

## Full Paper

# Identification of epistasis loci underlying rice flowering time by controlling population stratification and polygenic effect

Md. Asif Ahsan<sup>1</sup>, Md. Mamun Monir<sup>2</sup>, Xianwen Meng<sup>1</sup>,  
Md. Matiur Rahaman<sup>1,3</sup>, Hongjun Chen<sup>1</sup>, and Ming Chen<sup>1,2\*</sup>

<sup>1</sup>The State Key Laboratory of Plant Physiology and Biochemistry, Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China, <sup>2</sup>Institute of Bioinformatics, Zhejiang University, Hangzhou 310058, China, and <sup>3</sup>Department of Statistics, Faculty of Science, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj-8100, Bangladesh

\*To whom correspondence should be addressed. Tel. +86 571 88206612. Fax. +86 571 88206612.

Email: mchen@zju.edu.cn

Edited by Dr. Sachiko Isobe

Received 17 July 2018; Editorial decision 18 November 2018; Accepted 21 November 2018

## Abstract

Flowering time is an important agronomic trait, attributed by multiple genes, gene–gene interactions and environmental factors. Population stratification and polygenic effects might confound genetic effects of the causal loci underlying this complex trait. We proposed a two-step approach for detecting epistasis interactions underlying rice flowering time by accounting population structure and polygenic effects. Simulation studies showed that the approach used in this study performs better than classical and PC-linear approaches in terms of powers and false discovery rates in the case of population stratification and polygenic effects. Whole genome epistasis analyses identified 589 putative genetic interactions for flowering time. Eighteen of these interactions are located within 10 kilobases of regions of known protein–protein interactions. Thirty-seven SNPs near to twenty-five genes involve in rice or/and Arabidopsis (orthologue) flowering pathway. Bioinformatics analysis showed that 66.55% pairwise genes of the identified interactions (392 out of the 589 interactions) have similarity in various genomic features. Moreover, significant numbers of detected epistatic genes have high expression in different floral tissues. Our findings highlight the importance of epistasis analysis by controlling population stratification and polygenic effect and provided novel insights into the genetic architecture of rice flowering which could assist breeding programmes.

**Key words:** GWAS, epistasis analysis, population stratification, polygenic effect, rice flowering

## 1. Introduction

Rice (*Oryza sativa*) is one of the most important staple foods for a large part of the world's population and the main source of caloric intake. To meet the consumer demands for food by the growing

world population from 7.4 billion today to 9.1 billion by 2050, it is needed to increase the cereal production for the food security.<sup>1,2</sup> Grain production increases when crop plant flowers at the optimal time.<sup>3</sup> Flowering, the transition from vegetative stage to reproductive

stage, is controlled by complex internal genetic network and external factors (depends on different biotic and abiotic conditions).<sup>4</sup> Therefore, comprehensive understanding about the genetic control of flowering time is essential in crop breeding.

Flowering time is a complex trait, which is tightly governed by genetic factors, environmental cues as well as affected by population stratification.<sup>5,6</sup> One of the major goals of the modern genetics is identifying the genetic markers or factors those are associated with complex trait.<sup>7</sup> For revealing the genetic association of a trait, genome-wide association studies (GWAS) have emerged as one of the most powerful tool. In GWAS, single-nucleotide polymorphisms (SNPs) are typically examined for association across the genome with the trait of interest. Single-locus analyses are commonly used to estimate the marginal effects of individual SNPs. However, single-locus analysis could identify only the SNPs with relatively large effects, potentially miss the small effect SNPs.<sup>8,9</sup> Moreover, the loci identified by single-locus analysis collectively explain only a small fraction of genetic variation of complex trait, leading to the mystery of missing heritability.<sup>10,11</sup> Therefore, identification of causal genetic-interaction through epistasis analysis could possibly improve our understanding about genetic regulation of complex traits.<sup>12</sup>

For a pair of genes, each of them may have weak association or no association, but their interactions might have strong association with trait.<sup>13</sup> Moreover, epistasis effects could be affected by the additive effects of multiple genes and other environmental factors.<sup>14</sup> In addition, population stratification could confound the epistasis effect. As a consequence, the estimated effects could be biased upwards or downwards and standard error of the effects could be largely inflated. Simple epistatic models, those are not considering such important phenomenon may provide biased results.

In the past decades several methods and tools have been developed for studying epistasis. Among them PLINK,<sup>15</sup> FastEpistasis,<sup>16</sup> EpiGPU<sup>17</sup> are the widely applied tools for detecting epistasis underlying quantitative trait. These tools focused on parametric or nonparametric based linear regressions which do not control population stratification. In GWAS several methods have been proposed to control population stratification, including principal component analysis<sup>18</sup> and mixed linear model.<sup>5,19</sup> In epistasis analysis some studies also used principle components (PCs) as covariate for controlling population stratification<sup>20,21</sup> which could perform well when the pattern of population stratification is simple<sup>22</sup> but may perform poorly in presence of polygenic background effect or multiple level of relatedness.<sup>23</sup> However, mixed linear model could control both population stratification and polygenic effect in detecting epistasis which is considered as a best practice in GWAS but creates the additional computational burden and model complexity.

To reduce the computational cost and model complexity, the authors of GRAMMAR introduced a two-step approach as an alternative of mixed linear model for single-locus analysis.<sup>24</sup> Motivated from this idea, in this study we developed a two-step approach for whole genome epistasis analysis that can control additive polygenic background and population structure. Like GRAMMAR method, in the first step we calculated the breeding values and deducted from the trait values and finally used epistasis model instead of simple GWAS model. We have conducted extensive simulation studies to check the consequence of population stratification and polygenic effect in epistasis analysis and compared the performance of the method under different scenarios. Finally, we applied this method to analyse rice flowering time trait for detecting epistasis.

In epistasis analysis, detection of significant epistatic variants contribute to trait of interest is the primary and ultimate goal; however, further functional analysis of the interacting variants is necessary for characterizing the identified genes and constructing the biological link with phenotype.<sup>25</sup> Various functional assays such as gene ontology

(GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, protein–protein interactions (PPIs), subcellular location (SCL) of the gene/protein, tissue-specific gene expression and others biological information could help to find significant relationship among the candidate genes and phenotypes. With the rising availability of genomic data from different species, functional annotation of orthologous of the identified rice genes may help to reveal novel biological insights. A number of previous studies showed that a large per cent of interactions generally occur between proteins located in the same SCL or/and with a common functional assignment.<sup>26,27</sup> Localization and ontology of the candidate genes were used as important indicator for functional study and characterizing epistatic interactions. Moreover, epistasis interactions network could facilitate to reveal the underlying relations, biological mechanisms and important clustering information. We did literature curation to find previously reported flowering time-related genes, seek out various public databases for collecting PPIs and other biological information for intensive functional studies.

## 2. Materials and methods

### 2.1. Plant materials

Genotype and phenotype data used in this study were obtained from the rice diversity research platform ([www.ricediversity.org](http://www.ricediversity.org)). The population was recruited for a large-scale GWAS which included 413 diverse accessions of *O. sativa* at 36,901 SNPs of the Affymetrix Genome-Wide SNP Array after quality assurance screening.<sup>28</sup> All of the SNPs were selected by genotype call rate >70%, minor allele frequency >0.015. Individuals with missing phenotype were also removed from the study population. Missing genotypes were imputed with weighted *k*-Nearest-Neighbors method,<sup>29</sup> based on the five weighted nearest varieties present in the data set. Four hundred and thirteen *O. sativa* samples were used in the diversity panels that composed of six rice subpopulations, indica, aus, temperate japonica, tropical japonica, aromatic and admixed, which contains 87, 57, 96, 97, 14 and 62 accessions, respectively. Field data of flowering time were collected as the number of days until the inflorescence was 50% emerged from the flag leaf counted from the day of planting. The phenotype data used in this study for flowering time were measured at Faridpur, Bangladesh.

### 2.2. Statistical model

Multi-locus major and epistatic effects may control a complex trait. One of the major assumptions in GWAS is—complex traits are controlled by a large number of common variants with small effects. These genetic effects rarely exceed genome-wide significance threshold. Ignoring effects of the multiple common genetic variants may have large impact on epistasis analysis results. Controlling background genetic effects in time of testing main and epistatic effects were largely studied in QTL mapping era.<sup>30–32</sup> Significant background genetic markers selected by using stepwise regression are used as cofactors for controlling genetic background. Phenotype-adjusting method<sup>30</sup> was proposed for QTL analysis, and showed that adjusted approach (inclusive composite interval mapping, ICIM) has more power than cofactor model (composite interval mapping, CIM) approach. In GWAS, mixed model approach can control the effects of common variants, where background effects are considered as random. Best linear unbiased prediction using genetic relationship matrix of individual observations exhaust the effects of common variants<sup>33</sup> and the prediction generally highly correlate with phenotype.<sup>34</sup> We have used a two steps approach for epistasis analysis: (i) predicting the total additive genetic breeding value using mixed model approach and adjusting phenotypic data, and (ii) using the adjusted phenotypic data for whole genome epistasis analysis.

