated. ICIM achieved the lowest bias for 13 out of the 56 effects; Fixed-B, for 11; Random-B, for 27; and R/qtl2, for 5. For independent QTLs, ICIM provided better effect estimates, while for linked QTLs, Fixed-B, and Random-B provided better estimates. Considering the higher detection power and lower FDR of ICIM (except FDR for R/qtl2) and the unbiased estimation of QTL locations, ICIM is more efficient.

QTL Mapping of Flowering Time in Cowpea

The LOD profile from ICIM was displayed in Figure 5 for 2 phenotypic traits, namely, FTL and FTS, in the cowpea population. The LOD threshold obtained by permutation tests was 6.83. Seven QTLs were identified for FTL, explaining 68.23% of the phenotypic variance. The estimated positions and effects of the detected QTLs were summarized in Table 5. For the 7 FTL QTLs, 2 were located on chromosome 1, and 1 each was located on chromosomes 3, 4, 5, 9, and 11 (Table 5). The QTL located on chromosome 9 had the largest LOD score of 40.21 and the largest PVE of 25.53%. The alleles from parents IT84S-2049, CB27, and IT82E-18 reduced FTL (Table 5). The QTL with the second largest PVE was located on chromosome 11, explaining 13.57% of the total phenotypic variance. The alleles from parents CB27, Suvita2, IT00K-1263 and IT84S-2246 reduced FTL (Table 5). In summary, the alleles from the 8 parents may have different genotypic effects in different directions. Three QTLs were identified for FTS, explaining 24.34% of the phenotypic variance. One QTL each was located on chromosomes 4, 5, and 10 (Table 5). The QTL located on chromosome 4 had the largest LOD score of 11.62 and the largest PVE of 11.62%. The alleles from parents IT89KD-288, CB27, Suvita 2, and IT84S-2246 reduced FTS (Table 5). Lo et al. (2018) reported 2 candidate genes for flowering time called *CFt5* and *CFt9* located at SNP 2_05332 on chromosome 5 and 2_03945 on chromosome 9, close to the 2 FTL QTLs detected by ICIM. Genetic distance between the 2 pairs of linked QTLs were 2.90 and 1.34 cM, respectively (Table 5).

On chromosome 4, 1 QTL was detected for FTL and FTS separately with a distance at 0.9 cM; on chromosome 5, distance between the 2 QTLs affecting FTL and FTS was 5.2 cM. The correlation coefficient between FTL and FTS was 0.57 in the mapping population. The 2 pairs of QTLs genetically explained the observed phenotypic correlation between the 2 traits, and represented their genetic background independence and environmental stability. For the QTL on chromosome 4, the allele from parent CB27 reduced FTL by 4.71 days but reduced FTS by only 1.65 days. For the QTL on chromosome 5, the allele from parent IT84S-2246 delayed FTL by 4.04 days but delayed FTS by only 1.45 days. In general, the effects of FTL QTLs were stronger than those of FTS QTLs. Further investigation is needed to determine whether they are coincident or close-linked QTLs.

For convenience of comparison, the results from IM reported by Huynh et al. (2018) were given in Supplementary Table S6. The results from ICIM were also given in Supplementary Table S6, where the genotypic effects of individual parental genotypes were relative to IT93K-503-1. Four QTLs on chromosomes 4, 5, 9, and 11 were identified by IM for FTL, and 4 for FTS were located on chromosomes 1, 4, 5, and 9. ICIM resulted in a shorter confidence interval than did IM. The confidence intervals of the 4 FTL QTLs detected by IM overlapped with the 4 QTLs detected by ICIM, and the confidence intervals of 2 FTS QTLs detected by IM overlapped with 2 QTLs detected by ICIM. For the 6 common QTLs detected by both methods, the estimates of genotypic effects were not the same. However, for FTL, 22 out of the 28 effects had the same directions by both methods; for FTS, 8 out of the 14 effects had the same directions. In summary, most QTLs identified by IM were detected by ICIM as well; ICIM identified more QTLs explaining more phenotypic variation.

