

# Genome-wide dissection of the maize ear genetic architecture using multiple populations

Yingjie Xiao<sup>1\*</sup>, Hao Tong<sup>1\*</sup>, Xiaohong Yang<sup>2</sup>, Shizhong Xu<sup>3</sup>, Qingchun Pan<sup>1</sup>, Feng Qiao<sup>1</sup>, Mohammad Sharif Raihan<sup>1</sup>, Yun Luo<sup>1</sup>, Haijun Liu<sup>1</sup>, Xuehai Zhang<sup>1</sup>, Ning Yang<sup>1</sup>, Xiaqing Wang<sup>1</sup>, Min Deng<sup>1</sup>, Minliang Jin<sup>1</sup>, Lijun Zhao<sup>1</sup>, Xin Luo<sup>1</sup>, Yang Zhou<sup>1</sup>, Xiang Li<sup>1</sup>, Jie Liu<sup>1</sup>, Wei Zhan<sup>1</sup>, Nannan Liu<sup>1</sup>, Hong Wang<sup>1</sup>, Gengshen Chen<sup>1</sup>, Ye Cai<sup>2</sup>, Gen Xu<sup>2</sup>, Weidong Wang<sup>2</sup>, Debo Zheng<sup>2</sup> and Jianbing Yan<sup>1</sup>

<sup>1</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; <sup>2</sup>National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China; <sup>3</sup>Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA

## Summary

Author for correspondence:  
Jianbing Yan  
Tel: +86 27 87280110  
Email: yjianbing@mail.hzau.edu.cn

Received: 26 August 2015  
Accepted: 23 November 2015

*New Phytologist* (2015)  
doi: 10.1111/nph.13814

**Key words:** genome-wide association study (GWAS), joint linkage mapping, maize (*Zea mays*), multi-parent population, quantitative trait loci (QTLs), yield traits.

- Improvement of grain yield is an essential long-term goal of maize (*Zea mays*) breeding to meet continual and increasing food demands worldwide, but the genetic basis remains unclear.
- We used 10 different recombination inbred line (RIL) populations genotyped with high-density markers and phenotyped in multiple environments to dissect the genetic architecture of maize ear traits.
- Three methods were used to map the quantitative trait loci (QTLs) affecting ear traits. We found 17–34 minor- or moderate-effect loci that influence ear traits, with little epistasis and environmental interactions, totally accounting for 55.4–82% of the phenotypic variation. Four novel QTLs were validated and fine mapped using candidate gene association analysis, expression QTL analysis and heterogeneous inbred family validation.
- The combination of multiple different populations is a flexible and manageable way to collaboratively integrate widely available genetic resources, thereby boosting the statistical power of QTL discovery for important traits in agricultural crops, ultimately facilitating breeding programs.

## Introduction

Maize (*Zea mays*) was domesticated from its wild relative teosinte by ancient agriculturalists nearly 9000 yr ago (Matsuoka *et al.*, 2002). During this process, dramatic changes in the female inflorescence or ear made maize an important staple crop worldwide for food, feed and fuel. Maize grain yield has increased eight-fold in the past century (Duvick, 2005), with the majority of the gain being attributed to selection and hybrid breeding. Ear length, ear row number, ear weight and cob weight are important component traits of maize yield. The clarification of the genetic architecture of ear traits would allow breeders to more efficiently design breeding schemes to manipulate these traits.

The construction of a mapping population is a prerequisite for the identification of quantitative trait loci (QTLs) that influence a target trait. In plants, many different types of biparental population can be derived from the initial cross of two parental lines. Among these, the recombination inbred line (RIL) population is widely used to identify QTLs in many crop species (Yano & Tuberosa, 2009). However, the fine mapping and cloning of

genes underlying QTLs are resource- and time-consuming processes because large populations are required to achieve a sufficient map resolution. In addition, the complexity of the maize genome, with its abundance of transposons and repetitive sequences, further slows the progress of fine mapping (Salvi & Tuberosa, 2005; Mackay *et al.*, 2009; Schnable *et al.*, 2009). By contrast, genome-wide association study (GWAS) using genetically diverse inbred lines provides a tool that can fine map QTLs by taking advantage of historical recombinations (Flint-Garcia *et al.*, 2003). However, GWAS is often limited by the inherent population structure and has very low power for the detection of low-frequency variants (Flint-Garcia *et al.*, 2005; Yu *et al.*, 2006; Zhang *et al.*, 2010).

In recent years, the multi-parent design, which was originally used on a heterogeneous stock of mice (Valdar *et al.*, 2006), has emerged as an efficient way to identify QTLs for agriculturally important traits in plant species (Cavanagh *et al.*, 2008; Buckler *et al.*, 2009; Huang *et al.*, 2011). A classic example of multi-parent design is the maize nested association mapping (NAM) population (Yu *et al.*, 2008), the large diversity and clear population structure of which enabled it to unravel the genetic architecture of a wide range of complex traits (Buckler *et al.*, 2009;

\*These authors contributed equally to this work.

Brown *et al.*, 2011; Kump *et al.*, 2011; Tian *et al.*, 2011; Hung *et al.*, 2012; Peiffer *et al.*, 2014). However, the extremely unbalanced parental contributions might cause some statistical issues with regard to the low power of QTL detection. Conversely, the multi-parent advanced generation intercross (MAGIC) population has a balanced contribution from all founders (Kover *et al.*, 2009; Bandillo *et al.*, 2013). Recently, a maize MAGIC population was established with eight diverse founder lines that provided a useful resource for an understanding of the genetic basis of quantitative traits (Dell'Acqua *et al.*, 2015). Unfortunately, the development of NAM and MAGIC populations requires extensive field and laboratory effort, which greatly limits the application of the multi-parent population design to other plant species or for traits that have shown little variation in the currently available NAM and MAGIC populations.

In this study, we present a new design for a multi-parent population consisting of 10 independent RIL populations. This design provides researchers with a flexible and cost-effective method to combine publicly available genetic resources for the dissection of complex traits in plants collaboratively, instead of independently, not only potentially boosting statistical power, but also avoiding extensive efforts in *de novo* population development, such as NAM or MAGIC. In this study, we combined 10 RIL populations and three different, but complementary, statistical methods to identify QTLs or trait-associated single nucleotide polymorphisms (SNPs) for maize ear traits. The deep dissection of the mapping results provided insights into the genetic architecture of ear traits and led to informative clues for maize breeding.

## Materials and Methods

### Germplasm, trials and phenotypic data analysis

Ten RIL populations with nearly 200 lines per population were collected (Supporting Information Fig. S1). The 10 RIL populations had 14 parents that originated from the association panel of 508 genetically diverse maize inbred lines (AM508) reported previously (Yang *et al.*, 2011). The 10 RIL populations and the AM508 panel were planted in eight trials during the summer and winter of 2011 and 2012 in five locations with one random block replication per location in China. The association panel (AM508) and seven RIL populations (B73 × BY804, KUI3 × B77, K22 × CI7, DAN340 × K22, ZHENG58 × SK, YU87-1 × BK and ZONG3 × YU87-1) were planted in all eight trials (i.e. Hubei, Chongqing, Henan, Yunnan and Hainan during 2011 and 2012), whereas the remaining three RIL populations (DE3 × BY815, K22 × BY815 and BY815 × KUI3) were planted in four trials (i.e. Chongqing, Hubei, Henan and Yunnan during 2012) because of insufficient seeds for sowing in 2011 (Fig. S2). At least five well-pollinated ears in each row were harvested for phenotypic measurements of four ear traits by standard procedures, including ear length (EL), ear row number (ERN), ear weight (EW) and cob weight (CW). Some RILs had abnormal ear development and were discarded from the analysis. Each RIL population consisted of 165–207 RILs, resulting in 1887 RILs for further studies (Table S1).

