1 Title:

2 **Genome-wide association study of flowering time reveals complex genetic**
 3 **heterogeneity and epistatic interactions in rice**

4

5 Authors:

6 **Chang Liu, Yuan Tu, Shiyu Liao, Xiangkui Fu, Xingming Lian, Yuqing He,**
 7 **Weibo Xie (✉), Gongwei Wang (✉)**

8

9 National Key Laboratory of Crop Genetic Improvement and National Center of Plant
 10 Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, China

11

12 Corresponding author:

13 Weibo Xie(✉) E-mail: weibo.xie@mail.hzau.edu.cn Tel: 86-15327378537

14 Gongwei Wang (✉) E-mail: gwwang@mail.hzau.edu.cn Tel: 86-15827398206

15

Abstract Since domestication, rice has cultivated in a wide range of latitudes with different day lengths. Selection of diverse natural variations in heading date and photoperiod sensitivity is critical for adaptation of rice to different geographical environments. To unravel the genetic architecture underlying natural variation of rice flowering time, we conducted a genome wide association study (GWAS) using several association analysis strategies with a diverse worldwide collection of 529 *O. sativa* accessions. Heading date was investigated in three environments under long-day or short-day conditions, and photosensitivity was evaluated. By dividing the whole association panel into subpopulations and performing GWAS with both linear mixed models and multi-locus mixed-models, we revealed hundreds of significant loci harboring novel candidate genes as well as most of the known flowering time genes. In total, 127 hotspots were detected in at least two GWAS. Universal genetic heterogeneity was found across subpopulations. We further detected abundant interactions between GWAS loci, especially in *indica*. Functional gene families were revealed from enrichment analysis of the 127 hotspots. The results demonstrated a rich of genetic interactions in rice flowering time genes and such epistatic interactions contributed to the large portions of missing heritability in GWAS. It suggests the increased complexity of genetic heterogeneity might discount the power of increasing the sample sizes in GWAS.

Keywords Flowering time; GWAS; epistatic interactions; genetic heterogeneity; rice

Introduction

Cultivated rice (*Oryza sativa* L.), a major cereal especially in Asia, consists of two subspecies, *indica* and *japonica*. Since its domestication, rice has cultivated in a wide range of latitudes with different day lengths, from tropical to temperate regions. As a facultative short day (SD) plant, rice flowering can be promoted under SD conditions, whereas be repressed under long day (LD) conditions. A critical determinant for adaptation of rice to different geographical latitudes is the breeding selection of diverse natural variations in photoperiod sensitivity (Izawa 2007). For example, to ensure successful pollination and seed setting, rice cultivars grown in northern temperate regions need to flower very quickly under LD conditions during a short summer period, showing much less photoperiod sensitivity than those in other regions.

A complex genetic network controlling rice flowering time in response to photoperiod has been elucidated. It shows similarities as well as divergences to that in *Arabidopsis*. In rice, *OsGI*, *Hd1* and *Hd3a* are determined as orthologs of *Arabidopsis* *GI*, *CO* and *FT*, respectively (Hayama et al. 2003; Kojima et al. 2002; Yano et al. 2000). Although these components are conserved, their regulation mode has been modified. In *Arabidopsis* *CO* only activates transcription of the florigen *FT* and promotes flowering under LD conditions. In rice, *OsGI* regulates the expression of *Hd1* and *OsMADS51*. *Hd1* has dual function, repressing *Hd3a* expression and flowering under LD while promoting flowering under inductive SD condition (Hayama et al. 2003). *Hd1* functions as a flowering repressor through interaction with *Ghd7* (Nemoto et al. 2016). In contrast to *Arabidopsis*, rice can flower under non-inductive LD condition by inducing expression of a second florigen *RFT1* (Komiya et al. 2009). In addition, a unique *Ehd1* pathway, absent in *Arabidopsis*, was first discovered by Doi et al. (2004). *Ehd1* encodes a B-type response regulator and up-regulates florigen gene expression. It was shown that most LD repressors, such as

Ghd7, *DTH8/Ghd8* and *OsMADS56*, act through *Ehd1* by reducing its expression and that of its downstream florigen genes (Ryu et al. 2009; Wei et al. 2010; Xue et al. 2008; Yan et al. 2011). Under LD, *Hd1* is a strong repressor of *Ehd1* (Du et al. 2017; Goretti et al. 2017; Nemoto et al. 2016). *Ehd1* and *Ghd7* set critical day length for *Hd3a* florigen expression. *Ehd1* induction is mediated by *OsGI* and blue light at dawn, while *Ghd7* expression at dawn prevents *Ehd1* expression. However, under SD condition inducibility of *Ghd7* shifts to the dark phase and de-represses *Ehd1* (Itoh et al. 2010). Recently, several regulators which could promote *Ehd1* expression were also identified. *OsId1/RID1/Ehd2* acts as a master switch for flowering promotion under LD by up-regulating *Ehd1* expression (Matsubara et al. 2008; Park et al. 2008; Wu et al. 2008). *Ehd3* down-regulates *Ghd7* transcription, thus allowing *Ehd1* up-regulation under LD condition (Matsubara et al. 2011). *OsMADS50* strongly promotes *Ehd1* expression in LD condition, and *OsMADS50*, *DTH3* and *Hd9* are likely multiple alleles (Bian et al. 2011; Ryu et al. 2009). *Ehd4*, another activator of *Ehd1*, encodes a nucleus localized CCCH-type zinc finger protein unique to rice (Gao et al. 2013).

Although natural variation in flowering time has been studied extensively in rice, most of them was performed based on bi-parental quantitative trait locus (QTL) linkage mapping approach, with very limited range of allelic diversity and genomic resolution. In present study, using a diverse worldwide collection of 529 *O. sativa* accessions re-sequenced on the Illumina HiSeq 2000, the genetic architecture of natural variation in rice flowering time was characterized through GWAS with several association analysis strategies. Heading date and photosensitivity were investigated, and hundreds of significant association loci were identified. We detected 127 genomic hotspots associated with variation of rice flowering time by dividing the whole study population into subpopulations and using both linear mixed models and multi-locus mixed-models. We further analyzed interactions between loci which had been

identified by GWAS in at least one (sub)population. A rich of genetic heterogeneity and epistatic interactions between flowering time genes were revealed in rice.

