Briefings in Bioinformatics, 2026, 27, bbag273
https://doi.org/10.1093/bib/bbag273
Published: 31 May 2026
Problem Solving Protocol

OXFORD

QTLNetwork-MP: integrative mapping of additive, epistatic, and G × E effects for complex traits in multiparent advanced generation intercross populations

Tian- Neng Zhu[1,2,3,‡], Zhijun Tong[4,‡], Mingzhe Suo[1,2], Mingzi Xu[1,2], Asad Ullah[1,2], Feng Lin[1,2], Honghe Zhang[3,*], Bingguang Xiao[4,*], Hai-Ming Xu[1,2,*]

[1]Institute of Crop Science, College of Agriculture and Biotechnology, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058, China
[2]Institute of Bioinformatics, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058, China
[3]Department of Pathology, School of Medicine, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058, China
[4]Key Laboratory of Tobacco Biotechnological Breeding, National Tobacco Genetic Engineering Research Center, Yunnan Academy of Tobacco Agricultural Sciences, 33 Yuantong Road, Wuhua Distrct, Kunming, Yunnan 650021, China

*Corresponding authors. Honghe Zhang, Department of Pathology, School of Medicine, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058, China. E-mail: honghezhang@zju.edu.cn; Bingguang Xiao, Key Laboratory of Tobacco Biotechnological Breeding, National Tobacco Genetic Engineering Research Center, Yunnan Academy of Tobacco Agricultural Sciences, 33 Yuantong Road, Wuhua Distrct, Kunming, Yunnan 650021, China. E-mail: xiaobgsubmission@126.com; Hai-Ming Xu, Institute of Crop Science, College of Agriculture and Biotechnology, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang 310058, China. E-mail: hmxu@zju.edu.cn
‡Tianneng Zhu and Zhijun Tong contributed equally to this work and share the first authorship

Abstract

Quantitative trait locus (QTL) mapping is a powerful approach to uncover the genetic architecture of complex traits. Traditional biparental populations, though widely used, offer limited genetic diversity and resolution, making them less effective for robust genetic architecture dissection. In contrast, multiparent advanced generation intercross (MAGIC) populations possess more allelic diversity, enabling the detection of intricate gene interactions including dominance and epistasis in QTL mapping. However, QTL mapping in MAGIC populations encounters some great challenges, particularly in inferring parental origin of QTL and estimating their effects unbiasedly. We tackled these challenges by developing a novel QTL mapping method for pure-line MAGIC populations, accompanied by the software QTLNetwork-MP. This method extends the mixed linear model framework by integrating a Markov chain-based algorithm and orthogonal QTL effect decomposition. This allows for the estimation of QTL effects, including additive and additive–additive epistasis as well as environmental interactions for MAGIC populations derived from four-way or eight-way crosses. Simulation studies show that QTLNetwork-MP achieves high statistical power and well-controlled false discovery rates across varying heritability, population size, and parent number. We applied QTLNetwork-MP to map QTLs for seed size in an eight-parent recombinant inbred line MAGIC cowpea population, identifying three significant additive QTLs and accurately estimating their genetic effects, validating the value of the method in practical application.

Keywords pure-line MAGIC population, epistasis, QTL-environment interaction, Markov chain, mixed linear model

Key Points

- We developed a novel quantitative trait locus (QTL) mapping method named by QTLNetwork-MP for the pure-line MAGIC populations, based on a mixed linear model consisting of orthogonal parameters of QTL effects.
- Intensive simulations show the method possess high statistical power, low false discovery rates, and high accuracy in estimation of the position and genetic effect components of QTL.
- The practical utility of QTLNetwork-MP was demonstrated in a real eight-parent cowpea population, where it successfully identified both previously reported and novel QTLs for seed size and accurately estimated the QTL effects of parents.

Received: February 6, 2026. **Revised:** April 6, 2026. **Accepted:** May 4, 2026

Introduction

Crop production has long been a cornerstone of human civilization, providing not only essential food resources and industrial raw materials but also serving as a key driver of agricultural modernization. Despite the continuous increase in crop yields globally, rapid population growth and climate change pose major challenges to the genetic improvement of crops. Modern plant breeding now relies heavily on identification and utilization of quantitative trait locus (QTL) to improve most agronomic traits and breed new high-yield and stress-resistant cultivars. Over the past three decades, QTL mapping has been regarded as a powerful approach to uncover intricate genetic architecture of phenotypic traits. Early methods, such as the interval mapping (IM) [1], the composite interval mapping (CIM) [2, 3] and the mixed linear model-based composite interval mapping (MCIM) [4, 5] were developed for biparental mapping populations. These approaches marked a great milestone in discovering genes genome-wide. However, their effectiveness was inherently constrained because of limited genetic diversity in populations from two parents. This inherent limitation restricts the breeding utilization of the genetic diversity and valuable germplasm resources in a species [6].

To address these limitations, researchers developed multiparent advanced generation intercross (MAGIC) populations, an innovative mapping resource that integrates genetic constituents from multiple parents through successive inter-crossing and inter-selfing [7, 8]. The MAGIC populations exhibit broader genetic diversity and increased recombination events, making them suitable for exploring complex trait architectures. Since their development, MAGIC populations have been widely constructed in various crops, including rice, wheat, maize, and cowpea, demonstrating great potential for enhancing breeding efficiency [9–17]. While MAGIC populations possess significant advantages, they also present new challenges in analysis methods. The increased variety of alleles, genotypes and recombination events in MAGIC populations results in more complex genetic architectures, which can only be analyzed by specialized statistical models and computational tools. Early QTL analyses for MAGIC populations employed relatively simple fixed-effect models based on Haley–Knott (HK) regression [18] and the CIM framework. Tools such as R/happy [7], R/mpMap [19], and R/qtl [20] provided early solutions for MAGIC-based QTL mapping. With the advent of high-throughput genotyping based on sequencing, the size of datasets rapidly expanded, often consisting of tens of thousands of markers and hundreds or thousands of lines, rendering early methods computationally inefficient and statistically underpowered. To overcome these limitations, computationally highly efficient methods and software were developed, such as R/qtl2 [21] and GAPL [22]. However, despite their high computational speed, most existing frameworks are primarily optimized for additive effects. While interaction terms can theoretically be incorporated via interactive covariates, relying on these mathematical workarounds rather than a native unified full-model framework fundamentally limits their statistical efficiency and accuracy in dissecting complex epistasis or G × E interaction. Additionally, these methods often rely on fixed-effect models, which consumes excessive degrees of freedom in parameter estimation and significance testing when applied to MAGIC populations derived from eight or more parents, consequently limiting their ability to detect genetic signals. Furthermore, the lack of orthogonalization between QTL effect components has further complicated the interpretation of QTL effects in MAGIC populations. For high-order designs with eight or more parents, the definition and estimation of QTL effects in an interpretable and statistically robust manner becomes increasingly challenging. Although some studies [22, 23] have defined additive effects orthogonally based on Cockerham's genetic model [24], these decomposition approaches are difficult to extend to include epistasis or G × E interaction. Consequently, existing methods and tools are unable to fully exploit the genetic diversity of MAGIC populations, leaving a methodological gap compared to the well-established approaches for biparental mapping populations [3–5, 25–27] in the genetic dissection of complex traits.

To fill this gap, we proposed a novel statistical framework for mapping QTL in pure-line MAGIC populations, built on a mixed linear model employing our newly derived orthogonal decomposition of genetic effects. By mathematically defining the complex multi-parent matrices, this model integrates effects due to additive, additive-by-additive epistasis, and their interaction with environment within a unified and easily interpretable orthogonal framework. By treating some genetic components as random effects and employing parallelized computation algorithm, we achieve high analytical efficiency, enabling analysis of large-scale datasets with over 10 000 bin markers and thousands of individuals. Notably, our method overcomes the deficiency of degrees of freedom and enables unbiased estimation of genetic variances, heritabilities, and interaction effect components. To support practical application, we developed a software, named QTLNetwork-MP, by extending our QTL mapping software QTLNetwork for biparental mapping population. Extensive validation using both simulated datasets and real-world cowpea MAGIC data showed that our method achieves high power in QTL detection, accuracy in effect estimation and possess excellent scalability. We believe that this framework provides a valuable and new resource for geneticists and breeders to efficiently analyze the MAGIC populations and fully exploit their genetic diversity to promote genetic improvement of traits in crops.

Methods and materials

Orthogonal decomposition of genetic effects

Single-locus model

To explain our orthogonal decomposition of QTL effect, the four-parent (denoted by *AA*,*BB*,*CC*,*DD*) pure-line MAGIC population is used as an example. The genotypic value of the *i*-th parent ($i = 1, 2, 3, 4$) at a QTL is expressed as:

$$g_i = \mu + a_i = \mu + \varphi_{i1}\alpha_1 + \varphi_{i2}\alpha_2 + \varphi_{i3}\alpha_3 \tag{1}$$

Where $g_i\ (i = 1, 2, 3, 4)$ is the genotypic value of the QTL in the *i*-th parent, μ is the population mean, and $a_i\ (i = 1, 2, 3, 4)$ is the QTL additive effect of the *i*-th parent. Since only three parameters can be estimated based on four types of parental genotypes of the QTL, thus, we define other three additive parameters $\alpha_k\ (k = 1, 2, 3)$ to depict the genetic effect (a_i) of the QTL in the *i*-th parent, with coefficients φ_{ik} (Table 1).

To satisfy the constraint $\sum_i p_i a_i = \sum_i a_i = 0$ (due to equal expected genotype frequencies $p_i = 0.25$), the three additive effect parameters are defined as $\alpha_1 = \frac{a_1+a_2}{2}$, $\alpha_2 = \frac{a_1+a_3}{2}$, and $\alpha_3 = \frac{a_1+a_4}{2}$. The coefficients φ_{ik} is taken as shown in Table 1 to ensure orthogonality, i.e., $\sum_{i=1}^{4} \varphi_{ik} = 0$ and $\sum_{i=1}^{4} \varphi_{ik}\varphi_{il} = 0$ for $k \neq l$, enabling independent estimation of additive effects. Then, an additive orthogonal decomposition matrix

Table 1 Decomposition of QTL additive effects in four-way pure-line MAGIC population.

