An expression quantitative trait loci-guided co-expression analysis for constructing regulatory network using a rice recombinant inbred line population

Jia Wang, Huihui Yu, Xiaoyu Weng, Weibo Xie, Caiguo Xu, Xianghua Li, Jinghua Xiao and Qifa Zhang

National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research, Huazhong Agricultural University, Wuhan 430070, PR China

* To whom correspondence should be addressed. Email: qifazh@mail.hzau.edu.cn

Received 16 August 2013; Revised 26 October 2013; Accepted 28 November 2013

Abstract

The ability to reveal the regulatory architecture of genes at the whole-genome level by constructing a regulatory network is critical for understanding the biological processes and developmental programmes of organisms. Here, we conducted an eQTL-guided function-related co-expression analysis to identify the putative regulators and construct gene regulatory network. We performed an eQTL analysis of 210 recombinant inbred lines (RILs) derived from a cross between two indica rice lines, Zhenshan 97 and Minghui 63, the parents of an elite hybrid, using data obtained by hybridizing RNA samples of flag leaves at the heading stage with Affymetrix whole-genome arrays. Making use of an ultrahigh-density single-nucleotide polymorphism bin map constructed by population sequencing, 13 647 eQTLs for 10 725 e-traits were detected, comprising 5079 cis-eQTLs (37.2%) and 8568 trans-eQTLs (62.8%). The analysis revealed 138 trans-eQTLs hotspots, each of which apparently regulates the expression variations of many genes. Co-expression analysis of functionally related genes within the framework of regulator–target relationships outlined by the eQTLs led to the identification of putative regulators in the system. The usefulness of the strategy was demonstrated with the genes known to be involved in flowering. We also applied this strategy to the analysis of QTLs for yield traits, which also suggested likely candidate genes. eQTL-guided co-expression analysis may provide a promising solution for outlining a framework for the complex regulatory network of an organism.

Key words: Gene expression; network construction; phenotype; quantitative genomics; regulation.

Introduction

Expression of genes is co-ordinately regulated in space and time in an organism to deliver the gene products when needed, according to developmental programmes and in response to environmental cues (Schmid et al., 2005; Wang et al., 2010b). The ability to reveal the regulatory architecture of genes at the whole-genome scale by constructing a regulatory network is critical for understanding the biological processes and genome function of an organism (Davidson et al., 2002; Bar-Joseph et al., 2003; Shinozaki et al., 2003).

In the past, technical approaches have been developed in two major ways to characterize gene expression regulation. The first is the detection technology of gene expression, ranging from traditional analysis of one gene at a time using Northern blotting, reverse transcription-PCR and real-time quantitative PCR (qPCR), to large-scale analysis of many genes, such as microarrays composed of all the annotated genes in an entire genome (Schena et al., 1995; Tan et al., 2003), and more recently by RNA sequencing (Wang et al., 2009).
In parallel, there has also been the development of approaches to determine regulatory relationships of the genes. It is a common practice in the literature to identify the downstream targets of an altered gene using mutants or nearly isogenic lines (NILs) to analyse expression profiles for genes of interest, and then to expand the regulatory pathways stepwise using more mutants or NILs (Doi et al., 2004; Espinosa-Soto et al., 2004; Xue et al., 2008). Another approach involves co-expression analysis by calculating correlations between genes based on large volumes of expression data, which can also provide useful information on how likely the gene functions are to be related (van Noort et al., 2003; Zhang and Horvath, 2005; Wang et al., 2010b). Expression quantitative trait loci (eQTLs) are genomic regions that regulate the gene expression variations in a population. Analysis of eQTLs that treats the expression values of the genes as a quantitative trait (e-traits) to perform QTL analysis using a genetically segregating population appears to be a useful strategy for suggesting regulatory relationships between genes (Gilad et al., 2008; Kliebenstein, 2009). eQTL analysis can detect genetic elements that regulate the expression variation of the e-trait acting in cis if the QTL is located in the vicinity of an e-trait, or in trans if it is located in a distant position. Such analysis could suggest the existence of a potential regulator in a genomic region for a gene (Hansen et al., 2008; Kliebenstein, 2009). Moreover, one regulator could be the target of another regulator. Thus, connecting such regulator–target relationships revealed by multiple e-traits and eQTLs of different layers would lead to a net-like structure, providing basic elements for the construction of a regulatory network. More recently, an approach combining eQTL with co-expression analysis was proposed to identify regulator candidates underlying eQTLs, which may greatly refine the resolution of the analysis (Terpstra et al., 2010; Flassig et al., 2013).

We attempted to construct a regulatory network adopting an eQTL-guided co-expression analysis strategy using flag leaf, a tissue that contributes significantly to grain yield as a primary source of carbohydrate production, from a population of 210 recombinant inbred lines (RILs) in rice. We showed that the analysis could identify candidate regulators and their targets down to the level of single genes. We verified the usefulness of this approach using genes in the flowering pathways. The results suggest that this analysis may provide a promising solution for outlining the framework for the complex regulatory network of an organism.