If  $p$  individual loci with main effects and  $q$  pairs of loci with epistatic effects control a complex trait then statistical genetic model for additive and epistasis effects can be written as

$$y_j = \mu + \sum_{k=1}^p a_k x_{A_{kj}} + \sum_{k=1}^q i_{aa_k} x_{AA_{kj}} + \varepsilon_j,$$

where  $y_j$  is the phenotype of the  $j$ th individual observation,  $\mu$  is the general mean,  $a_k$  is the additive effect of  $k$ th locus;  $i_{aa_k}$  is the epistasis effect; individual level genotypes were coded for additive effects as  $x_{A_{kj}} = 2$  for  $QQ$ , 1 for  $Qq$  and 0 for  $qq$ , and for epistasis effects as  $x_{AA_{kj}} = 4$  for  $QQ \times QQ$ , 2 for  $QQ \times Qq$ , 1 for  $Qq \times Qq$  and 0 for  $QQ \times qq$ ,  $Qq \times qq$  and  $qq \times qq$ ; and  $\varepsilon_j$  is the random error.

In matrix notation the above equation can be written as

$$\mathbf{y} = \mathbf{X}_B \mathbf{b} + \mathbf{X}_A \mathbf{a} + \mathbf{X}_{AA} \mathbf{i} + \boldsymbol{\varepsilon},$$

where  $\mathbf{X}_A$  and  $\mathbf{X}_{AA}$  are the design matrices for additive and epistasis effects, respectively.

In real situation the number and chromosomal locations of causal loci are unknown, and need to use a powerful statistical approach to identify them from huge number of loci. By searching the causal loci for a complex trait, we can identify a few numbers of detectable causal loci and large number of causal loci with relatively small effects might remain undetectable.<sup>34,35</sup> Therefore estimating the total genetic effects for additive effect loci by searching significant loci is practically infeasible for GWAS. However, we can estimate the total additive genetic effects by utilizing additive kinship matrix via mixed model approach.<sup>19,35</sup> For example, the total genetic effects due to additive effects of loci can be predicted by the linear mixed model approach

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g} + \mathbf{e}.$$

In this case,  $\mathbf{X}_A \mathbf{a} \approx \bar{\mathbf{g}}$ ,  $\mathbf{g} \sim N(0, \mathbf{K}_A \sigma_A^2)$  and  $\mathbf{K}_A$  are additive kinship matrix. After predicting the total estimated effects for additive main effects, we can calculate the adjusted phenotype as

$$\Delta \mathbf{y} = \mathbf{y} - \bar{\mathbf{g}}.$$

The adjusted phenotypic data was used for epistasis analysis and the statistical model for epistasis analysis is

$$\Delta y_j = \mu + a_1 x_{A_{1j}} + a_2 x_{A_{2j}} + i_{aa} x_{AA_j} + \varepsilon_j.$$

We compared the approach with PLINK and PC-linear approaches in terms of power and false discovery rate (see Methods S1 and S2). We conducted Monte-Carlo simulation study for checking performance of the approaches. Details about of genotype and phenotype simulations were discussed in Methods S3 and S4.

### 2.3. Gene annotation of the flowering time associated epistatic SNPs

In total, 589 interactions comprised 499 SNPs were identified by whole genome epistasis analysis (Supplementary Data S1). We used MSU Rice Genome Annotation Project (RGAP) Release 7 (<http://rice.plantbiology.msu.edu/>) database to annotate the identified 499 SNPs (Supplementary Data S2). Among them 297 SNPs were annotated with protein-coding genes and rest of them were non-coding (Supplementary Table S1). For the non-coding SNPs the nearest genes were used for functional characterization. NCBI (<https://www.ncbi.nlm.nih.gov/>) database was used for gathering more information about identified genes (Supplementary Data S3).

### 2.4. GO and pathway enrichment analysis

To determine whether the annotated genes were enriched for biological or functional significance, GO enrichment analysis was performed using GO analysis toolkit implemented in CARMO.<sup>36</sup> A gene set was considered as significantly enriched for GO terms if  $P < 0.05$ . The GO treemap showing the biological process (BP) was generated using REVIGO.<sup>37</sup> We also performed KEGG pathway enrichment analysis. For KEGG enrichment analysis, Rice Information GetWay (RIGW) database was used.<sup>38</sup>

### 2.5. PPI search

A large number of rice PPIs were collected from three databases, PRIN,<sup>39</sup> RIGW<sup>38</sup> and RicePPINet.<sup>40</sup> We mapped our detected gene-gene (SNP-SNP) interactions to known PPIs and we hypothesized that some of the 589 epistasis interactions may arise from PPIs. We found 18 pairs of SNP interactions those were within 10 kb of known PPIs.

### 2.6. Orthologous gene and flowering pathway gene

Gene orthologous between rice and Arabidopsis were obtained from MSU RGAP Release 7 (<http://rice.plantbiology.msu.edu/>). Genes involved in flowering time, flower development and seed development pathway for rice and/or Arabidopsis were downloaded from the database.<sup>41,42</sup>

### 2.7. Subcellular localization prediction

Web-based integrative SCL predictor tool called plant subcellular localization integrative predictor (PSI)<sup>43</sup> was used to predict the SCL of the candidate epistatic genes. Finally the genetic network from SNP-SNP interactions projecting various biological information was visualized using Cytoscape 3.5.1.<sup>44</sup>

### 2.8. Tissue-specific expression

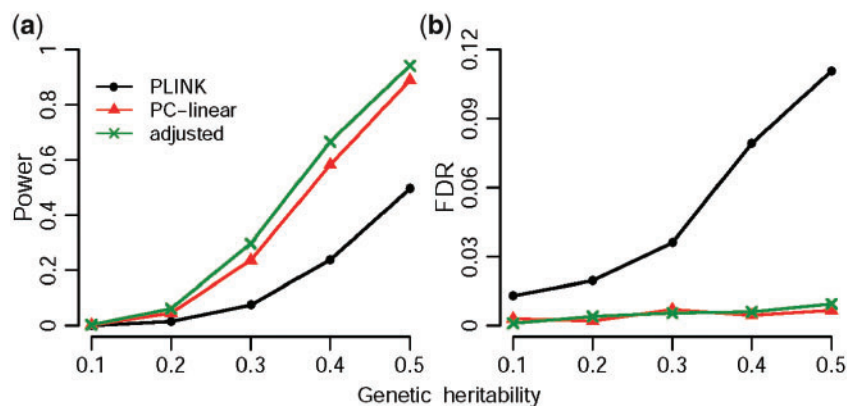
We explored the expression profile of the candidate genes in the six different tissues (i.e. post-emergence inflor, pre-emergence inflor, embryo-25DAP, anther, pistil and panicle). Tissue-specific expression of the epistatic genes were obtained from the comprehensive annotation of rice multi-omics data (CARMO)<sup>36</sup> annotation platform. The heatmap plot of the expression profile of the genes with dot plot of SCL was constructed using iTOL v3.<sup>45</sup> The clustering of the genes presented in the heatmap represents hierarchical clustering.

