**QTL mapping for agronomic traits using multi-parent advanced generation
inter-cross (MAGIC) populations derived from diverse elite *indica* rice lines**

Lijun Meng^{1,2}, Xiangqian Zhao¹, Kimberly Ponce¹, Guoyou Ye^{1,*}, Hei Leung¹

¹ Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Los
Baños, Laguna, Philippines

² Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences,
Shenzhen 518120, China

* Corresponding authors

E-mail: g.ye@irri.org

Abstract

In rice many QTL have been reported for yield and related traits using various bi-parental mapping populations. However, few of the well characterized genes/QTL have been successfully used in breeding for improved trait performance. One of the reasons is that the mapping populations used are irrelevant to breeding. Association mapping has also been used in identifying marker-trait associations that are effective in more complex background. However, the results from association mapping using diversity panels are difficult to be exploited in breeding, since most of the accessions had poor performance in many important agronomic traits. In this study, a genome-wide association study (GWAS) was performed using three multi-parent advanced generation inter-cross (MAGIC) populations derived from elite *indica* lines (DC1, DC2 and 8-way) to identify QTL for 14 traits including yield, yield components and other related traits. The three MAGIC populations were phenotyped in the dry season (DS) and wet season (WS) of 2014 at the headquarters of the International Rice Research Institute (IRRI) and genotyped with a Rice SNP chip containing 4,500 markers. A total of 26 QTL on all chromosomes except 7, 9 and 11 were identified for 10 traits in the DS or WS. Six, two, 12, 10 and 20 out of the 26 QTL were identified in the DC1, DC2, 8-way, DC12 (DC1+DC2) and RMPRIL (DC1+DC2+8-way) populations, respectively. Nine of the QTL corresponded to known QTL/genes, including *qFLW4* for FLW, *qTGW3* and *qTGW5* for TGW, *qFGN4* for FGN, *qSBN4* for SBN, *qPH1.1* and *qPH1.2* for PH, *qHD3* and *qHD6* for HD. All these nine QTL were identifiable using the RMPRIL population while only eight, seven, six and one could be identified using the 8-way, DC12, DC1 and DC2 populations, respectively. The 8-way population was more powerful than the DC1, DC2 and DC12 populations. The joint analysis, which combines different populations (e.g. DC12 and RMPRIL), increased the number of QTL identified and mapping resolution. The study showed that MAGIC populations derived from diverse elite parental lines can be used to detect QTL, which are ideal for linking gene identification and practical breeding. GWAS using such MAGIC populations had higher detection power (compared to assembled populations) and higher resolution (compared to biparental populations). The identified QTL are directly applicable, since the populations are good breeding populations as well.

Keywords: Agronomic traits, GWAS, MAGIC population, Rice (*Oryza sativa* L.)

45 **1. Introduction**

46 Rice is a major food crop worldwide, especially in developing countries and providing
47 more than 21% of the food for the world population and up to 76% of the calorific intake in
48 Southeast Asia (Fitzgerald et al., 2008; Miura et al., 2011). It has been estimated that a 40%
49 increase in rice production by 2030 will be needed to meet the predicted demand of the growing
50 world population (Khush, 2005). Rice yield is determined by several factors including the
51 number of panicles per plant (PN), the number of grains per panicle (GN) and grain weight.
52 These traits and the overall yield per plant are quantitatively inherited (controlled by many genes)
53 and influenced by many environmental factors. Progress of breeding for increased yield has been
54 slow in the last few decades in most of the rice production countries (Peng et al., 2004). Rice
55 breeders are facing the challenge of breaking the yield barrier. It is hoped that this can be
56 achieved by integrating well proven conventional breeding methods with the new options
57 provided by recent development in molecular biology and genomics (Liang et al 2015, 2016).

58 The potential benefits of using molecular markers linked to the genes of interest in
59 breeding programs, which have changed from phenotype-based toward a combination of
60 phenotype- and genotype-based selection, have attracted much attention for more than three
61 decades (Bernardo et al., 2008; Tester et al., 2010; Guo and Ye, 2014). The success of MAS for
62 traits of simple inheritance in many crops including rice has motivated rice breeders to search for
63 QTLs for complex traits, which account for a large proportion of phenotypic variation (major
64 QTL) (Guo and Ye, 2014). Many yield-related genes/QTL have been identified and some of
65 them are fine-mapped/cloned (Xing and Zhang, 2010). However, significant improvement of
66 grain yield in farm environments through the use of these well characterized genes has not yet
67 been reported (Guo and Ye, 2014; Liang et al., 2016).

