Mapping causal genes and genetic interactions for agronomic traits using a large F$_2$ population in rice

Laibao Feng,[1,2,3] Aimin Ma,[1,3] Bo Song,[1,3] Sibin Yu (iD),[4] and Xiaoquan Qi (iD) [1,3,*]

[1]Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China,
[2]University of Chinese Academy of Sciences, Beijing 100049, China,
[3]Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100049, China, and
[4]National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, Hubei, China

*Corresponding author: Institute of Botany, Chinese Academy of Sciences, No. 20 Nanxincun, Xiangshan, Beijing 100093, China. Email: xqi@ibcas.ac.cn

Abstract

Dissecting the genetic mechanisms underlying agronomic traits is of great importance for crop breeding. Agronomic traits are usually controlled by multiple quantitative trait loci (QTLs) and genetic interactions, and mapping the underlying causal genes is still labor-intensive and time-consuming. Here, we present a genetic tool for directly targeting the specific causal genes by using a single-gene resolution linkage map that was constructed from 3756 F$_2$ rice plants via targeted sequencing technology and Tukey-Kramer multiple comparisons test. Three large- and moderate-effect QTLs, *qHD6-2*, *qGL3-1*, and *qGW5-2*, were successfully mapped to their specific causal genes, *Hd1*, *GS3*, and *GW5*, respectively. A complex genetic interaction network containing 30 QTL–QTL interactions was constructed, revealing that the alternative allele of hub QTL, *qHD6-2*, can hide or release the genetic contributions of the alleles at interacting loci. Moreover, arranging genetic interactions in the models lead to more accurate phenotypic predictions. These results provide a community resource and new feasible strategy for deciphering the genetic mechanisms of complex agronomic traits and accelerating crop breeding.

Keywords: quantitative trait loci; large F$_2$ population; rice; causal gene; genetic interactions

Introduction

Agronomic traits are normally controlled by multiple quantitative trait loci (QTLs) and genetic interactions between QTLs and environmental factors. Dissecting the underlying genetic and molecular mechanisms is crucial for crop domestication and improvement. Isolating the causal genes underlying favorable agronomic traits has proved to be essential for crop breeding (Jiang *et al.* 2012). Linkage analysis and association mapping are two popular and efficient strategies to uncover the genetic basis and isolate causal genes in rice. However, traditional linkage analysis is limited by its low mapping resolution resulted from population size and phenotyping precision (Bai *et al.* 2012). Thus, linkage analysis usually requires several generations of backcrossing to construct the near isogenic lines for eliminating the genetic background noise, genotyping thousands of individuals for recombinants, and exhaustive field phenotyping, which together make it time-consuming and costly (Qi *et al.* 1998; Yu *et al.* 2002; Wang *et al.* 2011; Varshney *et al.* 2014). Although association mapping has been widely used for quickly identifying loci or genes underlying agronomic traits, the peak signals of association loci often appeared near (but not within) the known genes, the power to detect rare alleles is inherently limited, and assembling a proper collection with a population size and representative diversity is quite complicated and costly (Huang *et al.* 2010; Zhao *et al.* 2011; Crowell *et al.* 2016; Wang *et al.* 2018).

In the era of next-generation sequencing, several mapping by sequencing methods using bulked segregant analysis coupled with whole-genome sequencing (WGS), such as SHOREmap (Schneeberger *et al.* 2009), MutMap (Abe *et al.* 2012), next-generation mapping (Austin *et al.* 2011), QTL-seq (Takagi *et al.* 2013), QTG-seq (Zhang *et al.* 2019), and GradedPool-Seq (Wang *et al.* 2019), have been devised to dissect the genetic basis of both qualitative and quantitative traits. Although the QTL mapping resolution is improved dramatically, the underlying causal genes are still not identified directly. Therefore, a feasible strategy for rapid identification of causal genes underlying quantitative traits is still needed. Recently, study on a large segregating population derived from a highly inbred yeast cross-mapped hundreds of causal variants directly with the power of adequate meiotic recombination and without the need for any additional experimentation (She and Jarosz 2018). Given the multicellular organism, such as rice, the potential of such strategy for directly mapping causal genes is still unknown.

Despite the complex relationship between the genotype and phenotype in agronomic traits, a majority of genetic variance in a population can be captured using models that assume gene variants combine their effects in an additive manner (Hill *et al.* 2008). Based on this "additive model," new elite rice varieties have been developed by pyramiding superior alleles that significantly contribute to desirable traits (Zeng *et al.* 2017; Wu *et al.* 2018). However, many studies (Lee *et al.* 2010; Bloom *et al.* 2015; Forsberg

et al. 2017; McWhite *et al.* 2020) have shown that genetic interactions are common and affect many biological traits. Genetic interactions also play a more prominent role than dominance effects for grain-yield heterosis in rice and wheat (Yu *et al.* 1997; Hua *et al.* 2003; Jiang *et al.* 2017; Boeven *et al.* 2020). Especially, a recently study on rice breeding also indicated that the phenotype of designed rice lines cannot be predicted precisely when pyramiding multiple alleles underlying the same trait due to the complexities of epistasis interactions (Wei *et al.* 2021). Typically, genetic interaction mapping involves the pairwise perturbation of genes (*e.g.*, knockout, knockdown or overexpression) in order to elucidate how one gene modulates the phenotype of the other. However, the information of potential genetic interactions between genes underlying QTLs in the field is limited. Hence, precisely dissection of genetic components and molecular basis in complex quantitative traits is important to unlock the full potential for breeding elite varieties for sustainable agriculture by rational design (Qian *et al.* 2016).

Here, we constructed a single-gene resolution map using a large rice F_2 population containing 3756 individuals genotyped by sequencing. Larger population greatly enhanced the power to detect additive QTLs, dominant QTLs and QTL–QTL interactions, and significantly increased the proportion of captured genetic variance of seven yield-related traits. Attractively, three large- and moderate-effect QTLs were directly targeted to specific causal genes. The contributions of genetic interactions for phenotypic variance were also investigated in detail. Our method provides an efficient strategy for comprehensively dissecting genetic architecture and rapidly mapping causal genes underlying QTLs, and will accelerate crop breeding.

Materials and methods
Planting and phenotyping
A large F_2 population consisting of ~5000 plants were generated from a cross of two rice elite varieties, *Oryza sativa* subsp. *indica/xian* Zhenshan 97 (ZS97) and Minghui 63 (MH63). F_2 seeds were germinated and planted in the experimental fields in LingShui, China (at N 18.32°, E 110.01°) in the summer of 2015. All the F_2 plants were grown in the consecutive farmland with well-distributed soil status in uniform condition. Plant height (PH) was measured from the soil surface to the apex of the longest leaf. Heading date (HD) was recorded daily as the number of days from sowing to the observation of first inflorescences that emerged above the flag leaf sheath. The grain length (GL), grain width (GW), and length-width ratio (LWR), were measured from ~300 fully filled grains using SC-E software (Hangzhou Wanshen Detection Technology Co., Ltd.). Yield per plant was obtained by weighting all fully filled grains of each F_2 plant.

Sequencing, genotyping, and bin-map construction
Genomic DNA was extracted from the fresh leaf tissues using a modified SDS-based extraction method (Sika *et al.* 2015). Sequencing libraries were constructed as described by Elshire *et al.* (2011). Genomic DNA was digested using restriction endonuclease *Dpn*II or *Ape*KI. Every 96 samples were barcoded and pooled into one library. After size selection and purification, libraries were sequenced on Illumina platform.