## 3. Results

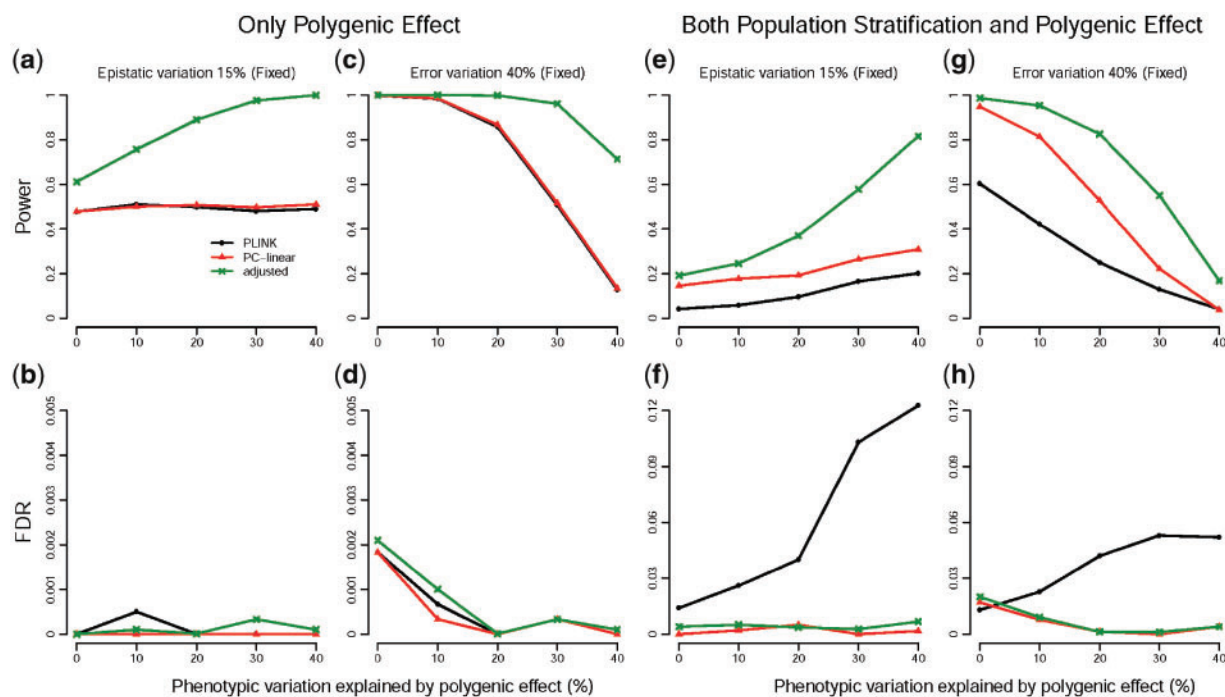
### 3.1. Simulation results

We used an adjusted statistical approach to detect epistasis in presence of complex polygenic background and population structure. The adjusted approach was compared with two methods: (i) the simple linear regression that implemented in PLINK which do not account population stratification and polygenic effect and (ii) principal component (PC) based linear regression where PCs are used as covariate to control population stratification (Methods S1 and S2). These two methods were abbreviated as PLINK and PC-linear, respectively. Simulations under different scenarios were conducted for comparing the methods in terms of statistical power and FDR (Methods S4). At each scenario, simulations were performed 1,000 times to estimate the average power and FDR.

To investigate the effect of population structure for detecting epistasis, we applied previously mentioned three methods to the simulated data that generated from heterogamous population. We calculated the power and FDR for the three methods with varying



**Figure 1.** Model comparisons under structured population. (a) Power and (b) FDR comparison at different genetic heritability for structured population and samples were considered from three different populations (1,000 sample size with 400, 300 and 300 sample for three populations, respectively). This figure is available in black and white in print and in colour at *DNARES* online.



**Figure 2.** Model comparisons under assumption of (i) polygenic effect (ii) both population structure and polygenic effect. Power and FRD for (a, b) 15% fixed epistatic variance and; (c, d) 40% fixed error variance under the assumption of polygenic effect. Power and FRD for (e, f) 15% fixed epistatic variance and; (g, h) 40% fixed error variance under the assumption of both population stratification and polygenic effect. Different proportions of polygenic effects (0–40%) were varied presented in x-axis. In all cases the additive variance and epistatic variance were equal. With fixed epistatic variance to 15%, the error variance was 70% when the polygenic variance was 0% and error variance was 30% when polygenic variance was 40% (Supplementary Table S8). Again with fixed error variance to 40%, the epistatic variance was 30% when polygenic variance was 0% and epistatic variance was 10% when polygenic variance was 40% (Supplementary Table S9). This figure is available in black and white in print and in colour at *DNARES* online.

per cent of genetic heritability and three subpopulation structures ( $k = 3$ ). Figure 1 shows the effect of population stratification on power and FDR for different genetic heritability. With increasing rates of genetic heritability, the powers of all three methods were increased. It was observed that detection power could be significantly decreased without controlling population stratification. The use of PC as covariate for PC-linear approach could increase the statistical power. Similar increasing rate of powers was also observed for the adjusted method. Besides, it was observed that with increasing rate of the genetic heritability, the FDR of the PLINK was increased,

however FDR for PC-linear and adjusted approach were under control. These results suggested that, both power and FDR of classical approach could be poor when the samples come from structured population (Supplementary Table S2). PC-linear and adjusted method could effectively control FDR as well as improve the detection power in presence of population stratification.

Next, the influence of polygenic effect on epistasis detection was assessed under homogeneous population (scenario-II). For this purpose, polygenic effects were added to the simulated phenotypes, where their contributions to the phenotypic variation were varied

(Methods S4). In this scenario, with increasing rate of variance due to polygenic effects (decreasing error variance) the powers of the adjusted method were significantly increased (Fig. 2a and Supplementary Table S3). It increased from 61% to 99% with respect to increasing rate of polygenic variation. Powers of PLINK and PC-linear were mostly constant (within 47–51%), because the epistatic variance was fixed. This result advocating that both PLINK and PC-linear methods cannot account polygenic variance and treated as error variance. For case (b) in scenario-II [fixed error variance (40%)], with increasing rate of polygenic variation, the powers were reduced due to decreasing the epistatic variance. However, in this case the adjusted method also performed better. The results in both cases of scenario-II, clearly suggest that, epistasis detection powers of the PLINK and PC-linear approaches could largely reduce due to increasing rate of polygenic effect variation (Fig. 2c and Supplementary Table S4). However, the FDR were almost similar for all of the methods (Fig. 2b and d).

Finally, we investigated the consequence of both polygenic effects and population stratification for epistasis analysis (scenario-III). The results for two different cases: (a) the epistatic variance fixed to 15% and (b) error variance fixed to 40%, were presented in (Fig. 2e and h, Supplementary Tables S5 and S6). The patterns of power comparisons of this scenario were very similar to scenario-II. The difference was that, there had impact of population stratification on the detection power in the cases of scenario-III. In this scenario, the powers were smaller for all methods as compared with scenario-II. Although PC-linear method cannot capture polygenic effect, the power of this method was larger than PLINK because of controlling population stratification. However, adjusted method had superior power due to capturing the polygenic variations and controlling population stratification. The FDR of PC-linear and adjusted methods were reasonable, but PLINK had inflated results (Fig. 2f and h).

### 3.2. Genome-wide epistasis analysis of rice flowering time

We analysed the flowering time trait to identify the epistatic loci influencing this complex trait. Previous study analysed this trait to identify only the underlying individual loci.<sup>28</sup> It was surprising that the previous analysis identified only 2 loci which explain 5% genetic variation, referring most of the variations may come from others types of genetic variants or environment factors. We were interested to identify the epistatic loci for flowering time trait. We used two-step adjusted approach for analysing this complex trait by controlling population stratification and polygenic effects of multiple individual loci (see Method). One of the assumptions about complex trait is that it could be controlled by multiple individual loci with relatively small effects. Before analysing epistatic effects our goal was to control those effects to improve detection power and reducing the FDR. With extensive simulation studies, we showed that adjusting population stratification and polygenic effects could improve the detection power and reduce FDR (Figs 1 and 2).

Since, the diverse accession of *O. sativa* is from admixed population, so it is expected that epistasis analysis using classical model could be bias due to population stratification. We conducted whole genome epistasis analysis using adjusted method by considering the polygenic effect and population structure. Due to small sample size a liberal significance threshold  $P < 9.98 \times 10^{-8}$  was used for summarizing the results.<sup>10</sup> We identified a total of 589 pairs of SNPs among the 680823450 possible SNP interactions (Supplementary Data S1). We also analysed the trait by using PLINK and PC-linear approaches. We constructed QQ-plot for whole genome *P*-value for three approaches. QQ-plot showed that the PC-linear and adjusted methods better fitted as compared with PLINK approach (Supplementary Fig. S1).

Altogether the discovered 589 epistatic interactions comprised 499 unique SNPs. As the number of unique SNPs are smaller than the number of interactions there should have some hub SNPs those are interacted with many other SNPs. The hub SNPs could be biologically important for the flowering time trait.<sup>46</sup> Of the discovered 499 unique epistatic SNPs, more than half (59.51%) were located within known annotated genes (Fig. 3a, d and e, Supplementary Table S1 and Data S2) and approximately one-third (31.86%) were located in chromosome 1 (Fig. 3d, Supplementary Data S2 and S3).

We observed that, among the identified SNPs, 9 (1.8%) had MAF less than 0.05 and more than 50% were in the range between 0.1 and 0.3 (Fig. 3b and Supplementary Fig. S2). Among the total identified interactions, 26.32% (155 out of 589) were *cis*-chromosomal interactions through five chromosomes (1, 2, 4, 5 and 11) and the rest 73.68% were *trans*-chromosomal interactions. Of the 155 *cis*-chromosomal interactions, 136 were in chromosome 1 (Fig. 3c and Supplementary Fig. S3).