GWAS results from FarmCPU were shown in Supplementary Table S7. A total of 9 QTLs were identified for the 2 traits. Three for FTL were located on chromosomes 1, 9, and 11, and 6 for FTS were located on chromosomes 4, 5, 8, 9, 10, and 11. Six of the 9 QTLs were close to the QTLs detected by ICIM. For the 3 FTL QTLs located on chromosomes 1, 9, and 11, difference on the estimated location between ICIM and FarmCPU was 7.2, 0.1, and 1.7 cM, respectively. For the 3 FTS QTLs located on chromosomes 4, 5, and
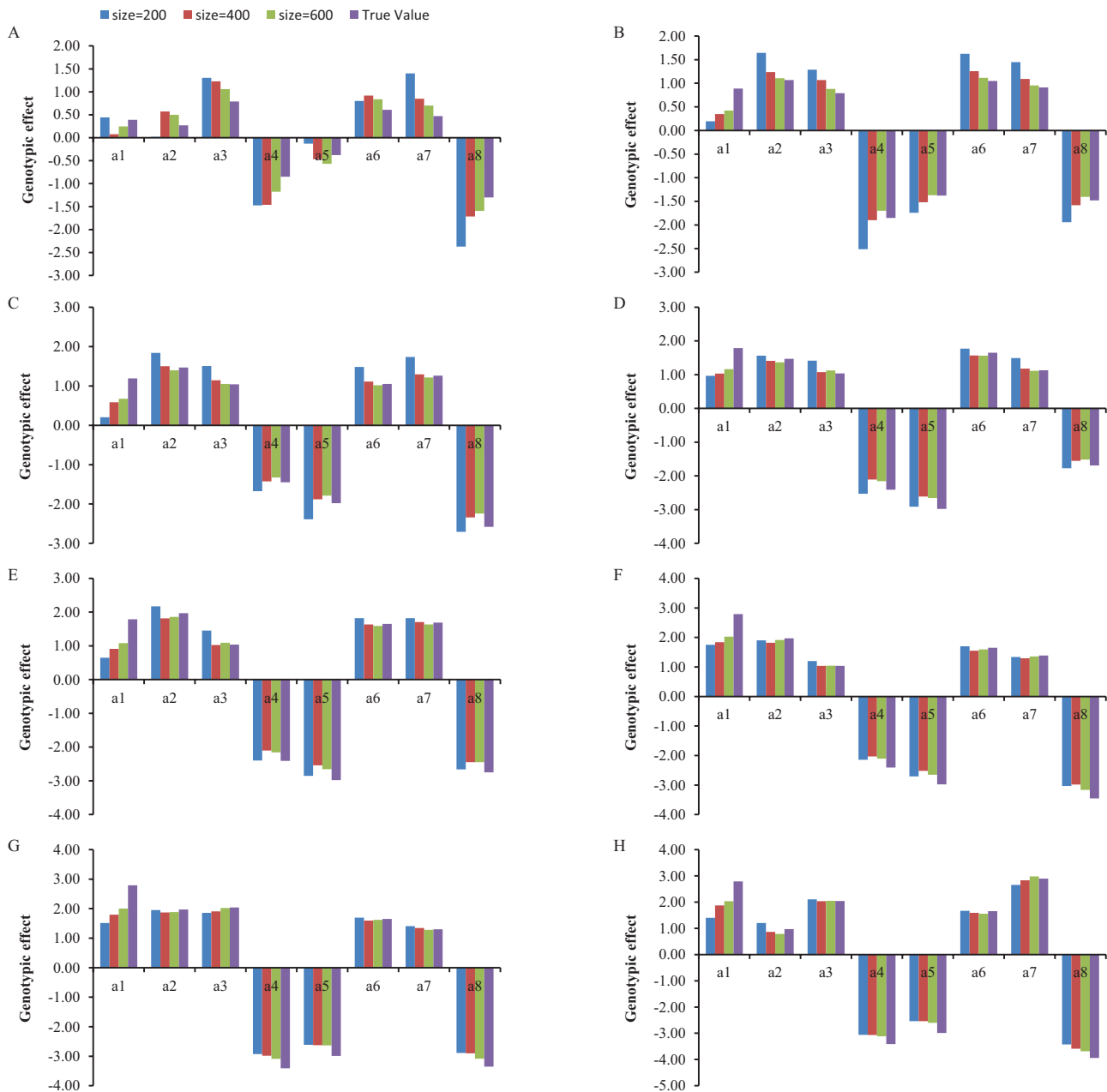
Figure 3. True and estimated genotypic effects for 8 simulated QTLs for model II. The estimated effects were the average from 1000 simulations.