When treating the location or year as a single environment, analysis of variance (ANOVA) for each trait was performed to evaluate the effect of genotype and environment on phenotypic variance in R function 'LM' (R Core Team, 2012). The line mean-based broad-sense heritability for each trait was calculated as:  $H^2 = \delta_g^2 / (\delta_g^2 + \delta_c^2/n)$ , where  $\delta_g^2$  is the genetic variance,  $\delta_c^2$  is the residual variance and  $n$  is the number of environments. The estimates of  $\delta_g^2$  and  $\delta_c^2$  were obtained by the mixed linear model, treating genotype and environment as random effects. To eliminate the influence of environmental effects on phenotypic variation ( $P < 0.001$ , ANOVA; Table S2), the best linear unbiased predictor (BLUP) value for each line within each RIL population was calculated across all environments using the mixed linear model with the fitting of both genotype and environment as random effects in the R package 'LME4' (R Core Team, 2012). The BLUP values for the RILs of 10 populations were combined to facilitate the following analyses, such as phenotypic description statistics, Pearson correlations and QTL analysis for the four ear traits.

### SNP genotyping, imputation and projection

The AM508 panel and the 10 RIL populations were characterized with 56 110 SNPs by an Illumina MaizeSNP50 BeadChip covering 19 540 maize genes (Ganal *et al.*, 2011). A total of 11 360–15 285 SNPs were polymorphic within each RIL population. A very high-density genetic map for each RIL population was constructed by our laboratory (Pan *et al.*, 2015), which captured 1979–3071 (Table S1) genetic blocks (a genomic region in which no recombination exists) per population, with an average block size of 859 kb (Fig. S3). For the regions in which the physical positions of the SNPs were not collinear with the genetic positions, we corrected the physical positions by the linear interpolation method according to the physical positions of the flanking collinear SNPs. The linear interpolation method was performed as:  $p = p_1 + (p_2 - p_1) \frac{g - g_1}{g_2 - g_1}$ , where  $g_1$  and  $g_2$  are the genetic positions of the flanking collinear markers,  $g$  is the genetic position of the in-collinear marker, and  $p_1$  and  $p_2$  are the physical positions of the flanking collinear markers. Considering the fact that our genetic maps captured the majority of recombination events that existed in the development of RIL populations, we directly imputed the missing marker genotypes using the closest flanking non-missing markers.

To facilitate joint linkage mapping (JLM) and GWAS, we projected the 1.03 million SNP genotypes of the 14 parental lines obtained by RNA-seq (Fu *et al.*, 2013; Li *et al.*, 2013; Wen *et al.*, 2014; Yang *et al.*, 2014) onto 1887 offspring RILs using a two-step imputation strategy. We first separately projected high-density SNPs from two parents onto offspring RILs based on the linkage map for each RIL population, similar to the aforementioned imputation procedure. We then mapped the projected genotypes of RILs to base pairs according to the parental genotypes in each RIL population, and merged the resulting 10 datasets together. Overall, there were 14 613 genetic blocks available for JLM and 185 212 blocks available for GWAS, where the additional blocks in GWAS were identified using the historical

recombination in the founders. The average block length implicitly indicated that GWAS achieved the highest resolution, followed by JLM and separate linkage mapping (SLM) (Fig. S3). Based on the minor allele frequency (MAF) value of each SNP in the merged dataset, the homozygote of the major allele was numerically coded '0', the homozygote of the minor allele was coded '2' and the heterozygote was coded '1'. The numerically coded genotypes were eventually used in the data analysis.

## SLM

Within each RIL population, a composite interval mapping procedure (Zeng, 1994) was used to perform SLM with the software Windows QTL CARTOGRAPHER v.2.5 (Wang *et al.*, 2012). We performed 500 permutations for each trait within each population to determine the threshold of the logarithm of odds (LOD) score for the significance test, and the resulting LOD score threshold ranged from 2.6 to 3.2 ( $\alpha = 0.05$ ). For simplicity, we chose the 3.0 LOD score as the global cut-off point. To avoid an overestimation of the number of QTLs, we declared adjacent peaks with nearby genetic positions ( $\leq 10$  cM) and identical effect directions as one QTL. A QTL support interval was defined as the two-LOD drop position ranging from the QTL peak. If the support intervals of QTLs detected by different RIL populations overlapped, the set of QTLs was integrated to a 'consensus QTL', the support interval of which was the union conjunction of overlapping QTL support intervals. If a QTL detected in one population could not be overlapped with QTLs in any population, the QTL was considered as a 'unique QTL'. To assess the influence of sample size and unbalanced environments on QTL detection with SLM, we performed a Monte Carlo resampling analysis based on empirical data to evaluate the magnitude of possible bias. The details are described in Methods S1. With the results of SLM, we performed meta-QTL analysis to integrate QTL information for 10 RIL populations, the technical details of which are described in Methods S2.

## JLM and GWAS

Combining the 10 RIL populations, JLM and GWAS were performed to further dissect the genetic determinants of ear traits. For JLM, a linear mixed model was built and the restricted maximum likelihood (REML) was used to test the significance of each recombination block, where the population mean and intercept term were fixed effects, and marker and polygenic effects were treated as random effects. The tested block contained, in total, 14 additive effects corresponding to the parental alleles of 14 founders. The covariance structure of the polygenic effects was inferred from the marker-inferred kinship matrix. We used a permutation test of 500 permuted samples to determine the threshold of likelihood ratio test (LRT) scores (Chen & Storey, 2006). For simplicity, we defined the physical position range delimited by the JLM threshold as JLM QTL support regions. Furthermore, we performed GWAS following the stepwise regression and resample methods reported previously (Valdar *et al.*, 2009; Tian *et al.*, 2011), with a minor modification.

Finally, a backward regression was employed to reduce the redundancy of the significant SNPs for each trait, leaving a set of candidate SNPs for gene annotation and validation. The permutation tests were used to determine the significance threshold for genome-wide SNPs in the stepwise regression, resample methods and final backward regression ( $\alpha < 0.05$ ). More details of JLM and GWAS are described in Methods S3 and S4, and the R scripts and the genotype and phenotype data of the 10 RIL populations are publicly available at the permanent website (<http://www.maizego.org/Resources.html>).

## Epistasis and QTL-by-environment interactions

Markers only significant in JLM or GWAS were used in the interaction analysis separately, because JLM employed parental allelic genotype data, whereas GWAS used biallelic genotype data. For simplicity, all the heterozygous genotypes ( $< 4\%$ ) were assigned as missing values to ensure that only homozygous allelic interactions were estimated and tested. A linear regression model including QTL main effects, family effects and pair-wise QTL interactions was used to estimate the epistatic effects. We tested all pair-wise interactions and used  $P < 0.01$  to declare significance.

The genotypic data used in the QTL-by-environment analysis were defined in a similar manner to the data used for epistatic analysis. The phenotypic data were not normalized within populations, which enabled an estimation of precise QTL-by-environment interaction. We used a linear mixed model including the family effects as fixed effects, and QTL, environment, QTL-by-environment and family-by-environment effects as random effects. The null model is that which excludes the QTL-by-environment effect. The REML method was used to estimate the variance components of QTL and QTL-by-environment interactions. We declared the significance of the QTL-by-environment interaction when the LRT value exceeded the threshold of JLM.

## Genetic validation with heterogeneous germplasm resources

The reliability of the identified candidate SNPs should be validated by other genetic populations with heterogeneous backgrounds. Here, we used two different germplasm resources to validate our results. (1) The maize NAM population. Previously, the NAM population detected a number of SNPs associated with ear length and ear row number using joint linkage and GWAS analyses (Brown *et al.*, 2011). If a candidate SNP was located within 1 Mb of any joint QTL peak or GWAS SNP detected in NAM, we defined that this candidate SNP co-localized with the previously reported peak in the maize NAM population. (2) The maize AM508 population. The legitimacy of the candidate SNP was considered to be greatly strengthened if it still showed a significant association with target traits in the maize AM508 population ( $P < 0.05$ ). The GLM model was used to correct the confounding impact of population structure, which has been thoroughly characterized previously (Yang *et al.*, 2011). To test whether there was an enrichment of genetic validation by

AM508, the observed proportion of validated candidate SNPs was compared with a null distribution, obtained by randomly selecting the same number of SNPs across the genome, calculating the validated proportion under the GLM model and repeating the process 500 times ( $P < 0.05$ ).