For calling single-nucleotide polymorphism (SNP) from parental genomes, we used MUMmer (Delcher *et al.* 2002) to align genomic sequence of ZS97 (ZS97RS2) and MH63 (MH63RS2) (Zhang *et al.* 2016), and used "delta-filter" command to filter the

alignment results. In total, 1,300,156 SNPs between ZS97RS2 and MH63RS2 were identified by using the "show-snp" command. For calling SNP from the F_2 population, the de-multiplexed raw sequencing reads were aligned against the parental genome ZS97RS2 and MH63RS2 by using of BWA software (Li and Durbin 2009), respectively. Potential SNPs were identified by using of SAMtools (Li *et al.* 2009) with the "mpileup" command. A candidate SNP site should be bi-allelic in F_2 population and satisfied with the following criteria: (i) only two type nucleotides presented on the position; (ii) the two nucleotides were already identified between parental genomes; (ii) minor allele frequency should be >5%; (iv) the SNP site was detected in more than 5 independent F_2 plants. A total of 404,643 high-confident SNPs were identified at last. For genotyping, we used an approach combining the "sliding window" (Huang *et al.* 2009) and Bayesian inference (Li 2011) method (Supplementary Figure S2). Finally, 3756 individuals remained after removing individuals with too few SNPs or too many ambiguous genotypes. Then the genotypic data of each F_2 individual were combined together for bin map construction with 100 kb intervals (Supplementary Figure S2C). The genetic linkage map was constructed from the recombination bins serving as genetic markers using Haldane's equation of the R/qtl package (Broman *et al.* 2003). The recombination rate (cM/Mb) was estimated using a 400 kb window size with a step size of 200 kb. The recombination hotspots were identified using 1000 times permutation test with the null hypothesis that every genomic region has an equal chance of recombination (Pan *et al.* 2016). The significance threshold (α) was obtained at false discovery rate (FDR) <5%.

Power calculations
We calculated statistical power (1-β) for population sizes of 100, 500, 1000, and 4000 individuals using the "power.t.test" function in R (R Core Team 2019). Power was estimated over a range of effect sizes, where effect size was calculated as the percentage of phenotypic variance explained by a single additive QTL, dominant QTL, or QTL–QTL interaction. To correct for multiple testing over thousands of markers, genome-wide significance threshold (α) of $P < 2.7 \times 10^{-3}$, $P < 1.6 \times 10^{-3}$, and $P < 9.7 \times 10^{-5}$ were used for declaring additive QTL, dominant QTL, and QTL–QTL interaction. These thresholds were chosen based on FDR <5%.

Decomposing genetic components
The relative contributions of additive, dominance, and genetic interaction effects in the F_2 population were estimated using a multivariate linear mixed model. The model can be written as:

$$y = \beta X + Za + Zd + Zaa + Zad + Zdd + e$$

where y is a vector of observed phenotypic values for n F_2 plants. β is a vector of estimated fixed effect coefficients, X is an incidence matrix of fixed effects. Z is an incidence matrix for random effects. a, d, aa, ad, and dd are the vectors of additive, dominance, additive-by-additive, additive-by-dominance, and dominance-by-dominance effects. e is a vector of residuals. These genetic effects are assumed to be normally distributed with mean zero and variance-covariance as follows:

$$a \sim N(0, \sigma_a^2 A), d \sim N(0, \sigma_d^2 D), aa \sim N(0, \sigma_{aa}^2 A{\circ}A), ad$$
$$\sim N(0, \sigma_{ad}^2 A{\circ}D), dd \sim N(0, \sigma_{dd}^2 D{\circ}D), and\, e \sim N(0, \sigma_{EV}^2 I_n)$$

The variance structure of the phenotypes is $V(y) = \sigma_a^2 ZAZ' + \sigma_d^2 ZDZ' + \sigma_{aa}^2 ZA{\circ}AZ' + \sigma_{ad}^2 ZA{\circ}DZ' + \sigma_{dd}^2 ZD{\circ}DZ' + \sigma_{EV}^2 I_n$. A and D

are additive and dominance relationship matrices. $A°A$, $A°D$, and $D°D$ are interaction relationship matrices. σ_a^2 and σ_d^2 are additive and dominant genetic variance captured by bin markers. σ_{aa}^2, σ_{ad}^2, and σ_{dd}^2 are additive-by-additive, additive-by-dominance, and dominance-by-dominance genetic variance captured by pairwise combinations of bin markers. I_n is $n \times n$ identity matrix, σ_{EV}^2 is the error variance. Variance components were estimated using "mmer2" function of R package Sommer (Covarrubias-Pazaran 2016) and custom R code. Standard errors of variance component estimates were calculated as the square root of the diagonal of the Fisher information matrix from the iteration.

Mapping additive and dominant QTL

First, genotypes were re-coded as follows:

$$\begin{bmatrix} G_2 \\ G_1 \\ G_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} a \\ d \end{bmatrix}$$

where G_k denotes the genotypic value of the bin marker. The value of k ($k = 0, 1, 2$) stands the genotype for homozygous MH63, heterozygous and homozygous ZS97, respectively. a and d denote the additive and dominance effects of the bin marker, respectively.

For each trait, the above multivariate linear mixed model was fitted. The additive QTL were mapped using a forward stepwise procedure described by Bloom *et al.* (2015). Such approach increases power to detect additive QTL by controlling for genetic contributions from additive effects of other loci, dominance effects, and interaction effects (Yang *et al.* 2014). The 1-dimensional LOD (logarithm of odds) score for each marker were calculated according to the equation LOD $= -n[\ln(1 - r^2)/2\ln(10)]$, where r is the Pearson correlation coefficient between the genotypes at the marker and BLUP residuals. BLUP residuals were calculated by subtracting the BLUPs for the additive effects of other loci, dominant effects, and pairwise interactions from the phenotypes. To estimate significance empirically, the thresholds were chosen based on FDR <5% by using of 1000 times permutation test. Confidence intervals for detected additive QTL were determined at 1.5 LOD drop using the "lodint" function in R/qtl package (Broman *et al.* 2003). The candidate gene underlying each QTL was inferred based on funRiceGenes database (Yao *et al.* 2018). To estimate the fraction of additive variance captured by detected additive QTLs, the model $y = \beta X + Za + e$ was fitted, where a was calculated from the relationship matrix of individuals only at the additive QTL peak markers (A_{QTL}) for the given trait, such that $a \sim N(0, \sigma_{A_{QTL}}^2 A_{QTL})$ and $e \sim N(0, \sigma_{EV}^2 I_n)$.

Similar approach was applied for mapping dominant QTL based on BLUP residuals for controlling remaining genetic effects. The fraction of phenotypic variance explained by each dominant QTL was also estimated as above. The fraction of dominant variance captured by detected dominant QTL was estimated by fitting the model $y = \beta X + Za + Zd + e$, where d was calculated from the relationship matrix of individuals only at the dominant QTL peak markers (D_{QTL}) for the given trait, such that $d \sim N(0, \sigma_{D_{QTL}}^2 D_{QTL})$.

Mapping QTL–QTL interactions

To increase computational efficiency, the marker set was reduced by selecting one marker per centimorgan on the genetic linkage map using the "pickMarkerSubset" function in R/qtl package (Broman *et al.* 2003). The genotypes of each pair of markers were coded as follows:

$$\begin{bmatrix} G_{22} \\ G_{21} \\ G_{20} \\ G_{12} \\ G_{11} \\ G_{10} \\ G_{02} \\ G_{01} \\ G_{00} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} aa_{12} \\ ad_{12} \\ ad_{21} \\ dd_{12} \end{bmatrix}$$