**Materials and methods**

**Plant material**

The population used in the study contained 210 RILs derived by single-seed descent from a cross between two indica rice lines, Zhenshan 97 and Minghui 63. The RILs and two parents were planted in a seed bed and transplanted to an experimental farm at Huazhong Agricultural University in the rice growing seasons in Wuhan, China. Ten seedlings were planted for each of the RILs in two different batches at 45 d apart as two biological replicates. Flag leaves from three random plants of each RIL per replicate were collected at the heading stage (day of the panicle emergence; Supplementary Fig. S1 available at JXB online) between 8:00 and 9:30 a.m.

RNA extraction, isolation, and Agilent Bioanalyzer quality testing were conducted for each of the two biological replicates. High-quality RNA samples of the two replicates were mixed (1:1) for cDNA synthesis, labelling, microarray hybridization and scanning. All processes were conducted by CapitalBio Corporation (Beijing, China).

NIL(mh7) and Zhenshan 97, which differ only by an introgressed genome segment, for the network verification were planted and the tissue samples collected in the same way as the RIL population.

**Microarray analysis**

R platform (http://www.R-project.org) and ‘Affy’ packages from bioconductor (http://www.bioconductor.org) were applied to manage all the CEL files. Robust multiarray average (RMA) analysis was used for background correction and probe set expression value collection. Probe sets identified by MAS 5.0 algorithm as ‘P’ or ‘M’ for at least one-third of the 210 RILs were employed as e-traits for eQTL analysis.

**eQTL and phenotypic QTL (pQTL) analyses**

eQTL and classical pQTL analyses were based on a genetic map consisting of 1619 recombinant bins constructed using high-quality single-nucleotide polymorphisms (SNPs) from resequencing of the 210 RILs (Yu et al., 2011). The R/qtl function cim (Broman et al., 2003) was employed for eQTL and pQTL mapping using a 10 cM scan window and five marker covariates (Potokina et al., 2008; Yu et al., 2011). The logarithm of the odds (LOD) threshold for eQTL identification was obtained based on a global permutation test that randomly selects 100 e-traits from all 21 929 e-traits to do 1000 permutations (West et al., 2007; Wang et al., 2010a). The eQTL additive effect and variation explained by the eQTL were determined using the linear QTL model described by Yu et al. (2011). With the ultrahigh-density SNP bin map, the LOD thresholds of pQTL analysis for 10 yield-associated traits at P=0.05 ranged from 4.76 to 5.10 using 1000 permutation test for each phenotypic trait, with an average LOD value of 4.97. We used 5.0 as the threshold for pQTL identification.

**Real-time PCR analysis**

RNA was isolated from NIL(mh7) and Zhenshan 97 rice flag leaves at the heading stage using TRizol reagent (Invitrogen). First-strand cDNA was synthesized from DNase I-treated total RNA using Superscript III reverse transcriptase (Invitrogen) according to the manufacturer’s instructions. qPCR was performed with gene-specific primers (Supplementary Table S1 available at JXB online) in Applied Biosystems Viia™ 7 using SYBR Premix Ex Taq (Takara) for six biological replicates of each sample. The program was 95 °C for 10 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 34 s, and relative expression levels were calculated by the 2\(^{-\Delta\Delta Ct}\) method (Livak and Schmittgen, 2001) compared with levels of expression of the ubiquitin gene.

**Accession codes**

The National Center for Biotechnology Information Gene Expression Omnibus microarray data have been submitted under accession number GSE49020.

**Results**

**Distribution of e-traits**

In total, 44 029 (76.7%) of the 57 194 probe sets on the GeneChip Rice Genome Array each had a locus support in
the Rice Genome Pseudomolecules (Release version MSU 6.1) representing 35 885 genes. In addition, 21 929 (49.8%) of the 44 029 probe sets corresponding to 17 934 genes were identified as ‘present’ in more than one-third of 210 RILs using the ‘mas5calls’ algorithm (Liu et al., 2002) and were thus regarded as e-traits. We mapped the e-traits to the ultrahigh-density SNP bin map consisting of 1619 bins (Yu et al., 2011). The distribution of e-traits was highly uneven across the genome (Supplementary Fig. S2 available at JXB online), and their numbers in the 1619 bins deviated significantly from the expectation based on physical sizes of the bins as shown by a $\chi^2$ test ($\chi^2=47 264$, $P=2.2E-16$) using the chisq.test function for a goodness-of-fit test. Further evaluation using the standardized residue (SR), which follows an asymptotical normal distribution, showed that the numbers of e-traits in 274 bins were significantly different from the expectations (Supplementary Table S2 available at JXB online). The density of the genes and e-traits in the centromeric and pericentromeric regions was lower than that in other chromosomal regions.

**Genome-wide eQTL mapping**

These e-traits were subjected to eQTL analysis based on the ultrahigh-density SNP bin map using composite interval mapping (Zeng, 1993, 1994). We applied an LOD threshold of 4.95 ($P=0.05$), which was determined by a global permutation test, to identify eQTLs. An eQTL was regarded as cis-acting if the gene (e-trait) was within the 1.5 LOD-drop support interval (SI) of the corresponding eQTL, or as trans-acting if the gene (e-trait) was not located within the 1.5 LOD-drop SI (Keurentjes et al., 2007). These processes eventually resulted in 13 647 eQTLs for 10 725 e-traits (Fig. 1 and Supplementary Table S3 available at JXB online), including 5079 cis-eQTLs (37.2%) and 8568 trans-eQTLs (62.8%) (Table 1). Only one eQTL was detected for 8258 (77.0%) of the 10 725 e-traits, and more than two eQTLs were detected for 407 (3.79%) of the e-traits. Most of the cis-eQTLs had greater LOD values and effects on gene expression than trans-eQTLs (Fig. 1 and Supplementary Fig. S3 available at JXB online). Overall, 69.2% of the cis-eQTLs each explained ≥20% of the expression variation for the corresponding e-traits, while only 8.79% of the trans-eQTLs could explain ≥20% of the expression variation for the corresponding e-traits.