Genotype	*AA* ($i=1$)	*BB* ($i=2$)	*CC* ($i=3$)	*DD* ($i=4$)
Genotypic value	$u+a_1$	$u+a_2$	$u+a_3$	$u+a_4$
Frequency	$p_1=0.25$	$p_2=0.25$	$p_3=0.25$	$p_4=0.25$
φ_{i1}	$\varphi_{11}=1$	$\varphi_{21}=1$	$\varphi_{31}=-1$	$\varphi_{41}=-1$
φ_{i2}	$\varphi_{12}=1$	$\varphi_{22}=-1$	$\varphi_{32}=1$	$\varphi_{42}=-1$
φ_{i3}	$\varphi_{13}=1$	$\varphi_{23}=-1$	$\varphi_{33}=-1$	$\varphi_{43}=1$

$\mathbf{D}_4^a=[\boldsymbol{\varphi}_1\vdots\boldsymbol{\varphi}_2\vdots\boldsymbol{\varphi}_3]$ is constructed using these coefficient vectors $\boldsymbol{\varphi}_i=(\varphi_{1i},\varphi_{2i},\varphi_{3i},\varphi_{4i})^T\ (i=1,2,3)$.

Two-locus model

For two QTLs (Q_1 and Q_2), the joint genotypic value $g_{ij}\ (i,j=1,2,3,4)$ of the i-th genotype (*AA*,*BB*,*CC*,*DD*) at Q_1 and the j-th genotype ($A'A'$,$B'B'$,$C'C'$,$D'D'$) at Q_2 can be expressed by:

$$g_{ij}=\mu+a_i+a'_j+aa_{ij} \tag{2}$$

where $a_i=\varphi_{i1}\alpha_1+\varphi_{i2}\alpha_2+\varphi_{i3}\alpha_3$ and $a'_j=\varphi_{j1}\alpha'_1+\varphi_{j2}\alpha'_2+\varphi_{j3}\alpha'_3$ are the additive effects of Q_1 and Q_2, respectively. The additive–additive epistatic effect aa_{ij} is decomposed into nine orthogonal components (parameters) as follows:

$$aa_{ij}=\sum_{s=1}^{3}\sum_{t=1}^{3}\varphi_{is}\varphi_{jt}\alpha\alpha_{st} \tag{3}$$

where $\alpha\alpha_{st}$ is the orthogonalized interaction parameter between the s-th additive orthogonal parameter of Q_1 (α_s) and the t-th additive orthogonal parameter of Q_2 (α'_t). The coefficients $\varphi_{is}\varphi_{jt}$ enforce orthogonality, ensuring $\sum_{i,j}p_ip_jaa_{ij}=0$. The orthogonal coefficients matrix $\mathbf{D}_4^{aa}$ for epistatic effects is constructed using Kronecker and Hadamard products of the coefficient vectors of additive effects (Table S1). The two loci collectively have 15 effective degrees of freedom, allowing estimation of six additive parameters ($\alpha_1,\alpha_2,\alpha_3,\alpha'_1,\alpha'_2,\alpha'_3$; see Table S1) and nine epistatic parameters ($\alpha\alpha_{st}$). The coefficient vectors for the six additive effect parameters are: $\boldsymbol{\omega}_1=\boldsymbol{\varphi}_1\bigotimes\mathbf{1}_4$, $\boldsymbol{\omega}_2=\boldsymbol{\varphi}_2\bigotimes\mathbf{1}_4$, $\boldsymbol{\omega}_3=\boldsymbol{\varphi}_3\bigotimes\mathbf{1}_4$, $\boldsymbol{\omega}_4=\mathbf{1}_4\bigotimes\boldsymbol{\varphi}_1$,$\boldsymbol{\omega}_5=\mathbf{1}_4\bigotimes\boldsymbol{\varphi}_2$, $\boldsymbol{\omega}_6=\mathbf{1}_4\bigotimes\boldsymbol{\varphi}_3$, where $\mathbf{1}_4=\begin{pmatrix}1&1&1&1\end{pmatrix}^T$, and $\bigotimes$ denotes the Kronecker product. The coefficient vectors for the nine epistatic effect parameters are: $\omega_7=\omega_1\bigodot\omega_4$, $\omega_8=\omega_1\bigodot\omega_5$, $\omega_9=\omega_1\bigodot\omega_6$, $\omega_{10}=\omega_2\bigodot\omega_4$, $\omega_{11}=\omega_2\bigodot\omega_5$, $\omega_{12}=\omega_2\bigodot\omega_6$, $\omega_{13}=\omega_3\bigodot\omega_4$, $\omega_{14}=\omega_3\bigodot\omega_5$, $\omega_{15}=\omega_3\bigodot\omega_6$, where $\bigodot$ denotes the Hadamard product. $\mathbf{D}_4^{aa}=[\omega_7\vdots\omega_8\vdots\omega_9\vdots\omega_{10}\vdots\omega_{11}\vdots\omega_{12}\vdots\omega_{13}\vdots\omega_{14}\vdots\omega_{15}]$ is the orthogonal coefficient matrix for additive–additive epistatic effects of the four-parent MAGIC population.

For MAGIC populations derived from eight parents, orthogonal decomposition matrices (e.g. $\mathbf{DD}_8^{aa}$) are recursively generated using Hadamard matrices [28]. This approach preserves orthogonality while enabling scalability to higher-dimensional genetic effect spaces.

Construction of the parental origin probability matrix for putative quantitative trait locus

If the putative QTL is not a marker, then the parental origin probability of the QTL will be determined by its flanking markers. Unlike standard procedures used for simple biparental crosses, calculating these probabilities in MAGIC populations requires tracking complex recombination events across multiple generations of intercrossing. To address this, we developed a novel Markov chain-based algorithm specifically tailored to compute the conditional probabilities of multi-parental origins. The detailed derivation based on the Markov chain method is provided in the Supplemental Information.

Construction of quantitative trait locus model

Suppose a MAGIC population of size n derived from four homozygous parents, is subjected to genetic experiments across p distinct environments, the phenotypic value of a target trait is controlled by s QTLs (denoted as $Q_1,Q_2,\ldots,Q_s$) and t pairs of epistatic interactions among them. By orthogonal decomposition of QTL effects and Markov chain principles, the phenotypic value y_{ij} of the i-th individual in the j-th environment can be modeled as the following mixed linear model:

$$\begin{aligned}y_{ij}=\mu+\textstyle\sum_{k=1}^{s}\mathbf{a}_k^T\mathbf{x}_{ki}+\sum_{k,h\in(1,2,\ldots,s),k\neq h}^{t}\mathbf{aa}_{kh}^T\left(\mathbf{x}_{ki}\bigotimes\mathbf{x}_{hi}\right)+e_j+\\ \textstyle\sum_{k=1}^{s}\mathbf{ae}_{kj}^T\mathbf{x}_{ki}+\sum_{k,h\in(1,2,\ldots,s),k\neq h}^{t}\mathbf{aae}_{khj}^T\left(\mathbf{x}_{ki}\bigotimes\mathbf{x}_{hi}\right)+\varepsilon_{ij}\end{aligned} \tag{4}$$

where, μ represents the population mean; $\mathbf{a}_k=(\alpha_{1k},\alpha_{2k},\alpha_{3k})^T$ is the additive effect vector at Q_k, derived from orthogonal decomposition (fixed effect), and $\mathbf{x}_{ki}=\left(\mathbf{D}_4^a\right)^T\mathbf{p}_k^i$ is the vector corresponding to the three orthogonal additive parameters $(\alpha_{1k},\alpha_{2k},\alpha_{3k})$ at Q_k of the i-th individual, wherein, $\mathbf{p}_k^i=\left(p_{ik}^1,p_{ik}^2,p_{ik}^3,p_{ik}^4\right)^T$ is the probability vector consisting of $p_{ik}^j\ (j=1\sim4)$ which is the probability of the allele deriving from j-th parent at the k-th QTL of the i-th individual. $\mathbf{aa}_{kh}=(\alpha\alpha_{1kh},\ldots,\alpha\alpha_{9kh})^T$ denotes the orthogonal parameter vector of the additive–additive epistatic effect between Q_k and Q_h (fixed effect). The random effects include e_j, the effect of the j-th environment; $\mathbf{ae}_{kj}=\left(\alpha e_{1kj},\alpha e_{2kj},\alpha e_{3kj}\right)^T$, the additive-environment interaction parameter vector of Q_k and the j-th environment; $\mathbf{aae}_{khj}=\left(\alpha\alpha e_{1khj},\alpha\alpha e_{2khj},\ldots,\alpha\alpha e_{9khj}\right)^T$, the interaction effect vector between $\mathbf{aa}_{kh}$ and e_j.