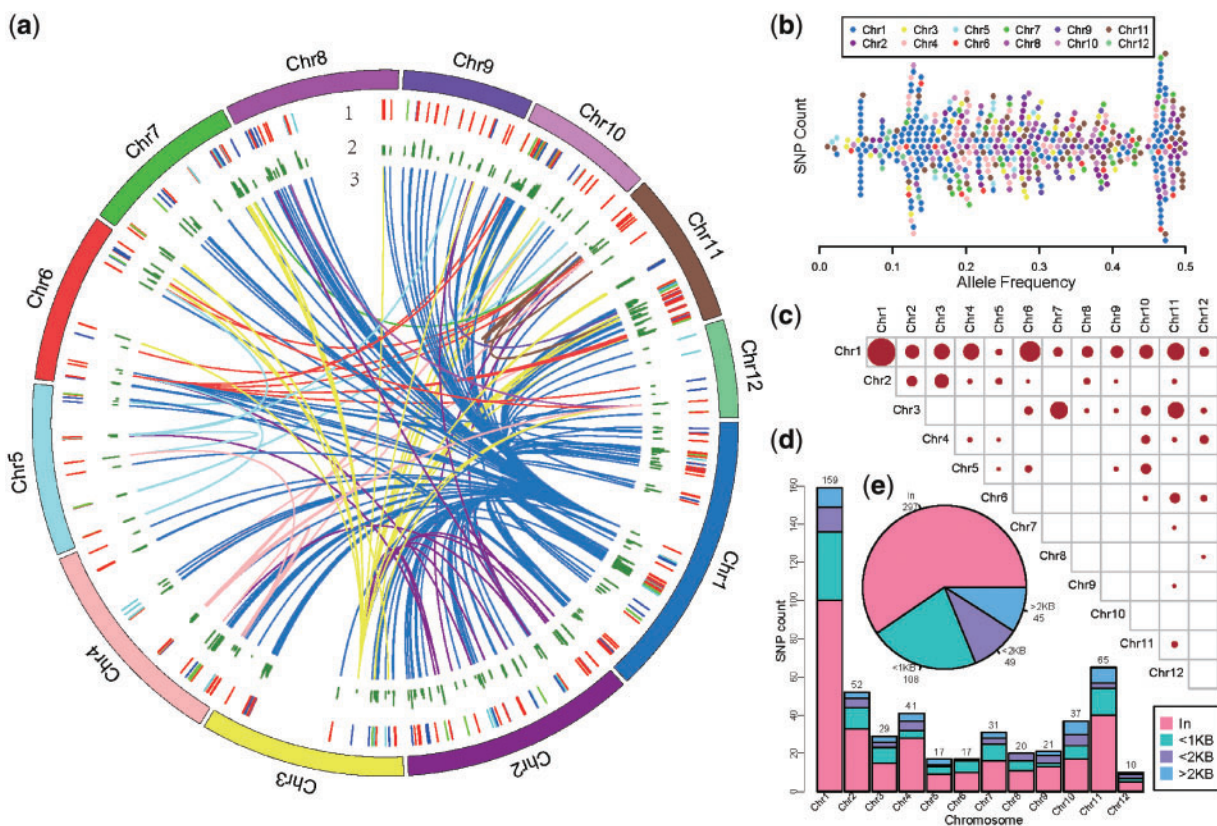
### 3.3. Candidate epistatic genes involved in different pathways

Gene enrichment analysis was performed to confirm the involvement and the potential contribution of candidate genes to flowering time and flower development. To do this, genes involved in different flower related pathways in rice and Arabidopsis were obtained and mapped to know whether the genes have relevant functions with flowering related traits (Table 1 and Supplementary Data S4). Among the detected genes nine were involved in rice flowering time pathway and fourteen were in rice seed development pathway. One gene (LOC\_Os07g41370, *MADS18*) was found which involved in all five pathways (Table 1). We have detected several genes which were not previously reported for rice flowering, however their Arabidopsis orthologues have biological function in regulating flowering time or flower development (Table 1). For example, the gene LOC\_Os08g42640 was not found in rice flowering time pathway, but its orthologue *RFI2* (AT2G47700) was found in Arabidopsis flowering time pathway. From Table 1 it is shown that, among the identified genes ten (Arabidopsis orthologue) were involved in flowering time pathway and four in flower development pathway. These results are suggesting the potential role of epistasis analysis in detecting novel genes.

### 3.4. Functional enrichment analysis of the epistatic genes

To characterize the identified epistatic genes various functional enrichment analyses were performed. First, we performed GO enrichment analysis and assessed whether the genes mapping to epistatic loci are enriched for GO terms. Significant overrepresentations ( $P < 0.01$ ) were observed for the BP terms ‘defense response’, ‘response to stress’, ‘apoptotic process’ and ‘signal transduction’ (Supplementary Data S5). Since only few BP terms were found as significant, treemap analysis was performed to represent the overall view of BP terms of the identified epistatic genes. As expected, many of the GO terms ‘flower development’, ‘multicellular organism development’, ‘reproduction’, ‘DNA metabolism’, ‘cellular process’ (Fig. 4) were found. We also compared the BP terms of the candidate genes participating in interaction for rice flowering time trait and gold standard genes those involved in flowering time pathway and found a large portion of common BP terms (Fig. 4, Supplementary Fig. S4 and Data S6).

We also investigated the overrepresentation of KEGG biochemical pathway enrichment analysis. The most representative pathway was



**Figure 3.** Overview of the identified epistatic loci for rice flowering time. (a) The Circos plot represents the interaction of the 499 unique SNPs that comprise the 589 epistatic interactions. The outer track shows 12 chromosomes levelled by different colours. The other tracks present (1) line plot of the location of the candidate genes of identified SNPs, (2) Line plot of the minor allele frequency (MAF) of the SNPs. The range of the MAFs are 0.015–0.5 and more than half of the SNPs are in the range between 0.1 and 0.3 (Supplementary Fig. S2). Track (3) presents the detected interactions through all chromosomes. (b) The distribution of the minor allele frequency. (c) The distribution of interactions through all chromosomes. Chromosome wise (d) and overall (e) distribution of location of the identified epistatic SNPs. This figure is available in black and white print and in colour at *DNARES* online.

observed for DNA replication (ko03030;  $P < 0.0015$ ). DNA replication is the first and vital process that occurred during cell cycle and cell proliferation and this process is highly related to plant cell growth and development. More recently some studies have established a connection of DNA replication, H3.1 and H3K27me3 to flowering time in Arabidopsis.<sup>47,48</sup> Because of the reduction of flow of canonical histone H3.1, propagation of H3K27me3 in the *FLC* (*FLOWERING LOCUS C*-a repressor of flowering) is affected during vernalization and flowering is delayed. Through vernalization the histone variant H3.1 facilitates H3K27me3 to take place at *FLC* locus by DNA replication<sup>47</sup> indicating towards the probable essential role of DNA replication in rice flowering. Another representative pathway term ‘phenylalanine metabolism’ was found and it involves in biosynthesis and metabolism of amino acids, including aromatic amino acids and these acids play important roles in plant growth, development, reproduction, defense and environmental stimuli.<sup>49,50</sup> We also found some KEGG pathway terms had  $P$ -values above the threshold of 0.05 [e.g. ko04144: endocytosis ( $P = 0.115$ ), ko00940: phenylpropanoid biosynthesis ( $P = 0.131$ ), ko00230: purine metabolism ( $P < 0.224$ )] (Supplementary Data S7). Ribosome biogenesis in eukaryotes and endocytosis; phenylpropanoid biosynthesis and purine metabolism pathway terms were reported for response to early chilling stress of rice<sup>49</sup> and response to vernalization of Oriental lily,<sup>51</sup> respectively.

### 3.5. Epistatic interactions reveal genetic network

A genetic network was constructed from the identified 589 epistatic interactions and integrated various biological features. By analysing the network, several genes were identified as hub genes based on node degree (Supplementary Table S7). The SNP rs18202417 located in the LOC\_Os01g39100 gene was the top hub node (gene) with degree 309 (Fig. 5 and Supplementary Data S1). This gene encodes a protein containing zinc finger CCH domain and involved in many BPs such as reproduction, embryo development and post-embryonic development. Another major hub SNP (rs351342502) was detected with node degree 66 that located in the gene LOC\_Os06g08520 (Fig. 5 and Supplementary Data S1). We also observed that, both hub SNPs had common interactions with other 36 SNPs (Fig. 5). Moreover, based on the topological structure of the network, five major groups were observed those connected with hub genes (Fig. 5). From the network we noticed that, majority of the gene interactions in G3 and G4 is responsible for delay flowering (dashed line), while all of the interactions in G2 and G5 were responsible for early flowering (solid line) as compared with average flowering time. In G1, only two interactions were found for delay flowering among the total of 309 interactions. Two hub SNPs (rs19539809 and rs19540019) in G4 are located in chromosome 3 and both the variants are near of the gene LOC\_Os03g15460. The gene is localized in cytosol and involved in the molecular function