Materials and Methods

The association panel

The association panel consisted of a diverse collection of 529 *O. sativa* accessions including both core/ mini core collections and parental lines in breeding program. The details about sequencing, SNP identification, and imputation were described in Xie et al. (2015). The population structure of the whole association panel was inferred using ADMIXTURE (Alexander et al. 2009). The set of 529 rice accessions was classified into 98 *indica I* (*IndI*), 105 *indica II* (*IndII*), 92 *indica* intermediate, 91 temperate *japonica* (*TeJ*), 44 tropical *japonica* (*TrJ*), 21 *japonica* intermediate, 46 *Aus*, 14 intermediate group (*VI*), and 18 intermediate. Information about the accessions, including accession name, country of origin, longitude and latitude origin, and subpopulation identity, has been reported previously (Wang et al. 2015; Xie et al. 2015) and is available at the RiceVarMap (<http://ricevarmap.ncpgr.cn>).

Field experiments and phenotyping

Field trials were carried out in three environments. The rice seeds were sown in the Experimental Station of Huazhong Agricultural University, Wuhan (central China, 30°28'N), in May of 2011 and 2012, and additionally in the Experimental Station of Lingshui County of Hainan Island (southern China, 18°48'N) in December of 2011. Seedlings about 25 days old were transplanted to the field. The field planting followed a randomized complete block design with two replications. Each plot consisted of four rows with 10 plants each. The planting density was 16.5 cm between plants in a row, and the rows were 26 cm apart. Field management, including irrigation, fertilizer

application and pest control, followed essentially the normal agricultural practice.

Heading dates were recorded as the number of days from sowing to the time when the first panicles emerged above the flag leaf sheathes for half of the individuals in an accession. Heading dates in three environments for the 529 accessions were collected during the summer of 2011 and 2012 in Wuhan under a long-day condition (13.5~14.2h) and during the spring of 2012 in Lingshui under a typical short-day condition (11.0~12.5h). The correlation coefficient between two years of heading dates in Wuhan is 0.86, while the correlation coefficients between Hainan and Wuhan were 0.41 and 0.43, respectively.

Genome-wide association analyses

Only SNPs with $MAF \geq 0.05$ and the number of accessions with the minor allele ≥ 6 in a (sub)population were used to carry out GWAS. There are 2,046,642, 2,671,688, 2,767,191, 1,041,514, 1,857,866, and 3,916,415 SNPs used in GWAS for subpopulations of *IndI*, *IndII*, *Indica* (including *IndI*, *IndII*, and *indica* intermediate), *TeJ*, *Japonica* (including *TeJ*, *TrJ*, and *japonica* intermediate) and the whole population, respectively. Totally, 4,634,871 SNPs were involved in at least one GWAS. We performed GWAS using the linear mixed model (LMM) and the simple linear regression model (LR) provided by FaST-LMM program (Lippert et al. 2011) and multi-locus mixed-model (MLMM) from an R script provided by Segura et al. (2012). Population structure was modeled as a random effect in LMM using the kinship (K) matrix and we found that it was enough to control spurious associations, for the genomic inflation factor was near one in all GWAS. The evenly distributed random SNP set for analyzing population structure was used to calculate K. The kinship coefficients were defined as the proportion of identity genotype for the 188,165 randomly selected SNPs for each pair of individuals (Zhao et al. 2007). To reduce the

computation burden, after obtaining association results of LR and LMM for each SNP, we split the genome into 50 kb regions and selected ten SNPs with the strongest association signals detected by LR and ten SNPs by LMM in the region. Ten SNPs in each 50 kb regions were usually included most of SNPs with P -value of LMM < 0.01 in our GWASs. Association studies using MLM were carried out using these SNPs. A uniform threshold $P = 5.0 \times 10^{-6}$ was set to identify suggestive significant association signals by LMM (Chen et al. 2014; Wang et al. 2015). To obtain independent association signals, multiple SNPs exceeding the threshold in a 5 Mb region were clustered by r^2 of LD ≥ 0.25 and SNPs with the minimum P -value in a cluster were considered as lead SNPs.

Gene nomenclature

Rice genes with CCT domain and their nomenclature were obtained from Cockram et al. (2012). The nomenclature of MADS box genes was according to the annotation version 6.1 of genomic pseudomolecules of *japonica* cv. Nipponbare from Michigan State University (MSU).

Epistatic interactions analysis

The analysis of two-locus interactions was carried out only between significant lead SNPs identified by LMM or MLM. We fitted the following linear mixed model to identify two-locus interactions:

$$Y = X\beta + M_1b_1 + M_2b_2 + (M_1 \odot M_2)b_3 + u + \varepsilon$$

Where Y is a vector of a phenotype. The X is a matrix of fixed effects excluding SNPs, M_1 and M_2 are the selected two SNPs, $u \sim N(0, \sigma_g^2 K)$ is a random effect, and ε is the noise term.

Gene ontology (GO) enrichment analysis

The GO classifications of rice genes were downloaded from Gramene (<http://www.gramene.org>). Only the terms in the biological process category were used for GO analysis. The R package topGO was used to carry out GO enrichment analysis using Fisher's exact test and the weight method (Alexa et al. 2006).

Gene family enrichment analysis of GWAS hotspots

The Pfam domain information of rice genes from the MSU annotation version 6.1 was used in this analysis. Only gene-Pfam hits with E-value $\leq 1 \times 10^{-10}$ were considered. A total of 472 Pfam domains owned by at least ten and at most 150 genes were used for enrichment analysis. A software INRICH was used to carry out interval-based enrichment analysis for GWAS hotspots (Lee et al. 2012).

Results

GWAS of heading date and photosensitivity in rice

Genome-wide association analyses were performed separately in the whole population and in the *IndI*, *IndII*, *indica* (consisting of *IndI*, *IndII* and *indica* intermediate), *TeJ*, and *japonica* (consisting of *TeJ*, *TrJ* and *japonica* intermediate) subpopulations for each environment. A total of 156 lead SNPs (the SNP with the lowest *P* value in a region) corresponding to 131 genomic clusters (adjacent lead SNPs in less than 100 kb were considered as a cluster) were detected in at least one population for the three datasets of heading date at linear mixed models (LMM). The analyses using multi-locus mixed-model (MLMM) based on the extended Bayesian information criterion (EBIC) detected a total of 234 lead SNPs with 198 genomic clusters. The details about these significant association signals are listed in Table S1

(for LMM) and Table S2 (for MLMM). The quantile-quantile plots and Manhattan plots for heading date of Wuhan_2012 in the whole population are illustrated in Fig. 1 as an example.