68 One of the main reasons for the ‘ineffectiveness’ of the well-characterized QTL in
69 breeding is that they were identified using biparental populations derived from parental lines
70 with contrast trait performance. The estimated effects are specific to the population and QTL
71 results are often not transferable to other populations (Bernardo, 2008; Holland, 2007).

72 An alternative method for identifying markers associated with trait of interest is
73 association mapping (AM). AM takes advantage of historical recombination that has occurred

within a germplasm collection by sampling numerous lines from a broader population not constrained by specific crossing. The associations between genotype and phenotype are dependent on linkage disequilibrium (LD), the nonrandom association of alleles, being broken down by many generations of recombination. AM has higher mapping resolution due to the decreased size of linkage blocks. This has important implications in applied breeding, since the significant markers will be more tightly linked to the QTL of interest, and therefore the likelihood of recombination between the markers and QTL will be reduced (Pauli and Torelli, 2015). However, the breeding and natural populations of inbred crops are not closed random-mating populations. The complex breeding history of breeding populations and the limited gene flow in most wild populations have created complex stratification within the germplasm, associations are common between unlinked loci. It is difficult to distinguish the true and false positive associations (Guo and Ye, 2014; Lipka et al., 2015). Another major drawback of AM is its reduced QTL detection power, which is linearly dependent on the LD between QTL and markers. QTL with small effects can only be detected if they are in strong LD with a marker (Van et al., 2011). In addition, incomplete LD between a marker and a QTL will lead to an underestimation of the variance explained by the QTL.

The past decade has seen the rise of multiparental populations as a study design offering great advantages for genetic studies in plants. Multi-parent mapping populations such as multi-parent advanced generation inter-cross (MAGIC) (Cavanagh et al., 2008) and nested association mapping (NAM) (Yu et al., 2006) populations are being developed for many crops. The genetic diversity of multiple parents, recombined over several generations, generates a genetic resource population with large phenotypic diversity suitable for high-resolution trait mapping (Huang et al., 2015). Such mapping populations is good for practical breeding as well, since outstanding lines selected from such populations of highly recombined lines can be released as varieties or utilized in different crossing programs to enhance the genetic base of cultivated gene pools. IRRI has developed four MAGIC populations, *indica* MAGIC (eight *indica* parents); MAGIC plus (eight *indica* parents with two additional rounds of 8-way F₁ inter-crossing); *japonica* MAGIC (eight *japonica* parents); and Global MAGIC (16 parents-eight *indica* and eight *japonica*). AM using a subset (200) of the *indica* MAGIC population and genotyping-by-sequencing identified many marker-trait associations (MTAs) including markers in the close regions of several known major genes/QTL and potentially novel loci associated with essential traits for rice improvement

(Bandillo et al., 2013). Li et al. (2014) developed a rice MAGIC population consisting of a set of 206 lines derived from 12 founder lines. These MAGIC populations were proven valuable for QTL identification and practical breeding (Bandillo et al., 2013). A multi-genotype variety developed using lines from Li et al. has been released in China (Li et al., 2014). Many promising lines from IRRI's MAGIC populations have been selected by breeders to be further tested (Bandillo et al., 2013).

We recently developed a set of three interconnected MAGIC populations (DC1, DC2 and 8-way) using eight elite *indica* breeding lines from a few large breeding programs worldwide. Genetic characterization of the three populations suggested that they are valuable for QTL identification (Meng et al 2016). In this study, 271 DC1, 268 DC2 and 532 8-way lines were used to identify QTL for 14 traits including grain yield per plant (GY), panicle number/per plant (PN), filled grain number per panicle (GN), thousand grain weight (TGW), panicle length (PL), primary branch number (PBN) and secondary branch number (SBN), total spikelet number per panicle (SN), spikelet fertility (SF) calculated as $100 \times \text{GN} / \text{SN} (\%)$, flag leaf length (FLL), flag leaf width (FLW), flag length and width ratio (FLR) calculated as $100 \times \text{FLL} / \text{FLW} (\%)$, plant height (PH) and heading date (HD). We also investigated whether the combined use of the MAGIC populations offer higher mapping power and resolution by conducting analysis using the DC12 (DC1+DC2) and RMPRIL (DC1+DC2+8-way) populations.