### Candidate genes of four QTLs for ear length

To further understand the genetic architecture of ear traits, we determined the candidate genes of four QTLs with greater effects or highly repeatable in multiple methods and populations. The regional association analysis was conducted per QTL to narrow down the QTL region and to identify the candidate gene based on the AM508 population. For the QTL detected by JLM or GWAS, the JLM peak block or GWAS candidate SNP was used to extend 500 kb upstream and downstream to infer the association region. For the QTL only detected by SLM, the support interval was directly referred to be the association region. The gene in the locus or closest to the most significant SNP was inferred as the candidate gene. To infer the functional mechanisms of candidate genes on the phenotype (i.e. mediated or not by expression variation), we used the previously published expression QTL (eQTL) data for 28 769 maize genes (Fu *et al.*, 2013) to evaluate the relationship between gene polymorphisms and expression, or between gene expression and phenotypic variations. In order to provide more evidence for the QTL or candidate gene, we obtained a near isogenic line (NIL) population of the target QTL by the heterogeneous inbred family (HIF) strategy, which is widely used in QTL fine mapping (Tuinstra *et al.*, 1997). The HIF-NIL population was employed to validate the existence of QTL using ANOVA.

### Overlap between linkage mapping and GWAS

To evaluate the overlap between QTLs identified via linkage mapping and SNPs identified via GWAS, we calculated the proportion of SNPs falling inside QTL regions detected by SLM and JLM, referred to as the observed proportion. The proportion of random SNPs in the genome falling inside QTL regions was referred to as the expected proportion. We used the 'binom.test' in the R package to test the significance of the difference between observed and expected proportions ( $P < 0.05$ ).

### Simulation studies

In order to appropriately interpret the empirical results showing that the three models had different capacities to detect a specific QTL, we simulated 25 QTLs evenly dispersed in a hypothetical genome with 10 chromosomes with different effect sizes and different directions of the effects, and only 7–11 of the simulated QTLs segregated within each of the 10 simulated biparental populations (Fig. S4). The same three models (SLM, JLM and GWAS) were used to estimate the QTL parameters and to detect QTLs. The simulation was replicated 500 times to evaluate the statistical power of each model. In addition, we built a hypothesis test with the simulated data to estimate the power and false

discovery rate (FDR) for declaring a QTL at low frequency when it was detected by SLM, but not by JLM or GWAS. Details of the simulation studies are described in Methods S5.

### Co-localization of QTLs and inflorescence candidate genes

Most maize inflorescence genes were isolated by transposon and chemical mutagenesis. We collected information on 20 cloned inflorescence genes from a previous study (Brown *et al.*, 2011) and tested these ear trait candidate genes for co-localization with QTLs identified by the three models in the current study. The support intervals of QTLs and 500-kb flanking regions of GWAS SNPs were used to evaluate the overlaps of the QTLs identified here with the candidate genes. To test whether there was a significant enrichment, we compared the number of candidate genes overlapping the QTLs identified in this study with a null distribution, obtained by randomly selecting 20 genes across the maize genome, evaluating the co-localization with the QTLs detected here and repeating the process 1000 times ( $P < 0.05$ ).

## Results

### Genetic and phenotypic diversity of the 10 RIL populations

The 10 independent RIL populations with 14 genetically diverse parental founders (Fig. S1) showed a broad spectrum of genetic divergence of offspring RIL lines, revealed by principal component analysis (Fig. S5). The distributions of ear length, ear row number, ear weight and cob weight were approximately normal for each RIL population, but exhibited clear variations among the 10 RIL populations (Fig. S6). The broad-sense heritability was generally high ( $H^2 = 0.76$ – $0.87$ ; Table 1), which was similar to a previous report on the maize female inflorescence traits (Brown *et al.*, 2011). All ear traits exhibited moderate correlations with each other, except for ear weight, which showed strong correlations with all other traits (Fig. S7; Table S2), indicative of the central role of ear weight in the ear development of maize.

**Table 1** Summary of quantitative trait loci (QTLs) or single nucleotide polymorphisms (SNPs) via three model-based approaches for four ear traits in maize

	Ear length	Ear row number	Ear weight	Cob weight
$H^2$ <sup>a</sup>	0.87	0.86	0.76	0.83
SLM <sup>b</sup>	43/8,26	46/5,36	33/2,29	41/2,36
JLM	41	67	74	61
GWAS <sup>c</sup>	202/32	173/34	103/17	122/24

<sup>a</sup>Average broad-sense heritability among families on the line mean basis.

<sup>b</sup>Total numbers of QTLs detected in 10 recombination inbred line (RIL) populations (before slash), consensus QTL across populations (bold after slash) and unique QTL in specific population (underline after slash).

<sup>c</sup>Number of significant SNPs detected by the GWAS procedure (before slash) and candidate SNPs by final backward regression (after slash).

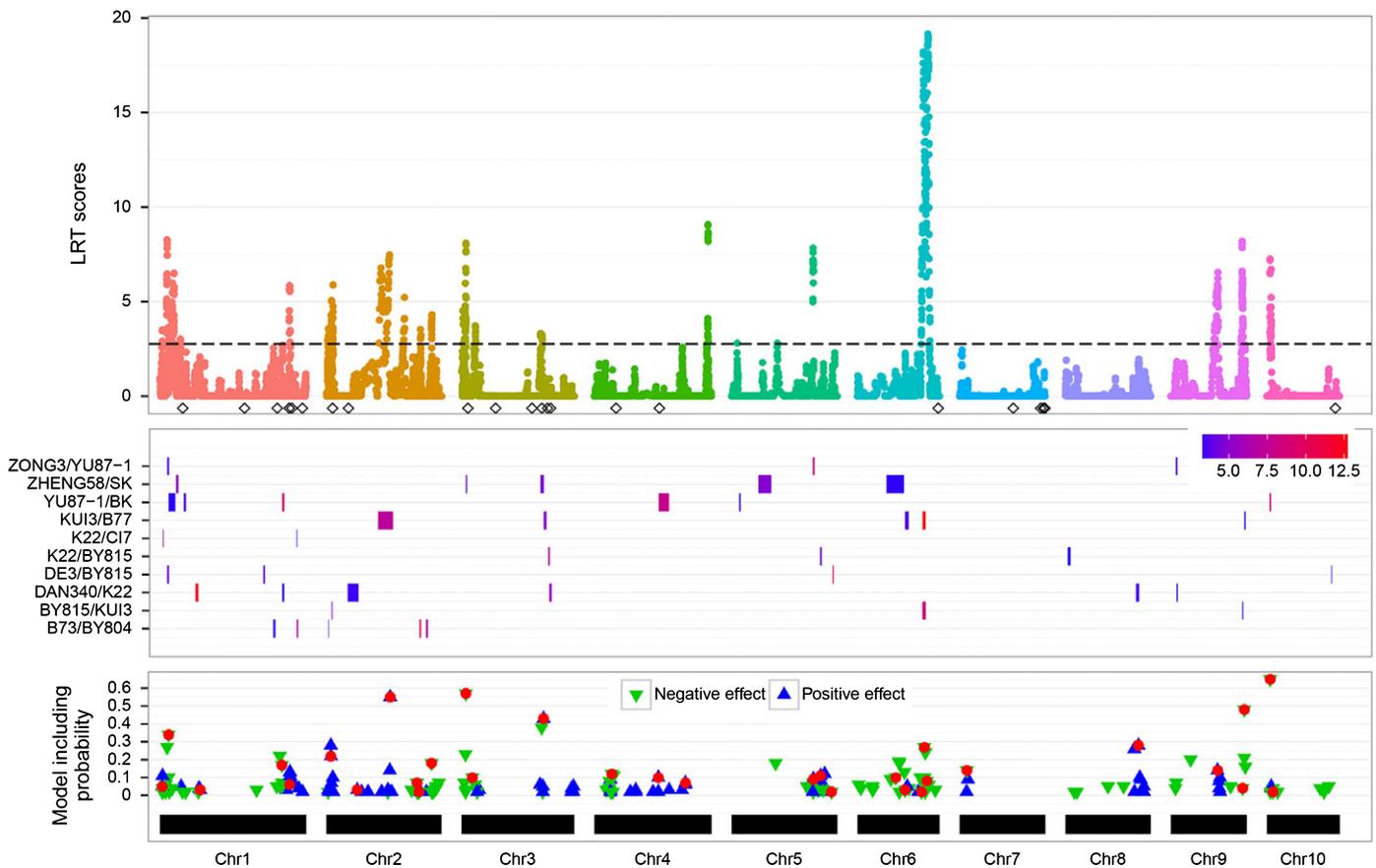
$H^2$ , broad-sense heritability; SLM, separate linkage mapping; JLM, joint linkage mapping; GWAS, genome-wide association study.