To search for candidate genes, the significant GWAS signals were compared with the positions of known or putative genes involved in flowering time pathway. We found that some known genes associated with rice flowering time, such as *Ghd7*, *Hd9* (*DTH3*), *Ehd1*, *OsMADS51*, and *SRT5*, were located within 50 kb adjacent to the identified lead SNPs (Table 1). Additional loci near known genes *Ghd8* and *Hd9* (*DTH3*), which were not significant at LMM model, could be detected using MLMM. *Hd9* (*DTH3*) was detected by both LMM and MLMM in short-day condition using the whole population or the *IndIII* subpopulation, but only be detected by MLMM when using all *indica* accessions (Table 1). There were three independent lead SNPs for the three detections, which suggested allelic heterogeneity in *Hd9* (*DTH3*) in different subpopulations. Genes with CCT (CONSTANS, CO-like, and TOC1) or MADS box domains are usually associated with variations of rice flowering time. We found that three CCT genes (*OsPRR59*, *OsP* and *OsCMF10*) and two MADS genes (*OsMADS87* and *OsMADS30*) were also located close to the identified lead SNPs. These genes have not been reported yet but should be regarded as important candidates for flowering time in rice (Murakami et al. 2005; Tsuji et al. 2011). Lead SNPs near *OsMADS87* and *OsMADS30* passed the suggestive threshold in short-day condition in *indica* accessions by LMM, but they were only detected by MLMM in the whole population (Table 1), which suggests that LMM and MLMM might have complementary properties in GWAS. Other candidate genes for GWAS loci of heading date were also suggested in Tables S1 and S2.

Furthermore, the photosensitivity of accessions was investigated and used as a derived trait for GWAS. The differences of heading date between Wuhan and

Lingshui were used to measure the photosensitivity. Totally 103 lead SNPs corresponding to 86 genomic clusters were detected by LMM (Table S3) and 25 of them were located near lead SNPs detected for heading date (<100kb). The analyses using MLM detected a total of 184 lead SNPs with 168 genomic clusters (Table S4), 37 of which were located near lead SNPs detected for heading date (<100kb). Additional significant loci near known genes were detected for photosensitivity (Table 1), which suggests an important role of environment interactions in rice heading date. The CCT gene *OsPRR95* and three MADS box domain genes (*OsMADS75*, *OsMADS29* and *OsMADS68*) were important candidates for the photosensitivity in rice. Intriguingly, we did not detect association signal for *Hd6*, but signals near a gene (LOC_Os07g02350) similar to *Hd6* were detected by LMM in *IndI* and MLM in *IndII* for difference of heading date between Wuhan and Lingshui in 2011 (Table 1). Compared with *Hd6*, LOC_Os07g02350 has 147 more amino acids at the N-terminal, and only 7 amino acids are different in the homologous region (Fig. S1). Other candidate genes for GWAS loci of photosensitivity were also suggested in Tables S3 and S4. The quantile-quantile plots and Manhattan plots for the differences of heading date between Wuhan_2012 and Lingshui in the whole population are illustrated in Fig. 2 as an example.

Besides known flowering time genes encoding proteins, we also observed that a lot of microRNAs were located near lead SNPs. miR156 (Xie et al. 2006), miR172 (Aukerman and Sakai 2003), miR159 (Achard et al. 2004), and miR399 (Kim et al. 2011), which have been shown to play important roles in flowering time in rice or *Arabidopsis*, were located near lead SNPs (Tables S1, S2, S3, and S4).

Taken together, there was a total of 248 lead SNPs detected by LMM (around ten lead SNPs were detected in average for each analysis) corresponding to 194 genomic clusters. Among them, thirty lead SNPs reoccurred in at least two GWASs. If taking

different SNPs in a cluster as a common causal gene, there were 54 clusters detected in at least two GWASs, containing 108 lead SNPs. To estimate the genome-wide error rate in the GWASs, we performed 100 permutations for each analysis, with the details of the threshold shown in Table S5. The permutation results suggest that there were 2.37 false positives in one GWAS. The analyses using MLM detected a total of 416 lead SNPs corresponding to 333 genomic clusters, of which 143 SNPs were near lead SNPs detected by LMM (<100kb). The selected optimal model of MLM for each GWAS contained fourteen SNPs in average, ranging from three for some GWAS using *TeJ* to nineteen for the whole population. We observed that lead SNPs detected by LMM were enriched with lower MAF comparing with MLM (Fig. S2). When merging the SNPs detected by LMM and MLM together, there were a total of 611 SNPs forming 429 genomic clusters. Among them, 127 clusters were detected in at least two GWASs, and were thus referred to as hotspots. The 127 hotspots contained 309 lead SNPs, and 95 of them were detected by at least one LMM (Table S6). Because the less stringent threshold used might increase the risk of including false positives, we then focused on loci located in the 127 hotspots for further analysis.

Universal genetic heterogeneity across subpopulations revealed by GWAS of rice flowering time

Most of the 127 hotspots contained multiple SNPs detected from different GWASs. We examined 151 SNPs located in the hotspots and detected by LMM in at least one GWAS, and found that only 44 SNPs showed polymorphic in all five subpopulations, indicating that a lot of loci were differentiated only in some of the subpopulations. Only 40 of the 151 SNPs exceeded the arbitrary threshold of $P_{LMM} \leq 5 \times 10^{-6}$ for more than one subpopulation, and no SNP exceeded the threshold in both *indica* (*IndI*, *IndII* and *indica* intermediate) and *japonica* (*TeJ* and *japonica* intermediate)

subpopulations simultaneously, suggesting the presence of universal genetic heterogeneity in rice flowering time. However, 102 hotspots were detected in more than one subpopulation, illustrating that there were multiple functional haplotypes in most of the hotspots. Interestingly, there were multiple independent lead SNPs, which were tightly close to each other and detected in one GWAS, in five clusters (Table S7). We found that such close lead SNPs detected in a GWAS were with opposite effects and resulted from independent variations (Table S7). Since the conditional P values for two SNPs decreased dramatically when they were both included in a model, they could not be detected by MLM. Further researches were needed for distinguishing whether these were variations in different but tightly linked genes or independent variations of common genes.

Abundant epistatic interactions revealed by GWAS of rice flowering time

Since the *indica* population for GWAS contains both *IndI* and *IndII* while *japonica* includes *TrJ* and *TeJ*, we examined if loci detected in a subpopulation could be also detected in the union populations. Interestingly, for lead SNPs located in the 127 hotspots, we observed that there were 27 SNPs with $P_{LMM} \leq 5 \times 10^{-5}$ in *TeJ* and 24 of them were also with $P_{LMM} \leq 5 \times 10^{-5}$ in *japonica*. In contrast, there were 27 and 30 SNPs with $P_{LMM} \leq 5 \times 10^{-6}$ in *IndI* and *IndII*, but only two and six of them were with $P_{LMM} \leq 5 \times 10^{-5}$ in *indica*, respectively. These observations indicated that the increased complexity of genetic background might neutralize the effect through increasing the number of samples in GWAS.