where G_{kl} denotes the genotypic value of the genotype whose first marker is coded as k and the second as l (k, $l = 0, 1, 2$). aa, ad, and dd denote the additive-by-additive, additive-by-dominance, and dominance-by-dominance interaction effects. The similar forward stepwise procedure was used for mapping QTL–QTL interactions by controlling for genetic contributions from the additive and dominant genetic variance (Bloom *et al.* 2015). The QTL–QTL interactions were detected either between all pairs of markers (full-genome scan) or only between pairs where one marker corresponds to a significant additive or dominant QTL (marginal scan). Theoretically, full-genome scan can detect a wider range of interactions, but marginal can have higher power due to a reduced search space. LOD scores for interactions were calculated for all pairs of markers as LOD $= [-n(\ln(1-r^2)/2\ln(10))]$, where n is the number of F_2 individuals with phenotypic values, and r is the Pearson correlation coefficient between the genotypes at pairs of markers separated by at least 10 cM and the BLUP residuals from the additive and dominant model as phenotypes. FDR at different LOD thresholds was calculated by dividing the number of peaks obtained from 1000 times permutations of F_2 individual identities by the number of peaks observed in the real data. The statistical significant QTL–QTL interactions were extracted at a threshold of FDR <5%. To estimate the fraction of interaction variance captured by significant QTL–QTL interactions, we fitted the model $y = \beta X + Za + Zd + Zaa + Zad + Zdd + e$, where $aa \sim N(0, \sigma_{A_{QTL}*A_{QTL}}^2 A_{QTL}°A_{QTL})$, $ad \sim N(0, \sigma_{A_{QTL}*D_{QTL}}^2 A_{QTL}°D_{QTL})$, and $dd \sim N(0, \sigma_{D_{QTL}*D_{QTL}}^2 D_{QTL}°D_{QTL})$. $\sigma_{A_{QTL}*A_{QTL}}^2$, $\sigma_{A_{QTL}*D_{QTL}}^2$ and $\sigma_{D_{QTL}*D_{QTL}}^2$ are fraction of phenotypic variance captured by various additive-by-additive, additive-by dominant and dominance-by-dominance interactions, respectively.

Targeted sequencing and mapping causal genes

A candidate region covering ~300 kb intervals around the QTL peak was selected for sequencing and genotyping (Supplementary Figure S7). First, the targeted loci of recombinants were amplified by using primary PCR with a pair of locus-specific primers (Supplementary Table S2) with common linker sequences (5'-ggagtgagtacggtgtgc-3' and 5'-gagttggatgctggatgg-3') added at the 5' end. Then secondary PCR was performed to add individual-specific barcode and common adaptor sequences. SNP calling and genotyping at each candidate locus were accomplished as above (Supplementary Figure S8) All phenotypic values were normalized to mean 0 and variance 1. For each candidate region, mapping causal genes were performed by pairwise comparison of candidate loci with all other loci using a modified MATLAB scripts (She and Jarosz 2018). For each comparison between loci i and j, we used one-way ANOVA analysis followed by Tukey-Kramer multiple comparisons test to estimate the null hypothesis (H_0) that locus$_i$ co-segregated with causal gene and the locus$_j$ did not (Supplementary Figure S9, A and B). We also estimated the alternative hypothesis (H_1) that locus$_j$ co-segregated with causal gene and the locus$_i$ did not. If the candidate locus$_i$ co-segregated with causal gene, swapping the genotypes of candidate locus$_i$ would lead to significant phenotypic change (such as

first column *vs* sixth column, third column *vs* seventh column in Supplementary Figure S9A). Otherwise, if the candidate locus$_j$ co-segregated with causal gene, significant phenotypic change would be observed between first column and fourth column, or between third column and fifth column in Supplementary Figure S9A. For each candidate locus, the probability of the most likely alterative hypothesis [max(H$_1$)] constituted the primary metric for significance (Supplementary Figure S9C). Here, we defined causal score $=-\log_{10}[\max(\text{H}_1)]$ for each candidate locus. The candidate locus with the largest causal score indicated most likely co-segregated causal gene for the mapped QTL (Supplementary Figure S9D).

Inferring and analyzing genetic interaction networks

By visualizing the 40 epistatic loci as nodes and the interactions between them as edges, a complex genetic interaction networks were constructed (Figure 4A and Supplementary Figure S12A). Here, QTLs interacting with four or more other loci were defined as hub QTLs. We estimated the difference in the phenotypic variance between individuals that carry alternative alleles at the interacted loci using a double generalized linear model (DGLM) (Smyth 1989). We fitted the DGLM with linear predictors for both mean and variance as $y \sim N(\mu_1 + X\beta_1, e^{\mu_2 + X\beta_2})$ using the R package dglm where y is the phenotypic value, X is the genotype (coded as 0 and 1), β_1 is the effect on the mean and β_2 is the effect on the variance. β_1 describes difference in mean, and β_2 describes the fold difference in variance between the individuals with alternative alleles. We estimated the capacitating effects of the hub QTLs using the method described by Forsberg *et al.* (2017). Capacitating effects are defined as a locus in the genetic interaction networks can hide, or reveal, the genetic effects of their interactors. For each network containing a hub QTL, we split the F$_2$ individuals into three groups according to their genotype at the hub QTL. The narrow-sense heritability in each group was calculated using the formula: $h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_d^2 + \sigma_{aa}^2 + \sigma_{ad}^2 + \sigma_{dd}^2 + \sigma_{EV}^2)$. We performed 1000 times permutation test to obtain the significance of the difference $(h_{ZS97}^2 - h_{MH63}^2)$ between the two groups. The significant difference was determined at a Bonferroni-corrected multiple-testing threshold of $0.05/3 \approx 0.0167$. Hub QTL with significant difference in h^2 was defined as a genetic capacitor. The average phenotypes were estimated for each of the $3^4 = 81$ possible combinations of alleles for three four-locus interaction networks. Each of these four-locus networks had a hub QTL connected with three other interacting loci by pairwise interactions. If the hub QTLs were connected to more than three other loci in the interaction network, we only kept the loci with the strongest statistical interaction with the hub QTL. For modeling the phenotypes of individual subpopulation, we fitted two different models, including (i) additive effects only and (ii) additive effects and pairwise interactions. The bias (estimation errors) of each model was evaluated using tenfold cross-validation. Within each of the 81 subpopulations, estimation errors were calculated as $e = y - \hat{y}$. Here, y is the actual phenotype, $\hat{y}$ is the estimated phenotype in each round cross-validation. *t*-test was used to determine whether e was significantly deviated from 0 at a Bonferroni-corrected multiple-testing threshold of $0.05/81 \approx 6.17 \times 10^{-4}$. The replicated 10-fold cross-validation (Cui *et al.* 2020) was used for phenotypic predictions with different combinations of various genetic components by use of QTL peak markers or all markers in genome, respectively.

Results

Construction of a genetic linkage map with single-gene resolution

A large F$_2$ population from a cross between two elite rice ZS97 and MH63, was generated. In order to genotyping this large population cost-effectively and quickly, the genotyping-by-sequencing (GBS) method was applied. All F$_2$ individuals were sequenced using Illumina platform with $\sim$0.35$\times$ coverage for each (Supplementary Figure S1A). A total of 404,643 high-confident SNPs ($>$1 SNP per 10 kb on average) that densely distribute across the whole genome were identified (Supplementary Figure S1B). Using a "sliding-window" approach, highly reliable genotypes from 3756 F$_2$ individuals were called (Supplementary Figure S2). The proportion of three genotypes (ZS97, heterozygote, and MH63) was consistent with the expectation ratio of 1: 2: 1 across the whole genome. The total distance of the constructed genetic linkage map was 1586.2 cM, $\sim$0.46 cM per bin marker. The length of bin markers ranged from 100 kb to 2.3 Mb, with a mean of 111.1 kb. The recombination rate was $\sim$3.65 cM/Mb, and recombination hotspots were observed on chromosomes 1, 2, 3, 4, 7, 9, 11, and 12 (Supplementary Figure S3B). There are 107,206 crossovers in total, $\sim$29.46 crossovers per F$_2$ individual on average (Supplementary Figure S3A). Scanning the 12 chromosomes with a 100-kb window revealed a median value of 27.86 crossovers and 15.54 genes per 100 kb (equal to 75.62 crossovers and 41.79 genes per cM), indicating that the constructed genetic linkage map reaches to a single-gene resolution ($27.86/15.54 \approx 1.79$ crossovers per gene) (Figure 1, A and B).