### Genetic dissection of ear traits via three models

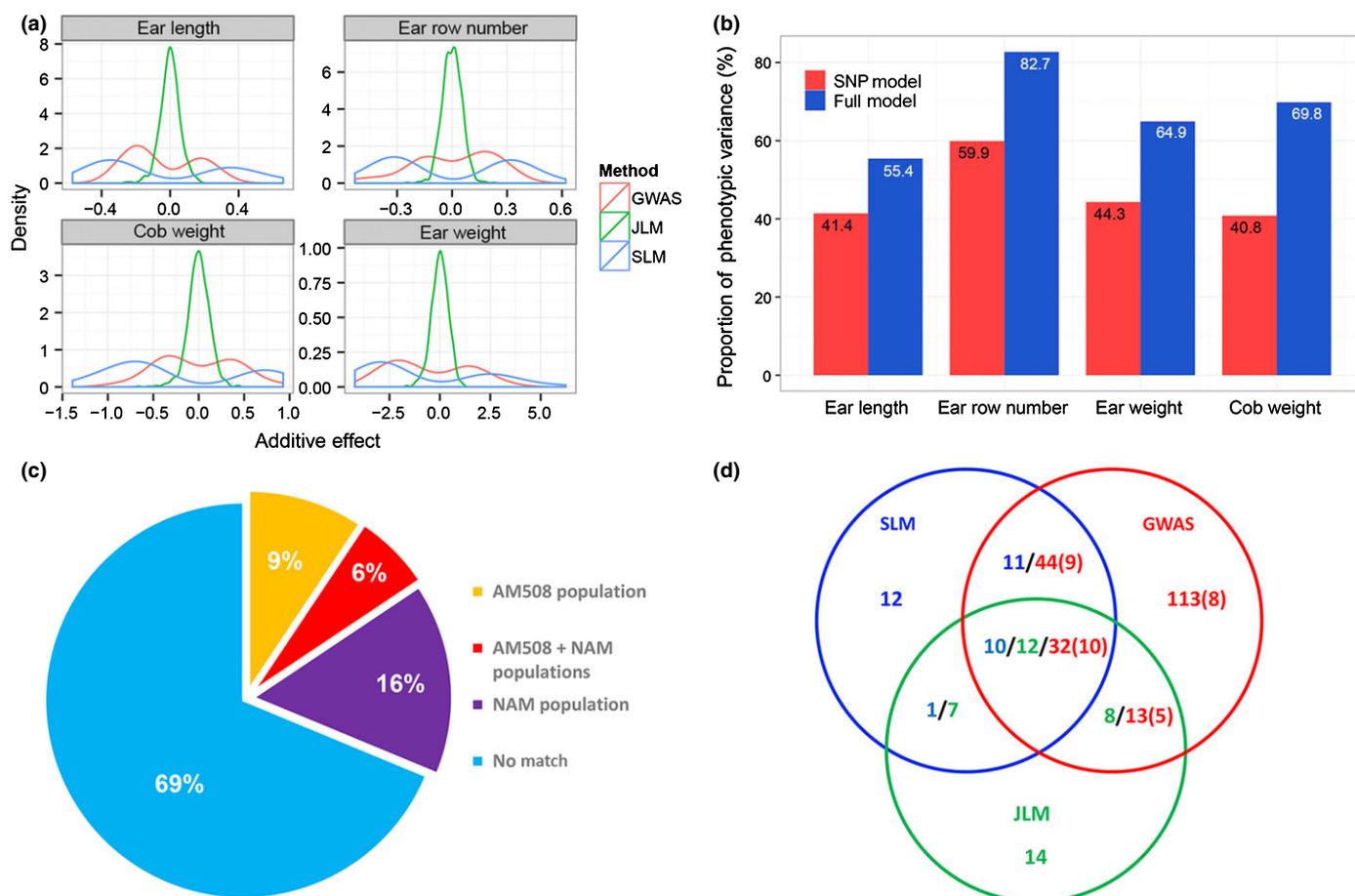
Three model-based approaches were used to systematically dissect the genetic bases of ear traits. In SLM, 33–46 QTLs were identified for each ear trait (Table 1; Figs 1, S8). Detailed information on the QTLs across the whole genome in the 10 RIL populations is given in Fig. S9. Most detected QTLs had small to moderate additive effects (Fig. 2a), whereas 25.2% had effects that could explain > 10% of the phenotypic variance per QTL (Notes S1). According to the physical overlap of the QTL support intervals, we integrated two or more co-localized QTLs detected in different genetic backgrounds into a single consensus QTL, resulting in two to eight consensus QTLs for each trait (Table 1), leaving most QTLs uniquely identified in specific RIL backgrounds (60.5–87.9%). We also performed meta-analysis to integrate the SLM results of the 10 populations, and found that there were 3–12 meta-QTLs for each ear trait, most of which were overlapped with SLM QTLs (58–100%; Notes S2). However, the QTL number detected by SLM was much greater than the number of meta-QTLs (Notes S1), which might be attributed to the heterogeneous backgrounds of

the 10 populations (Fig. S5), and is congruent with the aforementioned finding of large amounts of uniquely detected QTLs within populations. Although it was a bit of unbalanced for the current data structure of the 10 RIL populations in sample size and environment for collecting phenotypes (Table S1; Fig. S2), but the resampling analysis revealed that this situation may lead to a small fraction of type II errors in QTL detection with SLM (i.e. 3.2% or 7.4%), and which is incapable to make a significant statistical bias in the estimation of the proportion of unique QTLs ( $P=0.321$  or  $P=0.15$ ; Fig. S10). In addition, we found that a small fraction of QTLs were shared among traits, possibly explaining the observed weak phenotypic correlations (Fig. S7).

In JLM analysis, 41–74 QTLs were identified for each trait, and these QTLs had small estimated effects with each explaining a small percentage of the phenotypic variance (Table 1; Notes S3; Figs 1, S8, S11). A small proportion of QTLs identified by JLM showed significant QTL  $\times$  environment interactions, and the effects were much smaller than the QTL main effects (Notes S3). Pair-wise epistatic effects between JLM-identified QTLs were tested via linear model analysis. We found that the inclusion of



**Fig. 1** Overview of quantitative trait locus (QTL) results for ear length in maize. Top panel (Manhattan plot): the colored dots show the significance of genome-wide blocks estimated by the joint linkage mapping (JLM) method. Blank diamonds indicate the physical positions of the 20 maize inflorescence genes. Middle panel: colored rectangles indicate separate linkage mapping (SLM) QTL regions across the 10 recombination inbred line (RIL) populations. The color density of the rectangles indicates the magnitude of the logarithm of the odds (LOD) values. Bottom panel: triangles indicate significant single nucleotide polymorphisms (SNPs) identified by genome-wide association study (GWAS), where the blue upward triangles indicate that the minor allele increases ear length relative to the major allele, whereas the green downward triangles indicate the opposite effect. Red dots indicate the candidate SNPs identified by the final backward model. LRT, likelihood ratio test.