2. Materials and Methods

2.1. MAGIC populations

Eight genetic diverse parental lines selected from different breeding programs were used to develop three MAGIC RILs populations at the International Rice Research Institute (IRRI). They were SAGC-08 (A), HHZ5-SAL9-Y3-Y1 (B), BP1976B-2-3-7-TB-1-1 (C), PR33282-B-8-1-1-1-1 (D), FFZ1 (E), CT 16658-5-2-2SR-2-3-6MP (F), IR 68 (G), IR 02A127 (H). The eight founders were pairwise crossed to produce four two-way hybrids. These four two-way hybrids were intercrossed in a diallel fashion leading to six 4-way crosses. Two of the 4-way crosses, ABCD and EFGH, were advanced by single seed decent (SSD) to develop the DC1 and DC2 recombinant inbred lines (RILs) populations. An 8-way cross was made by intercrossing 100 F₁ plants of the 4-way cross ABCD and 100 F₁ plants of the 4-way cross EFGH. One thousand 8-

way cross F₁ plants were advanced by SSD to develop the 8-way RILs population. In short, the DC1 population was derived from the double cross A/B//C/D, the DC2 population was from the double cross E/F//G/H, and the 8-way population was derived from the crosses had all eight parental lines involved.

2.2. Phenotyping

The field trials were conducted at the headquarters of IRRI (Los Banos, Laguna, Philippines) in the dry season (DS) and wet season (WS) of 2014. Seeds were sown in the seedling nursery on Jan 03, 2014 (DS) and May 26, 2014 (WS). Twenty days after germination seedlings were transplanted into the field. An augmented randomized complete block design (Augmented RCBD) was used to layout the trials. One hundred and twenty check genotypes (the eight parental lines and 112 genotypes from a biparental RILs population) were replicated three times. The randomizations were generated using the PBTools software (bbi.irri.org) developed by IRRI. The plot size was six rows of six plants. The planting density was 20 x 20 cm. HD was recorded as the number of days from sowing to the date when 50% of the plants headed. At the maturity, three representative plants in the middle of a plot were sampled for measuring PH. At the maturity, three representative plants in the middle of each plot were sampled for measuring PN, FLL, FLW, FLR, GN, PBN, SBN, PBN, SBN, SN, SF, TGW and GY.

2.3. Genotyping

Genomic DNAs of 271 (DC1) and 268 4way (DC2), 532 8-way RILs and eight parents were isolated from lyophilized leaves of 5-week-old seedlings using Tissue Lyser and DNA extraction using Thermo KingFisher Flex 96 system. The Illumina Infinium rice 6K SNP chip developed by Susan McCouch's team at Cornell University was used for SNP fingerprinting. Quality checking was performed according to a three steps filtering strategy with custom R script (R Development Core Team, 2014). First, all heterozygous genotypes were set to missing and lines with more than 10% missing markers were discarded. Second, all markers with minor allele frequencies less than 3% were removed. Third, some markers were discarded to avoid the presence of marker pairs with very high correlations (0.95). The number of remaining markers was 907, 892, 1329, 1202 and 1267 for the DC1, DC2, 8-way, DC12 and RMPRI populations, respectively.

2.4. Statistical analysis

2.4.1. Phenotypic analysis

Phenotypic analysis was conducted using the PBTools (bbi.irri.org) to obtain adjusted trait value for each of the lines. Correlation coefficients among different traits were determined by R software (<http://www.r-project.org>). Correlation graph was produced using R software package, ggplot2 (Wickham and Chang, 2013).