Dissection of genetic components and mapping QTLs for agronomic traits

Grain yield and its related agronomic traits of the large F$_2$ population were investigated, showing a distribution of typical complex quantitative traits (Supplementary Figure S4). The multivariate linear mixed model was used to estimate additive effects, dominance effects, and pairwise interactions for trait variation. We found that moderate to large proportion of trait variation (41.55% to 73.04%), contributed from genetic components, were detected for PH, HD, GL, GW, and grain LWR, and about 30% for grain yield per plant (YD), but only 7.76% for tiller number (TN) (Figure 1C). Additive genetic variance was the most abundant genetic component, ranged from 4.58% (TN) to 67.81% (LWR), with an average of 44.36% for all seven agronomic traits. Dominant genetic variance was the secondary abundant genetic component, ranged from 0.80% (TN) to 6.00% (GL), with an average of 2.46%. Pairwise genetic interactions were relatively rare, with an average of 2%. HD has the largest additive-by-additive effects (2.57%), while PH and TN have larger dominance-by-dominance effects (2.93% and 2.39%). Additive-by-dominance effect was only detected in YD (1.18%) and GL (0.92%). These results showed that additive effects are predominance component and genetic interactions are significant but much minor component in explaining variance of rice grain yield and its related traits.

Our simulations suggested that larger population sizes provide higher statistical power and more accuracy to detect additive QTL, dominant QTL, and QTL–QTL interactions, especially for QTL with small effect size (Supplementary Figure S5A). For instance, we have 90% power to detect an additive QTL that explains 0.24% of phenotypic variation, a dominant QTL that explains 0.28% of phenotypic variation and a pairwise QTL–QTL interaction that explains 0.36% phenotypic variation with a population size of 4000. Indeed, we identified 162 additive QTLs, with
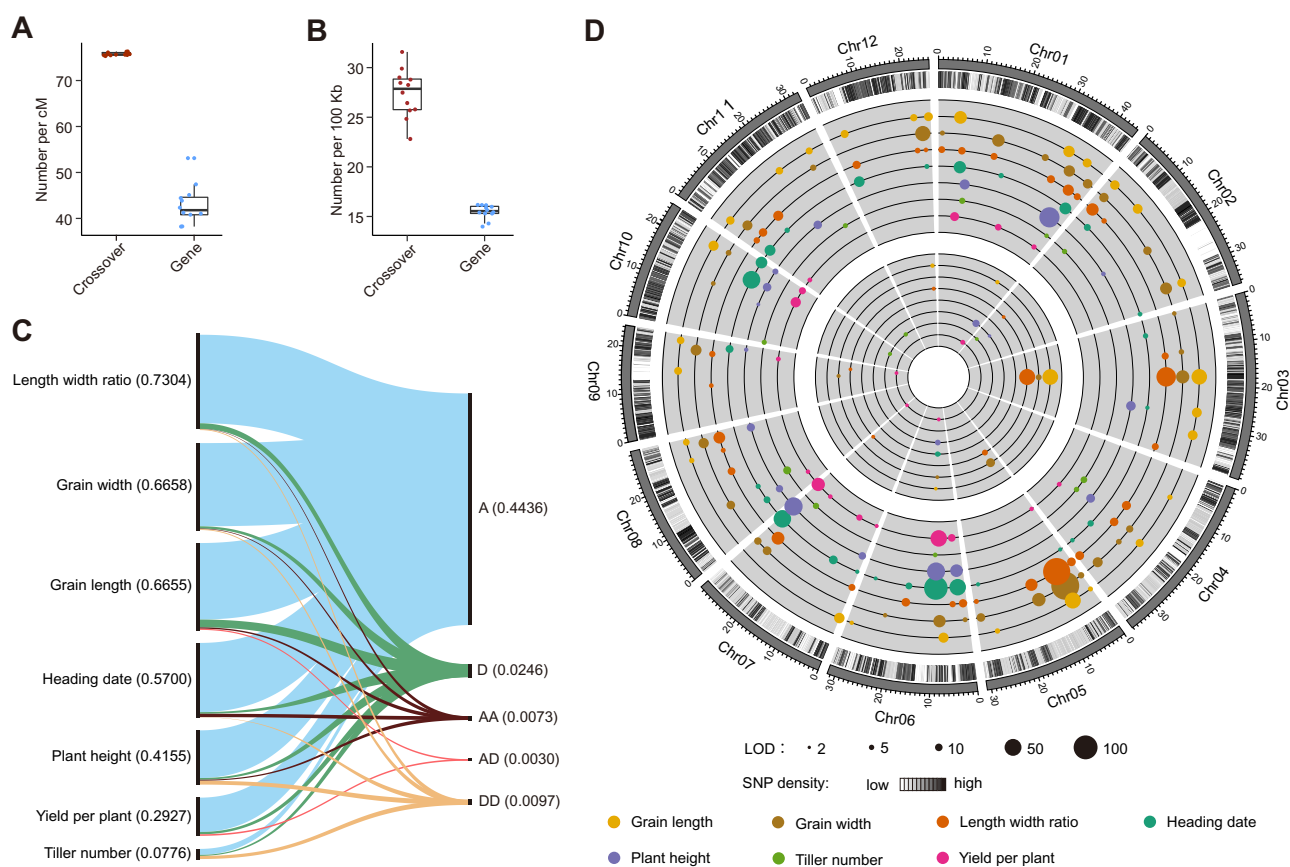
Figure 1 Genetic components dissection and QTL mapping. (A and B) The average density of crossover and gene on each chromosome based on genetic (window size of 1 cM) and physical (window size of 100 kb) maps. (C) Relative contributions of various genetic components to trait variance. The genetic components include additive variance A, cyan, dominance variance D, green, additive-by-additive interaction variance AA, brown, additive-by-dominance interaction variance AD, red and dominance-by-dominance interaction variance DD, yellow. The numbers in the brackets on left and right stand for total contribution of genetic components to each trait and average contribution of each genetic component to all traits, respectively. (D) Landscape of identified additive and dominant QTL in seven agronomic traits. Tracks from outside to inside indicate chromosomes of Minghui 63 genome, SNP density (window size of 100 kb), additive QTL, and dominant QTL, respectively.

an average of 23.14 additive QTLs per trait (ranged from 10 to 34) (Figure 1D, Supplementary Table S1). A clear genetic architecture of a few large- to moderate-effect QTLs plus many minor-effect QTLs was observed. For example, the largest QTL on chromosome 3 explained 46.90% of phenotypic variance of GL, while the others 27 QTLs totally explained 19.32% of phenotypic variance (ranged from 0.05% to 4.61%) (Supplementary Figure S6A). One or two moderate-effect QTLs (explained 8.29%–22.51% phenotypic variance) plus many minor-effect QTLs (explained 1.57% phenotypic variance on average) were identified for HD, PH, and YD. Except for TN, only a few minor-effect QTLs were detected, explained 0.28%–0.98% of phenotypic variance. These additive QTL captured an average of 94.45% of the estimated additive genetic variation (Figure 2A). Twenty-six significant QTL with dominance effects were identified, explained 0.54% phenotypic variance per locus on average (Figure 1D, Supplementary Figure S6B). Interestingly, all of them overlapped with the identified additive QTLs. These 26 loci explained 61.26% of the estimated dominance genetic variation (Figure 2B). Moreover, we identified 30 pairwise genetic interactions (margin scan, "see *Materials and Methods*") (Supplementary Table S1), including 16 additive-by-additive interactions, 2 additive-by-dominance interactions, and 12 dominance-by-dominance interactions. These genetic interactions explained few phenotypic variance with 0.19% on average. One relative strong interaction was identified on chromosome 6,

explained 0.90% of the phenotypic variance for HD (Supplementary Figure S6C). The pairwise interactions captured 84.51% of the estimated genetic interactions in total (Figure 2C). In summary, genetic architecture for seven agronomic traits in our study consist of a few large- to moderate-effect additive QTLs, many minor-effect additive and dominant QTLs, and minor-effect QTL–QTL interactions. The identified QTLs and genetic interactions captured majority of the estimated heritability in this large F_2 population.