The matrix form of the model is expressed as:

$$\begin{aligned}\mathbf{y}&=\boldsymbol{\mu}+\mathbf{X}_A\mathbf{b}_A+\mathbf{X}_{AA}\mathbf{b}_{AA}+\mathbf{U}_E\mathbf{e}_E+\textstyle\sum_{k=1}^{s}\mathbf{U}_{A_kE}\mathbf{e}_{A_kE}\\&\quad+\textstyle\sum_{v=1}^{t}\mathbf{U}_{AA_vE}\mathbf{e}_{AA_vE}+\mathbf{e}_\varepsilon\\&=\left[1\vdots\mathbf{X}_A\vdots\mathbf{X}_{AA}\right]\left[\boldsymbol{\mu}^T\vdots\mathbf{b}_A^T\vdots\mathbf{b}_{AA}^T\right]^T+\textstyle\sum_{u=1}^{r}\mathbf{U}_u\mathbf{e}_u+\mathbf{I}\mathbf{e}_\varepsilon\\&=\mathbf{X}\mathbf{b}+\textstyle\sum_{u=1}^{r+1}\mathbf{U}_u\mathbf{e}_u\end{aligned} \tag{5}$$

Here, $\mathbf{y}$ is a $n\times1$ vector of phenotypic values; $\mathbf{b}_A=\left[\mathbf{a}_1{}^T,\mathbf{a}_2{}^T,\ldots,\mathbf{a}_s{}^T\right]^T$ and $\mathbf{b}_{AA}=\left[\mathbf{aa}_1{}^T,\mathbf{aa}_2{}^T,\ldots,\mathbf{aa}_t{}^T\right]^T$ are orthogonal parameter vectors for additive and additive–additive epistatic effects, respectively. The

random effect components follow multivariate normal distributions: $\mathbf{e}_E = (e_1,\ldots,e_p)^T \sim MVN\left(0,\mathbf{I}\sigma_E^2\right)$ for environmental effects, $\mathbf{e}_{A_kE} = (\mathbf{aae}_{k1}{}^T,\mathbf{aae}_{k2}{}^T,\ldots,\mathbf{aae}_{kp}{}^T)^T \sim MVN\left(0,\mathbf{I}\sigma_{A_kE}^2\right)$ for additive-environment interactions, and $\mathbf{e}_{AA_vE} = (\mathbf{aae}_{v1}{}^T,\mathbf{aae}_{v2}{}^T,\ldots,\mathbf{aae}_{vp}{}^T)^T \sim MVN\left(0,\mathbf{I}\sigma_{AA_vE}^2\right)$ for additive–additive epistasis by environment interactions. The design matrices $\mathbf{X_A}$,$\mathbf{X_{AA}}$,$\mathbf{U}_E$,$\mathbf{U}_{A_kE}$, and $\mathbf{U}_{AA_vE}$ are constructed based on orthogonal decomposition principles. For example, $\mathbf{X_A} = [\mathbf{X}_1^a\vdots\mathbf{X}_2^a\vdots\ldots\vdots\mathbf{X}_s^a]$, where $\mathbf{X}_\mathbf{k}^a = [\mathbf{x}_{k1}\vdots\mathbf{x}_{k2}\vdots\ldots\vdots\mathbf{x}_{kn}]^T$. The residual vector $\mathbf{e}_\varepsilon \sim MVN\left(0,\mathbf{I}\sigma_\varepsilon^2\right)$.

One-dimensional and two-dimensional quantitative trait locus scanning

In brief, the scanning strategy in this study involves, first identifying QTL contributing significant additive or additive-environment interaction effects through genome-wide one-dimensional scanning. Subsequently, the paired epistatic QTLs contributing significant additive–additive epistasis or their interaction with environment using two-dimensional genome scanning. The stepwise regression for model selection in scanning was employed to eliminate false-positive QTLs. This one- and two-dimensional scanning approach has been successfully implemented in the QTLNetwork [4, 5], a QTL mapping software for biparental mapping populations.

One-dimensional scanning

The process follows the framework of the CIM. When testing a putative QTL in a specific genomic region, the selected marker intervals are used as covariates to control background genetic effects of the QTLs located outside the tested region. Let $I_1,I_2,\ldots,I_c$ denote the marker intervals selected as covariates (background controls) when testing a specific region. Each interval I_l is determined by its left and right flanking markers $\left(M_l^-,M_l^+\right)$. The one-dimensional scanning model is as follows:

$$y_{ij} = \mu_j + \mathbf{a}_{kj}^T\mathbf{x}_{ki} + \sum_{l=1}^{c}\left(\left(\mathbf{a}_{lj}^-\right)^T\boldsymbol{\xi}_{il}^- + \left(\mathbf{a}_{lj}^+\right)^T\boldsymbol{\xi}_{il}^+\right) + \varepsilon_{ij} \tag{6}$$

Here, μ_j is the population mean of the j-th environment; $\mathbf{a}_{kj} = (\alpha_{1kj},\alpha_{2kj},\alpha_{3kj})^T$ denotes the additive effect vector of the tested QTL in the j-th environment, with coefficient vector $\mathbf{x}_{ki}$, like that in Equation 4. Similarly, $\mathbf{a}_{lj}^-$ and $\mathbf{a}_{lj}^+$ are the additive effect vectors of the left and right flanking markers of the l-th interval in the j-th environment, with corresponding coefficients vectors $\boldsymbol{\xi}_{il}^-$ and $\boldsymbol{\xi}_{il}^+$, just like $\mathbf{x}_{ki}$, determined by the parental origin probability vector $\mathbf{p}$ and the orthogonal decomposition matrix $\mathbf{D}$. The term ε_{ij} represents the residual error.

This model can be expressed in a simplified matrix form:

$$\mathbf{y} = \mathbf{W}_Q\mathbf{b}_Q + \mathbf{W}_M\mathbf{b}_M + \mathbf{e}_\varepsilon \tag{7}$$

where $\mathbf{y}$ and $\mathbf{e}_\varepsilon$ have the same definitions as that in Equation 5. $\mathbf{b}_Q$ is the vector of effects for the tested QTL, with design matrix $\mathbf{W}_Q$. $\mathbf{b}_M$ is the effect vector of the markers to control background genetic effects, and $\mathbf{W}_M$ is the corresponding design matrix. According to Henderson Method III [29], the expectation of the extra sum of squares of $\mathbf{b}_Q$ is derived as:

$$\begin{aligned}\left[\mathrm{SSR}\left(\mathbf{b}_Q \mid \mathbf{b}_M\right)\right] &= E\left(\mathbf{y}^{\mathrm{T}}\mathbf{W}\mathbf{W}^{+}\mathbf{y} - \mathbf{y}^T\mathbf{W}_M\mathbf{W}_M^{+}\mathbf{y}\right)\\ &= \mathrm{tr}\left[\mathbf{W}_Q^{\mathrm{T}}\mathbf{A}_M\mathbf{W}_QE\left(\mathbf{b}_Q\mathbf{b}_Q^T\right)\right] + \sigma_\varepsilon^2\left(r_{\mathbf{W}} - r_{\mathbf{W}_M}\right)\end{aligned} \tag{8}$$

where $\mathbf{W} = [\mathbf{W}_Q\vdots\mathbf{W}_M]$, $\mathbf{A}_M = \mathbf{I} - \mathbf{W}_M\left(\mathbf{W}_M^T\mathbf{W}_M\right)^{+}\mathbf{W}_M^T$, and $r_{\mathbf{W}}$, $r_{\mathbf{W}_M}$ denote the ranks of $\mathbf{W}$ and $\mathbf{W}_M$, respectively. Under the null hypothesis $H_0 : \mathbf{b}_Q = \mathbf{0}$, the F-statistic

$$F = \frac{\mathrm{SSR}\left(\mathbf{b}_Q \mid \mathbf{b}_M\right)/\left(r_{\mathbf{W}} - r_{\mathbf{W}_M}\right)}{\mathrm{SSE}/\left(n - r_{\mathbf{W}}\right)} \tag{9}$$

follows an F-distribution. Genome-wide one-dimensional scanning is performed using this F-statistic. A candidate QTL is identified if its F-value exceeds the threshold determined by the permutation test method [30]. Stepwise regression is subsequently applied to refine candidate QTLs and obtain a set of QTLs with significant individual effects. Notably, in a four-parent homozygous MAGIC population, the one-dimensional scan for each environment tests the additive effects in specific environment, with the null hypothesis $H_0 : \alpha_{1kj} = \alpha_{2kj} = \alpha_{3kj} = 0$. Due to orthogonal decomposition constraints (e.g. $\alpha_{1kj} = \frac{a_{1kj}+a_{2kj}}{2}$, $\alpha_{2kj} = \frac{a_{1kj}+a_{3kj}}{2}$, $\alpha_{3kj} = \frac{a_{1kj}+a_{4kj}}{2}$) and the Cockerham genetic model restriction $a_{1kj} + a_{2kj} + a_{3kj} + a_{4kj} = 0$, the null hypothesis is equivalent to the $H_0 : a_{1kj} = a_{2kj} = a_{3kj} = a_{4kj} = 0$. A similar method was applied to test additive-by-additive epistatic effects, where the null hypothesis for nine orthogonal effect parameters is equivalent to that for all 16 original epistatic effects.

To select marker intervals for background control, the model:

$$y_{ij} = \mu_j + \left(\mathbf{a}_{lj}^-\right)^T\boldsymbol{\xi}_{il}^- + \left(\mathbf{a}_{lj}^+\right)^T\boldsymbol{\xi}_{il}^+ + \varepsilon_{ij} \tag{10}$$

is first employed before one-dimensional QTL scanning. Parameters and corresponding coefficients align with those in Equation 6. The same statistical framework (Henderson Method III) and permutation testing method are used to evaluate candidate intervals, followed by stepwise regression to control false positives.