**Trans-eQTL hotspots**

Four chromosomes (1, 3, 6, and 9) had higher numbers of eQTLs than expected based on both physical and genetic maps (Table 1), suggesting possible eQTL hotspots, as indicated by the horizontal bands (trans-eQTL hotspots) in Fig. 1. Comparison of observed and expected numbers of trans-eQTLs showed that trans-eQTLs detected in bins on some chromosomes (1, 3, 6, 7, and 9) were obviously more than the e-traits with cis-eQTLs (Supplementary Table S4 available at JXB online). We also performed a $\chi^2$ test for goodness of fit between observed and expected trans-eQTLs followed by calculating the SR, and a hypergeometric test for observed trans-eQTLs and e-traits with cis-eQTLs in the bins. It was found that the numbers of trans-eQTLs in 138 bins were significantly more than the expectations based on the sizes of the bins ($P<0.01$) and the numbers of e-traits with cis-eQTLs ($P<0.01$) (Supplementary Table S4 available at JXB online), indicating both trans-eQTL hotspots and coldspots. In particular, the top five bins of the trans-eQTL hotspots (Bin851 on chromosome 6; Bin857 on chromosome 6, Bin37 on chromosome 1, Bin 394 on chromosome 3, and Bin1012 on chromosome 7) ranging from 20 to 50kb harboured 439, 190, 169, 124, and 111 trans-eQTLs, respectively.

**Identifying candidate master regulators**

A regulator at the high-node level in the network, or master regulator, frequently regulates many genes involved in multiple biological processes. We thus reasoned that: (1) such a regulator should correspond to a trans-eQTL hotspot, and (2) it should be co-expressed (positively or negatively) with the genes (targets) it regulates. To validate such an assertion and find candidate master regulators, we performed a co-expression analysis by calculating correlations between e-traits with cis-eQTLs and e-traits whose trans-eQTLs are located in the 1.5 LOD-drop SI of the cis-eQTLs (Fig. 2A).
Table 1. Statistics of the eQTLs on each chromosome

<table>
<thead>
<tr>
<th>Chr.</th>
<th>eQTLs on chr.</th>
<th>Cis-eQTL</th>
<th>Trans-eQTL</th>
<th>Chr. size (Mb)</th>
<th>Expected</th>
<th>SR</th>
<th>Chr. size (cM)</th>
<th>Expected</th>
<th>SR</th>
<th>cM/Mb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2563</td>
<td>943</td>
<td>1620</td>
<td>43.3</td>
<td>1588</td>
<td>24.5</td>
<td>201</td>
<td>1684</td>
<td>21.4</td>
<td>4.6</td>
</tr>
<tr>
<td>2</td>
<td>929</td>
<td>463</td>
<td>466</td>
<td>35.9</td>
<td>1316</td>
<td>–10.7</td>
<td>175</td>
<td>1473</td>
<td>–14.2</td>
<td>4.9</td>
</tr>
<tr>
<td>3</td>
<td>1855</td>
<td>579</td>
<td>1276</td>
<td>36.4</td>
<td>1335</td>
<td>14.3</td>
<td>188</td>
<td>1574</td>
<td>7.1</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>706</td>
<td>379</td>
<td>327</td>
<td>35.3</td>
<td>1294</td>
<td>–16.4</td>
<td>127</td>
<td>1068</td>
<td>–11.1</td>
<td>3.6</td>
</tr>
<tr>
<td>5</td>
<td>608</td>
<td>373</td>
<td>235</td>
<td>29.9</td>
<td>1096</td>
<td>–14.7</td>
<td>116</td>
<td>974</td>
<td>–11.7</td>
<td>3.9</td>
</tr>
<tr>
<td>6</td>
<td>2453</td>
<td>410</td>
<td>2043</td>
<td>31.2</td>
<td>1144</td>
<td>38.7</td>
<td>144</td>
<td>1212</td>
<td>35.6</td>
<td>4.6</td>
</tr>
<tr>
<td>7</td>
<td>1135</td>
<td>290</td>
<td>845</td>
<td>29.7</td>
<td>1089</td>
<td>1.4</td>
<td>135</td>
<td>1137</td>
<td>–0.1</td>
<td>4.6</td>
</tr>
<tr>
<td>8</td>
<td>610</td>
<td>326</td>
<td>284</td>
<td>28.4</td>
<td>1041</td>
<td>–13.4</td>
<td>120</td>
<td>1010</td>
<td>–12.6</td>
<td>4.2</td>
</tr>
<tr>
<td>9</td>
<td>1093</td>
<td>356</td>
<td>737</td>
<td>23</td>
<td>843</td>
<td>8.6</td>
<td>107</td>
<td>900</td>
<td>6.4</td>
<td>4.7</td>
</tr>
<tr>
<td>10</td>
<td>509</td>
<td>332</td>
<td>177</td>
<td>23.1</td>
<td>847</td>
<td>–11.6</td>
<td>85</td>
<td>716</td>
<td>–7.1</td>
<td>3.7</td>
</tr>
<tr>
<td>11</td>
<td>739</td>
<td>380</td>
<td>359</td>
<td>28.5</td>
<td>1045</td>
<td>–9.5</td>
<td>117</td>
<td>980</td>
<td>–7.8</td>
<td>4.1</td>
</tr>
<tr>
<td>12</td>
<td>447</td>
<td>248</td>
<td>199</td>
<td>27.5</td>
<td>1008</td>
<td>–17.7</td>
<td>109</td>
<td>918</td>
<td>–15.5</td>
<td>4.0</td>
</tr>
<tr>
<td>13647</td>
<td>5079</td>
<td>8568</td>
<td>372.2</td>
<td>13647</td>
<td>1625</td>
<td>4.4</td>
<td>13647</td>
<td>1625</td>
<td>4.4</td>
<td></td>
</tr>
</tbody>
</table>