For example, we found that *Ghd7*, a gene suppresses flowering under long-day conditions by suppressing the expression of *Ehd1*, was detected in *IndI* (Fig. 3a) but not in *indica* (Fig. 3b) in Wuhan_2011. Since the functional variation of *Ghd7* in *indica* is mainly a complete deletion of GHD7 (Xue et al. 2008), all lead SNPs should

be resulted from indirect associations with the deletion. We examined the lead SNP sf079137593 near *Ghd7* detected by LMM in *IndI* in Wuhan_2011, and found that the SNP could only represent the deletion variation of GHD7 in *IndI* but not in other *indica* accessions ($r^2 = 0.70$ and 0.31 in *IndI* and *indica*, respectively. See Table S8 for details), suggesting the gains of GWAS in subpopulations. However, another SNP sf079157719, which was detected by MLMM and represent the deletion of GHD7 well in *indica* ($r^2 = 0.93$ and 0.77 in *IndI* and *indica*, respectively), still couldn't be significantly detected in GWAS of *indica* either.

As many epistasis between loci controlling heading date in rice have been discovered, we tried to examine whether such epistatic interactions could be detected in our data and might lead to missing of detecting *Ghd7* in *indica*. To reduce computational requirements and false positives, only the 611 lead SNPs detected by LMM or MLMM methods in our GWASs were used to analyze interactions. The interactions of any two loci were examined in *indica* population if the minimum numbers of individuals of all four genotype combinations (denotes as N_c) were no less than five. When scanning the interactions of *Ghd7* towards other lead SNPs, we found that a lead SNP sf1017003447 near *Ehd1* had the most significant P value of epistatic interaction (Fig. 3c), which was consistent with previous genetic researches (Xue et al. 2008). When only considering single locus, both lead SNPs were not significant ($P=0.084$ for sf079157719 and $P=0.236$ for sf1017003447). But when considering interaction, not only the interaction itself was extremely significant ($P = 2.13 \times 10^{-4}$) but also the two loci were significant ($P=0.027$ for *Ghd7* and $P=0.029$ for *Ehd1*) (Fig. 3d). We also investigated the situation in *IndI*, and found that the lead SNP sf1017003447 in *IndI* was nearly fixed (Fig. 3e), therefore, the effect of interaction did not cause trouble for the identification of *Ghd7* in *IndI*.

We further inspected the ten most significantly interacted SNPs in *japonica* for

each of the 24 SNPs with $P_{LMM} \leq 5 \times 10^{-5}$ in both *TeJ* and *japonica*. These associations are likely not affected by population differentiation. Indeed, there were 112 interactions with $P \leq 0.01$ in *japonica*. Ninety of them (80.4%) were with $N_c \geq 5$ in *TeJ* and 31 (27.7%) were also significant with $P \leq 0.01$ in *TeJ*. However, for the 25 SNPs with $P_{LMM} \leq 5 \times 10^{-6}$ in *IndI* and $P_{LMM} \geq 5 \times 10^{-5}$ in *indica* (i.e., detected in *IndI* only, These associations are likely affected by population differentiation), there were 191 interactions with $P \leq 0.01$ in *indica* but only 36 of them (18.9%) were with $N_c \geq 5$ in *IndI* and none of them was significant with $P \leq 0.01$ in *IndI*. The distribution of the frequencies of N_c for these SNPs was shown in Fig. 3f, which indicated that for interactions significant in *indica*, a lot of genotype combinations did not exist in *IndI*. Similar situations were also observed in *IndII*. But for interactions significant in *japonica*, many of them were also found in *TeJ*. We thus propose that the new emerging interactions in *indica* might account for the missing of the significant lead SNPs that were able to be detected in *IndI* or *IndII*. Such results also suggest that epistatic interaction is an important part of GWAS and should be considered as the large component of missing heritability in GWAS (Brachi et al. 2011).

Functional gene families revealed from enrichment analysis of GWAS hotspots for rice flowering time

Based on the 127 GWAS hotspots, we assessed whether genes with certain conservative domains enriched in genomic regions near lead SNPs using INRICH (Lee et al. 2012). We found that CCAAT binding factors (PF00808), MADS-box transcription factors (PF00319), Myb-like DNA-binding domain proteins (PF00249), and CCT domain proteins (PF06203), which are usually harbored by known rice flowering time genes, were ranked in the top of enrichment analysis with $P \leq 0.1$ (Table 2). The top significant enriched Pfam protein domains contained a lot of

transcription factors, suggesting that transcript regulation might be the main way to regulate flowering in rice. Intriguingly, we found that besides *Ghd8* which encodes a putative HAP3 subunit of the trimeric HAP2/HAP3/HAP5 complex (also known as CCAAT binding factor), three genes (LOC_Os02g07450, LOC_Os03g63530, and LOC_Os06g45640) encoding HAP5 subunit were also located near lead SNPs. It is proposed that CCT domain proteins share similarity to HAP2 and might replace that component to form the trimeric complex (Laloum et al. 2012; Wenkel et al. 2006). It seems that our results provided evidences for this supposal, for all three components were significantly enriched in GWAS results. Genes encoding histones (PF00125) were enriched, which might indicate the important roles of histone modifications for regulating flowering time in rice (He et al. 2003). Genes encoding kinesin (PF00225) were also enriched. A recent study showed that a kinesin OsGDD1 acted as a transcription factor for the synthesis of the phytohormone gibberellin in rice (Li et al. 2011), suggesting a possible role of kinesin genes for regulating flowering time in rice. Further experiments were needed to elucidate whether the genes encoding the enriched Pfam domain were true positive.