Two-dimensional scanning

Two-dimensional scanning is employed to detect paired epistatic QTLs within selected pair of marker interval. Assuming c significant marker intervals and s QTLs are identified from the one-dimensional scanning, each pair of marker intervals denoted by I_l^A and,I_l^B will be tested and selected if they reach significance using following model:

$$\begin{aligned}y_{ij} = \mu_j &+ \left(\boldsymbol{aa}_{lj}^{A^-B^-}\right)^T\left(\boldsymbol{\xi}_{lj}^{A^-}\otimes\boldsymbol{\xi}_{lj}^{B^-}\right) + \left(\boldsymbol{aa}_{lj}^{A^+B^+}\right)^T\left(\boldsymbol{\xi}_{lj}^{A^+}\otimes\boldsymbol{\xi}_{lj}^{B^+}\right)\\ &+ \textstyle\sum_{k=1}^{c}\left(\left(\mathbf{a}_{kj}^-\right)^T\boldsymbol{\xi}_{ik}^- + \left(\mathbf{a}_{kj}^+\right)^T\boldsymbol{\xi}_{ik}^+\right) + \varepsilon_{ij}\end{aligned} \tag{11}$$

Here, $\boldsymbol{aa}_{lj}^{A^-B^-}$ and $\boldsymbol{aa}_{lj}^{A^+B^+}$ represent the additive–additive epistatic effect vectors between flanking markers $\left(M_l^{A^-},M_l^{A^+}\right)$ and $\left(M_l^{B^-},M_l^{B^+}\right)$ in the j-th environment. Coefficient vectors $\boldsymbol{\xi}_{il}^{A^-}\otimes\boldsymbol{\xi}_{il}^{B^-}$ and $\boldsymbol{\xi}_{il}^{A^+}\otimes\boldsymbol{\xi}_{il}^{B^+}$ are constructed as described in Equation 4. The null hypothesis $H_0 : \alpha\alpha_{1lj}^{A^-B^-} = \alpha\alpha_{1lj}^{A^+B^+} = \cdots = \alpha\alpha_{9lj}^{A^-B^-} = \alpha\alpha_{9lj}^{A^+B^+} = 0$ is tested using Henderson Method III. Genome-wide scanning and model selection via stepwise regression are employed to select significant paired marker intervals with potential epistatic QTLs.

Suppose f interval pairs $\left(I_1^A,I_1^B\right),\left(I_2^A,I_2^B\right),\ldots,\left(I_f^A,I_f^B\right)$ have been selected for mapping epistatic QTLs, then the two-dimensional scanning model to test paired epistatic QTLs will be conducted based on the following model:

$$\begin{aligned}y_{ij} = \mu_j &+ \mathbf{aa}_{khj}^T\left(\mathbf{x}_{ki}\otimes\mathbf{x}_{hi}\right) + \textstyle\sum_{p=1}^{s}\mathbf{a}_{pj}^T\mathbf{x}_{pi}\\ &+ \textstyle\sum_{l=1}^{f}\left[\left(\boldsymbol{aa}_{lj}^{A^-B^-}\right)^T\left(\boldsymbol{\xi}_{il}^{A^-}\otimes\boldsymbol{\xi}_{il}^{B^-}\right) + \left(\boldsymbol{aa}_{lj}^{A^+B^+}\right)^T\left(\boldsymbol{\xi}_{il}^{A^+}\otimes\boldsymbol{\xi}_{il}^{B^+}\right)\right] + \varepsilon_{ij}\end{aligned} \tag{12}$$

where, $\mathbf{aa}_{khj}^{T} = (\alpha\alpha_{1khj}, \alpha\alpha_{2khj}, \ldots, \alpha\alpha_{9khj})^{T}$ denotes the orthogonalized epistatic effect vector between QTLs Q_k and Q_h in the j-th environment, with coefficient vector $\mathbf{x}_{ki} \otimes \mathbf{x}_{hi}$. The null hypothesis $H_0 : \alpha\alpha_{1khj} = \alpha\alpha_{2khj} = \cdots = \alpha\alpha_{9khj} = 0$ is tested using Henderson Method III and the significance threshold is specified by permutation method.

Estimation and statistical testing of quantitative trait locus effects

For a standard mixed linear model, the fixed effect of QTL can be estimated using generalized least squares; while the QTL-by-environment effects can be predicted using best linear unbiased prediction or adjusted unbiased prediction through restricted maximum likelihood estimation or minimum norm quadratic unbiased estimation [31]. The significance of random effect is assessed using Gibbs sampling [27] method. Detailed information on the algorithm is provided in the Supplemental Information.

Simulation

To evaluate the performance of the new mapping method, simulations were conducted using the R package mpMap2, which facilitates both the construction of linkage maps and the generation of genotype data of the MAGIC population. A simulated MAGIC population of size n was generated, consisting of three linkage groups and each spanning genetic distance of 200 cM with 201 markers. In the first linkage group, three QTLs denoted as Q1, Q2, and Q3 were positioned at genetic distances of 20 cM, 100 cM, and 180 cM, respectively. In the second linkage group, three QTLs (Q4, Q5, and Q6) were placed at 10, 100, and 180 cM. The third linkage group contained three QTLs Q7, Q8, and Q9 at 10, 90, and 170 cM, respectively. For individual-effect QTLs, Q1 and Q4 were modeled with additive effects, Q2 with additive-by-environment interaction effects and Q3 with both additive and additive-by-environment interaction effects. Regarding epistatic QTL pairs, EQ1 (Q4–Q5) contributed additive–additive epistatic effects; EQ2 (Q6–Q7) contributed additive–additive epistasis-by-environment interactions; EQ3 (Q8–Q9) contributed both additive–additive epistatic effects and their interaction effects with the environment (Fig. S1). Simulations were performed for four-parent and eight-parent recombinant inbred lines (RIL) MAGIC populations with sizes of 500, 750, 1000 and 1250 individuals, respectively. Genetic variances of 0.25, 0.50, and 0.75 were set. A total of 500 repetitions were conducted to calculate the mean and standard deviation (SD) of QTL parameters in each scenario. The support intervals (SI) for each QTL were estimated using the method proposed by Lander and Botstein. If the true genetic position of a simulated QTL fell within its corresponding SI [1], the QTL was considered to be successfully mapped. To systematically benchmark the performance of QTLNetwork-MP, comparative simulations were conducted using the widely adopted R package qtl2. Because qtl2 does not natively support repeated phenotypic measurements of the same individual across multiple environments within a single data column, independent genome scans were performed for each simulated environment. Locus scanning was executed using the scan1 function based on the HK regression model. To determine significance thresholds ($\alpha = 0.05$), permutation tests were conducted using the *scan1perm* function with 1000 permutations. Significant QTLs were subsequently identified using the *find_peaks* function (prob = 0.95) combined with the peak-drop algorithm. Finally, the additive effects for the detected QTLs in each environment were estimated using the *scan1coef* function.

Data description

To evaluate the effectiveness of QTLNetwork-MP on a real dataset, we applied the method to a widely studied cowpea MAGIC population, which has been extensively used in multi-environment and multi-year field trials. Our model is especially well-suited for QTL mapping in such environmental contexts due to its ability to accommodate complex environmental interaction. Notably, given the relatively small size of the cowpea dataset, the analysis was limited to single-locus scanning.

Cowpea population

The cowpea MAGIC population was developed by Huynh et al. through a series of intercrosses involving eight parents selected according to their diverse agronomic characteristics [16]. The parents (IT89KD-288, IT84S-2049, CB27, IT82E-18, SuVita 2, IT00K-1263, IT84S-2246, and IT93K-503-1) were intercrossed through three successive generations (diallel, four-way, and eight-way crosses) to maximize recombination events. A total of 305 F_8 recombinant inbred lines (RILs) were generated through single-seed descent to reach near-complete homozygosity. Genomic DNA was extracted from young leaf tissue of both the RILs and parental lines, and genotyped using the Illumina Cowpea Consortium Array [32]. The array initially assayed 51 128 SNPs, which were then subjected to stringent quality control: markers with >10% missing data or a minor allele frequency <5% were removed. After filtering, 4361 high-confidence SNPs were retained for linkage map construction. SNPs were aligned to the cowpea reference genome (IT97K-499-35 v1.1), and genetic distances were estimated using the R/mpMap package. Marker order was refined through iterative inspection of recombination rate heatmaps. Redundant markers with no detectable recombination were collapsed into bins using the *mpcollapse* function. The resulting linkage map spanned 1246 cM across 11 linkage groups, with an average marker density of 3.5 SNPs per cM.

Results

Mapping results of the simulated recombinant inbred lines multiparent advanced generation intercross population

The simulation results of QTL mapping for RIL MAGIC populations are summarized in Fig. 1 and Table 2 for QTLs with individual effect (hereafter referred to as individual-effect QTL), and in Fig. 2 and Table 3 for epistatic QTLs (more detailed results are provided in Extended Tables 1–8 for QTLNetwork-MP, and Extended Tables 9–12 for qtl2 benchmarking). Each scenario was simulated 500 repetitions. Figure 1 illustrates the performance of mapping QTLs with individual effect across various scenarios on heritability and population size, for the MAGIC populations derived from four and eight parents. Comparative analysis clearly revealed that QTLNetwork-MP consistently outperformed qtl2 in both statistical power and FDR control across almost all simulated scenarios (Fig. 1). A critical limitation of qtl2 was observed when mapping QTLs with environmental interactions (e.g. Q3). Due to its inability to jointly analyze repeated measures across environments in a unified model, qtl2 failed to leverage the full dataset simultaneously. Consequently, qtl2 could only detect Q3 within the first environment (E1), artificially capping its statistical mapping power at ∼50%. In contrast, QTLNetwork-MP successfully integrated data across environments, maintaining high detection power (>0.75) with low false discovery rates (FDR <0.1) in most scenarios. However, under more challenging conditions such as