Mapping causal genes

In this study, three large- and moderate-effect QTLs (phenotypic variance explained > 15%) was mapped to 300–500 kb when the 1.5 delta logarithm of the odds (ΔLOD] was used to define the statistical boundaries (Supplementary Table S1). But such confidence intervals in rice genome contain at least dozens of genes, causing failure in identification of the causal gene directly. This is likely due to a relative low marker density by using GBS method and very rough determination of recombinant breakpoints by using the 100-kb sliding window for constructing linkage map. To resolve these, we developed a multiplex high-throughput targeted sequencing method to increase local marker density and determine the recombinant breakpoints more precisely (Supplementary Figures S7 and S8). The most likely causal gene was identified by using one-way ANOVA analysis followed
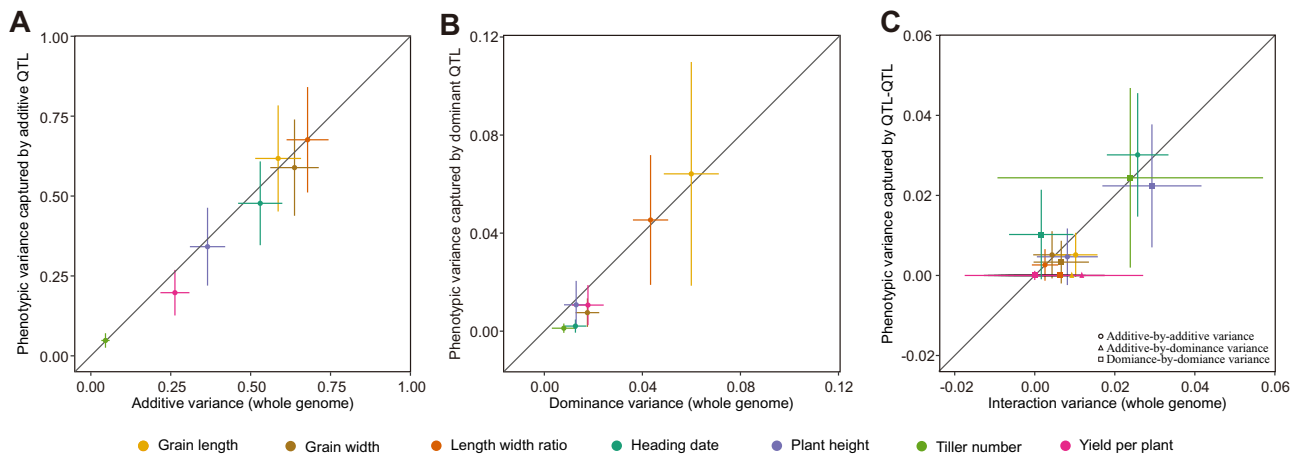
Figure 2 Additive, dominance, and interaction variance captured by detected loci. (A) Total variance captured by detected additive QTL for each trait is plotted against the whole-genome estimate of additive genetic variance. (B) Total variance captured by detected dominant QTL for each trait is plotted against the whole-genome estimate of dominant genetic variance. (C) Total variance captured by detected QTL–QTL interactions for each trait is plotted against the whole-genome estimate of interaction genetic variance. The diagonal lines represent variance captured by detected additive QTL = additive genetic variance, variance captured by dominant QTL, dominant genetic variance; variance captured by detected QTL–QTL interactions = interaction genetic variance, respectively. Error bars show ± SE.

by Tukey-Kramer multiple comparisons test in the confidence interval (Supplementary Figure S9).

The pleiotropic QTL, *qHD6-2*, was mapped to 400-kb on chromosome 6, and explained 22.51%, 8.86%, and 8.29% phenotypic variation for HD, PH, and YD, respectively (Figure 3A, Supplementary Table S1). Among the 3756 F_2 plants, 73 F_2 plants harbored meiotic crossovers in this 400-kb region. Eleven markers relative evenly distributed in this candidate gene region were designed and genotyped in these 73 recombinants by using targeted sequencing method (Supplementary Figure S8 and Table S2). Pairwise comparisons between each marker locus with all other loci within this 400-kb region following by the Tukey-Kramer multiple comparisons test showed that the marker *M9065137* had the highest causal score for HD (Figure 3B and Supplementary Figure S9), indicating that *M9065137* most likely co-segregated with causal gene. Indeed, the MH63 allele of *M9065137* contains a 4-bp deletion in the second exon of *Hd1* gene, and significant phenotypic differences were observed among the three alleles at the *Hd1* locus (Figure 3C). *Hd1* which highly expressed in leaf and inflorescence, was first mapped and cloned by using a backcrossed population (BC_3F_3) from a cross between varieties Nipponbare and Kasalath (Yano *et al.* 2000). The 2-bp deletion in the second exon in Kasalath allele resulted in a premature stop codon, causing later flowering under long-day conditions. The 4-bp deletion of MH63 allele locates at the CCT domain and causes a frame-shift mutation of *Hd1*, likely resulting in loss of *Hd1* function and later flowering compared to the ZS97 allele containing functional *Hd1*. Interestingly, the 4-bp deletion in MH63 allele is identical to that of Teqing, the parental line of BC_4F_2 mapping population for cloning the HD QTL, *Ghd6*, in recent study (Zhang *et al.* 2017) and shown that *Hd1* was also the causal gene of *Ghd6*. Hence, *qHD6-2* is directly mapped to the *Hd1* gene by using our single-gene resolution map.

We have also directly targeted the causal genes of another two large-effect QTLs, *qGL3-1* and *qGW5-2* (Supplementary Figure S10). *qGL3-1* explained 46.90% phenotypic variation of GL (Supplementary Figure S10A and Table S1) and was mapped to the *GS3* gene (Supplementary Figure S10C). The peak marker *M17605373* happened to be a single nucleotide substitution of C in ZS97 by A in MH63 in second exon of *GS3* gene. This is a

common SNP in rice varieties, causing premature termination of the predicted Gγ protein and resulting in long grain, which was observed in several previous studies (Fan *et al.* 2006; Takano-Kai *et al.* 2009; Mao *et al.* 2010; Wang *et al.* 2016). *qGW5-2* explained 31.94% phenotypic variation of GW (Supplementary Figure S10B and Table S1). The marker *M5431559*, giving the highest causal score, was close to the *GW5* gene (Supplementary Figure S10D). The ZS97 allele has a 950-bp deletion at the upstream of *GW5*, is a common haplotype in *indica* varieties associated with decreased expression of *GW5*, resulting in wide grain through brassinosteroid signaling pathway (Duan *et al.* 2017; Liu *et al.* 2017). To test whether our approach can further resolving minor-effect QTL, *qGL3-3* explained 4.61% phenotypic variation of GL was analyzed (Supplementary Figure S11). But the marker with high causal score was not observed.