aExpected number of eQTLs based on chromosome size from physical map. χ² = 3690.2 (P < 2.2e–16) for the test of goodness of fit between the observed and expected numbers of eQTLs on the 12 chromosomes.

bSR: for physical map, standardized residue [=(observed – expected)/√expected], which follows a normal distribution asymptotically. An absolute SR value >2.33 indicates statistical significance at P<0.01. A positive value indicates that the observed number is greater than expected.

cExpected number of eQTLs based on chromosome size from genetic map. χ² = 2801.18 (P < 2.2e–16).

dSR: standardized residue for genetic map.

The threshold for significant correlations (0.60 and –0.52, P<0.01) was appraised based on 1 000 000 pair-wise correlations from two randomly selected groups each containing 1000 e-trait. Six genes seemed to be putative master regulators, as each was associated with more than 100 genes as the targets (Table 2), and other 27 genes obtained 20–100 candidate targets (Supplementary Table S5 available at JXB online).

A gene ontology analysis based on biological processes of the associated targets for the six genes revealed that the targets for two of the putative regulators (Os.12583.1.S1_at annotated as expressed protein and Os.53449.1.A1_at annotated as OMP85 family protein) seemed to be highly enriched in certain functional categories (Supplementary Table S6 available at JXB online). No obvious enrichment was detected for the other four putative master regulators, indicating that these putative master regulators might be involved in diverse regulatory pathways.

Network construction for functionally related genes

Gene regulatory networks usually regulate the expression of arrays of genes in specific spatial and temporal patterns. To explore the feasibility of using the expression data to identify regulatory networks, we performed co-expression analysis of genes known to be involved in the long-day flowering pathway. Many studies have revealed that, as a short-day plant, rice flowering is regulated by both a short-day activation pathway and a long-day suppression pathway (Hayama et al., 2003). Our RIL populations were planted in the rice growing season (May–September) in Wuhan, China, in which the natural day length is 13–14.5 h (nearly a long day).

An iterative group analysis (iGA) algorithm (Breitling et al., 2004; Keurentjes et al., 2007) was used to analyse the e-trait expression profile to find master regulators of the flowering genes, based on the assumption that a concerted expression variation was controlled by the same regulator and that the functions of the genes were physiologically relevant. In this process (Fig. 2B, C), a putative local regulator gene that has a cis-eQTL co-located with a trans-eQTL was assigned a group name, and an e-trait with a trans-eQTL that co-mapped with the group was assigned as a member of the group. We first sorted the putative targets (members) according to the expression correlations with potential regulators (groups) in descending order, and moved along the list counting the targets of each candidate regulator. When a new target was encountered each time, we inquired: How likely is it that we would observe this many targets of the regulator that high up in the list by chance? For each group, we defined the probability of change (PC) value, which indicated how likely a regulator with cis-eQTL is to show a strong correlation with several members with trans-eQTLs of a selected group of genes. A regulatory group with a significant PC value (P<0.05) would be regarded as a regulator controlling the members contributing to this PC value.

We illustrated this procedure using the flowering regulation genes as an example. We identified 577 unique loci annotated as known or predicted genes in the category GO:0009908 (flower development), as well as ones known to be involved in flower development and the flowering time pathway from the literature (Supplementary Table S7 available at JXB online). There were 280 eQTLs associated with 177 of the 577 unique probe sets including 109 cis-eQTLs and 171 trans-eQTLs (Supplementary Table S8 available at JXB online). Sixty-one of the 109 genes with cis-eQTLs had a SI overlapping the SI of trans-eQTLs. We calculated expression correlations between the 61 genes (potential regulators) and 171 genes with trans-eQTLs (potential targets), which amounted to 7808 gene pairs when disregarding the genes that had both cis- and trans-eQTLs (Supplementary Table S9 available at JXB online). The correlation coefficients were rank ordered.
iGA (Breitling et al., 2004) was applied to select the most significant groups and the significant change targets contributing to PC value, which indicate how likely a regulator with cis-eQTL is to observe a strong correlation with several members with trans-eQTLs of a selected group of genes.