10, difference on the estimated location between between ICIM and FarmCPU was 0.6, 3.5, and 0 cM, respectively. No information for PVE and confidence interval of the identified QTL was provided by FarmCPU. In addition, only one allelic effect relative to the minor allele was estimated by FarmCPU, which could not be separated into genotypic effects of the 8 founder parents. Therefore, source of the most favorable allele in parental lines cannot be determined when using GWAS on multiparental populations.

Discussion

ICIM has been proven to be an efficient QTL mapping method and widely used in biparental (e.g., Yin et al. 2015, 2017; Wu et al. 2018), 4-way-cross pure-line (e.g., Ning et al. 2017), and clone F_1

and 4-way-cross F_1 populations (e.g., Ding et al. 2015; Chen et al. 2016). Simulation studies and application in a cowpea MAGIC population in this study validated its efficiency with 8-parental pure-line populations. Simulation results in this study also provided a reference for the mapping population size that is probably needed for detecting QTLs with different genotypic variances. Compared with Fixed-B, Random-B, and R/qtl2, ICIM provides higher detection power, a relatively lower FDR, and unbiased estimation of QTL locations. Bias was observed between the estimated QTL effects and the true values (Supplementary Figure S1), but the bias was reduced as the increase in population size.

Stepwise regression was used to reduce the model complexity in this study. In 8-way crosses, alleles at each locus in parents may be different and therefore have different effects. To handle multiple
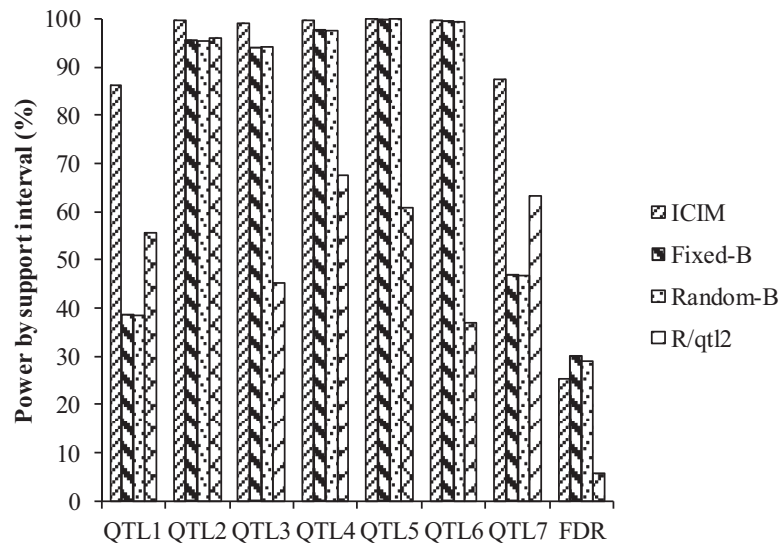
Figure 4. Power analysis from 1000 simulated populations for model III and a population size of 458. Support interval for each predefined QTL was set to 10 cM. The last group of bars represented FDR.