2.4.2. Population structure

Principal component analysis (PCA) that summarizes the major patterns of variation in a multi-locus dataset was performed for each of the five populations using TASSEL v5.0 (Bradbury et al., 2007). The first three principal components (PCs) were used to visualize the dispersion of the lines within a population in a graph. Pairwise similarity coefficients were calculated for all pairwise combinations of the lines within a population according to the method developed by Nei and Li (Nei and Li, 1979) implemented in the software TASSEL v5.0. The neighbor-joining (NJ) tree based on this genetic similarity matrix was constructed accordingly. To display the familial relationship, a heatmap of the kinship matrix was built up in R.

2.4.3. Analysis of linkage disequilibrium

The LD patterns of the five populations were evaluated by computing the r^2 values between pairs of SNP markers using Tassel v5.0 on all chromosomes basis (Bradbury et al., 2007). Significant LD locus pairs were selected ($p < 0.001$), the rest of r^2 values were not considered as informative. Two approaches were considered as the estimate of the range of LD: (a) LD decay analysis calculated to 50 % of its initial value, according to Zhao et al. (2011) and (b) critical value for r^2 was estimated by square root transforming of unlinked r^2 values to obtain a normally distributed random variable, and the parametric 95th percentile of that distribution was taken as a critical r^2 value (Brescaglio and Sorrells, 2006), beyond which LD was likely to be caused by genetic linkage. The intersection of the loess curve fit to synthetic r^2 with this baseline was considered as the estimate of the extent of LD in the chromosome. Unlinked r^2 refers to the r^2 between the marker loci with a physical distance greater than 30 Mb or on independent chromosomes. Significant r^2 values of intra-chromosomal LD were plotted against physical distance and a second-degree LOESS curve (Cleveland, 1979) was drawn using

R software (<http://www.r-project.org>) package, ggplot2 (Wickham and Chang, 2013). The intersection of the critical value baseline with the LOESS curve was considered to be the estimate of the extent of LD in the chromosome (Bresnahan and Sorrells, 2006).

2.4.4. Association analysis

Six models comprising both of general linear models (GLM) and mixed linear models (MLM) were compared to choose the best model. They were 1) Naïve: GLM without any correction for population structure; 2) P: GLM with population structure correction using PCs; 3) Q: GLM with Q matrix derived from Structure analysis using the optimal number of subpopulations for controlling population structure; 4) K: MLM with the K matrix estimated as identity by state (IBS) using the method implemented in TASSEL v5.0 being used for correcting unequal relationships between genotypes ; 5) QK: MLM with Q and K being used for correcting population structure and unequal relationships between genotypes; 6) PK: MLM with PCs and K for correcting population structure and unequal relationships between genotypes. All analysis was conducted using the software TASSEL v5.0 (Bradbury et al., 2007). The p values were adjusted with the Benjamini and Hochberg (1995) method to obtain q values (false discovery rate). The q-values were calculated for each of the trait-season combination separately using the R package QVALUE R 3.1.0 employing the smoother method (Storey and Tibshirani, 2003). A q value cut off of 0.05 was used as the threshold to determine whether a significant association existed. When the QTL profile displayed more than one peak exceeding the significance threshold on a chromosome, multiple QTL were declared if peaks were separated by more than 2.5 Mb. The MATLAB (The Math Works, Natick, USA) package was used for visualizations. Horizontal solid line on the Manhattan plot indicates the genome-wide significance threshold $q \leq 0.05$.

3. Results

3.1. Phenotypic variation

Large phenotypic variation was observed for all traits in all the three MAGIC populations in the two growing seasons. The means and ranges in the DS and WS for all traits were presented in Table 1 and their frequency distributions were illustrated in Suppl. Fig.1. The eight parental lines were different for all 14 traits in both of the DS and WS. Transgressive segregations in both

directions were observed for all traits in all three populations in the two seasons. GY was higher in the DS than in the WS while GN and SN were higher in the WS than in the DS. For the remaining 11 traits the differences between the DS and WS were small. The 8-way population had larger trait variations than the DC1 and DC2 populations. The variation for GY was much lower in the DS compared to the WS while the opposite was true for most of other traits.