Construction of QTL–QTL interaction networks

Abundant genetic interactions were observed for seven agronomic traits (Figure 1C). Most epistatic QTLs interacted with one or two QTLs, whereas several epistatic QTLs interacted with more than three QTLs (Supplementary Table S1). Such epistatic QTLs including *qHD6-2*, *qHD10-1*, and *qPH1-2*, were defined as hub QTLs. A genetic interaction network containing 13 statistically significant epistatic loci (QTLs) for HD was constructed using epistatic QTLs as nodes and the interactions between them as edges (Figure 4A). The hub QTL, *qHD6-2* (*Hd1*), contributed the most abundant additive genetic variance (22.51%) in HD. The phenotypic variance was twofold higher for individuals with the homozygous ZS97 allele than for individuals with the homozygous MH63 allele at this locus and its genetic variance heterogeneity is statistically significant ($P < 3.9 \times 10^{-6}$; dglm, two-sided test). By estimating the narrow-sense heritability (h^2) separately among individuals with the homozygous ZS97 and homozygous MH63 alleles, we showed that much of the difference in phenotypic variation was due to genetic effects ($h^2_{MH63} = 0.1185$ *vs* $h^2_{ZS97} = 0.3990$; $P < 0.001$, one-sided permutation test). Such highly connected hub QTL was earlier described as genetic capacitor where one allele suppresses genetic contributions of other interacting loci and the other allele uncovers them (Forsberg *et al.* 2017). The estimation of genetic variance heterogeneity at other
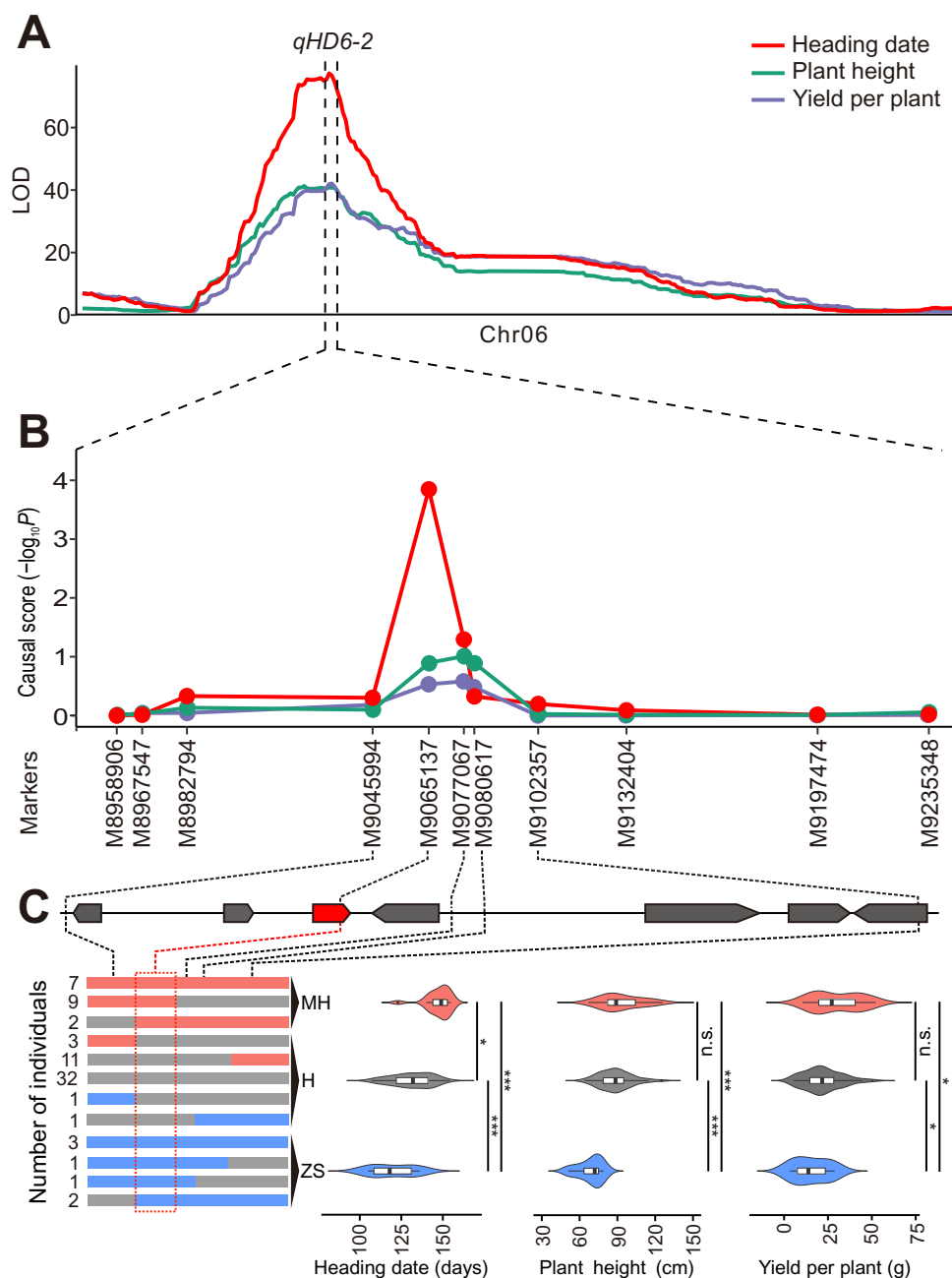
Figure 3 Mapping causal gene for *qHD6-2*. (A) The traditional LOD (logarithm of odds) score plot for *qHD6-2* controlling HD, plant height and yield per plant on chromosome 6. (B) The causal score is calculated at each candidate locus by using pairwise comparisons. (C) The up panel shows the gene annotation of candidate gene region. The bottom panel shows the numbers, genotypes and the phenotypes of recombinants at causal gene locus. MH, H, and ZS represent the genotype of Minghui 63, heterozygous, Zhenshan 97, respectively. n.s., not significant, *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

two hub QTLs (*qHD10-1* and *qPH1-2*) for HD and PH did not show such capacitation effects.

Our large segregating population provides abundant genotype and phenotype data for confidently estimating sufficient individuals of four-locus genotype classes (subpopulations) (see *Materials and Methods*). For the four-locus interaction network, the 3756 individuals were grouped into 81 (3^4) subpopulations, and their phenotypic means and variances were estimated. At the genetic capacitor (*Hd1*) locus, all subpopulations with canalizing MH63 allele had consistent longer HD regardless of how many flowering-suppressing alleles they had at the interacting loci. In contrast, the subpopulations with capacitating ZS97 allele

increased HD as the number of flowering-suppressing alleles increasing at the interacting loci (Figure 4B).

The additive variance contributed by the four-locus network at *Hd1* locus amounted to 28.52% of the total phenotypic variance in HD. In the subpopulations where the MH63 or ZS97 alleles at *Hd1* locus were fixed, the four-locus network instead contributed 2.21% and 13.58% of the total phenotypic variance, respectively. These results indicated that the phenotypic variance explained by an epistatic locus not only depend on its own effect size and allele frequency but also the allele frequencies at the interacting loci in the network. Indeed, phenotypic prediction models containing pairwise interaction effects captured more phenotypic
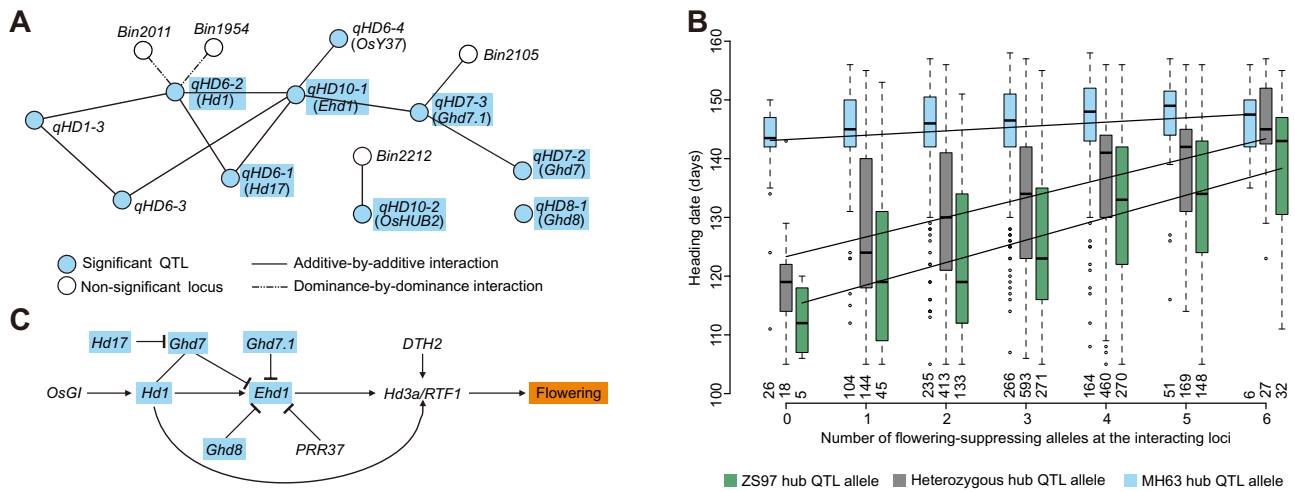
Figure 4 Genetic interaction network. (A) Each solid circle represents a significant QTL, hollow circle represents nonsignificant locus. Solid lines stand for additive-by-additive interactions, dashed lines stand for dominance-by-dominance interactions. The name of QTLs and bin-markers are given. The genes in the brackets are inferred candidate genes for QTLs. (B) The four-locus network contains hub QTL (*Hd1*) capacitor allele that regulates phenotypic effects of segregating alleles at the interacting loci in HD. The MH63 allele suppresses (canalizes) the phenotypic effects of segregating alleles at the interacting loci, while the ZS97 allele uncovers (capacitates) them. The individuals are divided and colored on the basis of genotype at the hub QTL. Each box shows the phenotype distribution for each group of individuals that have the same number of flowering-suppressing alleles at the interacting loci. The x-axis gives the number of flowering-suppressing alleles at the interacting loci and the number of individuals in each group. The black lines through the boxes illustrate the fit for linear additive model. (C) The simplified interaction network was inferred from reported experimental evidences. The overlapped genes with the genetic interaction network of this study (A) are labeled with cyan color.

variance and had less significant bias than models containing additive effects only in four-locus network (Supplementary Figures S12B, S13A, and S14) and whole-genome level (Supplementary Figure S15).