Discussion

Heading date is a very important and complex trait that controls adaptation of rice varieties to their local environment and yield performance, and flowering time loci are often the targets of both natural and artificial selection. The trait is strongly affected by population structure, and the loci often exhibit complex forms of allele sharing and admixture in diverse germplasm. Previous GWAS studies in rice yielded only a few known loci associated with heading date, with some or even a large portion of known functional genes failed to be identified (Huang et al. 2012; Zhao et al. 2011). In present study, to unravel the genetic architecture underlying natural variation of

rice flowering time, using a diverse worldwide collection of 529 *O. sativa* accessions as the GWAS platform, we adopted several association analysis strategies, which enabled us to detect hundreds of significant association loci. Heading date in three environments was recorded during the summer seasons of 2011 and 2012 in Wuhan (central China, 30°28'N) under a long-day condition (13.5~14.2h) and during the spring season of 2012 in Lingshui (southern China, 18°48'N) under a typical short-day condition (11.0~12.5h). Photosensitivity of rice accessions was evaluated (differences of heading date between long-day condition in Wuhan and short-day condition in Lingshui) and used as a derived trait for GWAS. Previous studies discovered that additional association signals can be detected when GWASs are performed on subpopulations and suggested the existence of genetic heterogeneity across subpopulations in rice (Huang et al. 2010; Huang et al. 2012; Zhao et al. 2011). In addition to the whole population, we also divided it into several subpopulations and performed GWAS. The whole genome scanning of all SNPs with MAF $\geq$ 0.05 in the target population was first carried out using both a simple linear regression (LR) and a linear mixed model (LMM) (Lippert et al. 2011). A recent study showed that the multi-locus mixed-model (MLMM) outperforms the single-locus model (Segura et al. 2012). We also performed stepwise mixed-model regressions to construct multi-locus models based on preliminary LMM and LR results. Our results suggest that LMM and MLMM might have complementary properties in GWAS. We revealed hundreds of significant loci harboring novel candidate genes as well as most of the known flowering time genes. According to literature, there are eight rice flowering time genes identified by map-based cloning, and five of them (*DTH3*, *Hd3a*, *Ghd7*, *Ghd8*, and *Ehd1*) were detected in at least two GWASs in our study. Although GWAS is a powerful tool, due to the limitation of population size, GWAS will show its limitations, which may lead to the occurrence of false-positive. How to reduce the

occurrence of false-positive and improve the accuracy of detection remains to be further studied.

Genetic heterogeneity across subpopulations for flowering time was analyzed in depth in this study. We found that a lot of significant loci were differentiated only in some of the subpopulations. No lead SNPs exceeded the threshold in both *indica* (*IndI*, *IndII*, and *indica* intermediate) and *japonica* (*TeJ*, *TrJ*, and *japonica* intermediate) subpopulations simultaneously. Some close lead SNPs detected in a GWAS were even with opposite effects and resulted from independent variations. These results suggest the presence of universal genetic heterogeneity in rice flowering time. Further, it is interesting to note that, especially in *indica*, the increased complexity of genetic background might neutralize the effect through increasing the number of samples in GWAS. For epistatic interactions between flowering time loci significant in *indica*, a lot of genotype combinations did not exist in *IndI*, either in *IndII*. But for those interactions significant in *japonica*, many of them were also found in *TeJ*. We propose that the new emerging interactions in *indica* might account for the missing of the significant lead SNPs that were able to be detected in *IndI* or *IndII*.

In present study, besides known genes and microRNAs in flowering time regulation located around lead SNPs, a lot of candidate genes were also suggested for the GWAS loci. We found that CCAAT binding factors, MADS-box transcription factors, Myb-like DNA-binding domain proteins, and CCT domain proteins, which are usually harbored by known rice flowering time genes, was ranked in the top of enrichment analysis with $P \leq 0.1$ in the 127 GWAS hotspots. Many genes with known domains associated with flowering time were first reported in our study and should be good candidates for further studies. We noticed that *Hd1*, the first map-based cloned flowering time gene in rice, which was observed to be corresponding to a ‘mountain range’ in GWAS by Zhao et al (2011), could not be even detected in our study. Instead,

we observed a ‘mountain range’ in *Ehd1* region in several GWASs. This might be due to different rice lines and environments in different studies. For example, *Ghd7* was cloned by our laboratory and could be detected clearly in different seasons and populations but did not display signal in Zhao et al (2011). Such observations suggest calls for collaborations that a common set of rice lines should be planted in different places around the world and such collaborations would lead to detection of more functional genes.

Authors’ contributions

WX and GW performed data analysis and wrote the manuscript; CL performed the final data analysis; YT, SL, XF, XL, and YH conducted field trials and collected the phenotypic data.

Funding information

This work was supported by grants from the Ministry of Agriculture of China (2016ZX08009002), National 863 Project (No.2014AA10A600), the earmarked fund for the China Agriculture Research System (CARS-01-03) of China, and the Fundamental Research Funds for the Central Universities (Program No. 2662016PY065).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Table 1 A subset of significant lead SNPs near known genes identified by GWAS for rice flowering time in Wuhan and Lingshui