Hd3a and RFT1, essential for flowering in rice, share 91% identity in their deduced amino acid sequences and are located only 11.5 kb apart on chromosome 6 (Komiya et al., 2008). A cis-eQTL was detected for RFT1, but the expression variation of Hd3a mapped no significant locus. To avoid the obscuring effect of sequence identity between the genes, the expression levels of RFT1 and Hd3a were recalculated using a subset of four probes selected from the 11 probes in the chip for each gene that had more than five SNPs between RFT1 and Hd3a. With this analysis, the effect of the cis-eQTL for RFT1 was greatly elevated with the LOD increasing from 7.33 to 76.02 (Supplementary Table S10 and Supplementary Fig. S4 available at JXB online), and the amount of variation explained from 9.9 to 12.3%. A trans-eQTL for Hd3a (LOD=5.2, R²=9.2%), associated with Bin1006 (chromosome 7), emerged from this analysis (Supplementary Table S10 and Supplementary Fig. S5 available at JXB online). At the cis-eQTL for RFT1 on chromosome 6, the Minghui 63 allele had a positive effect on RFT1 expression. However, RFT1 also had a trans-eQTL on chromosome 1, at which the Zhenshan 97 allele had a positive effect on RFT1 expression. Consequently, the expression levels of RFT1 did not seem to be very different in the two parents (Supplementary Fig. S6 available at JXB online). In contrast, the two parents showed large difference in the expression level of Hd3a, apparently due to the trans-eQTL on chromosome 7.

**Fig. 2.** Schematic diagram of the procedure for identification of master regulators and construction of regulatory network for functionally related genes. (A) Identification of master regulators in the trans-eQTL hotspots by eQTL mapping. (B) eQTL map of functionally related genes. (C) The process of seeking master regulators and their targets according to the expression correlation between putative regulators and candidate targets for constructing a regulatory network (see main text for details).
Eight obvious regulatory groups were identified ($P<0.05$) (Table 3) based on the iGA analysis, and the targets for each regulator are shown in Fig. 3, with the expression correlations displayed in Supplementary Table S11 available at JXB online. Ghd7, Hd1, and RFT1 were identified as master regulators in this analysis, which were the trans-eQTLs for several flowering-related genes. The trans-eQTL for Hd3a appeared to be Ghd7 (cor$=-0.30, P=8.63E-6$). Several trans-eQTLs for flowering-related genes seemed to be RFT1, of which OsMADS14 and OsMADS15 were identified as targets of RFT1 in a previous work (Komiya et al., 2008, 2009). Meanwhile, OsGF14b and OsGF14c, two 14-3-3 proteins, act as intracellular receptors for rice Hd3a florigen (Taoka et al., 2011), and were identified as the targets of RFT1 in our results. This is in agreement with the conclusion that RFT1 was the major regulator for long-day flowering (Komiya et al., 2009). Ehd1 and OsMADS50, confirmed as upstream regulators of RFT1/Hd3a, both had trans-eQTLs co-located with Hd1, with negative expression correlations of Hd1 with Ehd1 (cor$=-0.54, P=4.06E-17$) and OsMADS50 (cor$=-0.27, P=7.51E-05$) in 210 RILs indicated that they both were downregulated by Hd1, suggesting Hdl as a regulator of Ehd1 and OsMADS50. Others have reported that Hd1 and Ehd1 are independent in flowering regulation (Doi et al., 2004; Takahashi et al., 2009), and a more recent study showed that Ehd1 appears to be negatively regulated and Ghd7 was positively regulated by Hd1 under long-day conditions (Song et al., 2012), which seems to substantiate our network. We obtained the 2 kb genomic sequence upstream of the ATG site of Ehd1 and OsMADS50, and found no difference between Zhenshan 97 and Minghui 63, suggesting that this 2 kb sequence may not be the cause of the expression difference. In contrast, OsMADS50 was positively regulated by Ghd7 (cor$=0.25$).

Table 2. Information on the six putative master regulators genes, associated with more than 100 genes as targets

The candidate regulators were suggested to be the gene with cis-eQTLs that co-mapped with trans-eQTLs of many co-expressed genes ($P<0.01$).

<table>
<thead>
<tr>
<th>Regulator</th>
<th>N¹</th>
<th>Bin</th>
<th>Chr.</th>
<th>LOD²</th>
<th>Inf. Mb³</th>
<th>Sup. Mb⁴</th>
<th>Var⁵</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Os.3539.1.S1_at</td>
<td>412</td>
<td>Bin851</td>
<td>6</td>
<td>5.61</td>
<td>2.89</td>
<td>4.50</td>
<td>7.45%</td>
<td>Vesicle-associated membrane protein, putative, expressed</td>
</tr>
<tr>
<td>Os.12583.1.S1_at</td>
<td>213</td>
<td>Bin389</td>
<td>3</td>
<td>14.77</td>
<td>8.66</td>
<td>9.53</td>
<td>22.27%</td>
<td>Outer-membrane protein, OMP85 family, putative, expressed</td>
</tr>
<tr>
<td>Os.53449.1.A1_at</td>
<td>204</td>
<td>Bin395</td>
<td>3</td>
<td>6.80</td>
<td>8.75</td>
<td>10.04</td>
<td>11.39%</td>
<td>Expressed protein</td>
</tr>
<tr>
<td>Os.9960.1.S1_at</td>
<td>145</td>
<td>Bin857</td>
<td>6</td>
<td>23.56</td>
<td>3.29</td>
<td>4.02</td>
<td>35.33%</td>
<td>WD-40 repeat family protein, putative, expressed</td>
</tr>
<tr>
<td>Os.16410.1.S1_s_at</td>
<td>127</td>
<td>Bin853</td>
<td>6</td>
<td>4.96</td>
<td>2.89</td>
<td>4.50</td>
<td>4.61%</td>
<td>BSD domain containing protein, expressed</td>
</tr>
<tr>
<td>Os.27513.1.A1_a_at</td>
<td>103</td>
<td>Bin36</td>
<td>1</td>
<td>5.45</td>
<td>4.75</td>
<td>6.04</td>
<td>5.17%</td>
<td>Phosphoenoehydrate carboxylase, putative, expressed</td>
</tr>
</tbody>
</table>