The correlation coefficients between traits between each other were given in Suppl. Table 1 and Suppl. Fig. 2. The correlations between traits measured in the three populations were similar. Although the correlations between traits were in the same directions in the two seasons, the correlation coefficients (magnitudes) were greatly affected by the growing season. GY was strongly and positively correlated with GN and SN, and moderately with PH, SF, SBN, PL and PBN. The correlation between TGW was weak and not significant in most of the season-population combinations. TGW was negatively and moderately correlated with PN, GN and SN. PH was moderately and positively correlated with all traits except SF in the DC2 and 8-way populations for which the correlation was moderate and positive. HD was moderately and positively correlated with GY, FLL, FLR, PBN, PL and PH. FLR and FLL were highly and positively correlated. Similarly, the correlation between GN and SN was high and positive. The correlation between SBN and PBN was moderate and positive.

3.2. Population structure and linkage disequilibrium

PCA and NJ tree suggested that DC1, DC2 and 8-way populations showed no clear population structure, DC12 and RMPRIL had two and three subgroups corresponding to their component MAGIC populations (Fig. 1 and Fig. 2). DC1 showed high kinship between about 10 lines while the DC2 showed high kinship between about 20 lines. The DC12 and RMPRIL had clearly kinship block (Suppl. Fig. 3).

Extensive variability in the magnitude of r^2 at a given physical distance was detected reflecting the wide local variation in the extent of LD across the chromosomes. The average of r^2 for all marker pairs was 0.21, 0.20, 0.13, 0.11 and 0.09 for the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. In average, 31.8%, 29.7%, 37.5%, 45.5% and 74.9% of the intra-chromosomal marker pairs showed a significant level of LD ($p < 0.001$) for the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. A scatter diagram of r^2 against physical distances showed a clean pattern of LD decay in all five populations (Fig. 3, Suppl. Fig. 4). LD

decayed more rapidly and cleanly in the DC12, 8-way and RMPRIL populations than in the two DC1 and DC2 populations. A critical value of r^2 or basal LD, beyond which LD is likely to be caused by genetic linkage in the absence of population structure, was estimated at 0.06, 0.08, 0.12, 0.14 and 0.2 for RMPRIL 8-way, DC12, DC1 and DC2, respectively. A critical value of $r^2 > 0.2$ was determined to be the appropriate threshold for LD due to physical linkage for all populations. On the other hand, LD declined to 50 % of its initial value below $r^2 = 0.2$ at about 1.25 Mb in RMPRIL, 8-way and DC12 populations, below $r^2 = 0.2$ at about 2.5 Mb in DC1 and DC2 populations. Given that our average inter-marker distance was about 100 kb we expected to have reasonable power to identify common variants of large effects associated with our traits of interest. Among all significant loci in LD, most of the marker pairs with above the critical r^2 value of 0.2 were within 2.5 Mb in all populations. Similarly, most of the marker pairs in complete LD ($r^2 > 0.9$) were spaced at physical distance < 2.5 Mb for all populations.

3.3. Comparison of association models

We tested six models including the naïve (no correction for population structure and unequal familiar relationship), P (using PCs to control population structure), Q (using Q matrix derived from STRUCTURE to correct population structure), K (using relative kinship to account for unequal familiar relationship between lines), PK (PCs and kinship) and QK (Q and kinship) models to detect associations between SNP markers and 14 agronomic traits in the DS. Q-Q plot was used to check the fit of observed expected $-\text{Log}_{10}(p)$ values and the expected $-\text{Log}_{10}(p)$ values for 14 traits. For the three GLM models (Naive, P and Q models), considerable deviations from the expected values were observed. The models were characterized by the excess of small p values which is tantamount to an abundance of spurious associations (Suppl. Fig. 5). For the three MLM models (K, PK and QK models), Q-Q plots showed a good fit for p values (Suppl. Fig. 5). The three models performed very similarly. We also found that for the trait PH p value diverged from the expected value in all models and all populations except in the DC2 population (Fig. 4). Therefore, for all the five populations, MLM model was selected as the best model for association analysis. For all traits only results from the PK-model will be presented.

3.4. QTL identified by GWAS

Significant MTAs were not identified for GY, FFL, FLR and SN in the five populations in

the two seasons.

3.4.1. Panicle number

One marker and thus one QTL (*qPN2*) was identified for PN in the DC12 in the WS (Fig. 5). *qPN2* peaked at 25.1 Mb on chromosome 2 and explained 3.3% of the total phenotypic variance. The allele from parental lines B, F and H increased PN by an average of 1.9 (Table 2).