Table 4 Estimated QTL locations (cM) by mapping methods ICIM, Fixed-B, Random-B, and R/qtl2 for model III

Method	QTL1	QTL2	QTL3	QTL4	QTL5	QTL6	QTL7
ICIM	41.24 (1.87)[a]	21.25 (1.55)	61.46 (2.19)	64.29 (2.04)	27.31 (1.19)	41.47 (1.26)	28.61 (1.86)
Fixed-B	41.16 (1.32)	21.28 (1.10)	60.29 (2.35)	64.81 (1.66)	27.25 (0.70)	41.39 (0.73)	28.72 (1.33)
Random-B	41.16 (1.32)	21.28 (1.10)	60.29 (2.35)	64.81 (1.66)	27.25 (0.70)	41.39 (0.73)	28.71 (1.32)
R/qtl2	41.21 (1.74)	21.23 (1.29)	59.72 (2.03)	64.76 (1.77)	27.31 (1.04)	41.29 (1.16)	28.76 (1.63)
True pos.	41.35	21.16	58.79	65.18	27.42	41.19	28.65

Each value was the average from 1000 simulations.

[a]The number in parentheses is the standard error.

allele effects, each marker has to be defined with 8 effects. Therefore, a very large number of marker variables were included in the regression model (Equation 7), which complicated the model selection procedure in stepwise regression. The incompleteness of the linear model built from stepwise regression may not have a large effect on LOD score and QTL position estimation, but may have some impact on the estimation of QTL effects. Further investigation is needed to determine suitable probability levels for variables to enter into and leave from the stepwise regression model. Other model selection algorithms which may provide better estimates for the linear model based on other criteria may be considered in future.

The method proposed in this study was specifically designed for pure-line populations from an 8-way crosses, that is, $[(A \times B) \times (C \times D)] \times [(E \times F) \times (G \times H)]$. However, the method can also be directly applied to pure-line populations from fewer than 8 parents. For example, the cross between 2 top crosses, that is, $[(A \times (C \times D)] \times [E \times (G \times H)]$, where 6 parents are involved, can be treated as 1 special case of 8-way cross, that is, $[(A \times A) \times (C \times D)] \times [(E \times E) \times (G \times H)]$, where parent A is the same as parent B, and parent E is the same as parent F.

To provide more recombination events and higher mapping resolution, some MAGIC-like populations have also been developed. For example, Bandillo et al. (2013) reported an *indica* rice 8-parental MAGIC population with 2 more cycles of intercrossing for enhancing recombination before selfing. A 16-parental MAGIC population (including 8 *indica* and 8 *japonica* parents) was developed to capture broader genotypic diversity (Bandillo et al. 2013). Currently, the MAGIC-like populations mentioned above are analyzed by GWAS.
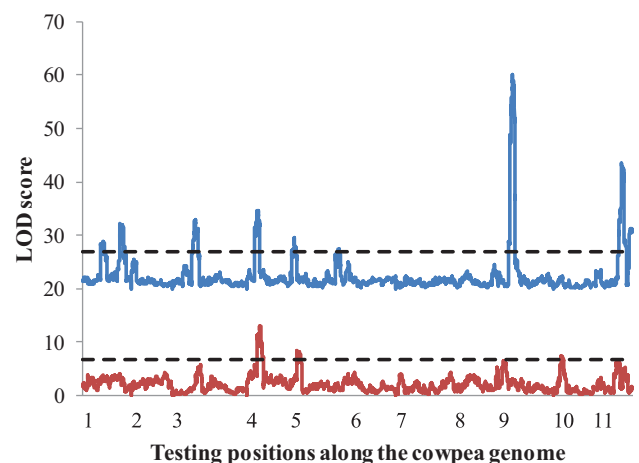

Figure 5. LOD score of flowering time under long-day conditions (top) and short-day conditions (bottom) obtained by ICIM for the real cowpea MAGIC population consisting of 305 RILs. Twenty was added to LOD score of flowering time under long-day conditions (top). The horizontal dashed lines represented the threshold calculated by permutation tests.