¹N, the number of co-expression probe sets with trans-eQTLs located in the 1.5 LOQ-drop interval of corresponding cis-eQTL of the e-trait.
²LOD, the LOD value for the cis-eQTL for the regulator.
³Inf Mb, the inferior position for the 1.5 LOD-drop interval of cis-eQTLs.
⁴Sup Mb, the superior position for the 1.5 LOD-drop interval of cis-eQTLs.
⁵Var, the expression variation explained by the cis-eQTL for the regulator.

Table 3. The eight regulatory groups identified in the analysis

<table>
<thead>
<tr>
<th>Group¹</th>
<th>Members²</th>
<th>Number changed³</th>
<th>P value changed⁴</th>
<th>Gene</th>
<th>MUS6.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Os.15230.1.S1_at</td>
<td>24</td>
<td>19</td>
<td>1.14E-05</td>
<td>RFT1</td>
<td>LOC_Os05g06300</td>
</tr>
<tr>
<td>Os.10204.1.S1_at</td>
<td>10</td>
<td>7</td>
<td>1.08E-04</td>
<td>Os01g09590</td>
<td>LOC_Os01g09590</td>
</tr>
<tr>
<td>Os.12795.1.S1_at</td>
<td>2</td>
<td>2</td>
<td>7.68E-03</td>
<td>OsPRRS9</td>
<td>LOC_Os11g05930</td>
</tr>
<tr>
<td>OsAffx.28467.1.S1_at</td>
<td>11</td>
<td>11</td>
<td>0.016</td>
<td>Ghd7</td>
<td>LOC_Os07g15770</td>
</tr>
<tr>
<td>Os.12674.1.S1_s_at</td>
<td>2</td>
<td>2</td>
<td>0.022</td>
<td>Os03g11340</td>
<td>LOC_Os03g11340</td>
</tr>
<tr>
<td>Os.1189.1.S1_at</td>
<td>6</td>
<td>6</td>
<td>0.032</td>
<td>Hdl</td>
<td>LOC_Os06g16370</td>
</tr>
<tr>
<td>Os.30077.2.S1_at</td>
<td>5</td>
<td>3</td>
<td>0.033</td>
<td>Os03g16210</td>
<td>LOC_Os03g16210</td>
</tr>
<tr>
<td>Os.2720.1.S1_at</td>
<td>3</td>
<td>1</td>
<td>0.042</td>
<td>OsPEP</td>
<td>LOC_Os10g41440</td>
</tr>
</tbody>
</table>

¹Group, All local regulatory genes were classified as group names.
² Members, the number of e-traits with trans-eQTLs co-mapped with the corresponding group.
³Number changed, the number of candidate targets that may be controlled by the regulator.
⁴P value changed, the probability of change, indicating the likelihood of observing the members of the group that high up in the list by chance.
$P=2.05E^{-4}$ with a trans-eQTL (LOD=14.1, $R^2=20\%$) in the genomic position of $Ghd7$.

To confirm the network, we checked the transcript abundance of the candidate downstream targets for modulator of $Ghd7$ using flag leaves at heading date from Zhenshan 97 and NIL (mh7), which have an identical genetic background except for the introgressed segment (Xue et al., 2008). The qPCR results showed that the transcript levels of OsFTL10, OsFTL12, and $Hd3a$ were lower in NIL (mh7) than in Zhenshan 97, while the reverse was the case for OsMFT1, OsMADS50, OsMADS55, and OsMADS16 (Fig. 4). This was consistent with the result presented in the network analysis (Supplementary Table S11 available at JXB online). However, the situation for $Hd3a$ was different from the analysis of the population data (Fig. 3). Moreover, OsSPL17 and OsFTL1, putative targets of $Ghd7$ that were identified with very low correlations (0.19, $P=0.007; 0.24, P=4.8E^{-4}$), did not show significantly different gene expression variation due to sequence polymorphisms ($P=0.007$; $RFT1$, $Os01g05950$).