**Table 1.** Detected epistatic genes involved in rice and Arabidopsis (orthologue) flowering time or related pathways

SNP	Gene	OS_FT <sup>a</sup>	OS_SD <sup>b</sup>	AT_FT <sup>c</sup>	AT_FD <sup>d</sup>	AT_FL_ID <sup>e</sup>	AT orthologue <sup>f</sup>	Gene symbol <sup>g</sup>
rs53491160	LOC_Os01g10504	√	√		√		AT4G18960	MADS3/AG
rs350793833	LOC_Os01g12890			√	√	√	AT5G11530	EMF1
rs350793833	LOC_Os01g12900		√				NA	RAC
rs348030366	LOC_Os01g13740					√	AT2G20570	GLK1,GPPI1
rs350255043	LOC_Os01g49690	√		√			AT1G50370	FYPP3
rs18797452	LOC_Os01g49830		√	√			AT1G13260	RAV1
rs18797497								
rs350507540	LOC_Os01g51300	√		√			AT2G19520	FVE,MSI4
rs351226054	LOC_Os01g73580		√				NA	CIN4
rs347877916								
rs348843117								
rs352509847								
rs348111382								
rs350433325	LOC_Os01g73770	√					NA	DREB1E
rs18768624	LOC_Os02g02290	√					NA	SNF2L
rs348828354								
rs348442835	LOC_Os02g02380					√	AT5G23730	RUP2,EF02
rs18770180								
rs352920030								
rs350623682	LOC_Os02g05030		√				NA	SPP2
rs350780688								
rs17921993								
rs353005292	LOC_Os03g08460		√				NA	EBP89
rs352497391	LOC_Os03g08754	√	√	√		√	AT2G22540	MADS47/SVP
rs19214075	LOC_Os03g09310	√		√		√	AT5G16320	FRL1,SUF8
rs351154822	LOC_Os04g55560	√		√	√	√	AT4G36920	AP2
rs19734267	LOC_Os05g33570		√				NA	PPDKB
rs351342502	LOC_Os06g08530		√				NA	bip110
rs350150232	LOC_Os07g41370	√	√	√	√	√	AT5G60910	MADS18/FUL,AGL8
rs351408615								
rs352265155								
rs20215256	LOC_Os07g42300		√				NA	EF-1-d1
rs347572446	LOC_Os08g42640			√		√	AT2G47700	RF2
rs348156805	LOC_Os11g16470		√				NA	MLA10
rs21765814	LOC_Os10g30100					√	AT5G35910	RRP6L2
rs349850721	LOC_Os11g32110		√				NA	ARF1
rs20841419	LOC_Os11g41820		√				NA	U2 snRNP
rs348908455								
Total = 37	25	9	14	9	4	10		

<sup>a</sup>Genes involved in rice flowering time pathway.

<sup>b</sup>Genes involved in rice seed development pathway.

<sup>c</sup>Genes involved in Arabidopsis flowering time pathway.

<sup>d</sup>Genes involved in Arabidopsis seed development pathway.

<sup>e</sup>Genes involved in Arabidopsis flowering time gene network collected from FLOR-ID.

<sup>f</sup>Arabidopsis orthologous of the corresponding rice gene.

<sup>g</sup>Gene symbol of the rice and Arabidopsis identifier.

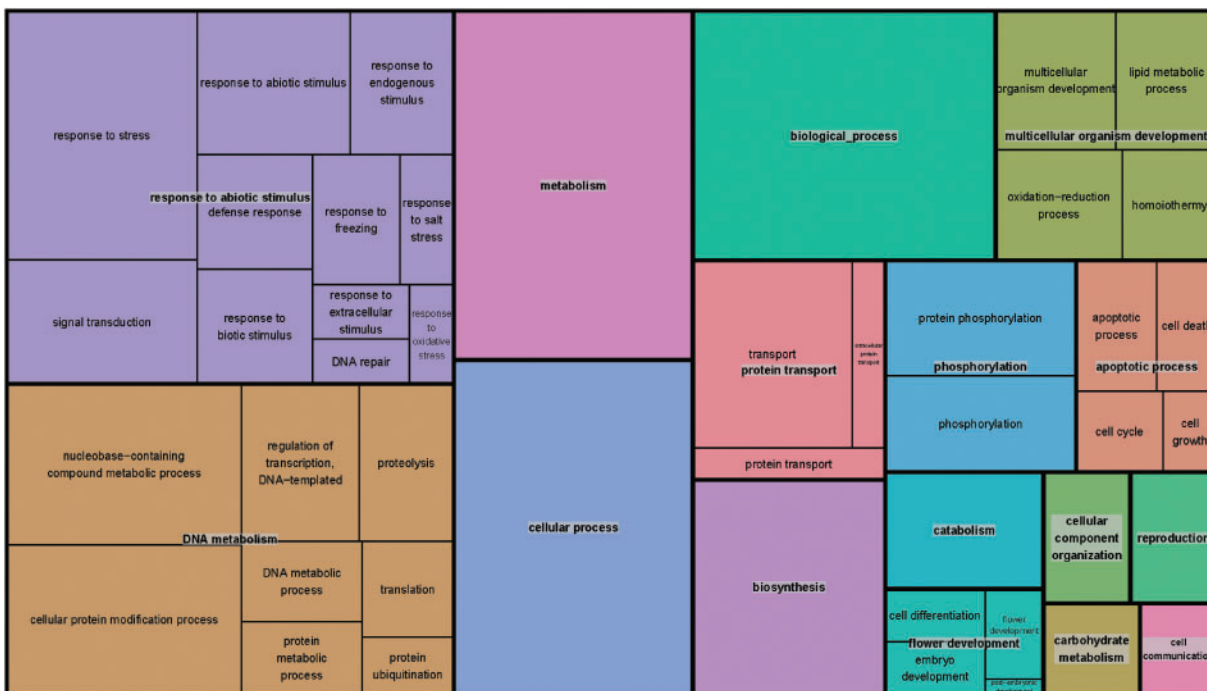
(MF) phospholipase A2 activity and orthologue to Arabidopsis phospholipase A2 (*PLA2*) gene. The MF of *PLA2* genes in rice is still poorly known. However, its orthologue plays role in jasmonic acid (JA) biosynthesis, pollen maturation, anther dehiscence, and flower opening in Arabidopsis.<sup>52,53</sup> G3 contains 38 SNPs, which are near or within 28 genes, and among the genes 67.85% localize in plastid (Supplementary Fig. S5a), suggesting that perhaps most of the genes of the group involve in photosynthesis. No nuclear gene was found in that group.

GO enrichment analysis on the basis of genes containing in the five groups revealed that different groups genes are involved in different BP (Supplementary Data S5). As for example, the genes in G1 were significantly enriched for chloride transport, apoptotic process and others, while genes in G2 enriched for wounding and oxidative

stress-related process. One and three GO BPs were enriched for G3 and G4, respectively. At significance threshold  $P < 0.05$ , we found no enriched BP and other GO terms for G5. The lack of overrepresentation for G5 is likely to be only few genes was involved in this group.

### 3.6. Co-location and functional similarity of interacting genes

Interacting genes or proteins are expected to have same biological function or located in same compartment. We therefore used GO functional similarity and colocalization as important indicator for further evidence of detected epistatic genes. For assessing whether the interacting genes share similar BP, cellular component (CC) and MF,



**Figure 4.** Gene ontology biological process treemap. Gene ontology biological process treemap of the annotated genes for rice flowering time. Sizes of the rectangles are adjusted on the basis of the frequency of the GO terms (Supplementary Data S6). This figure is available in black and white in print and in colour at *DNARES* online.

47, 96 and 115 pairs of interacting genes were found with similar BP, MF and CC, respectively (Fig. 5c and Supplementary Data S8). Among them, many of the interacting genes have shared more than one domain. As for example, 19 pairs of gene interactions have shared both BP and CC, 13 pairs shared both BP and MF and 28 pairs shared CC and MF (Fig. 5c). Moreover, 61 pairs of interactions were found those have shared both SCL and GO terms and most of them were for CC. From our results we found that, many genes localized to multiple compartments (Fig. 6, Supplementary Figs S6 and S7) and many of the interactions of genes were found located in cytosol and plastid (Supplementary Fig. S8).

### 3.7. Expression profile of the candidate genes

To further characterize and confirm the involvement of the epistatic genes in rice flowering, expression analysis was performed. For this purpose gene expression of six tissues, i.e. post-emergence inflor, pre-emergence inflor, embryo-25DAP, anther, pistil and panicle were obtained and plotted using heatmap (Fig. 6 and Supplementary Fig. S6). Expression analysis was conducted separately for the genes those clustered into five groups in the network (Fig. 5a). We applied hierarchical clustering to cluster the genes. From the heatmap plot it was observed that, in every group some genes highly express across all six tissues. We also noticed a large per cent of candidate genes in every group were preferentially expressed in panicle (Fig. 6 and Supplementary Fig. S6). These results support the evidence of the involvement of the identified epistatic genes in rice flowering.