**Fig. 2** Effect distribution of quantitative trait loci (QTLs) and significant single nucleotide polymorphisms (SNPs) and cross-validation in maize. (a) Separate linkage mapping (SLM) always shows relatively higher additive effects in both directions than the genome-wide association study (GWAS), followed by joint linkage mapping (JLM). (b) Candidate SNPs jointly explain the majority of the phenotypic variation. The SNP model fits only the SNP effects, whereas the full model fits both family and SNP effects. (c) Cross-validation results for candidate SNPs of ear length using nested association mapping (NAM) and AM508 populations. (d) Overlaps in QTL results identified by the three models. The colored numbers show the QTL counts detected by the different models. The numbers in parentheses are the candidate SNP counts identified by backward regression following GWAS.

epistatic effects in the additive model explained little additional phenotypic variance (2.5–8.4%; Table S3), suggesting that additive effects play more important roles than epistatic effects in the genetic variation of ear traits.

Overall, 122–202 significant SNPs were detected for each trait by GWAS (Table 1; Figs 1, S8). To address the problem of redundancy of the significant SNPs because of the strong linkage disequilibrium among physically close SNPs, we performed a backward regression on the significant SNPs. Eventually, 17–34 significant SNPs – referred to as candidate SNPs – were retained in the model for each trait after backward elimination. Each of the candidate SNPs explained a small fraction of the phenotypic variation, with a maximum of 3.4% for ear length (Notes S4), congruent with the findings of the JLM analysis (Notes S3). However, all candidate SNPs jointly explained the majority of phenotypic variance ( $R^2 = 55.4$ – $82.7\%$ ) and genetic variance ( $R^2 = 63.7$ – $96.2\%$ ) for each trait ('Full model'; Fig. 2b). Similar to QTLs identified by JLM, the candidate SNPs from GWAS rarely exhibited significant SNP  $\times$  environment interactions or

pair-wise epistatic effects on ear traits (Notes S4; Table S3), further confirming that epistasis is unimportant relative to the additive effects.

Furthermore, we found that the exclusion of family effect from the model caused a significant reduction in the explained phenotypic variance ( $R^2 = 40.8$ – $59.9\%$ , 'SNP model'; Fig. 2b), demonstrating the large influence of population structure on the phenotypic variation in the 10 RIL populations. In both JLM and GWAS, we controlled the family effect in the models to reduce false positives. However, the effects of causal QTLs are possibly masked by the confounding of population structure in cases in which QTLs segregate among RIL populations, but not within RIL populations (Flint-Garcia *et al.*, 2005). Further expansion of the 10 populations would enhance the QTL detection power by breaking the connection between QTL distribution and population structure. Fortunately, the design and analysis framework enabled us to easily expand the populations via direct inclusion of existing segregating populations whenever necessary.

## Genetic validation of associations and the determination of candidate genes

We employed two diverse germplasm resources, NAM and AM508 populations, to validate the final set of candidate SNPs. Overall, there were 22% and 32% candidate SNPs for ear length and ear row number validated in NAM (Figs 2c, S12; Notes S5). This inconsistency of QTL detection was probably caused by the very distinct genetic backgrounds of the present 10 RIL populations and the 25 RIL populations of NAM, which have been documented previously (Fu *et al.*, 2013). However, 15–21% of the candidate SNPs for ear traits were validated in the AM508 population, excluding cob weight, in which only 8% of the candidate SNPs were validated (Figs 2c, S12; Notes S5). It was only marginally more significant than random at a statistical level of 0.05 ( $P=0.008$  for ear row number,  $P=0.049$  for ear length and  $P=0.074$  for ear weight; Fig. S13). The AM508 population contains more diverse lines (and thus more historical recombination events) and more low-frequency SNPs than do the 10 RIL populations. This causes a reduction in statistical power for the AM508 population to identify all but the most closely linked SNPs, and only if they have a balanced allele frequency.

To gain further insights into maize ear traits, we attempted to determine the candidate genes of four major QTLs by jointly utilizing multiple approaches (Table S4). For example, in the KUI3/B77 population, a QTL was mapped to the region of 133.3–139.5 Mb on chromosome 6 with a peak LOD of 12, whereas, in the BY815/KUI3 population, the QTL was mapped to the region of 132.9–139.9 Mb with a peak LOD of 9 (Fig. 3a; Notes S1). In addition, an HIF originating from a RIL offspring of the KUI3/B77 population validated this QTL ( $P=4.3 \times 10^{-3}$ ; Fig. 3b). When analyzing the 10 populations simultaneously, JLM found the same QTL with an LRT value equal to 14, whereas GWAS identified a candidate SNP at 137.7 Mb within the QTL region (Fig. 3c; Notes S3, S4). We further conducted a regional association analysis using SNPs from the 500 kb flanking the candidate SNP in the AM508 population (as linkage disequilibrium (LD), extended to 200–500 kb in the 10 populations; Fig. S14). Gene *GRMZM5G864815* is the most likely candidate for the QTL on chromosome 6, based on the physical position of the most significant SNP (Fig. 3d). *GRMZM5G864815*, a homolog of *thiamine pyrophosphokinase 1 (TPK1)* in *Arabidopsis*, is capable of producing thiamine pyrophosphate, which is involved in the major carbohydrate metabolic pathways (Rapala-Kozik *et al.*, 2009) and thus may potentially affect ear development (Table S4). This hypothesis was confirmed by the negative correlation between the expression level of *GRMZM5G864815* and ear length ( $r=-0.16$ ,  $P=1.9 \times 10^{-3}$ ; Fig. 3e). The eQTL and GWAS analyses also revealed that the most significant SNP (i.e. PZE-106080641) simultaneously influenced the expression level of *GRMZM5G864815* ( $P=6.0 \times 10^{-4}$ ,  $n=340$ ; Fig. 3f) and ear length ( $P=3.2 \times 10^{-4}$ ,  $n=339$ ; Fig. 3f). These findings suggested that *GRMZM5G864815* might be a candidate gene for ear length and that the phenotypic difference may be caused by gene transcriptional regulation, but further studies are required

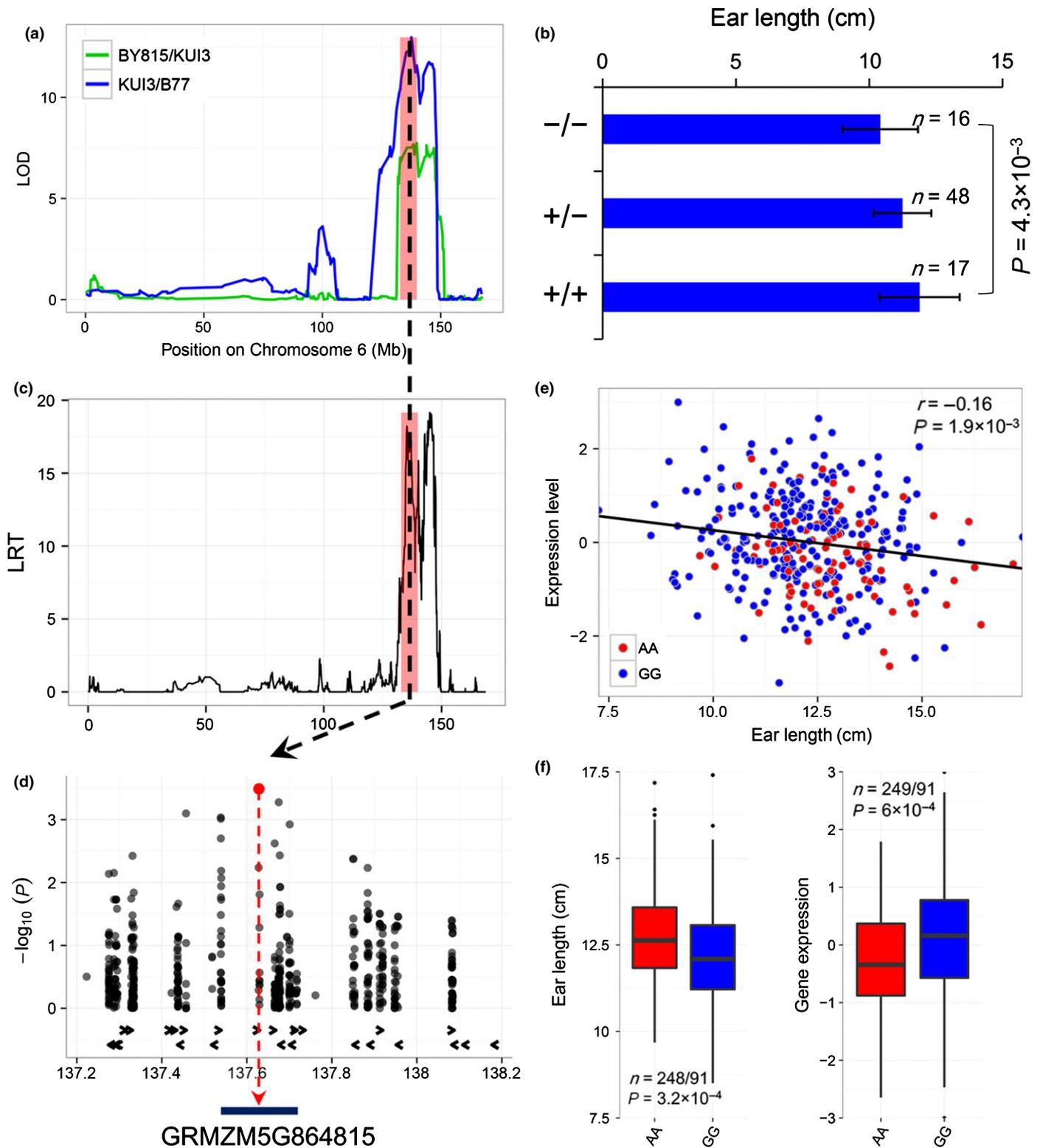
to validate this hypothesis. Three other QTLs with large effects were similarly analyzed and candidate genes were identified (Table S4; Figs S15–S17).