Some linkage analysis methods and tools can be used for these populations, such as R/qtl2 (Broman et al. 2019), R/MagicQTL (Wei and Xu 2016), and mixed-model methods. However, the efficiency of these methods for more general MAGIC populations has not been

Table 5. Mapping results for flowering time in the real cowpea MAGIC population by ICIM

Trait	Chr.	Pos. (CI)[a]	Flanking markers	LOD	PVE (%)[b]	IT89KD-288[c]	IT84S-2049	CB27	IT82E-18	Suvita 2	IT00K-1263	IT84S-2246	IT93K-503-1
FTL	1	36.8 (36.45, 36.95)	2_20277–2_38654	8.75	4.25	3.56	-3.50	-1.45	0.10	0.31	2.51	1.95	-3.49
	1	66.60 (66.55, 68.45)	2_10794–2_24198	12.18	6.13	0.79	-0.74	-2.02	-0.67	-4.21	4.51	-2.85	5.18
	3	40.7 (39.85, 40.75)	2_51003–2_23897	12.84	6.80	6.44	4.21	-0.22	4.21	-6.71	-1.94	-4.22	-1.77
	4	19.9 (19.45, 20.05)	2_25338–2_34933	14.61	7.39	0.31	2.95	-4.71	-1.75	4.61	4.40	-3.44	-2.37
	5	6.8 (6.05, 7.35)	1_0790–2_01044	9.54	4.56	4.48	1.66	-3.73	-3.09	-2.04	-1.67	4.04	0.35
	9	24.70 (24.65, 24.75)	2_10720–2_14698	40.21	25.53	0.05	-2.95	-8.08	-9.32	0.52	5.06	8.01	6.72
	11	51.20 (50.85, 52.55)	2_18085–2_54622	23.45	13.57	5.45	3.47	-3.37	4.81	-1.69	-2.23	-9.42	2.98
FTS	4	20.6 (20.35, 21.25)	2_50486–2_42838	11.62	11.62	-1.09	1.10	-1.65	1.28	-1.03	2.40	-1.78	0.77
	5	12 (11.55, 12.55)	2_36891–2_15997	8.27	7.00	-0.27	-0.02	-1.28	-0.26	-1.09	-0.51	1.45	1.97
	10	18.2 (17.85, 18.85)	2_47423–2_00495	6.83	5.72	-0.72	-0.70	0.81	1.11	-1.10	-0.82	-0.02	1.45

[a]Position in centi-Morgans and 1-LOD confidence interval (CI).
[b]Percentage of phenotypic variance explained by individual QTLs.
[c]Genotypic effects of individual parental genotypes.

explored. We are considering extending the ICIM-based mapping method to accommodate more kinds of multiparental populations.

Epistasis plays a crucial role in the genetic variation underlying many complex traits in plants. However, the detection of epistasis and effect estimation are still difficult because of the complex interacting patterns, insufficient sample sizes of mapping populations and lack of efficient statistical methods (Zhang et al. 2012). Identifying genome-wide epistasis is a high-dimensional multiple regression problem that also requires the application of dimensionality reduction techniques (Ehrenreich 2017). High collinearity among markers and computational complexity are always obstacles to epistasis detection. ICIM has been applied to detect epistasis in biparental populations (Li et al. 2008; Zhang et al. 2012). Recently, a Bayesian-based method was proposed to identify epistasis for flowering time in a barley MAGIC population. However, this method has some disadvantages. High collinearity among markers created inconsistency in the results from different Markov chain Monte Carlo (MCMC) chains. Some epistatic interactions cannot be detected owing to the coding system used in this study, especially for signals whose marginal associations are not large enough (Mathew et al. 2018). Generally speaking, studies on epistasis in MAGIC populations are rare. The extension of ICIM to epistatic QTLs in multiparental populations requires further investigation.

Supplementary Material

Supplementary data are available in *Journal of Heredity* online.

Funding

Conflict of Interest

The authors declare that they have no conflicts of interest associated with the content of this article.