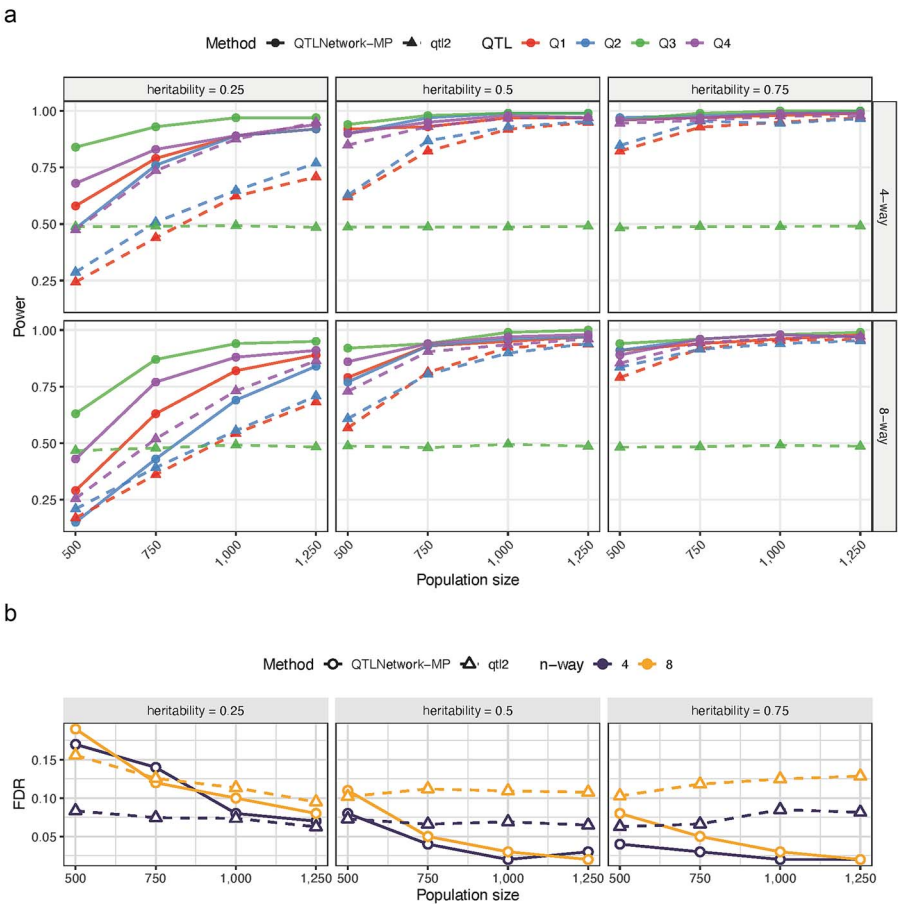

Figure 1 Power and false discovery rate of mapping individual-effect QTL using QTLNetwork-MP and qtl2 in simulated pure-line MAGIC population under different heritability and population size. Notes: (a) Q1 exhibits only additive effect; Q2 exhibits only additive-by-environment effect; Q3 exhibits both additive and additive-by-environment effects; and Q4 exhibits both additive and additive-by-additive epistatic effects, though epistatic mapping results are not shown here (they will be presented below). (b) FDR performance across each simulated scenario. The performance of QTLNetwork-MP and qtl2 are distinguished by different line types as indicated in the legend.

in an eight-parent population with 500 individuals and a low trait heritability of 0.25 (where the genetic contribution of an individual-effect QTL is only 0.024), the detection power of QTLNetwork-MP was generally maintained, except for Q4, which showed a power <0.25. Notably, its FDR remained acceptably low (∼0.2). Table 2 reports the accuracy of QTLNetwork-MP under different scenarios in the four-parent population, indicating that QTLs were generally localized accurately with small standard deviations. For mapping individual-effect QTL, a sample size of 500 is generally sufficient for four-parent MAGIC (Table 4). However, for complex eight-parent MAGIC population, increasing the population size is necessary to achieve reliable mapping power (Extended Table 5). Furthermore, because existing software frameworks, including qtl2, currently lack the capability to perform multi-dimensional epistatic scanning in MAGIC populations, benchmarking for epistatic QTLs was uniquely conducted within the QTLNetwork-MP framework.

Figure 2 presents the mapping results of QTLNetwork-MP for epistatic QTL pairs. The results clearly indicate that epistatic QTLs are substantially more challenging to detect than individual-effect QTLs. To achieve adequate statistical power (>0.5) while controlling the FDR <0.2, a minimum population size of ∼750 is required for a four-parent population and 1000 for an eight-parent population, provided that the trait heritability is not too low (e.g. ∼0.25 or higher). Table 3 reflects the accuracy of mapping epistatic QTLs in the four-parent population (with more comprehensive details available in Extended Tables 2 and 6 for both four-parent and eight-parent population). Although the statistical power for mapping epistatic QTLs is generally lower than that for individual-effect QTLs, once detected, their genomic positions can still be estimated precisely with relatively small standard deviations. Overall, the proposed method exhibits strong statistical power, low false discovery rates, and high accuracy in detecting both individual-effect QTLs and epistatic QTL pairs in MAGIC populations.

Additionally, within the QTLNetwork-MP results, we observed that Q1, contributing only additive effect, exhibited higher mapping power than Q2 which contributed solely additive-by-environment interaction effect, even when both explained equal proportions of the phenotypic variance. Furthermore, although Q1 and Q4 both have additive effects with equivalent genetic contribution, Q4 consistently demonstrated slightly higher mapping power across all scenarios. This difference may be related to the involvement of Q4 in additive–additive interaction with Q5 (EQ1). Consequently, a small proportion of the additive–additive epistatic effect may be absorbed into the additive effect, empirically demonstrating the robustness of our 1D scanning model even under the misspecification caused by unobserved interacting loci.

Table 2 Simulation on mapping QTL with additive and additive-by-environment interaction using QTLNetwork-MP in four-way pure-line MAGIC population.

Scenario	$h^2 = 0.25$				$h^2 = 0.50$				$h^2 = 0.75$			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
RC (%)	2.42	2.42	4.84	2.42	4.84	4.84	9.68	4.84	7.26	7.26	14.52	7.26
Pos (cM)	20	100	180	10	20	100	180	10	20	100	180	10
Est (I)	19.88	100.09	180.06	9.97	19.90	100.00	180.06	10.11	19.94	99.97	179.98	10.05
SD (I)	2.13	2.03	1.53	2.05	1.32	1.36	0.94	1.34	0.97	0.96	0.85	0.87
Power (I)	0.58	0.48	0.84	0.68	0.92	0.90	0.94	0.90	0.96	0.97	0.96	0.96
FDR (I)	0.17				0.08				0.04			
Est (II)	20.07	99.89	180.01	10.12	20.02	99.97	180.05	9.97	19.95	99.99	180.03	10.01
SD (II)	1.80	1.84	1.28	1.73	1.17	1.13	0.78	1.06	0.82	0.79	0.58	0.68
Power (II)	0.79	0.76	0.93	0.83	0.93	0.97	0.98	0.95	0.97	0.98	0.99	0.97
FDR (II)	0.14				0.04				0.03			

Notes: Q1–Q4 denote the names of simulated QTLs; Pos indicates the QTL position relative to the first marker in the same linkage group; RC represents the relative contribution; Est and SD refer to the estimates of parameters and their standard deviations, with I and II in parentheses corresponding to sample sizes of 500 and 750, respectively; Power and FDR represent statistical power and false discovery rate (For more complete results, see Extended Table 1).

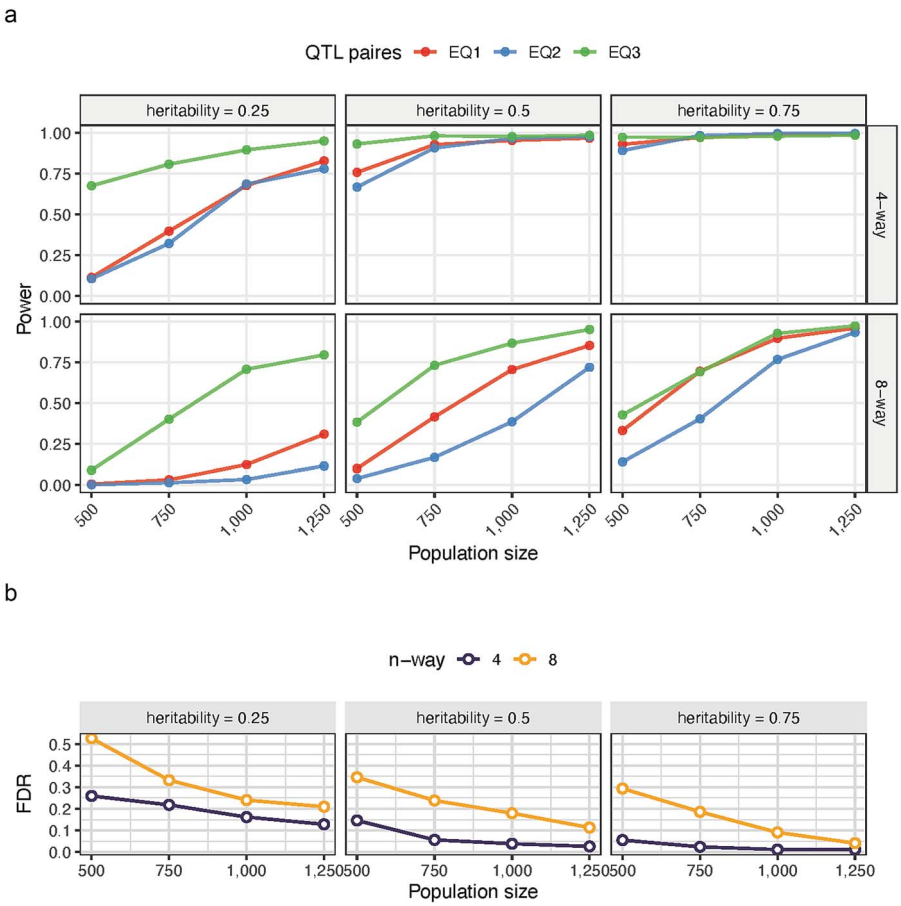

Figure 2 Power and false discovery rate of mapping epistatic QTLs using QTLNetwork-MP in simulated pure-line MAGIC population under different heritability and population size. Notes: (a) EQ1 exhibits only additive–additive epistatic effect; EQ2 exhibits only additive–additive -by-environment epistatic effect; EQ3 exhibits both additive–additive and additive–additive-by-environment epistatic effects. (b) FDR performance across each simulated scenario.