## Model preference for the identification of QTLs with different features

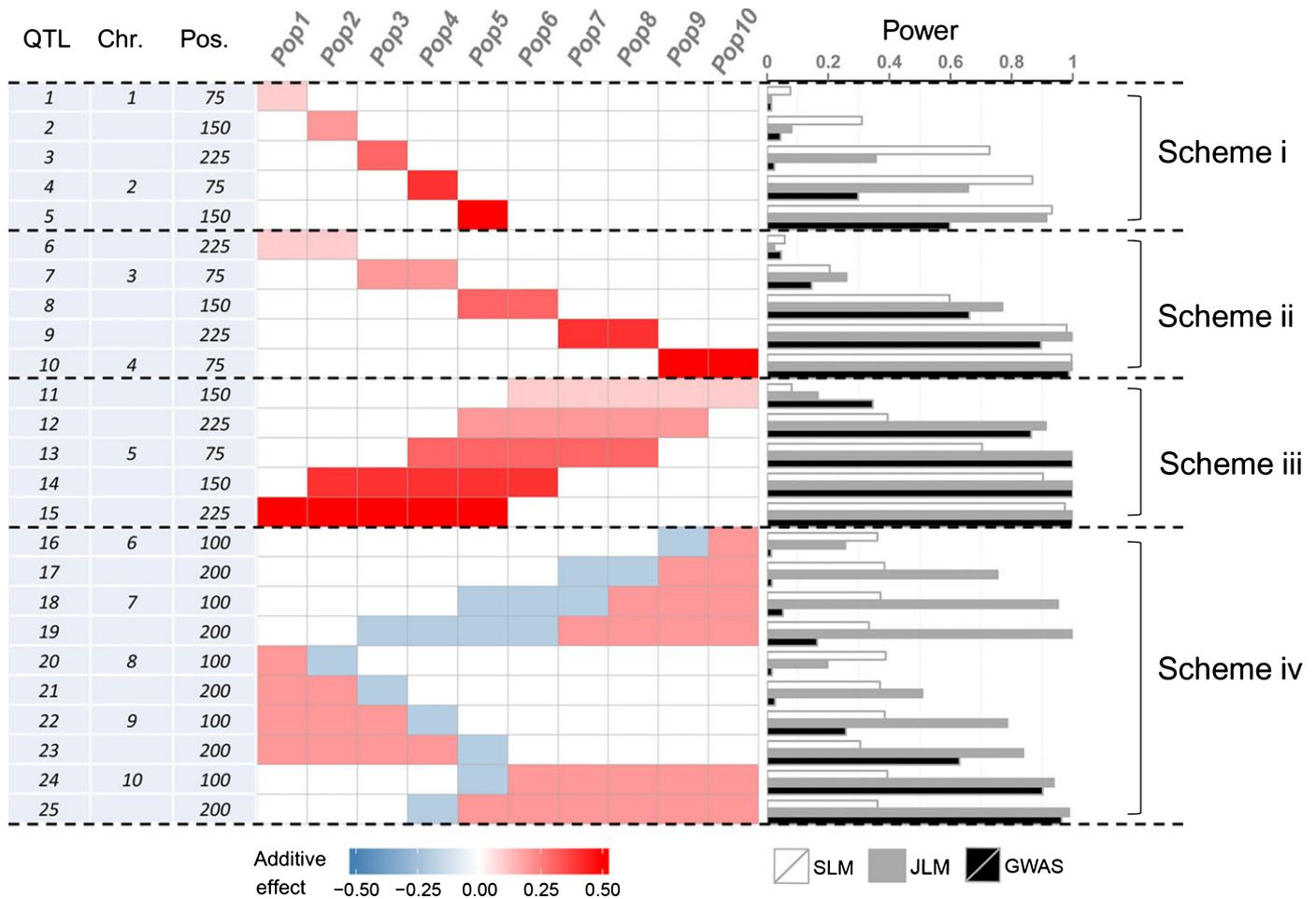
Although the SNPs identified by GWAS were significantly enriched in the QTL regions ( $P<0.001$ ; Fig. S18), there were a considerable number of QTLs or SNPs that were identified by only one model (Figs 2d, S19). Different results of QTL mapping from SLM and JLM have been reported previously in the NAM population, with a plausible explanation of a low frequency of significant SNPs segregating within families (Buckler *et al.*, 2009). However, this difference may also be caused by a lack of coincident segregation of SNPs and QTLs within one family – that is if the SNP detected is not the causal variant for the QTL. This has often been a problem with the use of flanking markers to infer the presence of a QTL, especially in marker-assisted selection. To determine the cause of the inconsistent mapping results, we simulated a series of QTLs across 10 pseudo-chromosomes with varying effect sizes and frequencies of QTLs across different families. The same three statistical methods were used to detect QTLs at  $\alpha<0.05$ . These simulated QTLs were classified into four schemes (Fig. 4): (i) low-frequency QTLs, where all three methods show an increased power as the QTL effect increases (but more so with SLM); (ii) modest-frequency QTLs, where the three methods show similar powers; (iii) high-frequency QTLs, where GWAS has a significantly higher power than JLM, followed by SLM; (iv) QTLs with effects in opposite directions, where JLM has the highest power, followed by GWAS and SLM. The difference in power reflects the difference in the genetic architectures assumed by different methods. In SLM, the QTL model is based on the biallelic QTL assumption within a biparental cross-population. It has the highest power to detect QTLs, as the biallelic markers (i.e. SNP markers) enable QTL alleles to be represented in a perfect way. In JLM, the QTL model assumes that the parents are independent and each parent carries a different QTL allele in the multi-parent population (as the multi-allelic QTL assumption). In GWAS, the QTL model assumes that the parent carrying the same marker allele has the same QTL allele in the multi-parent population (also as the biallelic QTL assumption). In multi-parent populations, it is impossible for the genetic architectures for all QTLs along the entire genome to be completely consistent; there is therefore no ‘perfect method’ available to identify all QTLs with different architectures. Overall, methodological complementarity appears to be critical for the systematic interpretation of complex traits.