### 3.8. PPIs corresponding to identified epistatic interactions

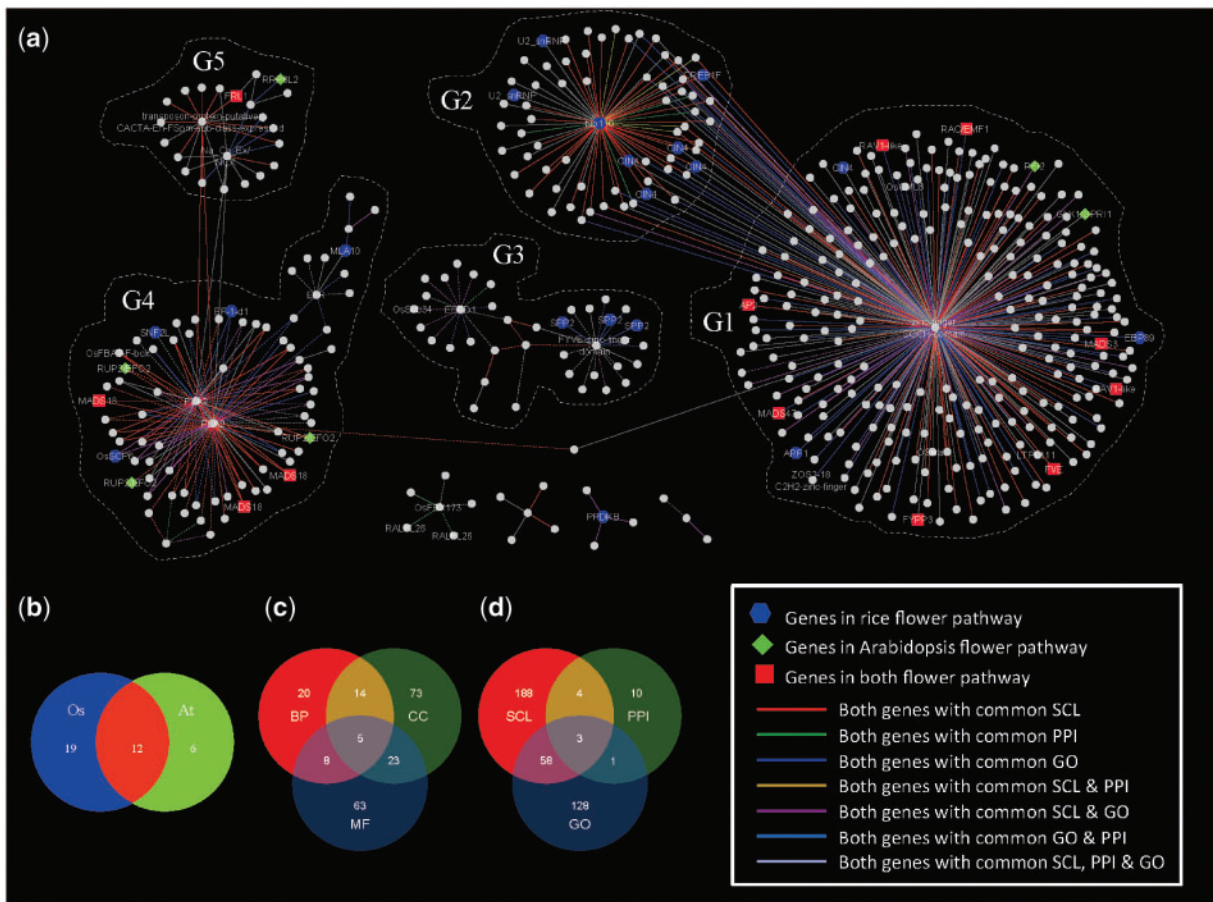
We further investigated whether there have overlap between detected epistasis and known (PPIs) in the chromosomal location. Known rice

PPIs obtained from three databases, PRIN,<sup>39</sup> RIGW<sup>38</sup> and RicePPINet<sup>40</sup> were used (see Method) to search overlap between detected epistasis and the biological interaction. We found several pairs of genes harbouring SNPs interactions which were also predicted for PPI (Fig. 5). Eighteen pairs of SNP interaction comprise to 12 pairs of gene interactions were overlapped to 3 different databases. Of these interactions, three, five and six pairs of interactions were overlapped with the PPIs of three databases RicePPINet, RIGW and PRIN, respectively (Supplementary Data S9). Among all overlapped interactions two pairs were common between RicePPINet and PRIN and only chromosome 11 had four pairs of *cis*-chromosomal interactions (Supplementary Data S9). RicePPINet also provided other additional information such as interaction probability and co-expression between two interacting proteins (Supplementary Data S9). The PPI of the two genes LOC\_Os11g07120 and LOC\_Os11g36190 showed higher probability (0.884) and the remaining two interactions have moderate probability (>0.5). In the interaction LOC\_Os11g07120 × LOC\_Os11g36180, the level of co-expression is more than 0.5 (Supplementary Data S9). These results clearly reinforced the validity of our findings.

## 4. Discussion

For a complex trait like flowering time, it is crucial to discover how multiple genes regulate the trait. Epistasis analysis is recognized as powerful technique for identifying novel genes and improves our understanding in the genetic regulation of complex trait.<sup>8,34,55,56</sup> However, some reports argue the relative importance of epistasis because of the low contribution of genetic variation.<sup>12,57,58</sup> The lack of success to dig out big amount epistatic variance does not imply non-existence of the epistatic gene action.<sup>59</sup> Low contribution of epistasis could be due to considering epistasis model for only the loci with





**Figure 5.** Network of epistasis interaction for rice flowering time. (a) A Cytoscape network generated from SNPs interactions inferred following proposed method for the rice flowering time. The nodes are represented to the SNPs and the shape of the nodes are symbolised on the basis of annotated genes those involve in the rice or Arabidopsis flower related pathway: octagon (involved in rice flowering time or seed development pathway), diamond (involved in Arabidopsis flowering time or flower development pathway) and rectangle (involved in both rice and Arabidopsis flower related pathway). The edges are coloured on the basis of the connected genes located in similar subcellular location or overlap with same GO terms and/or predicted for PPI and/or overlap of the three genomic features (Supplementary Fig. S7, Data S8 and S9). Solid line for the edge indicating early flowering and dash line indicating delay flowering. (b) Rice and Arabidopsis genes involved in flower pathway. (c) The overlap of GO domain and epistatic genes (d) the overlap of interacting genes in subcellular location, gene ontology and PPI. This figure is available in black and white in print and in colour at *DNARES* online.

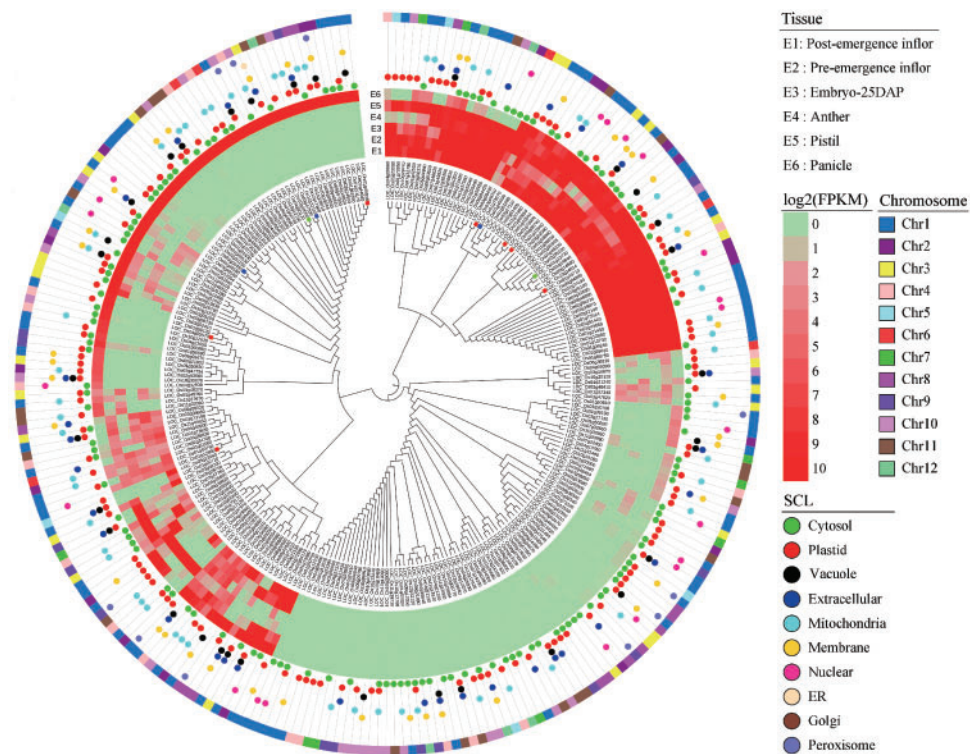
moderate or high marginal single-locus effects. Another important reason could be avoiding the confounding due to population stratification and polygenic effect in detecting epistatic interaction that leading to false-positive or false-negative results.<sup>14,19</sup> Therefore, exhaustive epistasis analysis by controlling population stratification and polygenic effect is needed to uncover the underlying structure, genetic pathways and understanding the function of complex genetic system.<sup>12</sup>

In this study, we used an adjusted epistasis analysis approach to identify significant epistasis interaction under the condition of polygenic effects and population stratification. Extensive simulation studies showed that adjusted approach could improve epistasis analysis results over classical approach in terms of power and FDR in the case of polygenic effects and population stratification (Figs 1 and 2, Supplementary Figs S11–S13, and Tables S1–S5). Epistasis detection powers were relatively low under all simulation scenarios for classical model implemented in PLINK. Under all scenarios, the FDR of PLINK method was higher than other two methods except the scenario considering only polygenic effect (Fig. 2b and d). In this scenario no significant difference was observed in FDR between models. In previous GWAS also reported that the tests could control the Type I error rates satisfactorily.<sup>60</sup> These results refer that the effect of

polygenicity is responsible for repress the main effect as well as epistasis effect which decrease the detection power if not controlled.