Application of cis-eQTLs for gene discovery of phenotypic traits

Genetic variations for gene expression and phenotypic quantitative traits are mostly the results of DNA sequence polymorphisms. Combining pQTL mapping with eQTL analysis may facilitate investigation of the molecular regulatory networks underlying quantitative traits. A gene with cis-eQTLs usually has functional polymorphisms in promoter and/or gene-coding regions, which may cause variations in gene expression levels and phenotypic effects. Indeed, phenotypic variations in a number of QTLs for phenotypic traits result from gene expression variation due to sequence polymorphisms (Chu et al., 2006; Weng et al., 2008; Li et al., 2011). Flag leaves at the heading stage, the tissue used in this analysis, have a crucial role for rice yield. We thus used data for yield and yield-related traits for QTL mapping to identify candidate genes controlling the variation of the traits. With the ultrahigh-density SNP bin map, the LOD thresholds at $P=0.05$ ranged from 4.76 to 5.10 for 1000 permutation test, with an average LOD value of 4.97, and we selected 5.0 as the threshold for cis-eQTLs (Supplementary Table S12 available at JXB online). We then calculated correlations between phenotypic trait scores and gene expression levels of e-traits with cis-eQTLs with SIs overlapping the corresponding pQTLs. The genes with significant correlations with the trait scores are shown for each trait in Supplementary Table S13 (available at JXB online). Interestingly, a pleiotropic QTL controlling the number of grains per panicle, plant height, heading date, yield per plant, and flag leaf width was shown to be associated with Bin 1006 on chromosome 7. Three genes ($LOC_Os07g16030$, $LOC_Os07g15670$, and $LOC_Os07g15770$) with larger-effect cis-eQTLs were regarded as putative candidate genes, and exhibited significant correlations with the phenotypes of the four traits (Supplementary Table S14 available at JXB online).
The cloned gene Ghd7 (LOC_Os07g15770) has been shown to be a pleiotropic QTL for number of grains per panicle, plant height, heading date, and yield per plant (Xue et al., 2008), and Zhenshan 97 lacked the Ghd7 locus while Minghui 63 had a functional allele. This confirmed a cis-eQTL with a remarkable effect on the expression level of Ghd7 (LOD= 67.6, $R^2$=68.5%).

Discussion

Construction of a regulatory network is the basis for depicting regulation of gene expression at the whole-genome level, which is fundamental for an understanding of biological processes and the functions of both individual genes and the genome. A classical gene expression network is constructed based on one gene perturbation at a time, making use of genetic materials such as mutants and NILs. Although such analysis could provide reliable information on the immediate downstream target(s) of the perturbed gene, it may not be able to provide useful information for regulators and/or targets that are in more distant positions or upstream of the perturbed gene in the network. eQTL mapping seeks to understand the genetic basis of gene expression variation and identify causal relationships between gene expression and genetic variation by analysing multifactorial natural perturbations (Jansen and Nap, 2001; Logsdon and Mezey, 2010). However, like pQTLs, an eQTL corresponds to a genomic region of variable size, usually ranging from hundreds of kilobases to megabases, which still requires huge efforts to identify possible candidate genes.

In this study, we have presented a simple, yet effective method for gene regulatory network reconstruction using an eQTL-guided function-related co-expression analysis. We combined data for gene expression from a rice RIL population, gene functional annotation of prior knowledge and iGA with the co-expression level to identify candidate regulators. Such analysis is based on the hypothesis that a regulator in a high node of the network regulates many genes and would show trans-acting effects in the form of trans-eQTLs for many e-traits that show a correlated expression profile at the population level, and the targets and regulator would be involved in the same biological process.

We used genes involved in the flowering regulation as an example to demonstrate the utility of the method in constructing a gene regulatory network, and identified and validated with high confidence a number of elements and the
pathways known to be involved in flowering regulation. Our results demonstrated that the strategy adopted here could precisely identify a putative regulator down to the single gene, representing a great improvement compared with ordinary eQTL analysis, which detected the trans-eQTL as an interval harbouring many genes. Compared with co-regulation analysis by simple correlations, the strategy here was able to provide a causal relationship between the co-expressed genes, which also represents a major improvement. In addition, the analysis revealed other genes in the network that have not been identified previously as elements in flowering regulation. Whether they really represent genes in the system that are yet to be discovered remains to be confirmed, but these may be used as clues for future experiments.

We further expanded such eQTL-guided co-expression analysis to the identification of candidate genes for pQTLs. The results also demonstrated the promise of this approach, which may lead to the identification of candidate genes controlling quantitative traits.

Another point to be discussed is that the tissue used in this study was flag leaf at the heading stage, at which time the plant is already flowering, well after flowering induction. This raises a question about the functions of these genes and the regulatory network, suggesting that they may also have functions in other biological processes besides flowering regulation. In fact, it has been shown that flowering-related genes often exhibit pleiotropic effects that cannot be explained by the extended length of the vegetative stage alone (Yan et al., 2011, 2013). For example, Ghd7 affects grain number, plant height, and heading date under long-day conditions (Xue et al., 2008), and was also reported as controlling flag leaf area (Tan et al., 2012). Combination of Hdl and Ehdl in rice can reduce the number of primary branches in a panicle, resulting in smaller spikelet numbers per panicle (Endo-Higashi and Izawa, 2011). Pleiotropic effects of flowering-related genes have also been reported in other plants. Arabidopsis LHP1 controls developmental pathways involved in organ and cell size (Gaudin et al., 2001). The potato tuberization transitions are controlled by two different FT-like paralogues (StSP3D and StSP6A) that respond to independent environmental cues (Navarro et al., 2011). Another observation is that flowering-related genes are not only expressed in the panicle, but also in the leaf, sheath, stem, hull, shoot, seedling, and even roots (Supplementary Fig. S7 available at JXB online) (Wang et al., 2010b). All this evidence suggests that these firmly regarded ‘flowering genes’ have more functions than just flowering induction. However, whether the compositions and architectures of the regulatory networks in terms of the genes and pathways are the same in different processes remain to be determined.