## Discussion

The majority of the genetic polymorphisms (>50%) are rare and many alleles are even private in specific maize lines; however, the functional importance of these rare variants remains unclear (Myles *et al.*, 2009; Fu *et al.*, 2013). From our QTL results, we found that many QTLs were exclusively identified



**Fig. 3** Quantitative trait locus (QTL) dissection for ear length on chromosome 6 in maize. (a) Two recombinant inbred line (RIL) populations detected a major QTL. (b) QTL validation using a heterogeneous inbred family analysis. The error bars indicate the standard deviation of ear length for each genotypic group in the QTL peak. (c) Joint linkage mapping (JLM) analysis further dissects the QTL. Red shading indicates the QTL region; black dashed line indicates the candidate single nucleotide polymorphism (SNP) position. LRT, likelihood ratio test. (d) Regional association for candidate gene identification. Right and left arrows indicate genes on positive and negative DNA strands. The red dotted-dashed line indicates the most significant SNP and underlying candidate gene. (e) Correlation between gene expression and ear length. (f) Genetic impact of significant SNPs on gene expression and ear length. In the box plots, the horizontal line within the box indicates the median value; the bars of the box indicate the limits as 1.5 times the interquartile range from the box; the dots outside the bars indicate the most extreme data points or possible outliers.



**Fig. 4** Different models exhibiting a dynamic statistical power for quantitative trait locus (QTL) detection in a multi-parent design. Ten recombination inbred line (RIL) populations were simulated using 200 RILs for each population. Left panel: 25 QTLs were simulated across 10 pseudo-chromosomes (to mimic the maize genome) with an even distribution. Middle panel: each colored box indicates the presence of a QTL in a specific population. The color density of the box indicates the additive genetic effects of the QTL in the RIL population, with a deeper shade denoting a greater effect. Right panel: the horizontal bars represent the power of separate linkage mapping (SLM), joint linkage mapping (JLM) and genome-wide association study (GWAS) methods.

by one specific model, but not by others (Fig. 2d). The results of the simulation studies suggested two possible explanations: (1) for high-frequency QTLs, JLM and GWAS boost the statistical power compared with SLM because of pooling of the 10 RIL populations; (2) low-frequency QTLs are more easily detected by SLM in one specific background where the allele frequency of QTLs may be more balanced within a specific population (Fig. 4). Motivated by the phenomenon that the majority of QTLs detected by SLM were unique to a specific population (Table 1; Fig. 1), we intuitively proposed a hypothesis to explain why there are a considerable number of QTLs which are method-specifically detectable: the QTLs that were only detected by SLM, but not by JLM or GWAS, are probably low-frequency QTLs, or at least to some extent. The reliability of this hypothesis and the frequency of the QTL allele could not be determined directly in this experiment, but the simulation data provided an opportunity to evaluate the robustness of making such a hypothetical claim for the identified QTLs, that is, the statistical power and FDR. According to the

simulation, the truth of this claim depends on the distribution of QTL additive effects, where the QTLs must have the approximate effect size (e.g.  $a \approx 0.3$  times the phenotypic standard deviation) to reach the highest power (Table S5). The FDR is high when the effects are too small, because both SLM and JLM have no power to detect QTLs, whereas the power is low when the effects are sufficiently large, because both SLM and JLM have comparable power to detect QTLs (Table S5). For the empirical data of ear traits, the additive effects of QTLs detected by SLM were actually enriched at 0.30–0.33 times the phenotypic standard deviation (Fig. 2a; Notes S1), which suggests that the hypothetical claim that the QTLs detected only by SLM, but not by JLM or GWAS, are probably low-frequency QTLs in the present experiment is probably true.

A number of maize inflorescence genes were cloned through mutagenesis (Table S6). These mutations produce dramatic effects on maize ear development, which are probably deleterious and therefore rare in natural germplasm. These genes were candidates for ear traits and were tested for co-localization with QTLs

identified by the three models. The mutation genes were significantly enriched among QTLs by SLM ( $P=0.015$ ), but not by JLM and GWAS ( $P=0.698$  and  $P=0.328$ , respectively; Table S6; Fig. S20). This finding implies that SLM is more likely to detect variants with reasonable effects that always underlie low-frequency loci, whereas JLM and GWAS are more likely to detect regulatory variants with allelic series, which provides indirect evidence for the aforementioned hypothetical claim of low-frequency QTLs in maize germplasm. In rice, a similar phenomenon was observed: seven of the 10 cloned QTLs affecting yield-related traits were found at a frequency of  $<0.1$  in diverse germplasm (Table S7). A similar hypothesis, that genome-wide causal variants are enriched for low-frequency alleles, was also proposed for human GWAS (Gusev *et al.*, 2013). Overall, the low-frequency or rare allelic variants seem to be important for complex traits across diverse species, which deserves attention in applications in medicine, agricultural and other areas (Schork *et al.*, 2009). Based on the present data analysis, we estimate that there are 59–82 QTLs or genes (most may be low-frequency variants) involved in the four ear traits surveyed in the diverse 508 maize inbred collection, which represents the majority of the modern breeding program diversity worldwide (Table S8) (Yang *et al.*, 2011).

Linkage mapping and GWAS are very efficient methods to unravel the genetic architecture of complex human and agricultural traits, but, nowadays, only a few heritabilities of important traits have been accounted for, such as human height ( $<5\%$ ) (Visscher, 2008). A low diversity of the mapping population, small effect size of QTLs and low frequency of the causal variants are the main factors that impede the comprehensive dissection of the genetic basis of complex traits in linkage mapping or GWAS (Manolio *et al.*, 2009). Here, we have proposed a new multi-parent designed population which permits the direct integration of the currently existing population resources into a large-scale genetic analysis. It provides a good alternative to improve genetic resolution and boost power for the identification of minor-effect and low-frequency variants via efficient utilization of the large population size, high genetic diversity and multiple statistical approaches. To fully explore the architecture of complex quantitative traits, such as yield, it is necessary to integrate a large number of biparental populations into a large-scale analysis. Many different types of biparental population are currently available worldwide, and the time is ripe to integrate them in a single large-scale analysis to improve the power of QTL detection in the era of low-cost next-generation sequencing or SNP array technology (Ozsolak & Milos, 2011; Grada & Weinbrecht, 2013). Thus, the present design provides a very flexible and manageable way to integrate available genetic resources in the research community, facilitating the comprehensive interpretation of the genetic architecture of complex quantitative traits in plants.

## Acknowledgements

We thank Dr Marilyn Warburton for helpful comments on the manuscript. This research was supported by the National Research and Development Project of Transgenic Crops of the

Ministry of Agriculture of China (2014ZX0800944B), the National Hi-Tech Research and Development Program of China (2012AA10A307), the National Natural Science Foundation of China (31525017, 31222041, 31401389) and the National Youth Top-notch Talent Support Program.

## Author contributions

J.Y. designed and supervised the study. Y.X., H.T., X.Y., Q.P., F.Q., M.S.R., Y.L., H.L., X.Z., N.Y., X.W., M.D., M.J., L.Z., X.Luo, Y.Z., X.Li, J.L., W.Z., N.L., H.W., G.C., Y.C., G.X., W.W. and D.Z. performed the experiments. Y.X., H.T., H.L. and N.Y. analyzed the data. Y.X. improved the genome-wide association method and performed simulation analysis. H.T. and S.X. developed the joint linkage mapping method. H.T. performed the epistasis analysis and meta analysis. Y.X., H.T., S.X. and J.Y. wrote the manuscript.