Population	Chr	Pos (bp)	P_{LMM}	P_{MLMM}	q^2	MAF	Dis(kb)	Candidate locus	Symbol
Associated with heading date in Wuhan in 2011, long-day condition									
<i>All</i>	3	13139351	5.4×10^{-7}	-	0.0079	0.074	12	LOC_Os03g22770	<i>OsP(CCT)</i>
<i>All</i>	11	2752748	3.6×10^{-7}	6.2×10^{-18}	0.0642	0.068	32	LOC_Os11g05930	<i>OsPRR59</i>
<i>indica</i>	1	40348638	3.5×10^{-6}	1.1×10^{-25}	0.0497	0.054	0	LOC_Os01g69850	<i>OsMADS51</i>
<i>Indl</i>	2	20725893	9.3×10^{-7}	5.1×10^{-11}	0.3773	0.143	10	LOC_Os02g34560	<i>OsCYT-INV1/SRT5</i>
<i>Indl</i>	7	9157719	2.9×10^{-6}	2.5×10^{-17}	0.3677	0.177	4	LOC_Os07g15770	<i>Ghd7</i>
<i>Indl</i>	8	4294440	5.8×10^{-3}	9.1×10^{-16}	0.0758	0.082	38	LOC_Os08g07740	<i>DTH8/Ghd8</i>
Associated with heading date in Wuhan in 2012, long-day condition									
<i>All</i>	3	13139351	1.4×10^{-9}	-	0.0107	0.074	12	LOC_Os03g22770	<i>OsP(CCT)</i>
<i>All</i>	7	9177474	2.4×10^{-4}	4.3×10^{-7}	0.0252	0.434	23	LOC_Os07g15770	<i>Ghd7</i>
<i>indica</i>	7	9140779	1.6×10^{-6}	-	0.1772	0.068	11	LOC_Os07g15770	<i>Ghd7</i>
<i>indica</i>	8	4319611	3.1×10^{-6}	-	0.1012	0.075	13	LOC_Os08g07740	<i>DTH8/GHD8</i>
<i>Indl</i>	7	9140475	1.5×10^{-6}	2.0×10^{-23}	0.3006	0.146	11	LOC_Os07g15770	<i>Ghd7</i>
<i>japonica</i>	7	9202668	3.4×10^{-10}	3.3×10^{-20}	0.2891	0.058	48	LOC_Os07g15770	<i>Ghd7</i>
<i>japonica</i>	6	17917224	4.3×10^{-7}	-	0.3072	0.141	27	LOC_Os06g30830	<i>OsMADS76</i>
<i>japonica</i>	2	24944723	3.8×10^{-3}	1.3×10^{-7}	0.0505	0.494	28	LOC_Os02g41550	<i>OsCRY2</i>
<i>TeJ</i>	7	9202668	3.2×10^{-7}	2.2×10^{-18}	0.2740	0.099	48	LOC_Os07g15770	<i>Ghd7</i>
Associated with heading date in Lingshui in 2012, short-day condition									
<i>All</i>	10	17119647	1.8×10^{-9}	1.8×10^{-8}	0.0162	0.291	31	LOC_Os10g32900	<i>OsCMF10 (CCT)</i>
<i>All</i>	3	1256780	2.6×10^{-6}	5.2×10^{-13}	0.0534	0.267	12	LOC_Os03g03070	<i>DTH3/Hd9</i>
<i>All</i>	3	21446165	1.3×10^{-5}	6.1×10^{-9}	0.0260	0.145	18	LOC_Os03g38610	<i>OsMADS87</i>
<i>All</i>	6	27634041	2.6×10^{-4}	1.2×10^{-8}	0.0212	0.074	3	LOC_Os06g45650	<i>OsMADS30</i>
<i>indica</i>	3	1274292	1.1×10^{-4}	1.1×10^{-19}	0.0280	0.191	4	LOC_Os03g03070	<i>DTH3/Hd9</i>
<i>indica</i>	3	21446165	1.4×10^{-6}	-	0.1535	0.240	18	LOC_Os03g38610	<i>OsMADS87</i>
<i>indica</i>	6	27634041	2.6×10^{-6}	1.1×10^{-17}	0.0997	0.132	3	LOC_Os06g45650	<i>OsMADS30</i>
<i>indica</i>	7	9177499	4.7×10^{-6}	-	0.1579	0.099	23	LOC_Os07g15770	<i>Ghd7</i>
<i>Indl</i>	1	40341860	3.1×10^{-3}	6.5×10^{-7}	0.0054	0.179	1	LOC_Os01g69850	<i>OsMADS51</i>

<i>IndII</i>	3	1244905	2.9×10^{-7}	7.5×10^{-24}	0.3712	0.365	24	LOC_Os03g03070	<i>DTH3/Hd9</i>
<i>IndII</i>	3	4554342	5.4×10^{-4}	5.0×10^{-9}	0.0586	0.423	30	LOC_Os03g08754	<i>OsMDP1/OsMADS47</i>
<i>japonica</i>	10	17003447	6.4×10^{-7}	1.0×10^{-28}	0.1065	0.062	1	LOC_Os10g32600	<i>Ehd1</i>
Associated with the difference of heading date between Wuhan (2011) and Lingshui									
<i>All</i>	11	2752748	6.3×10^{-9}	-	0.1450	0.070	32	LOC_Os11g05930	<i>OsPRR59</i>
<i>All</i>	8	325967	3.0×10^{-7}	-	0.0327	0.149	50	LOC_Os08g01420	<i>Ehd3</i>
<i>All</i>	2	23947664	2.8×10^{-3}	2.2×10^{-7}	0.0330	0.155	36	LOC_Os02g39710	<i>OsCOL4</i>
<i>indica</i>	6	2912060	8.0×10^{-8}	-	0.3374	0.057	14	LOC_Os06g06300, LOC_Os06g06320	<i>RFT1, Hd3a</i>
<i>indica</i>	11	2438273	3.3×10^{-6}	-	0.1427	0.053	10	LOC_Os11g05470	<i>RCN1</i>
<i>indica</i>	9	20873760	4.0×10^{-4}	8.8×10^{-7}	0.0841	0.184	11	LOC_Os09g36220	<i>OsPRR95</i>
<i>indica</i>	6	17855567	1.3×10^{-3}	2.0×10^{-10}	0.0857	0.191	13	LOC_Os06g30810	<i>OsMADS75</i>
<i>IndI</i>	2	3880514	8.9×10^{-7}	2.4×10^{-9}	0.3736	0.453	43	LOC_Os02g07430	<i>OsMADS29</i>
<i>IndI</i>	7	866633	2.8×10^{-6}	-	0.0409	0.075	62	LOC_Os07g02350	<i>Similar with Hd6</i>
<i>IndII</i>	7	808440	6.9×10^{-3}	9.2×10^{-14}	0.1133	0.289	4	LOC_Os07g02350	<i>Similar with Hd6</i>
<i>japonica</i>	11	25964291	2.2×10^{-8}	6.8×10^{-23}	0.2513	0.062	17	LOC_Os11g43740	<i>OsMADS68</i>
<i>TeJ</i>	3	11042303	1.9×10^{-5}	4.0×10^{-28}	0.3319	0.082	15	LOC_Os03g19590	<i>phyB</i>
Associated with the difference of heading date between Wuhan (2012) and Lingshui									
<i>All</i>	10	17016459	1.5×10^{-10}	1.0×10^{-27}	0.0014	0.289	10	LOC_Os10g32600	<i>Ehd1</i>
<i>All</i>	8	26825697	8.3×10^{-8}	-	0.0085	0.062	31	LOC_Os08g42440	<i>OsO (CCT)</i>
<i>indica</i>	6	2912060	1.8×10^{-8}	-	0.3176	0.057	14	LOC_Os06g06300, LOC_Os06g06320	<i>RFT1, Hd3a</i>
<i>IndI</i>	2	590389	3.9×10^{-6}	-	0.4403	0.095	39	LOC_Os02g01990	<i>OsCMF2</i>

Only lead SNPs nearby genes (≤ 50 kb) known to associated with rice flowering time or with CCT or MADS box domain are listed. See Tables S1, S2, S3, and S4 for the complete set of significant loci. Chr, chromosome; Pos, position on rice genome assembly MSU version v6.1; P_{LMM} , P value from LMM; P_{MLMM} , conditional P value from MLMM; q^2 , variance explained by the single SNP effect; MAF, minor allele frequency; Dis, distance between the significant position and the target gene; Locus and symbol, the target genes and their symbols.