Effect estimation in simulated multiparent advanced generation intercross population

Unlike current QTL mapping methods for MAGIC population, our approach not only accurately identifies QTL positions but also estimates the genetic effects of each individual-effect QTL (Table 4) and each epistatic QTL pair (Table 5). The effect estimation in the eight-parent populations is lengthy and complex; therefore, the estimates of effects are provided in Extended Tables 7 and 8. Here, we focus on the estimates of effect for the four-parent MAGIC population with a sample size of 500. Table 4 summarizes the estimated additive and additive-by-environment effects for individual-effect QTLs. As

Table 3 Simulation on mapping QTL with additive–additive epistasis and interaction with environment using QTLNetwork-MP in four-way pure-line MAGIC population.

Scenario	$h^2 = 0.25$			$h^2 = 0.50$			$h^2 = 0.75$		
	EQ1	EQ2	EQ3	EQ1	EQ2	EQ3	EQ1	EQ2	EQ3
RC (%)	3.22	3.22	6.45	6.45	6.45	12.90	9.68	9.68	19.35
Pos (cM)	10–100	180–10	90–170	10–100	180–10	90–170	10–100	180–10	90–170
Est (I)	10.05–99.91	179.98–9.70	90.03–170.04	10.03–100.10	180.02–9.90	89.97–170.00	10.05–100.01	180.00–10	89.99–170.04
SD (I)	2.14–1.98	1.95–2.04	1.39–1.35	1.34–1.15	1.17–0.98	0.84–0.80	0.85–0.61	0.63–0.53	0.43–0.55
Power (I)	0.11	0.11	0.68	0.76	0.67	0.93	0.93	0.89	0.97
FDR (I)	0.26			0.15			0.06		
Est (II)	10.05–99.91	179.98–9.70	90.03–170.04	10.03–100.10	180.02–9.90	89.97–170.00	10.05–100.01	180.00–10.00	89.99–170.04
SD (II)	2.14–1.98	1.95–2.04	1.39–1.35	1.34–1.15	1.17–0.98	0.84–0.80	0.85–0.61	0.63–0.53	0.43–0.55
Power (II)	0.11	0.11	0.68	0.76	0.67	0.93	0.93	0.89	0.97
FDR (II)	0.22			0.06			0.02		

Notes: EQ1–EQ3 denote the simulated paired QTL involved in epistasis; other parameters have the same definitions as those in Table 2 (For more complete results, see Extended Table 2).

expected, Q1, which carries only additive effect, shows non-zero estimates exclusively for the additive component, while its additive-by-environment interaction effect is estimated as zero. Similarly, Q2, contributing only additive-by-environment effect, shows negligible additive estimates. For Q3, contributing both effect types, our method accurately estimates both the additive and additive-by-environment components without bias. When comparing effect estimation accuracy, the limitations of the separate environmental scanning strategy employed by qtl2 became highly apparent. For Q3, qtl2 inaccurately estimated the additive effect within the E1 environment to be approximately double the true simulated additive effect (or equivalent to double the absolute value of the interaction effect). This severe bias occurs because the environmental interaction effect is incorrectly absorbed into the marginal additive effect when separate environments are analyzed. Conversely, because QTLNetwork-MP simultaneously fits all genetic and environmental variance components within a comprehensive full mixed model, our estimates for both additive and interaction effects align perfectly with the simulated true values. Notably, although Q4 is involved in the epistatic interaction (EQ1), the estimation of its additive effect remains essentially stable. Table 5 presents the estimates of epistatic effects. For EQ1, which contains only additive-by-additive effects, the estimated additive–additive-by-environment effects are close to zero, as expected. Similarly, EQ2, which only exhibits additive–additive-by-environment effect, shows minimal estimates for the additive–additive epistatic component. In the case of EQ3, which express additive–additive epistasis and epistasis–environment interaction, both effect components are estimated accurately and without bias. Moreover, the additive–additive effects of EQ1 are estimated unbiasedly even in the presence of additive effect of Q4, indicating that our model can effectively separate additive and epistatic components in estimation.

Overall, our method could generate reliable effect estimates with relatively low standard deviations. Although detection power varies significantly between individual-effect QTLs and epistatic QTL pairs under the same simulation conditions, the accuracy of effect estimation remains consistently high for both. Across 500 repeated simulations, the average estimates of effects are very close to the true values, with minimal variation. These results demonstrate the high accuracy and effectiveness of our method in estimation of position and genetic effects of QTL.

Results of quantitative trait locus mapping for seed size traits in the eight-parent cowpea multiparent advanced generation intercross population

The genetic linkage map was constructed using 4361 high-quality SNP markers after filtering adjacent loci (within 100 kb) exhibiting high linkage disequilibrium ($r^2 > 0.99$) using PLINK2.0 [33]. The final map comprises 11 linkage groups (LGs), with genetic distances ranging from 69.40 cM (LG10) to 132.68 cM (LG3). Marker distribution was generally uniform across most LGs, with the exception of localized clustering observed near the distal end of LG5 and the proximal region of LG10 (Fig. 3A).

Our new mapping method successfully identified three additive QTLs associated with seed size located on LGs 6, 7, and 8. The QTLs on LG6 (80.0 cM, support interval: 76.5–80.3 cM) and LG8 (71.6 cM, single-marker interval) were consistent with previously reported findings, while a novel QTL was detected on LG7 (23.8 cM, support interval: 23.3–24.1 cM). Parental effect estimates revealed that IT89KD-288 contributed the largest additive effect (2.41, $P = 2.01 \times 10^{-7}$) at the LG6 QTL, aligning with its known superiority in grain size traits. Similarly, IT82E-18 and IT00K-1263 exhibited significant positive effects (3.24, $P = 1.75 \times 10^{-14}$; 3.02, $P = 9.18 \times 10^{-13}$) at the LG8 QTL, corroborating their documented influence on seed size. The newly identified LG7 QTL was predominantly influenced by CB27 (1.29, $P = 1.27 \times 10^{-3}$) and IT00K-1263 (1.36, $P = 6.71 \times 10^{-4}$), consistent with CB27's historical performance as a large-seeded parent (Table 6).

No epistatic effect or QTL-by-environment interaction was detected, probably due to the relatively small sample size ($n = 305$), which limits statistical power to identify complex genetic interactions. However, the detected additive QTLs explained a substantial proportion of phenotypic variation, underscoring the utility of MAGIC populations in dissecting polygenic traits. The concordance between the estimated parental effects and field agronomic performance of traits validates the robustness of our approach.

Discussion

The current QTL mapping methods tailored for MAGIC populations exhibit significant limitations, particularly in dissecting the genetic architecture of complex traits governed by multiple loci and their interactions with environmental factors. In this study, we proposed

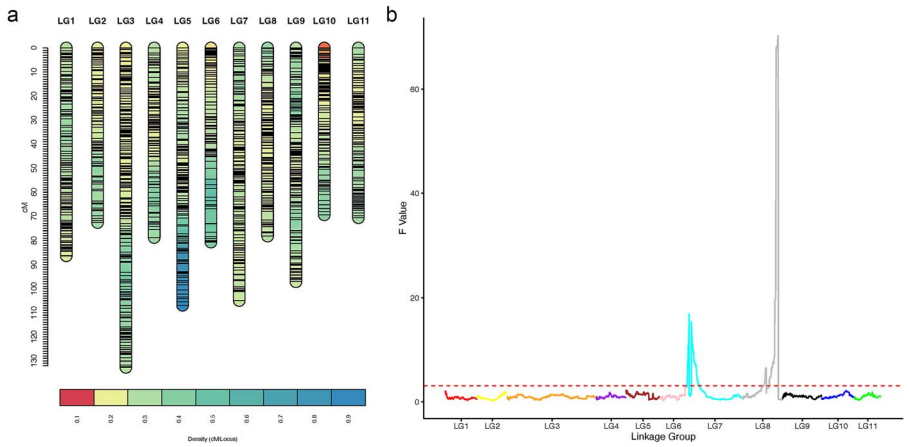

Figure 3 The linkage map and mapping results in the cowpea population. Notes: (a) linkage map of the cowpea eight-parent RIL MAGIC population. (b) 1-D genome scan for seed size traits in cowpea based on *F*-statistic.

a flexible and computationally efficient framework that integrates orthogonal decomposition of genetic effects with mixed linear models. By establishing the novel mathematical derivations for multi-parent interactions, this approach enables simultaneous estimation of additive effects, additive-by-additive epistatic effects, and their interactions with environments, thereby overcoming the limitations of traditional biparental models and fixed-effect methods, especially in MAGIC populations derived from multiple parents.

Extensive simulations of four and eight-parent MAGIC populations demonstrated the robustness, scalability, and high accuracy of our method. Statistical power was substantially improved when the heritability increased from 0.25 to 0.75 and the sample size from 500 to 1250, while the false discovery rate was well controlled (Fig. 1, Fig. 2). Importantly, although the eight-parent design poses greater challenges for QTL detection, our study showed that the power of detecting additive QTLs could achieve over 0.8 under a sufficiently large sample size (e.g. 1250 individuals, which is practically attainable) and moderate trait heritability (∼0.25 or above). By leveraging orthogonal decomposition of QTL effects, the number of estimated parameters was significantly reduced. For example, reducing 16 pairwise additive-by-additive terms to 9 orthogonal components in the four-parent case (Table S1), without sacrificing estimation precision (Tables 4 and 5). Furthermore, modeling epistatic and G × E effects as random components in the mixed model (especially in eight-parent populations) mitigated over-parameterization while accounting for polygenic background variation effectively. While shrinkage-based or Bayesian alternatives are highly effective for whole-genome prediction of highly polygenic traits, they often lack the rigorous hypothesis testing required for precise locus discovery. By contrast, our mixed linear model based framework effectively isolates significant, actionable QTLs for marker-assisted selection while robustly absorbing the collective variance of numerous unmapped small-effect loci into the random background. This is a notable advantage over existing methods based on fixed-effect modeling. Our method not only alleviates previous issues regarding over-parameterization in QTL mapping for MAGIC population, but also has great scalability to extend model for more complex MAGIC population from more parents.