For reducing the computation cost, most of the time epistasis analyses have been conducted for the SNPs, which are nominally significant in the single-locus analysis.<sup>61,62</sup> We performed whole genome pairwise epistasis analysis for rice flowering time and identified 589 epistatic interactions comprised 499 SNPs. By comparing to previous study result,<sup>28</sup> we observed that none of the identified SNPs in epistasis analysis were detected in the single locus analysis (Supplementary Fig. S9), advocating the necessity of whole genome epistasis analysis.

Different genomic features such as GO, biological pathway, PPIs, SCL and tissue-specific expression can be used to characterize the identified epistatic loci and these analyses could provide new insights into the biological function of the candidate genes. For example, GO and KEGG pathway enrichment analysis of the candidate genes showed that signal transduction, response to stress, DNA metabolic process are most significant GO BP terms, while DNA replication, phenylalanine metabolism and ribosome biogenesis are the most relevant pathways (Supplementary Data S5 and S7). Several genes were not statistically significant in enrichment analysis but are involved in



**Figure 6.** Subcellular location and tissue-specific expression. Subcellular location and tissue-specific expression of the candidate genes involved in G1 of the network. Colour strip represents for chromosome, dot plot for predicted subcellular locations (cytosol, plastid, vacuole, extracellular, mitochondria, membrane, nuclear, ER: endoplasmic reticulum, Golgi, golgi apparatus and peroxisome) and heatmap for six floral tissues (E1: Post-emergence inflor, E2: Pre-emergence inflor, E3: Embryo-25DAP, E4: Anther, E5: Pistil and E6: Panicle) specific expression. Dendrogram showing clustering (hierarchical clustering) of genes in heatmap is on the basis of tissue-specific expression. The colour scale bar of the figure represents  $\log_2$  transformed FPKM values. Square on the dendrogram represents the corresponding annotated genes of the SNPs in Table 1. This figure is available in black and white in print and in colour at *DNARES* online.

the BP cell cycle, cell growth, cell death, flower development, embryo development and metabolism (Fig. 4) which are closely related to rice flowering. The pathway DNA replication was found as significant which helps H3K27me3 to take place at FLC during vernalization and promotes flowering<sup>47,48</sup> while phenylalanine metabolism involves in biosynthesis and metabolism of amino acids those might play vital roles in plant growth, development and reproduction.<sup>49,50</sup> These enrichment analysis results indicate the involvement of the candidate genes in flowering development.

Epistatic interactions represented by network can reveal the global picture of gene connections. In our result it was observed that, interacting genes formed hub sub-networks by connecting with hub genes (Fig. 5) and these hub genes might have vital role for interaction and may lead to important biological functions.<sup>63</sup> As for example, the hub gene *bip110* were found for seed development<sup>42</sup> and the hub gene LOC\_Os01g39100 localized to nucleus, encodes zinc finger CCCH domain and might involve in photoperiodic control of flowering time in rice like *Ehd*.<sup>64</sup>

Gene expression profiles have long plays fundamental role in evolution<sup>65</sup> and gene's expression across tissue could help to reveal the function and important role of those genes.<sup>66</sup> We used the expression profile of candidate epistatic genes across different floral tissues and found majority of them expressed in panicle and other floral tissues (Fig. 6 and Supplementary Fig. S6), supporting the genetic evidence for the identified epistatic genes involving in rice flowering.

It is hypothesized that, to interact proteins or genes, the interactome tend to be located in the same SCL, or in physically adjacent SCLs.<sup>67</sup> Moreover, similarity between two genes ontology have been employed as an additional criteria of confidence for a predicted interaction.<sup>68</sup> Different genomic features and biological information such as PPI, SCL and GO terms were combined in the network to validate our findings. A total of 392 (66.55%) interactions were found those either located in the region of known PPI or/and overlapped with SCL or/and GO terms (Fig. 5). Moreover, in different circumstances proteins migrate between compartments and therefore could have interaction partners in both locations.<sup>69</sup> For instances, among the identified epistatic genes, 11 were found those involved in the BP protein ubiquitination (Fig. 4 and Supplementary Data S6). It is well known that protein ubiquitination alter their cellular location, affect their activity and promote or prevent protein interactions.<sup>70,71</sup> A significant number of interacting genes were found between nucleus–cytosol, plastid–cytosol and nucleus–plastid (Supplementary Fig. S8) suggesting the proteins may have been switched to the cytosol from other compartments.

We also found 37 SNPs harbouring 25 genes those involved in rice or/and Arabidopsis (orthologue) flower related pathway among them 12 genes were common (Fig. 5b). Six genes were found in Arabidopsis floral pathway, indicating novel genes for rice and might have association with rice flowering time. According to our epistasis and functional analysis results, we may conclude the identified genes might be biologically plausible for flowering time.

## Acknowledgements

The authors would like to thank the rice diversity research platform for sharing the data used in this study. We thank to Dr. Imrul Mosaddek Ahmed for helpful discussions.

## Funding

This work was supported by the National Natural Science Foundation of China [31571366, 31771477]; National Key Research and Development Program of China [2016YFA0501704, 2018YFC0310602]; the Chinese Government Scholarship for foreign students, Jiangsu Collaborative Innovation Center for Modern Crop Production, the Fundamental Research Funds for the Central Universities.

## Conflict of interest

None declared.

## Supplementary data

Supplementary data are available at DNARES online.

## References

- Furbank, R.T., von Caemmerer, S., Sheehy, J. and Edwards, G. 2009, C-4 rice: a challenge for plant phenomics, *Funct. Plant Biol.*, **36**, 845–56.
- Rahaman, M.M., Chen, D.J., Gillani, Z., Klukas, C. and Chen, M. 2015, Advanced phenotyping and phenotype data analysis for the study of plant growth and development, *Front. Plant Sci.*, **6**, 619.
- Xu, L., Hu, K., Zhang, Z., et al. 2016, Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (*Brassica napus* L.), *DNA Res.*, **23**, 43–52.
- Roux, F., Touzet, P., Cuguen, J. and Le Corre, V. 2006, How to be early flowering: an evolutionary perspective, *Trends Plant Sci.*, **11**, 375–81.
- Yu, J., Pressoir, G., Briggs, W.H., et al. 2006, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, *Nat. Genet.*, **38**, 203–8.
- Huang, X., Zhao, Y., Wei, X., et al. 2011, Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm, *Nat. Genet.*, **44**, 32–9.
- Cordell, H.J. 2009, Detecting gene–gene interactions that underlie human diseases, *Nat. Rev. Genet.*, **10**, 392–404.
- Marchini, J., Donnelly, P. and Cardon, L.R. 2005, Genome-wide strategies for detecting multiple loci that influence complex diseases, *Nat. Genet.*, **37**, 413–7.
- Yang, J., Benyamin, B., McEvoy, B.P., et al. 2010, Common SNPs explain a large proportion of the heritability for human height, *Nat. Genet.*, **42**, 565–9.
- Huang, W., Richards, S., Carbone, M.A., et al. 2012, Epistasis dominates the genetic architecture of *Drosophila* quantitative traits, *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 15553–9.
- Manolio, T.A., Collins, F.S., Cox, N.J., et al. 2009, Finding the missing heritability of complex diseases, *Nature*, **461**, 747–53.
- Phillips, P.C. 2008, Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems, *Nat. Rev. Genet.*, **9**, 855–67.
- Li, J., Zhong, W., Li, R. and Wu, R. 2014, A fast algorithm for detecting gene–gene interactions in genome-wide association studies, *Ann. Appl. Stat.*, **8**, 2292–318.
- Xu, S. 2013, Mapping quantitative trait loci by controlling polygenic background effects, *Genetics*, **195**, 1209–22.
- Purcell, S., Neale, B., Todd-Brown, K., et al. 2007, PLINK: a tool set for whole-genome association and population-based linkage analyses, *Am. J. Hum. Genet.*, **81**, 559–75.
- Schupbach, T., Xenarios, I., Bergmann, S. and Kapur, K. 2010, FastEpistasis: a high performance computing solution for quantitative trait epistasis, *Bioinformatics*, **26**, 1468–9.
- Hemani, G., Theodoridis, A., Wei, W. and Haley, C. 2011, EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards, *Bioinformatics*, **27**, 1462–5.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. 2006, Principal components analysis corrects for stratification in genome-wide association studies, *Nat. Genet.*, **38**, 904–9.
- Kang, H.M., Zaitlen, N.A., Wade, C.M., et al. 2008, Efficient control of population structure in model organism association mapping, *Genetics*, **178**, 1709–23.
- Li, D. and Won, S. 2016, Efficient strategy to identify gene–gene interactions and its application to type 2 diabetes, *Genomics Inform.*, **14**, 160–5.
- Zhu, S. and Fang, G. 2018, MatrixEpistasis: ultrafast, exhaustive epistasis scan for quantitative traits with covariate adjustment, *Bioinformatics*, **34**, 2341–8.
- Segura, V., Vilhjalmsson, B.J., Platt, A., et al. 2012, An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations, *Nat. Genet.*, **44**, 825–30.
- Zhao, K., Aranzana, M.J., Kim, S., et al. 2007, An Arabidopsis example of association mapping in structured samples, *PLoS Genet.*, **3**, e4.
- Aulchenko, Y.S., de Koning, D.J. and Haley, C. 2007, Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis, *Genetics*, **177**, 577–85.
- Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. and Snyder, M. 2012, Linking disease associations with regulatory information in the human genome, *Genome Res.*, **22**, 1748–59.
- Gandhi, T.K.B., Zhong, J., Mathivanan, S., et al. 2006, Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets, *Nat. Genet.*, **38**, 285–93.
- Schwikowski, B., Uetz, P. and Fields, S. 2000, A network of protein–protein interactions in yeast, *Nat. Biotechnol.*, **18**, 1257–61.
- Zhao, K., Tung, C.W., Eizenga, G.C., et al. 2011, Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*, *Nat. Commun.*, **2**, 467.
- Schwender, H. 2012, Imputing missing genotypes with weighted k nearest neighbors, *J. Toxicol. Environ. Health A*, **75**, 438–46.
- Li, H., Ye, G. and Wang, J. 2007, A modified algorithm for the improvement of composite interval mapping, *Genetics*, **175**, 361–74.
- Kao, C.H., Zeng, Z.B. and Teasdale, R.D. 1999, Multiple interval mapping for quantitative trait loci, *Genetics*, **152**, 1203–16.
- Zeng, Z.B. 1994, Precision mapping of quantitative trait loci, *Genetics*, **136**, 1457–68.
- VanRaden, P.M. 2008, Efficient methods to compute genomic predictions, *J. Dairy Sci.*, **91**, 4414–23.
- Endelman, J.B. 2011, Ridge regression and other kernels for genomic selection with R package rrBLUP, *Plant Genome*, **4**, 250–5.
- Yang, J.A., Lee, S.H., Goddard, M.E. and Visscher, P.M. 2011, GCTA: a tool for genome-wide complex trait analysis, *Am. J. Hum. Genet.*, **88**, 76–82.
- Wang, J.W., Qi, M.F., Liu, J. and Zhang, Y.J. 2015, CARMO: a comprehensive annotation platform for functional exploration of rice multi-omics data, *Plant J.*, **83**, 359–74.
- Supek, F., Bosnjak, M., Skunca, N. and Smuc, T. 2011, REVIGO summarizes and visualizes long lists of gene ontology terms, *PLoS One*, **6**, e21800.
- Song, J.M., Lei, Y., Shu, C.C., et al. 2018, Rice Information GateWay (RIGW): a comprehensive bioinformatics platform for indica rice genomes, *Mol. Plant*, **11**, 505–7.
- Gu, H.B., Zhu, P.C., Jiao, Y.M., Meng, Y.J. and Chen, M. 2011, PRIN: a predicted rice interactome network, *BMC Bioinform.*, **12**, 161.
- Liu, S., Liu, Y., Zhao, J., et al. 2017, A computational interactome for prioritizing genes associated with complex agronomic traits in rice (*Oryza sativa*), *Plant J.*, **90**, 177–88.
- Bouche, F., Lobet, G., Tocquin, P. and Perilleux, C. 2016, FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*, *Nucleic Acids Res.*, **44**, D1167–D71.