Gene expression is a dynamic process depending on development stages, and different tissues or cells, as well as genetic backgrounds and growth environments (Jiao et al., 2009; Wang et al., 2010b). Thus, the results and interpretations of the eQTL analysis are also valid in terms of the tissues from which the RNA samples were collected. Whether the regulatory network and/or co-expression relationships between the genes identified in a single tissue study are tissue specific, or are to some extent also valid in other tissues, remains to be investigated.

It should also be noted that the outcome of the eQTL detection is critically dependent on the genetic composition of the mapping population defined by the polymorphisms between the parents. For example, Komiya et al. (2009) confirmed that the OsMADS50–Ehd1–RFT1 pathway is involved in floral activation under long-day conditions. However, no cis-eQTL was detected for either OsMADS50 or Ehd1, probably because of no sequence difference in the promoter regions of these genes between the parents of our population. Consequently, neither OsMADS50 and Ehd1 was identified as a trans-eQTL for other e-trait, and hence were not suggested as regulators by the eQTL analysis. Clearly, the parental genotypes pose a limiting factor for the extent to which the components in the regulatory network can be discovered in the analysis. Multiple populations of diverse parents may be used to provide complementary information for the network construction.

A further challenge is to identify the cis-acting element in each case that regulates the expression variation of the e-trait, or the causal agent of the cis-eQTL. The definition of cis-effect in eQTL analyses is somewhat arbitrary, varying from 5 Mb (Morley et al., 2004) to 10 kb of the gene (Yvert et al., 2003), or a 1.5 LOD-drop SI of the QTL (Keurentjes et al., 2007). In principle, a cis-eQTL should be caused by polymorphisms within a neighbouring range of the gene, including polymorphisms within the gene itself that alter the cis-acting elements or sites for mRNA stability, or polymorphisms between parents in the promoter region (Heap et al., 2009; Smith et al., 2011). The above definition does not exclude the possibility that the regulatory element may act in trans, as in the case of the AMN1 gene in Saccharomyces cerevisiae that was found to be a local but trans-acting through feedback regulation due to a coding polymorphism in its transcript (Ronald et al., 2005). In any case, it is essential to find sequence polymorphisms between the parents and associate such polymorphisms with the expression level variation of the gene for identifying the cis-acting elements of the e-trait.

The large number of regulators and the regulatory network identified in the study only provide a starting point for understanding the system. Although we have illustrated that some are supported by results from previous studies, huge effort is needed experimentally to confirm any of the master regulators and the regulatory network.

Supplementary data

Supplementary data is available at JXB online.

Supplementary Fig. S1. An example of a flag leaf harvested at the heading stage.

Supplementary Fig. S2. The distribution of e-trait and probe sets with locus support in the genome.

Supplementary Fig. S3. Distribution of R² values for cis-eQTLs and trans-eQTLs individually.

Supplementary Fig. S4. LOD curve for expression variation that was calculated with the full set of probe sets and modified probe sets of RFT1 and Hd3a on chromosome 6.
Supplementary Fig. S5. LOD curve for expression variation, which was calculated with the full set of probe sets and modified probe sets of RFT1 and Hd3a on chromosome 7.

Supplementary Fig. S6. Gene expression levels for RFT1 and Hd3a in Zhenshan 97 and Minghui 63 of three biological replicates.

Supplementary Fig. S7. Heatmap of the global development stage of the flowering-related genes identified in our network.

Supplementary Table S1. Primers used in qPCR analysis for the candidate targets expression in NIL(mh7) and Zhenshan 97.

Supplementary Table S2. Numbers of observed and expected e-trait in 1619 bins.

Supplementary Table S3. Details of 13 647 significant gene expression quantitative trait loci (eQTLs) \((P \geq 0.05)\), affecting the expression of 10 725 rice unique probe sets.

Supplementary Table S4. Information on observed trans-eQTLs, expected trans-eQTLs, and e-trait with cis-QTLs in 1619 bins.

Supplementary Table S5. The significant master candidate regulators.

Supplementary Table S6. The enrichment gene ontology (GO) terms for significantly associated targets for six master regulators.

Supplementary Table S7. Information on 577 unique loci involved in flowering time and development.

Supplementary Table S8. Cis and trans-eQTL identification of flowering-associated genes.

Supplementary Table S9. Total number of unique correlation pairs: 7808 for 61 candidate regulators and 171 potential targets.

Supplementary Table S10. Gene expression for RFT1 and Hd3a was recalculated with the full set of probe sets and modified probe sets with probes with less than five SNPs removed between RFT1 and Hd3a.

Supplementary Table S11. Expression correlation for regulator and corresponding targets.

Supplementary Table S12. The statistics of 25 pQTLs identified using the ultrahigh-density SNP bin map.

Supplementary Table S13. The most significant candidate gene for each trait.

Supplementary Table S14. Correlation between phenotypic traits and e-trait with cis-eQTL in the SI of QTL controlling corresponding traits.

Acknowledgements

This research was supported by grants from the National 863 Project (2012AA10A304) and the National Natural Science Foundation (31330039) of China, and the Bill & Melinda Gates Foundation.

References