Data Availability

The authors affirm that all data necessary for confirming the conclusions of this article are represented fully within the article and its tables and figures. Data files of simulated and actual populations are available at http://www.isbreeding.net/8-waycross/.

References

Bandillo N, Bandillo N, Raghavan C, Muyco PA, Sevilla MA, Lobina IT, Dilla-Ermita CJ, Tung CW, McCouch S, Thomson M, *et al.* 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice (N Y)*. 6:11.

Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 57: 289–300.

Broman KW. 2012. Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. *Genetics*. 190:403–412.

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA. 2019. R/qtl2: software for mapping quantitative trait

loci with high-dimensional data and multiparent populations. *Genetics*. 211:495–502.

Cavanagh C, Morell M, Mackay I, Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol*. 11:215–221.

Chen J, Zhang L, Liu S, Li Z, Huang R, Li Y, Cheng H, Li X, Zhou B, Wu S, *et al*. 2016. The genetic basis of natural variation in kernel size and related traits using a four-way cross population in maize. *PLoS One*. 11:e0153428.

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, *et al*.; Complex Trait Consortium. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet*. 36:1133–1137.

Darvasi A, Soller M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*. 141:1199–1207.

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, *et al*. 2015. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biol*. 16:167.

Dempster A, Laird N, Rubin D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39: 1–38.

Ding J, Zhang L, Chen J, Li X, Li Y, Cheng H, Huang R, Zhou B, Li Z, Wang J, *et al*. 2015. Genomic dissection of leaf angle in maize (*Zea mays* L.) using a four-way cross mapping population. *PLoS One*. 10:e0141619.

Efroymson MA. 1960. In: Ralston A, Wilf HS, editors. *Multiple regression analysis, mathematical methods for digital computers*. Vol. 1. New York: Wiley. p. 191–203.

Ehrenreich IM. 2017. Epistasis: searching for interacting genetic variants using crosses. *Genetics*. 206:531–535.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet*. 11:446–450.

Huang BE, George AW. 2011. R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics*. 27:727–729.

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR. 2012. A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J*. 10:826–839.

Huang X, Paulo MJ, Boer M, Effgen S, Keizer P, Koornneef M, van Eeuwijk FA. 2011. Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. *Proc Natl Acad Sci U S A*. 108:4488–4493.

Huynh BL, Ehlers JD, Huang BE, Muñoz-Amatriaín M, Lonardi S, Santos JRP, Ndeve A, Batieno BJ, Boukar O, Cisse N, *et al*. 2018. A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). *Plant J*. 93:1129–1142.

Jansen RC. 1994. Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*. 138:871–881.

Kao CH, Zeng ZB, Teasdale RD. 1999. Multiple interval mapping for quantitative trait loci. *Genetics*. 152:1203–1216.

King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ. 2012. Genetic dissection of a model complex trait using the *Drosophila* synthetic population resource. *Genome Res*. 22:1558–1566.

Korte A, Farlow A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*. 9:29.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. 2009. A Multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet*. 5:e1000551.

Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 121:185–199.

Li H, Hearne S, Bänziger M, Li Z, Wang J. 2010. Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity (Edinb)*. 105:257–267.

Li H, Ribaut JM, Li Z, Wang J. 2008. Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet*. 116:243–260.

Li H, Ye G, Wang J. 2007. A modified algorithm for the improvement of composite interval mapping. *Genetics*. 175:361–374.

Liu X, Huang M, Fan B, Buckler ES, Zhang Z. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet*. 12:e1005767.

Lo S, Muñoz-Amatriaín M, Boukar O, Herniter I, Cisse N, Guo YN, Roberts PA, Xu S, Fatokun C, Close TJ. 2018. Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Sci Rep*. 8:6261.

Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R, Howells R, O'Sullivan DM, *et al*. 2014. An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 (Bethesda)*. 4:1603–1610.

Mackay I, Powell W. 2007. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci*. 12:57–63.