With a large number of the cloned genes (QTLs) for HD in rice (Supplementary Table S1), we observed that the constructed QTL–QTL interaction network tended to be consistent with the gene/protein interaction network established by the previously studies (Yano *et al.* 2000; Song *et al.* 2012; Nemoto *et al.* 2016; Zhang *et al.* 2019) (Figure 4C). For example, *Hd1* physically interacts with *Ghd7 in vivo* through binding of the CCT domain of *Ghd7* to the transcription-activating domain of *Hd1*, then this complex specifically binds to a *cis*-regulatory region in *Ehd1* and represses its expression only under long-day conditions. Thus, the constructed genetic interaction networks could not only dissect genetic architecture but also provide evidences and candidate gene loci for further exploring the underlying molecular mechanism of complex quantitative trait.

Discussion

The statistical resolution of mapping genotype to phenotype is fundamentally limited by genetic linkage between genes and adjacent genes. The phenotypic effect of causal gene and adjacent genes underlying QTLs can be distinguished in theory if a sufficient number of individuals contain meiotic recombination between the genes. Traditional QTL cloning methods depending on the genetic linkage of QTLs to visible makers are labor-intensive and time-consuming. With the prevalence of NGS technologies, several new methods have been established to accelerate the works of genetic mapping and gene cloning. In this study, we developed a new genetic tool and successfully implemented this tool to target specific causal genes of three large- and moderate-effect QTLs rapidly and directly. Our tool has several notable advantages compared with other methods. First, large population

size provided high detection power for mapping various genetic components and QTLs, especially for minor-effect QTLs and QTL–QTL interactions (Figure 2 and Supplementary Figure S5). By comprehensively analyzing seven agronomic traits, we concluded that additive effects contribute most of the phenotypic variation, while dominance effects and genetic interactions have smaller but significant contributions (Figure 1C). Second, combining a larger number of informative recombinants with targeted sequencing technology, the specific causal genes underlying QTLs were identified directly (Figure 3 and Supplementary Figure S10). Whereas the mapping resolution of QTL-seq is $\sim$2 Mb (Takagi *et al.* 2013), the mapping resolution of QTG-seq is $\sim$150 kb (Zhang *et al.* 2019), the mapping resolution of GradedPool-Seq is $\sim$400 kb (Wang *et al.* 2019). Third, our method only requires F_2 generation from the first cross of two parent lines, reducing a lot of time and works needed for constructing genetic population. In addition, the cost-effectiveness of the entire process is relatively high compared with traditional WGS method. Low sequencing coverage of GBS combined with "sliding window" approach guarantee high-density genetic linkage map in a cost-effective way. Moreover, our tool can map QTLs underlying multiple traits simultaneously in one population. For instance, a leaf sampled once can be used multiple times. Overall, we present a time- and cost-saving tool for dissecting the genetic and molecular mechanisms underlying agronomic quantitative traits.

It is noteworthy that our strategy could be easily extended to other species, such as maize and wheat. Although larger genome size correlates with reduced recombination rate, most plant species maintain a relatively constant number of protein-coding genes (Tiley and Burleigh 2015; https://phytozome.jgi.doe.gov/). Given the recombination rate and genome size of typical crops, such as *Zea mays*, *Triticum aestivum*, *Sorghum bicolor*, *Solanum lycopersicum*, *Solanum tuberosum*, and *Glycine max*, 3000–5000 F_2 plants could provide sufficient recombination to obtain a single-gene revolution linkage map. Taken maize as an example, 3000 F_2

plants could provide 75,000 crossovers in total, *i.e.*, approximately 34 crossovers per Mb, giving nearly two folds of gene density (18.7 genes per Mb) in maize genome (Chen *et al.* 2014).

Our tool shows a relatively wide range of applications in QTL mapping and cloning, but some limitations also exist. First, our tool may not work very well for targeting causal genes of minor-effect QTLs (Supplementary Figure S11). Second, the interaction between QTL and environmental factors is widely present in crops and other species. F_2 population cannot distinguish phenotypic variation contributed from environmental changes. However, these issues can be partially solved by transforming the F_2 population to permanent recombinant inbred (RI) population using single seed descent method. The statistical mapping power will be enhanced by removing partial dominance issue. Because the phenotypic distinction of heterozygous genotypes to homozygous genotypes is weaker than the phenotypic distinction of two parental homozygous genotypes. Besides, QTL by environment interaction can be evaluated precisely by planting RI population in different environments (Groen *et al.* 2020).

Our study also suggests another useful strategy for precisely characterizing the epistatic interactions in complex quantitative traits. We note that a small contribution of genetic interactions to phenotypic variance does not imply that interactions do not exist (Figure 1C), or that they are not important for exploring the complete genetic basis of specific traits (Mackay 2014; Zan and Carlborg 2020). We showed that highly connected hub QTL could acted as genetic capacitors, which can buffer or release cryptic genetic effects of the interacting loci in the epistatic networks (Figure 4B). The genes underlying epistatic QTLs also act epistatically at the molecular level (Figure 4). Cryptic genetic variance affects the prediction accuracy of traits in agricultural or medical applications and long-term selection responses in populations (Carlborg and Haley 2004; Carlborg *et al.* 2006). These epistatic interactions are non-negligible for rational design in crop breeding which aims to pyramid multiple superior alleles into one ideal plant (Qian *et al.* 2016; Wei *et al.* 2021). Moreover, several studies already showed that epistatic interactions play a more important role than dominance effects for grain-yield heterosis in wheat (Jiang *et al.* 2017; Boeven *et al.* 2020).

ZS97 and MH63 are the parents of the elite hybrid Shanyou 63, which has been the most widely cultivated hybrid in China. Heterotic pattern investigations of the detected 162 QTLs underlying seven yield-related traits showed that numerous loci, mostly with partial dominance effects for heterozygous genotypes and more proportion of superior alleles from male parental line (Supplementary Figure S16 and Table S1), are the major causes of strong heterosis in hybrid Shanyou 63. These results are consistent with previous study in rice that the accumulation of superior alleles with partial dominance is an important contributor to the heterosis (Huang *et al.* 2015, 2016).

With the decline in high-throughput sequencing, studying large segregating population for QTL mapping and cloning will be more and more feasible in crops. We believe our genetic tool will promote deciphering the genetic mechanisms of complex quantitative traits and accelerating crop breeding.

Data availability

DNA sequencing data have been deposited to BIG Data Center (https://bigd.big.ac.cn/gsa/) under BioProject Accession number: PRJCA003774. Supplementary materials are available at figshare: https://doi.org/10.25387/g3.15106053.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, *et al.* 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol. 30:174–178.

Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, *et al.* 2011. Next-generation mapping of *Arabidopsis* genes. Plant J. 67:715–725.

Bai X, Wu B, Xing Y. 2012. Yield-related QTLs and their applications in rice genetic improvement. J Integr Plant Biol. 54:300–311.

Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, *et al.* 2015. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nat Commun. 6:8712.