Table 2 The top fifteen enriched Pfam protein domains detected from GWAS hotspots for rice flowering time

Pfam domain	Total gene	No. in hotspot	<i>P</i> value	Description
PF00808	24	4	0.017	Histone-like transcription factor (CBF/NF-Y)
PF02364	14	3	0.017	Glucan_synthase
PF01398	15	3	0.023	JAB1/Mov34/MPN/PAD-1 ubiquitin protease
PF00702	55	6	0.026	haloacid dehalogenase-like hydrolase
PF00319	55	6	0.029	SRF-type transcription factor (DNA-binding and dimerisation domain)
PF01167	15	3	0.030	Tubby protein
PF00225	48	6	0.033	Kinesin motor domain
PF02170	21	3	0.041	PAZ domain
PF00125	48	5	0.045	Core histone H2A/H2B/H3/H4
PF00249	146	11	0.049	Myb_DNA-binding
PF00501	45	5	0.057	AMP-binding
PF07762	31	3	0.063	DUF1618
PF00118	20	3	0.066	TCP-1/cpn60 chaperonin family
PF06203	39	4	0.077	CCT
PF00010	50	5	0.086	Helix-loop-helix DNA-binding domain

Figure

Fig. 2 GWAS for the differences of heading date between Wuhan_2012 and Lingshui in the whole association population. (a-b) The heatmap (a) and histogram (b) distribution of the derived trait in 529 accessions. (c-d) Q-Q plot of the expected null distribution and the observed P -value using the linear mixed model (c) and the simple linear regression model (d). (e-g) Genome-wide P -values for the linear mixed model (e), simple linear regression model (f), and multi-locus mixed-model (g). The horizontal dashed line indicates the significance thresholds set as $P=5.0 \times 10^{-6}$ by LMM. The SNP positions of representative peak signals were denoted.

Fig. 3 The GWAS results near *Ghd7* region for rice flowering time and epistatic interaction between *Ghd7* and *Ehd1*. (a-b) Association results near *Ghd7* for heading date in Wuhan_2011 in *IndI* (a) and *indica* (b). The blue point denotes the lead SNP sf079137593. The colors of the other points represent the linkage disequilibrium for the lead SNP. The arrow represents the position of *Ghd7*. (c) The result of scanning the interactions of *Ghd7* towards other lead SNPs. (d-e) epistatic interaction between *Ghd7* and *Ehd1*. A lead SNP sf1017003447 near *Ehd1* had the most significant epistatic interaction with *Ghd7* in *indica* (d). In contrast, sf1017003447 in *IndI* was nearly fixed (e). The distribution of the minor allele frequencies for the interacted

481 SNPs. Each column represents the ten most significant SNPs interacted with a certain
 482 SNP. Before the vertical line: twenty-five SNPs with $PLMM \leq 5.0 \times 10^{-6}$ in *IndI* and
 483 $PLMM \leq 5.0 \times 10^{-5}$ in *indica*. The inverted triangles represent the minor allele
 484 frequencies of SNPs showing significant interactions in *indica*, while the points in
 485 purple is the minor allele frequencies of these interacted SNPs in *IndI* (obviously
 486 lower than in *indica*). After the vertical line: twenty-four SNPs with $PLMM \leq 5.0 \times$
 487 10^{-5} in both *TeJ* and *japonica*. The inverted triangles represent the minor allele
 488 frequencies of SNPs showing significant interactions in *japonica*, while the points in
 489 purple is the minor allele frequencies of these interacted SNPs in *TeJ*. (f).

References

- Achard P, Herr A, Baulcombe DC, Harberd NP (2004) Modulation of floral development by a gibberellin-regulated microRNA. *Development* 131:3357-3365. <https://doi.org/10.1242/dev.01206>
- Alexa A, Rahnenfuhrer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600-1607. <https://doi.org/10.1093/bioinformatics/btl140>
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19:1655-1664. <https://doi.org/10.1101/gr.094052.109>
- Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* 15:2730-2741. <https://doi.org/10.1105/tpc.016238>
- Bian XF, Liu X, Zhao ZG, Jiang L, Gao H, Zhang YH, Zheng M, Chen LM, Liu SJ, Zhai HQ, Wan JM (2011) Heading date gene, *dth3* controlled late flowering in *O. glaberrima* Steud. by down-regulating *Ehd1*. *Plant Cell Report* 30:2243-2254. <https://doi.org/10.1007/s00299-011-1129-4>
- Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology* 12:232. <https://doi.org/10.1186/gb-2011-12-10-232>
- Chen W, Gao Y, Xie W, Gong L, Lu K et al (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet* 46:714-721. <https://doi.org/10.1038/ng.3007>
- Cockram J, Thiel T, Steuernagel B, Stein N, Taudien S, Bailey PC, O'Sullivan DM (2012) Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. *PLoS ONE* 7:e45307. <https://doi.org/10.1371/journal.pone.0045307>
- Du A, Tian W, Wei M, Yan W, He H, Zhou D, Huang X, Li S, Ouyang X (2017) The DTH8-Hd1 module mediates day-length-dependent regulation of rice flowering. *Mol Plant* 10:948-961. <https://doi.org/10.1016/j.molp.2017.05.006>
- Gao H, Zheng X, Fei G, Chen J, Jin M, Ren Y, Wu W, Zhou K, Sheng P, Zhou F, Jiang L, Wang J, Zhang X, Guo X, Wang J, Cheng Z, Wu C, Wang H, Wan J (2013) *Ehd4* encodes a novel and *Oryza*-genus-specific regulator of photoperiodic flowering in rice. *PLoS Genet* 9:e1003281. <https://doi.org/10.1371/journal.pgen.1003281>
- Goretti D, Martignago D, Landini M, Brambilla V, Gómez-Ariza J, Gnesutta N, Galbiati F, Collani S, Takagi H, Terauchi R, Mantovani R, Fornara F (2017) Transcriptional and post-transcriptional mechanisms limit Heading Date 1 (Hd1) function to adapt rice to high latitudes. *PLoS Genet* 13:e1006530. <https://doi.org/10.1371/journal.pgen.1006530>
- Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422:719-722. <https://doi.org/10.1038/nature01549>
- He Y, Michaels SD, Amasino RM (2003) Regulation of flowering time by histone acetylation in *Arabidopsis*. *Science* 302:1751-1754. <https://doi.org/10.1126/science.1091109>
- Huang X, Kurata N, Wei X, Wang ZX, Wang A et al (2012a) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497-501. <https://doi.org/10.1038/nature11532>
- Huang X, Wei X, Sang T, Zhao Q, Feng Q et al (2010) Genome-wide association