## References

- Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R. 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* 6: 11.
- Brown PJ, Upadaya Y, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES *et al.* 2011. Distinct genetic architectures for male and female inflorescence traits of maize. *PLoS Genetics* 7: e1002383.
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC *et al.* 2009. The genetic architecture of maize flowering time. *Science* 325: 714–718.
- Cavanagh C, Morell M, Mackay I, Powell W. 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology* 11: 215–221.
- Chen L, Storey JD. 2006. Relaxed significance criteria for linkage analysis. *Genetics* 173: 2371–2381.
- 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 Biology* 16: 167.
- Duvick DN. 2005. Genetic progress in yield of United States maize (*Zea mays* L.). *Maydica* 50: 193–202.
- Flint-Garcia SA, Thornsberry JM, Buckler SE IV. 2003. Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357–374.
- Flint-Garcia SA, ThUILlet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant Journal* 44: 1054–1064.
- Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z *et al.* 2013. RNA sequencing reveals the complex regulatory network in the maize kernel. *Nature Communications* 4: 2832.
- Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J *et al.* 2011. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE* 6: e28334.
- Grada A, Weinbrecht K. 2013. Next-generation sequencing: methodology and application. *Journal of Investigative Dermatology* 133: e11.
- Gusev A, Bhatia G, Zaitlen N, Vilhjalmsón BJ, Diogo D, Stahl EA, Gregersen PK, Worthington J, Klareskog L, Raychaudhuri S *et al.* 2013. Quantifying missing heritability at known GWAS loci. *PLoS Genetics* 9: e1003993.
- Huang XQ, 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. *Proceedings of the National Academy of Sciences, USA* 108: 4488–4493.
- Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF *et al.* 2012. *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences, USA* 109: 1913–1921.
- 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 Genetics* 5: e1000551.
- Kump KL, Bradbury PJ, Wissler RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D *et al.* 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genetics* 43: 163–168.
- Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N *et al.* 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genetics* 45: 43–50.
- Mackay TF, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565–577.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A *et al.* 2009. Finding the missing heritability of complex diseases. *Nature* 461: 747–753.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences, USA* 99: 6080–6084.
- Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. 2009. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21: 2194–2202.
- Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* 12: 87–98.
- Pan Q, Li L, Yang X, Tong H, Xu S, Li Z, Li W, Muehlbauer GJ, Li J, Yan J. 2015. Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytologist*. doi: 10.1111/nph.13810.
- Peiffer JA, Romay MC, Gore MA, Flint-Garcia SA, Zhang Z, Millard MJ, Gardner CA, McMullen MD, Holland JB, Bradbury PJ *et al.* 2014. The genetic architecture of maize height. *Genetics* 196: 1337–1356.
- R Core Team. 2012. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing [WWW document] URL <http://www.R-project.org/>. [accessed 17 December 2015].
- Rapala-Kozik M, Golda A, Kujda M. 2009. Enzymes that control the thiamine diphosphate pool in plant tissues. Properties of thiamine pyrophosphokinase and thiamine-(di)phosphate phosphatase purified from *Zea mays* seedlings. *Plant Physiology and Biochemistry* 47: 237–242.
- Salvi S, Tuberosa R. 2005. To clone or not to clone plant QTLs: present and future challenges. *Trends in Plant Science* 10: 297–304.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.* 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.
- Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development* 19: 212–219.
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES. 2011. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics* 43: 159–162.
- Tuinstra MR, Ejeta G, Goldsborough PB. 1997. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. *TAG. Theoretical and Applied Genetics* 95: 1005–1011.
- Valdar W, Holmes CC, Mott R, Flint J. 2009. Mapping in structured populations by resample model averaging. *Genetics* 182: 1263–1277.
- Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, Taylor MS, Rawlins JN, Mott R, Flint J. 2006. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics* 38: 879–887.
- Visscher PM. 2008. Sizing up human height variation. *Nature Genetics* 40: 489–490.
- Wang S, Basten CJ, Zeng Z. 2012. *Windows QTL Cartographer 2.5*. Raleigh, NC, USA: Department of Statistics, North Carolina State University. [WWW document] URL <http://statgen.ncsu.edu/qtcart/WQTLCart.htm>. [accessed 17 December 2015].
- Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J *et al.* 2014. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications* 5: 3438.
- Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. 2014. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genetics* 10: e1004573.
- Yang X, Gao S, Xu S, Zhang Z, Prasanna B, Li L, Li J, Yan J. 2011. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Molecular Breeding* 28: 511–526.
- Yano M, Tuberosa R. 2009. Genome studies and molecular genetics—from sequence to crops: genomics comes of age. *Current Opinion in Plant Biology* 12: 103–106.
- Yu J, Holland JB, McMullen MD, Buckler ES. 2008. Genetic design and statistical power of nested association mapping in maize. *Genetics* 178: 539–551.
- Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203–208.
- Zeng Z. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.
- Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordoñas JM *et al.* 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 44: 355–360.

## Supporting Information

Additional supporting information may be found in the online version of this article.

**Fig. S1.** Construction of the 10 diverse maize recombination inbred line populations.

**Fig. S2** Field experimental distribution across China in two consecutive years.

**Fig. S3** Distribution of the length of genetic blocks exhibiting mapping resolution of three models.

**Fig. S4** Principal component analysis (PCA) plot of 10 simulated recombination inbred line (RIL) populations.

**Fig. S5** Principal component analysis (PCA) plot exhibiting genetic differentiation among 10 recombination inbred line (RIL) populations.

**Fig. S6** Phenotypic distribution of maize ear traits across 10 recombination inbred line (RIL) populations.

**Fig. S7** Phenotypic and genetic correlations between four ear traits.

**Fig. S8** Overviews of quantitative trait locus (QTL) results for three traits.

**Fig. S9** Results of quantitative trait locus (QTL) mapping in each recombination inbred line (RIL) population for ear traits.

**Fig. S10** Empirical distribution of the proportions of unique quantitative trait loci (QTLs) in an unbalanced data situation.

**Fig. S11** Heatmaps of allelic series of quantitative trait loci (QTLs) in joint linkage mapping.

**Fig. S12** Summary of genetic validation via two heterogeneous populations.

**Fig. S13** Enrichment of candidate single nucleotide polymorphisms (SNPs) validated by AM508 relative to random SNPs.

**Fig. S14** Linkage disequilibrium (LD) decay in the 10 maize recombination inbred line (RIL) populations.

**Fig. S15** Fine mapping of a quantitative trait locus (QTL) for ear length spanning 22–25 Mb on chromosome 1.

**Fig. S16** Fine mapping of a quantitative trait locus (QTL) for ear length spanning 249–254 Mb on chromosome 1.

**Fig. S17** Fine mapping of a quantitative trait locus (QTL) for ear length spanning 279–282 Mb on chromosome 1.

**Fig. S18** Enrichment of genome-wide association study (GWAS) and candidate single nucleotide polymorphisms (SNPs) in the quantitative trait locus (QTL) intervals for maize ear traits.

**Fig. S19** Venn plots showing overlaps of quantitative trait loci (QTLs) detected by the three models.

**Fig. S20** Null distributions for the proportion of randomly selected genes matched by the three models.

**Table S1** Information on genetic linkage maps for 10 recombination inbred line (RIL) populations in maize

**Table S2** Analysis of variance (ANOVA), descriptive statistics and correlation analysis of four ear traits in 10 recombination inbred line (RIL) populations

**Table S3** Phenotypic variance explained by additive and epistatic models in maize ear traits

**Table S4** Information on four candidate quantitative trait loci (QTLs) for ear length in further genetic dissection

**Table S5** Power and false discovery rate (FDR) for the low-frequency quantitative trait locus (QTL) interpretations

**Table S6** Genetic overlaps between cloned maize inflorescence genes and quantitative trait loci (QTLs) of ear traits identified by the three models

**Table S7** Information list of natural variations underlying the cloned genes of rice yield and quality traits

**Table S8** Estimate of expected quantitative trait locus (QTL) numbers with low-frequency variants for ear development

**Methods S1** Resampling analysis.

**Methods S2** *Z* score-based meta-quantitative trait locus (QTL) analysis.

**Methods S3** Joint linkage mapping.

**Methods S4** Genome-wide association study.

**Methods S5** Simulation studies.

**Notes S1** Quantitative trait locus (QTL) information for four ear traits via separate linkage mapping (SLM) analysis in 10 recombination inbred line (RIL) populations.

**Notes S2** Meta-quantitative trait locus (QTL) analysis of the separate linkage mapping (SLM) results in 10 recombination inbred line (RIL) populations.

**Notes S3** Quantitative trait locus (QTL) and QTL-by-environment information of four ear traits via joint linkage mapping (JLM) analysis.

**Notes S4** Candidate single nucleotide polymorphisms (SNPs) and SNP-by-environment information for four ear traits via genome-wide association study (GWAS) analysis.

**Notes S5** Genetic validation of candidate single nucleotide polymorphisms (SNPs) via AM508 and nested association mapping (NAM) populations.

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.