The example analysis of cowpea population of 305 RILs derived from eight parents further validated the practical applicability of our method. We identified three significant additive QTLs of grain size (*qSS6*, *qSS7*, *qSS8*) on linkage groups 6, 7, and 8 (Fig. 3), *qSS6* and *qSS8* were consistent with previous findings (Table 6), *qSS7* is a novel QTL. Notably, the alleles contributing the largest positive effects at *qSS7* originate from CB27 and IT00K-1263. This aligns well with the trait performance of parents of the population, as CB27 is known to consistently perform the largest seed and highest yield among the parents, and IT00K-1263 is a high-yielding line selected for its grain quality. This correspondence provides strong biological validation for this newly identified QTL. No QTL involved in epistasis or interaction with environment were detected. This may be due to relatively small population size and is consistent with the results of simulation study on the method, that population size of at least 500 is required to detect epistatic effects in eight-parent populations.

The discovery of a new QTL on linkage groups 7 (*qSS7*) emphasizes the importance of methods designed specifically for MAGIC populations. The proposed framework of QTL mapping is very flexible and particularly suitable for populations with eight or more parents, where conventional fixed-effect models become impractical. For MAGIC populations from larger number of parents, such as the 16-parent design, our method remains applicable. Preliminary simulations for mapping individual-effect QTL in 16-parent pure-line MAGIC populations have confirmed the effectiveness of our approach. However, exhaustive simulation for mapping epistatic QTLs still needs more time due to high computational demands of larger matrix operations; therefore, the results for the 16-parent design are not included in this study. Our future work will focus on refining the method and designing new software based on combined computation of GPU and CPU to tackle these computational challenges.

By treating many additive–additive epistatic effects of QTLs as random, estimation of epistatic effects is avoided and only the variance of the effects is required to be estimated using mixed linear model approach; thus, our method effectively balances statistical power and model complexity, and is easy to extend for other MAGIC populations of arbitrary structure and generation. Our study found that population size is a key factor affecting the reliability of the result of QTL mapping for MAGIC population, we recommend at least 500 lines for mapping individual-effect QTL and 1000 lines for detecting epistatic QTLs. It is also obvious that a larger population size is required for QTL mapping in MAGIC population derived from a larger number of original parents. In the future, this framework can be applied in

Table 4 Estimates of additive and additive-by-environment interaction effects of QTL using QTLNetwork-MP in four-way pure-line MAGIC population (population size: 500).

Scenario	$h^2 = 0.25$				$h^2 = 0.50$				$h^2 = 0.75$				Par
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
RC (%)	2.42	2.42	4.84	2.42	4.84	4.84	9.68	4.84	7.26	7.26	14.52	7.26	—
a_1 (SD)	4.91 (0.73)	***0.11 (1.08)***	4.56 (0.92)	5.10 (0.92)	4.49 (0.51)	***0.07 (0.57)***	4.45 (0.54)	4.51 (0.54)	4.45 (0.30)	***0.11 (0.34)***	4.42 (0.32)	4.47 (0.31)	4.5
a_2 (SD)	−1.68 (0.89)	***−0.07 (0.97)***	−1.56 (1.01)	−1.69 (0.91)	−1.52 (0.52)	***−0.00 (0.52)***	−1.49 (0.57)	−1.50 (0.50)	−1.51 (0.29)	***0.03 (0.30)***	−1.47 (0.32)	−1.50 (0.30)	−1.5
a_3 (SD)	−1.67 (0.87)	***0.00 (1.00)***	−1.51 (0.96)	−1.68 (0.91)	−1.49 (0.51)	***−0.04 (0.54)***	−1.50 (0.51)	−1.49 (0.53)	−1.48 (0.29)	***−0.09 (0.31)***	−1.48 (0.30)	−1.48 (0.31)	−1.5
a_1e_1 (SD)	***0.08 (0.46)***	4.57 (0.72)	4.32 (1.03)	***−0.01 (0.51)***	***0.02 (0.21)***	4.27 (0.51)	4.38 (0.56)	***0.00 (0.25)***	***0.01 (0.13)***	4.30 (0.30)	4.46 (0.32)	***−0.01 (0.15)***	4.5
a_2e_1 (SD)	***−0.03 (0.48)***	−1.64 (0.82)	−1.49 (0.89)	***0.04 (0.46)***	***−0.00 (0.23)***	−1.50 (0.50)	−1.47 (0.52)	***0.01 (0.22)***	***−0.01 (0.13)***	−1.52 (0.29)	−1.51 (0.31)	***0.01 (0.14)***	−1.5
a_3e_1 (SD)	***−0.02 (0.42)***	−1.51 (0.88)	−1.41 (0.83)	***−0.04 (0.48)***	***−0.00 (0.19)***	−1.39 (0.51)	−1.44 (0.50)	***−0.02 (0.23)***	***0.00 (0.12)***	−1.37 (0.28)	−1.47 (0.29)	***−0.01 (0.14)***	−1.5
a_1e_2 (SD)	***−0.08 (0.46)***	−4.58 (0.73)	−4.31 (1.03)	***0.01 (0.51)***	***−0.01 (0.21)***	−4.37 (0.53)	−4.33 (0.55)	***−0.00 (0.25)***	***−0.00 (0.13)***	−4.49 (0.32)	−4.39 (0.31)	***0.00 (0.15)***	−4.5
a_2e_2 (SD)	***0.03 (0.48)***	1.66 (0.82)	1.47 (0.89)	***−0.04 (0.46)***	***0.00 (0.23)***	1.48 (0.50)	1.45 (0.52)	***−0.01 (0.22)***	***0.01 (0.13)***	1.45 (0.29)	1.46 (0.30)	***0.00 (0.14)***	1.5
a_3e_2 (SD)	***0.02 (0.42)***	1.49 (0.88)	1.40 (0.84)	***0.04 (0.48)***	***0.00 (0.19)***	1.49 (0.51)	1.43 (0.50)	***0.02 (0.23)***	***−0.00 (0.11)***	1.57 (0.29)	1.45 (0.28)	***0.00 (0.14)***	1.5

Notes: a_1-a_3 (SD) denote estimates and standard deviations of additive effects for parent 1 to parent 3; a_1e_1-a_3e_2 (SD) denote estimates and standard deviations for the additive effects and additive-by-environment interaction effects for parent 1 to parent 3; Par indicates the true values of the simulated parameters, with parameter values equal to zero denoted in bold italics. Due to space limitations only the estimated effects of QTLs in the first three parents are shown (see Extended Table 3 for more details), Other parameter definitions are the same as those in Table 2.

Table 5 Estimates of effects due to additive–additive epistasis and its interaction with environment of QTL using QTLNetwork-MP in four-way pure-line MAGIC population (population size: 500).

Scenario	$h^2 = 0.25$			$h^2 = 0.50$			$h^2 = 0.75$			Par
	EQ1	EQ2	EQ3	EQ1	EQ2	EQ3	EQ1	EQ2	EQ3	
RC (%)	3.22	3.22	6.45	6.45	6.45	12.90	9.68	9.68	19.35	—
aa_{11} (SD)	10.05 (1.15)	***−0.20 (1.82)***	9.01 (1.53)	8.82 (0.84)	***−0.07 (0.96)***	8.80 (0.89)	8.79 (0.54)	***−0.04 (0.54)***	8.87 (0.53)	9
aa_{12} (SD)	−3.65 (1.71)	***0.15 (1.88)***	−2.94 (1.68)	−2.90 (0.97)	***−0.03 (0.96)***	−2.94 (0.91)	−2.95 (0.57)	***−0.04 (0.54)***	−2.95 (0.53)	−3
aa_{13} (SD)	−3.02 (1.94)	***−0.09 (1.79)***	−2.96 (1.70)	−2.99 (0.97)	***0.10 (0.94)***	−2.95 (1.02)	−2.94 (0.56)	***0.08 (0.52)***	−3.02 (0.58)	−3
$aa_{11}e_1$ (SD)	***0.03 (0.69)***	9.00 (1.37)	7.41 (1.59)	***0.00 (0.27)***	8.35 (0.88)	8.19 (0.91)	***0.00 (0.19)***	8.70 (0.56)	8.63 (0.53)	9
$aa_{12}e_1$ (SD)	***0.13 (0.68)***	−2.96 (1.48)	−2.56 (1.34)	***−0.00 (0.27)***	−2.74 (0.85)	−2.77 (0.86)	***0.00 (0.17)***	−2.83 (0.51)	−2.92 (0.51)	−3
$aa_{13}e_1$ (SD)	***−0.08 (0.67)***	−3.08 (1.75)	−2.49 (1.35)	***−0.00 (0.27)***	−2.80 (0.89)	−2.68 (0.87)	***−0.00 (0.18)***	−2.96 (0.54)	−2.80 (0.53)	−3
$aa_{11}e_2$ (SD)	***−0.02 (0.68)***	−8.99 (1.37)	−7.42 (1.60)	***−0.00 (0.27)***	−8.29 (0.87)	−8.19 (0.92)	***0.01 (0.19)***	−8.60 (0.55)	−8.65 (0.54)	−9
$aa_{12}e_2$ (SD)	***−0.13 (0.68)***	2.95 (1.49)	2.52 (1.33)	***0.00 (0.27)***	2.78 (0.85)	2.78 (0.87)	***−0.02 (0.17)***	2.89 (0.52)	2.93 (0.52)	3
$aa_{13}e_2$ (SD)	***0.08 (0.67)***	3.06 (1.75)	2.53 (1.35)	***0.00 (0.27)***	2.75 (0.89)	2.77 (0.87)	***0.01 (0.18)***	2.87 (0.54)	2.95 (0.53)	3

Notes: aa_{11}-aa_{13} (SD) indicate estimates and standard deviations for the epistatic effects between parent 1 and parents 1 to 3; $aa_{11}e_1$-$aa_{13}e_2$ (SD) represents estimate and standard deviations for the epistatic effects and their interaction with the environment between parent 1 and parents 1 to 3. Par indicates the true values of the simulated parameters, with parameter values equal to zero denoted in bold italics. Due to space limitations, only the estimated interaction effects between the QTL in the first parent and the QTL in the first three parents are displayed (see Extended Table 4 for more details), Other parameters definitions are consistent with those in Table 3.