Mathew B, Léon J, Sannemann W, Sillanpää MJ. 2018. Detection of epistasis for flowering time using Bayesian multilocus estimation in a barley MAGIC Population. *Genetics*. 208:525–536.

Mott R, Talbot CJ, Turri MG, Collins AC, Flint J. 2000. A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci USA*. 97:12649–12654.

Ning H, Zhuang X, Yang C, Fang Y, Wang P, Li W, Xue H, Li W. 2017. Mapping stable QTLs for growth period procedure based on four-way recombination inbred line population. *Soybean Sci* 36:692–698.

Sannemann W, Huang BE, Mathew B, Léon J. 2015. Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol Breed*. 35:86.

Teuscher F, Broman KW. 2007. Haplotype probabilities for multiple-strain recombinant inbred lines. *Genetics*. 175:1267–1274.

Verbyla AP, Cullis BR, Thompson R. 2007. The analysis of QTL by simultaneous use of the full linkage map. *Theor Appl Genet*. 116:95–111.

Wang J. 2009. Inclusive composite interval mapping of quantitative trait genes. *Acta Agron Sin*. 35:239–245. In Chinese with English abstract.

Wang J, van Ginkel M, Podlich D, Ye G, Trethowan R, Pfeiffer W, DeLacy IH, Cooper M, Rajaram S. 2003. Comparison of two breeding strategies by computer simulation. *Crop Sci*. 43:1764–1773.

Wang J, van Ginkel M, Trethowan R, Ye G, DeLacy I, Podlich D, Cooper M. 2004. Simulating the effects of dominance and epistasis on selection response in the CIMMYT Wheat Breeding Program using QuCim. *Crop Sci*. 44:2006–2018.

Wang J, Li H, Zhang L. 2014. *Genetics mapping and breeding design*. Beijing: The Science Press. In Chinese.

Ward LD, Kellis M. 2012. Interpreting noncoding genetic variation in complex traits and human disease. *Nat Biotechnol*. 30:1095–1106.

Wei J, Xu S. 2016. A random-model approach to QTL mapping in multiparent advanced generation intercross (MAGIC) populations. *Genetics*. 202:471–486.

Wu Y, Zhao S, Li X, Zhang B, Jiang L, Tang Y, Zhao J, Ma X, Cai H, Sun C, *et al*. 2018. Deletions linked to *PROG1* gene participate in plant architecture domestication in Asian and African rice. *Nat Commun*. 9:4157.

Yin C, Li H, Li S, Xu L, Zhao Z, Wang J. 2015. Genetic dissection on rice grain shape by the two-dimensional image analysis in one *japonica × indica* population consisting of recombinant inbred lines. *Theor Appl Genet*. 128:1969–1986.

Yin C, Li H, Zhao Z, Wang Z, Liu S, Chen L, Liu X, Tian Y, Ma J, Xu L, *et al*. 2017. Genetic dissection of top three leaf traits in rice using progenies from a *japonica × indica* cross. *J Integr Plant Biol*. 59:866–880.

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

Zhang L, Li H, Ding J, Wu J, Wang J. 2015. Quantitative trait locus mapping with background control in genetic populations of clonal F1 and double cross. *J Integr Plant Biol*. 57:1046–1062.

Zhang L, Li H, Li Z, Wang J. 2008. Interactions between markers can be caused by the dominance effect of quantitative trait loci. *Genetics*. 180:1177–1190.

Zhang L, Li H, Wang J. 2012. The statistical power of inclusive composite interval mapping in detecting digenic epistasis showing common F2 segregation ratios. *J Integr Plant Biol*. 54:270–279.

Zhang L, Meng L, Wang J. 2019. Linkage analysis and integrated software GAPL for pure-line populations derived from four-way and eight-way crosses. *Crop J*. 7:283–293.

Zhang S, Meng L, Wang J, Zhang L. 2017. Background controlled QTL mapping in pure-line genetic populations derived from four-way crosses. *Heredity (Edinb)*. 119:256–264.