- studies of 14 agronomic traits in rice landraces. *Nature Genetics* 42:961-967.
<https://doi.org/10.1038/ng.695>
- Huang X, Zhao Y, Wei X, Li C, Wang A et al (2012b) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics* 44:32-39. <https://doi.org/10.1038/ng.1018>
- Itoh H, Nonoue Y, Yano M, Izawa T (2010) A pair of floral regulators sets critical day length for *Hd3a* florigen expression in rice. *Nature Genetics* 42:635-638.
<https://doi.org/10.1038/ng.606>
- Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *J Exp Bot* 58:3091-3097.
<https://doi.org/10.1093/jxb/erm159>
- Kim W, Ahn HJ, Chiou TJ, Ahn JH (2011) The role of the miR399-PHO2 module in the regulation of flowering time in response to different ambient temperatures in *Arabidopsis thaliana*. *Molecular Cells* 32:83-88. <https://doi.org/10.1007/s10059-011-1043-1>
- Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, Araki T, Yano M (2002) *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiology* 43:1096-1105. <https://doi.org/10.1093/pcp/pcf156>
- Komiya R, Yokoi S, Shimamoto K (2009) A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development* 136:3443-3450. <https://doi.org/10.1242/dev.040170>
- Laloum T, De Mita S, Gamas P, Baudin M, Niebel A (2012) CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci.*
<https://doi.org/10.1016/j.tplants.2012.07.004>
- Lee PH, O'Dushlaine C, Thomas B, Purcell SM (2012) INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 28:1797-1799. <https://doi.org/10.1093/bioinformatics/bts191>
- Li J, Jiang J, Qian Q, Xu Y, Zhang C, Xiao J, Du C, Luo W, Zou G, Chen M, Huang Y, Feng Y, Cheng Z, Yuan M, Chong K (2011) Mutation of rice BC12/GDD1, which encodes a kinesin-like protein that binds to a GA biosynthesis gene promoter, leads to dwarfism with impaired cell elongation. *Plant Cell* 23:628-640.
<https://doi.org/10.1105/tpc.110.081901>
- Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) FaST linear mixed models for genome-wide association studies. *Nature Methods* 8:833-835. <https://doi.org/10.1038/nmeth.1681>
- Matsubara K, Yamanouchi U, Nonoue Y, Sugimoto K, Wang ZX, Minobe Y, Yano M (2011) *Ehd3*, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *Plant Journal* 66:603-612.
<https://doi.org/10.1111/j.1365-3113X.2011.04517.x>
- Matsubara K, Yamanouchi U, Wang ZX, Minobe Y, Izawa T, Yano M (2008) *Ehd2*, a rice ortholog of the maize *INDETERMINATE1* gene, promotes flowering by up-regulating *Ehd1*. *Plant Physiology* 148:1425-1435.
<https://doi.org/10.1104/pp.108.125542>
- Murakami M, Matsushika A, Ashikari M, Yamashino T, Mizuno T (2005) Circadian-associated rice pseudo response regulators (*OsPRRs*): insight into the control of flowering time. *Biosci Biotechnol Biochem* 69:410-414.
<https://doi.org/10.1271/bbb.69.410>
- Nemoto Y, Nonoue Y, Yano M, Izawa T (2016) Hd1, a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot specific CCT-

- domain protein Ghd7. *Plant J* 86:221-233. <https://doi.org/10.1111/tpj.13168>
- Park SJ, Kim SL, Lee S, Je BI, Piao HL, Park SH, Kim CM, Ryu CH, Park SH, Xuan YH, Colasanti J, An G, Han CD (2008) *Rice Indeterminate 1 (OsId1)* is necessary for the expression of *Ehd1 (Early heading date 1)* regardless of photoperiod. *Plant Journal* 56:1018-1029. <https://doi.org/10.1111/j.1365-313X.2008.03667.x>
- Ryu CH, Lee S, Cho LH, Kim SL, Lee YS, Choi SC, Jeong HJ, Yi J, Park SJ, Han CD, An G (2009) *OsMADS50* and *OsMADS56* function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environment* 32:1412-1427. <https://doi.org/10.1111/j.1365-3040.2009.02008.x>
- Segura V, Vilhjalmsdottir BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* 44:825-830. <https://doi.org/10.1038/ng.2314>
- Tsuji H, Taoka K, Shimamoto K (2011) Regulation of flowering in rice: two florigen genes, a complex gene network, and natural variation. *Current Opinion in Plant Biology* 14:45-52. <https://doi.org/10.1016/j.pbi.2010.08.016>
- Wang QX, Xie WB, Xing HK, Yan J, Meng XZ, Li XL, Fu XK, Xu JY, Lian XM, Yu SB, Xing YZ, Wang GW (2015) Genetic Architecture of Natural Variation in Rice Chlorophyll Content Revealed by a Genome-Wide Association Study. *Molecular Plant* 8: 946-957. <http://doi.org/10.1016/j.molp.2015.02.014>
- Wei X, Xu J, Guo H, Jiang L, Chen S, Yu C, Zhou Z, Hu P, Zhai H, Wan J (2010) *DTH8* suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology* 153:1747-1758. <https://doi.org/10.1104/pp.110.156943>
- Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierc J, Samach A, Coupland G (2006) CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18:2971-2984. <https://doi.org/10.1105/tpc.106.043299>
- Wu C, You C, Li C, Long T, Chen G, Byrne ME, Zhang Q (2008) *RID1*, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. *Proceedings of the National Academy of Sciences, USA* 105:12915-12920. <https://doi.org/10.1073/pnas.0806019105>
- Xie K, Wu C, Xiong L (2006) Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiology* 142:280-293. <https://doi.org/10.1104/pp.106.084475>
- Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics* 40:761-767. <https://doi.org/10.1038/ng.143>
- Yan WH, Wang P, Chen HX, Zhou HJ, Li QP, Wang CR, Ding ZH, Zhang YS, Yu SB, Xing YZ, Zhang QF (2011) A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant* 4:319-330. <https://doi.org/10.1093/mp/ssq070>
- Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, Sasaki T (2000) *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene *CONSTANS*. *Plant Cell* 12:2473-2484. <https://doi.org/10.1105/tpc.12.12.2473>

640 Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H,
641 Dean C, Marjoram P, Nordborg M (2007) An *Arabidopsis* example of association
642 mapping in structured samples. PLoS Genetics 3:e4.
643 <https://doi.org/10.1371/journal.pgen.0030004>
644 Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam
645 MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011)
646 Genome-wide association mapping reveals a rich genetic architecture of complex
647 traits in *Oryza sativa*. Nature Communication 2:467.
648 <https://doi.org/10.1038/ncomms1467>