3.4.2. Grain number

In the DC1 population, three SNPs were significantly associated with GN in the DS. In the RMPRIL population, three and one SNPs were significantly associated with GN in the DS and WS, respectively (Fig. 6). No markers were significantly associated with GN in the DC2, 8-way and DC12 populations. By delineating the significant markers within a physical distance of 2.5 Mb into a single QTL, a total of three QTL (*qGN3*, *qGN4* and *qGN8*) on chromosomes 3, 4 and 8 were identified for GN. *qGN4*, peaked at 31.1 Mb on chromosome 4 where the known gene *qGN4-1* locates (Deshmukh et al., 2010), was identified in the DC1 and RMPRIL populations in the DS. It explained 6.3% and 1.4% of the phenotypic variance. The allele from parental line A increased GN by an average of 26.3 and 12.2 (Table 2). *qGN3*, peaked at 17.0 Mb on chromosome 3, was identified in the RMPRIL population in the WS. It explained 1.8% of the phenotypic variance. The allele from parental lines A and C increased GN by an average of 25.1 (Table 2). *qGN8*, peaked at 2.2 Mb on chromosome 8, was identified in the RMPRIL population in the DS. It explained 1.4% of the phenotypic variance. The allele from parental line C increased GN by an average of 16.0 (Table 2).

3.4.3. Thousand grain weight

In the DC1 population, one and two SNPs were significantly associated with TGW in the DS and WS, respectively. In the 8-way population, two and four SNPs were significantly associated with TGW in DS and WS. In the DC12 population, one and two SNPs were identified in the DS and WS, respectively. In the RMPRIL population, two and three SNPs were significantly associated with TGW in the DS and WS, respectively (Fig. 7). These markers were delineated into three QTL (*qTGW2*, *qTGW3* and *qTGW5*) on chromosomes 2, 3 and 5. No QTL was identified using the DC2 population. *qTGW3*, peaked at 17.0 Mb on chromosome 3 where

the known gene *GS3* locates (Fan et al., 2006), was identified in four populations in the DS and WS. The percentage of variance explained ranged from 2.7% (RMPRIL in the DS) to 8.0% (DC1 in the WS) (Table 2). The allele from parental lines A and C decreased TGW with the average effect being from 1.9g (RMPRIL in the DS) to 2.7g (8-way in the DS) (Table 2). *qTGW5*, peaked at 5.7 Mb on chromosome 5 where the known gene *qSW5* locates (Shomura et al., 2008), was identified in the DC1, 8-way, DC12 and RMPRIL populations. The percentage of variance explained ranged from 1.5% (RMPRIL in the DS) to 7.9% (DC1 in the DS). The allele from parental line A increased TGW with the average effect being from 1.5g (RMPRIL in the DS) to 2.6g (DC12 in the WS) (Table 2). *qTGW2*, peaked at 16.1 Mb on chromosome 2, was identified in 8-way population in the DS. It explained 4.6% of the phenotypic variance. The allele from parental line G decreased TGW by an average of 1.9g (Table 2).

3.4.4. Panicle length

Three MTAs delineated into two QTL (*qPL6* and *qPL8*) were identified for PL (Fig. 8). *qPL6*, peaked at 2.9 Mb on chromosome 6, was identified in the 8-way and RMPRIL populations in the WS with the percentage of variance explained being 4.3% and 2.2%, respectively (Table 2). The allele from parental lines D and H increased PL by an average of 18.7 mm and 15.5 mm, respectively (Table 2). *qPL8*, peaked at 26.0 Mb on chromosome 8, was identified in the 8-way population in the WS with the percentage of variance explained being 3.3%. The allele from parental line C increased PL by an average of 22.1 mm (Table 2).