Boeven PHG, Zhao Y, Thorwarth P, Liu F, Maurer HP, *et al.* 2020. Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. Sci Adv. 6:eaay4897.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19:889–890.

Carlborg O, Haley CS. 2004. Epistasis: too often neglected in complex trait studies? Nat Rev Genet. 5:618–625.

Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L. 2006. Epistasis and the release of genetic variation during long-term selection. Nat Genet. 38:418–420.

Chen Z, Wang B, Dong X, Liu H, Ren L, *et al.* 2014. An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F_2 maize population. BMC Genomics. 15:433.

Covarrubias-Pazaran G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One. 11:e0156744.

Crowell S, Korniliev P, Falcão A, Ismail A, Gregorio G, *et al.* 2016. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. Nat Commun. 7:10527.

Cui Y, Li R, Li G, Zhang F, Zhu T, *et al.* 2020. Hybrid breeding of rice via genomic selection. Plant Biotechnol J. 18:57–67.

Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 30:2478–2483.

Duan P, Xu J, Zeng D, Zhang B, Geng M, *et al.* 2017. Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. Mol Plant. 10:685–694.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, *et al.* 2011. A robust, simple genotyping-by-sequencing GBS approach for high diversity species. PLoS One. 6:e19379.

Fan C, Xing Y, Mao H, Lu T, Han B, *et al.* 2006. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and

thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet. 112:1164–1171.

Forsberg SK, Bloom JS, Sadhu MJ, Kruglyak L, Carlborg Ö. 2017. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat Genet. 49: 497–503.

Groen SC, Calić I, Joly-Lopez Z, Platts AE, Choi JY, et al. 2020. The strength and pattern of natural selection on gene expression in rice. Nature. 578:572–576.

Hill WG, Goddard ME, Visscher PM. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4:e1000008.

Hua JP, Xing YZ, Wu WR, Xu CG, Sun XL, et al. 2003. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA. 100:2574–2579.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, et al. 2009. High-throughput genotyping by whole-genome resequencing. Genome Res. 19:1068–1076.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 42:961–967.

Huang X, Yang S, Gong J, Zhao Q, Feng Q, et al. 2016. Genomic architecture of heterosis for yield traits in rice. Nature. 537:629–633.,

Huang X, Yang S, Gong J, Zhao Y, Feng Q, et al. 2015. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat Commun. 6:6258.

Jiang Y, Schmidt RH, Zhao Y, Reif JC. 2017. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nat Genet. 49:1741–1746.

Jiang YH, Cai ZX, Xie WB, Long T, Yu HH, et al. 2012. Rice functional genomics research: progress and implications for crop genetic improvement. Biotechnol Adv. 30:1059–1070.

Lee K, Thorneycroft D, Achuthan P, Hermjakob H, Ideker T. 2010. Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. Plant Cell. 22: 997–1005.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27: 2987–2993.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al.; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. Bioinformatics. 25: 2078–2079.

Liu J, Chen J, Zheng X, Wu F, Lin Q, et al. 2017. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat Plants. 3:17043.

Mackay TFC. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 15: 22–33.

Mao H, Sun S, Yao J, Wang C, Yu S, et al. 2010. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA. 107:19579–19584.

McWhite CD, Papoulas O, Drew K, Cox RM, June V, et al. 2020. A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell. 181:460–474.e14.

Nemoto Y, Nonoue Y, Yano M, Izawa T. 2016. Hd1,a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7. Plant J. 86:221–233.

Pan Q, Li L, Yang X, Tong H, Xu S, et al. 2016. Genome-wide recombination dynamics are associated with phenotypic variation in maize. New Phytol. 210:1083–1094.

Qi X, Niks RE, Stam P, Lindhout P. 1998. Identification of QTLs for partial resistance to leaf rust Puccinia hordei in barley. Theor Appl Genet. 96:1205–1215.

Qian Q, Guo L, Steven MS, Li J. 2016. Breeding high-yield superior quality hybrid super rice by rational design. Natl Sci Rev. 3: 283–294.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Australia: R Foundation for Statistical Computing. Available at: https://www.r-project.org/ (Accessed: 2019 August 29).

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, et al. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods. 6:550–551.

She R, Jarosz DF. 2018. Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. Cell. 172:478–490.e15.

Sika KC, Kefela T, Adoukonou-Sagbadja H, Ahoton L, Saidou A, et al. 2015. A simple and efficient genomic DNA extraction protocol for large scale genetic analyses of plant biological systems. Plant Gene. 1:43–45.

Smyth GK. 1989. Generalized linear models with varying dispersion. J R Stat Soc B. 51:47–60.

Song Y, Gao Z, Luan W. 2012. Interaction between temperature and photoperiod in regulation of flowering time in rice. Sci China Life Sci. 55:241–249.

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, et al. 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74:174–183.

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, et al. 2009. Evolutionary history of GS3, a gene conferring grain length in rice. Genetics. 182:1323–1334.

Tiley GP, Burleigh JG. 2015. The relationship of recombination rate, genome structure, and patterns of molecular evolution across angiosperms. BMC Evol Biol. 15:244.

Varshney RK, Thudi M, Nayak SN, Gaur PM, Kashiwagi J, et al. 2014. Genetic dissection of drought tolerance in chickpea Cicer arietinum L. Theor Appl Genet. 127:445–462.

Wang C, Tang S, Zhan Q, Hou Q, Zhao Y, et al. 2019. Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy. Nat Commun. 10:2982.

Wang H, Xu X, Vieira FG, Xiao Y, Li Z, et al. 2016. The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication. Mol Plant. 9:975–985.

Wang L, Wang A, Huang X, Zhao Q, Dong G, et al. 2011. Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theor Appl Genet. 122:327–340.

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, et al. 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature. 557:43–49.

Wei X, Qiu J, Yong K, Fan J, Zhang Q, et al. 2021. A quantitative genomics map of rice provides genetic insights and guides breeding. Nat Genet. 53:243–253.

Wu K, Xu X, Zhong N, Huang H, Yu J, et al. 2018. The rational design of multiple molecular module-based assemblies for

simultaneously improving rice yield and grain quality. J Genet Genomics. 45:337–341. S1673-8527(18)30085-7.

Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. 2014. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 46:100–106.

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, *et al.* 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell. 12:2473–2484.

Yao W, Li G, Yu Y, Ouyang Y. 2018. funRiceGenes dataset for comprehensive understanding and application of rice functional genes. Gigascience. 7:1–9.

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, *et al.* 1997. Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA. 94:9226–9231.

Yu SB, Li JX, Xu CG, Tan YF, Li XH, *et al.* 2002. Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. Theor Appl Genet. 104:619–625.

Zan Y, Carlborg Ö. 2020. Dynamic genetic architecture of yeast response to environmental perturbation shed light on origin of cryptic genetic variation. PLoS Genet. 16:e1008801.

Zeng D, Tian Z, Rao Y, Dong G, Yang Y, *et al.* 2017. Rational design of high-yield and superior-quality rice. Nat Plants. 3:17031.

Zhang B, Liu H, Qi F, Zhang Z, Li Q, *et al.* 2019. Genetic interactions among *Ghd7*, *Ghd8*, *OsPRR37* and *Hd1* contribute to large variation in heading date in rice. Rice (N Y). 12:48.

Zhang H, Wang X, Pan Q, Li P, Liu Y, *et al.* 2019. QTG-Seq accelerates QTL fine mapping through QTL partitioning and whole-genome sequencing of bulked segregant samples. Mol Plant. 12:426–437.

Zhang J, Chen L, Xing F, Kudrna DA, Yao W, *et al.* 2016. Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. Proc Natl Acad Sci USA. 113:E5163–E5171.

Zhang Z, Hu W, Shen G, Liu H, Hu Y, *et al.* 2017. Alternative functions of Hd1 in repressing or promoting heading are determined by Ghd7 status under long-day conditions. Sci Rep. 7:5388.

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali MH, *et al.* 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun. 2:467.

Communicating editor: P. Brown