42. Hanumappa, M., Preece, J., Elser, J., et al. 2013, WikiPathways for plants: a community pathway curation portal and a case study in rice and arabidopsis seed development networks, *Rice*, **6**, 14.
43. Liu, L., Zhang, Z., Mei, Q. and Chen, M. 2013, PSI: a comprehensive and integrative approach for accurate plant subcellular localization prediction, *PLoS One*, **8**, e75826.
44. Shannon, P., Markiel, A., Ozier, O., et al. 2003, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, **13**, 2498–504.
45. Letunic, I. and Bork, P. 2016, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Res.*, **44**, W242–W5.
46. Kogelman, L.J. and Kadarmideen, H.N. 2014, Weighted Interaction SNP Hub (WISH) network method for building genetic networks for complex diseases and traits using whole genome genotype data, *BMC Syst. Biol.*, **8**, S5.
47. Jiang, D.H. and Berger, F. 2017, DNA replication-coupled histone modification maintains Polycomb gene silencing in plants, *Science*, **357**, 1146–9.
48. Yang, H.C., Berry, S., Olsson, T.S.G., Hartley, M., Howard, M. and Dean, C. 2017, Distinct phases of Polycomb silencing to hold epigenetic memory of cold in Arabidopsis, *Science*, **357**, 1142–5.
49. Wang, Y.L., Jiang, Q.G., Liu, J.B., et al. 2017, Comparative transcriptome profiling of chilling tolerant rice chromosome segment substitution line in response to early chilling stress, *Genes Genomics*, **39**, 127–41.
50. Maeda, H. and Dudareva, N. 2012, The shikimate pathway and aromatic amino acid biosynthesis in plants, *Annu. Rev. Plant Biol.*, **63**, 73–105.
51. Li, W.Q., Liu, X.H. and Lu, Y.M. 2016, Transcriptome comparison reveals key candidate genes in response to vernalization of Oriental lily, *BMC Genomics*, **17**, 664.
52. Singh, A., Baranwal, V., Shankar, A., et al. 2012, Rice phospholipase A superfamily: organization, phylogenetic and expression analysis during abiotic stresses and development, *PLoS One*, **7**, e30947.
53. Ishiguro, S., Kawai-Oda, A., Ueda, J., Nishida, I. and Okada, K. 2001, The DEFECTIVE IN ANTHHER DEHISCENCE gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in Arabidopsis, *Plant Cell*, **13**, 2191–209.
54. Evans, D.M., Marchini, J., Morris, A.P. and Cardon, L.R. 2006, Two-stage two-locus models in genome-wide association, *PLoS Genet.*, **2**, 1424–32.
55. Hemani, G., Shakhbazov, K., Westra, H.J., et al. 2014, Detection and replication of epistasis influencing transcription in humans, *Nature*, **508**, 249–53.
56. Wei, W.H., Hemani, G. and Haley, C.S. 2014, Detecting epistasis in human complex traits, *Nat. Rev. Genet.*, **15**, 722–33.
57. Hill, W.G., Goddard, M.E. and Visscher, P.M. 2008, Data and theory point to mainly additive genetic variance for complex traits, *PLoS Genet.*, **4**, e1000008.
58. Wright, S. 1931, Evolution in Mendelian populations, *Genetics*, **16**, 97–159.
59. Mackay, T.F. 2015, Epistasis for quantitative traits in *Drosophila*. *Methods Mol. Biol.*, **1253**:47–70.
60. Pan, W., Chen, Y.M. and Wei, P. 2015, Testing for polygenic effects in genome-wide association studies, *Genet. Epidemiol.*, **39**, 306–16.
61. Molinaro, A.M., Carriero, N., Bjornson, R., Hartge, P., Rothman, N. and Chatterjee, N. 2011, Power of data mining methods to detect genetic associations and interactions, *Hum. Hered.*, **72**, 85–97.
62. Culverhouse, R.C. 2012, A comparison of methods sensitive to interactions with small main effects, *Genet. Epidemiol.*, **36**, 303–11.
63. Lu, X., Jain, V.V., Finn, P.W. and Perkins, D.L. 2007, Hubs in biological interaction networks exhibit low changes in expression in experimental asthma, *Mol. Syst. Biol.*, **3**, 98.
64. Gao, H., Zheng, X.M., Fei, G.L., et al. 2013, Ehd4 encodes a novel and oryza-genus-specific regulator of photoperiodic flowering in rice, *PLoS Genet.*, **9**, e1003281.
65. Fraser, H.B., Moses, A.M. and Schadt, E.E. 2010, Evidence for widespread adaptive evolution of gene expression in budding yeast, *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 2977–82.
66. Chen, J., Swofford, R., Johnson, J., et al. 2017, A quantitative model for characterizing the evolutionary history of mammalian gene expression, *bioRxiv*, <https://doi.org/10.1101/229096>.
67. Shin, C.J., Wong, S., Davis, M.J. and Ragan, M.A. 2009, Protein–protein interaction as a predictor of subcellular location, *BMC Syst. Biol.*, **3**, 28.
68. Mahdavi, M.A. and Lin, Y.H. 2007, False positive reduction in protein–protein interaction predictions using gene ontology annotations, *BMC Bioinform.*, **8**, 262.
69. Geisler-Lee, J., O’Toole, N., Ammar, R., Provart, N.J., Millar, A.H. and Geisler, M. 2007, A predicted interactome for Arabidopsis, *Plant Physiol.*, **145**, 317–29.
70. Hochstrasser, M. 1995, Ubiquitin, proteasomes, and the regulation of intracellular protein-degradation, *Curr. Opin. Cell Biol.*, **7**, 215–23.
71. Yang, R.L. and Liu, C.F. 2015, Chemical methods for protein ubiquitination, *Top. Curr. Chem.*, **362**, 89–106.