Table 6 Estimates of QTL effects on cowpea seed size trait.

QTL (cM)	*qSS6*		*qSS7*		*qSS8*	
Parent	**Estimate**	***P* value**	**Estimate**	***P* value**	**Estimate**	***P* value**
IT89KD-288	2.41	2.01E-7	−0.81	4.28E-2	−0.80	5.67E-02
IT84S-2049	−0.46	3.21E-1	0.29	4.71E-1	−1.12	8.13E-3
CB27	−1.14	1.42E-02	1.29	1.27E-03	0.22	5.94E-01
IT82E-18	−1.47	1.54E-03	−0.39	3.29E-01	3.24	1.75E-14
Suvita 2	0.87	5.96E-02	−0.046	9.09E-01	−1.36	1.29E-03
IT00K-1263	−1.42	2.22E-03	1.36	6.71E-04	3.02	9.18E-13
IT84S-2246	0.64	1.64E-01	−0.64	1.12E-01	−1.04	1.34E-02
IT93K-503-1	0.55	2.32E-01	−1.06	8.27E-03	−2.16	3.28E-07

Notes: *qSS6* is located at 80.0 cM on linkage group 6, with a support interval of 76.5–80.3 cM. *qSS7* is located at 23.8 cM on linkage group 7, with a support interval of 23.3–24.1 cM. *qSS8* is located at 71.6 cM on linkage group 8, with a support interval of 71.5–71.7 cM.

integration of multi-omics datasets and advanced machine learning approaches, so as to enhance its efficacy for uncovering the intricate genetic architecture of complex traits. For modern high-density sequencing datasets, two-dimensional epistatic scanning inherently scales at $O(M^2)$, which poses a substantial computational burden. Therefore, we strongly recommend utilizing a bin-marker strategy to compress the number of effective markers (M), restricting the search space to a computationally feasible scale. On the other hand, for extreme cases—such as an 8-parent MAGIC population with tens of thousands of bin markers and a sample size of several thousand lines—a full 2D scan on a high-performance computing cluster can still require several days. To address this, the computational efficiency of the software can be boosted by some techniques, such as bitwise computation and GPU acceleration, making it suitable for analyzing large-scale, high-density datasets involving many parents. It should be noted that the proposed model does not take into account dominance effects, as these variance components are mathematically unfeasible to estimate within the pure-line (homozygous inbred) MAGIC populations targeted in this study. However, expanding this orthogonal framework to accommodate heterozygous multi-parent designs (e.g. F2-based crosses) represents a critical future direction. While introducing multi-parent dominance exponentially increases the complexity of genotypic combinations and orthogonal effect matrices, this theoretical expansion will be vital for investigating complex phenomena such as heterosis.

Acknowledgements

This work was supported by the Major Project of Natural Science Foundation of Zhejiang Province (LD25C130001) and the grants from the China National Tobacco Company [110202101038 (JY-15)], Yunnan Tobacco Company (2022530000241009, 2024530000241003, 2025530000241005).

Author contributions

Tian-Neng Zhu (Conceive the study, Draft the manuscript, Developed algorithm, Conducted simulation and software design, Contributed to the manuscript and approved the submitted version), Zhijun Tong (Conceive the study, Draft the manuscript, Obtained the grants to support and manage the research, Contributed to the manuscript and approved the submitted version), Hai-Ming Xu (Conceive the study, Draft the manuscript, Obtained the grants to support and manage the research, Reviewed and revised the manuscript, Contributed to the manuscript and approved the submitted version), Bingguang Xiao (Obtained the grants to support and manage the research, Contributed to the manuscript and approved the submitted version), Mingzhe Suo (Undertook program test and data analysis, Contributed to the manuscript and approved the submitted version), Mingzi Xu (Undertook program test and data analysis, Contributed to the manuscript and approved the submitted version), Feng Lin (Undertook program test and data analysis, Contributed to the manuscript and approved the submitted version), Asad Ullah (Reviewed and revised the manuscript, Contributed to the manuscript and approved the submitted version), Honghe Zhan (Reviewed and revised the manuscript, Contributed to the manuscript and approved the submitted version)

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Conflict of interests

None.

Funding

This work was supported by the Major Project of Natural Science Foundation of Zhejiang Province (LD25C130001) and the grants from the China National Tobacco Company [110202101038 (JY-15)], Yunnan Tobacco Company (2022530000241009, 2024530000241003,2025530000241005).

Data availability

Public genetic datasets used in this study can be freely downloaded from the following URLs.

The rice dataset: data are publicly available at: https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Ftpj.13827&file=tpj13827-sup-0004-DataS1.xlsx.

The proposed algorithm has been realized in C++ called QTLNetwork-MP and available at https://github.com/Zhutn/QTLNetwork-MP.

References

1. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989;**121**:185–99.
2. Zeng Z-B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci* 1993;**90**:10972–6.
3. Zeng Z-B. Precision mapping of quantitative trait loci. *Genetics* 1994;**136**:1457–68.
4. Yang J, Zhu J, Williams RW. Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 2007;**23**:1527–36.
5. Yang J, Hu C, Hu H *et al.* QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 2008;**24**:721–3. https://doi.org/10.1093/bioinformatics/btm494
6. Huang BE, Verbyla KL, Verbyla AP *et al.* MAGIC populations in crops: current status and future prospects. *Theor Appl Genet* 2015;**128**:999–1017. https://doi.org/10.1007/s00122-015-2506-0
7. Mott R, Talbot CJ, Turri MG *et al.* A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci* 2000;**97**:12649–54.
8. Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* 2007;**12**:57–63.
9. Huang BE, George AW, Forrest KL *et al.* A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnol J* 2012;**10**:826–39. https://doi.org/10.1111/j.1467-7652.2012.00702.x
10. Mackay IJ, Bansept-Basler P, Barber T *et al.* An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. *G3 (Bethesda)* 2014;**4**:1603–10. https://doi.org/10.1534/g3.114.012963
11. Sannemann W, Huang BE, Mathew B *et al.* Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Mol Breed* 2015;**35**:86.
12. Pascual L, Desplat N, Huang BE *et al.* Potential of a tomato MAGIC population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era. *Plant Biotechnol J* 2015;**13**:565–77. https://doi.org/10.1111/pbi.12282
13. Dell'Acqua M, Gatti DM, Pea G *et al.* Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biol* 2015;**16**:167. https://doi.org/10.1186/s13059-015-0716-z
14. Liller CB, Walla A, Boer MP *et al.* Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. *Theor Appl Genet* 2017;**130**:269–81. https://doi.org/10.1007/s00122-016-2807-y
15. Raghavan C, Mauleon R, Lacorte V *et al.* Approaches in characterizing genetic structure and mapping in a rice multiparental population. *G3 (Bethesda)* 2017;**7**:1721–30. https://doi.org/10.1534/g3.117.042101
16. Huynh BL, Ehlers JD, Huang BE *et al.* A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (Vigna unguiculata L. Walp.). *Plant J* 2018;**93**:1129–42. https://doi.org/10.1111/tpj.13827
17. Kumar N, Boatwright JL, Brenton ZW *et al.* Development and characterization of a sorghum multi-parent advanced generation inter-cross (MAGIC) population for capturing diversity among seed parent gene pool. *G3: Genes, Genomes, Genetics* 2023;**13**:jkad037. https://doi.org/10.1093/g3journal/jkad037
18. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 1992;**69**:315–24.
19. Huang BE, George AW. R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* 2011;**27**:727–9.
20. Broman KW, Wu H, Sen Ś *et al.* R/QTL: QTL mapping in experimental crosses. *Bioinformatics* 2003;**19**:889–90.
21. Broman KW, Gatti DM, Simecek P *et al.* R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* 2019;**211**:495–502. https://doi.org/10.1534/genetics.118.301595
22. Zhang L, Meng L, Wang J. Linkage analysis and integrated software GAPL for pure-line populations derived from four-way and eight-way crosses. *Crop J* 2019;**7**:283–93.
23. Zhang S, Meng L, Wang J *et al.* Background controlled QTL mapping in pure-line genetic populations derived from four-way crosses. *Heredity* 2017;**119**:256–64.
24. Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 1954;**39**:859.
25. Li H, Ye G, Wang J. A modified algorithm for the improvement of composite interval mapping. *Genetics* 2007;**175**:361–74.
26. Wang D, Zhu J, Li Z *et al.* Mapping QTLs with epistatic effects and QTL× environment interactions by mixed linear model approaches. *Theor Appl Genet* 1999;**99**:1255–64.
27. Wang C, Rutledge J, Gianola D. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet Sel Evol* 1994;**26**:91–115.
28. Hedayat A, Wallis WD. Hadamard matrices and their applications. *Ann Stat* 1978;**6:**1184–1238.
29. Searle SR, Casella G, McCulloch CE. Variance components. *Wiley Series in Probability and Statistics* 1992.
30. Doerge RW, Churchill G. Permutation tests for multiple loci affecting a quantitative character. *Genetics* 1996;**142**:285–94.
31. Zhu J. Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics* 1995;**141**:1633–9.
32. Lonardi S *et al.* Assembly of eleven pseudomolecules representing the cowpea genome sequence. *Plant Animal Genome* 2017;**25**:P0688.
33. Chen Z-L, Meng JM, Cao Y *et al.* A high-speed search engine pLink 2 with systematic evaluation for proteome-scale identification of cross-linked peptides. *Nat Commun* 2019;**10**:3404. https://doi.org/10.1038/s41467-019-11337-z

Briefings in Bioinformatics, 2026, **27**, bbag273
https://doi.org/10.1093/bib/bbag273
Problem Solving Protocol