3.4.5. Primary branch number

In the RMPRIL population, two and one SNPs were significantly associated with PBN in the DS and WS, respectively (Fig. 9). These MTAs were delineated into two QTL (*qPBN3* and *qPBN4*) on chromosomes 3 and 4 (Table 2). *qPBN3*, peaked at 1.3 Mb on chromosome 3, was identified in the RMPRIL population in the DS and WS with the percentage of variance explained being 2.6% and 1.7%. The allele from parental lines A, B, C and G increased PBN by an average of 0.6 and 0.7 (Table 2). *qPBN4*, peaked at 15.9 Mb on chromosome 4, was identified in the RMPRIL population in the DS with the percentage of variance explained being 1.8%. The allele from parental line D decreased PBN by an average of 0.8 (Table 2).

3.4.6. Secondary branch number

A total of three MTAs were identified for SBN across two seasons. They were one in the DC2 population in the DS, one in the 8-way population in the WS and one in the RMPRIL population in the WS (Fig. 10). These significantly associated markers were delineated into two QTL (*qSBN4* and *qSBN12*) on chromosomes 4 and 12. *qSBN4*, peaked at 33.2 Mb on chromosome 4 where the known gene *SPIKELET NUMBER (SPIKE)* locates (Fujita et al., 2013), was identified in the 8-way and RMPRIL populations in the WS with the percentage of variance explained being 3.3% and 1.6%, respectively. The allele from parental line A increased SBN by an average of 6.2 and 6.0, respectively (Table 2). *qSBN12*, peaked at 3.1 Mb on chromosome 12, was identified in the DC2 population in the DS with the percentage of variance explained being 6.4%. The allele from parental lines B, E, G and H decreased SBN by an average of 5.1 (Table 2).

3.4.7. Spikelet fertility

Four markers were found to be significantly associated with SF in the RMPRIL population in the WS (Fig. 11). They were delineated into two QTL (*qSF3.1* and *qSF3.2*). *qSF3.1* peaked at 17.0 Mb on chromosome 3 and explained 2.3% of the phenotypic variation. The allele from parental lines A and C increased SF by an average of 5.8% (Table 2). *qSF3.2* peaked at 31.8 Mb on chromosome 3 and explained 1.9% of the phenotypic variation. The allele from parental lines G and H decreased SF by an average of 7.0% (Table 2).

3.4.8. Flag leaf width

In the 8-way population, four and one SNPs were significantly associated with FLW in the DS and WS, respectively. In the DC12 population, three and three SNPs were significantly associated with FLW in the DS and WS, respectively. In the RMPRIL population, 16 and 22 SNPs were identified in the DS and WS, respectively (Fig. 12). No markers were significantly associated with FLW in the DC1 and DC2 populations. These MTAs were delineated into three QTL (*qFLW4*, *qFLW6* and *qFLW10*) on chromosomes 4, 6 and 10 (Table 2). Among of them, one, one and two QTL were identified in the 8-way, DC12 and RMPRIL populations, respectively. *qFLW4*, peaked at 31.1 Mb on chromosome 4 where the known gene *Nall/qFLW4/LSCHL4* locates (Zhang et al., 2014; Chen et al., 2012), was identified in three populations in the DS and WS. The percentage of variance explained ranged from 3.1% in

RMPRIL in the WS to 6.2% in RMPRIL in the DS. The allele from parental line A increased FLW with the average effect being from 1.5 mm (RMPRIL in the WS) to 2.6 mm (8-way in the WS) (Table 2). *qFLW6*, peaked at 6.4 Mb on chromosome 6, was identified in the RMPRIL population in the DS. It explained 1.5% of the phenotypic variance. The allele from parental lines A, B, C, E, F and H increased FLW by an average of 0.8 mm (Table 2). *qFLW10*, peaked at 0.1 Mb on chromosome 10, was identified in the RMPRIL population in the WS. It explained 1.3% of the phenotypic variance. The allele from parental lines D and H decreased FLW by an average of 1.1 mm (Table 2).

3.4.9. Plant height

Markers associated with PH were identified in all populations except DC2 (Fig 13). In the DC1 population, six and eight SNPs were significantly associated with PH in the DS and WS, respectively. The number of significant markers was six (five) in the DS (WS) for the 8-way population, 11 (12) for the DC12 population, and 11 (10) for the RMPRIL population. These markers were delineated into five QTL (*qPH1.1*, *qPH1.2*, *qPH2*, *qPH6* and *qPH12*) on chromosomes 1, 2, 6 and 12 (Table 2). Among of them, two, three, four, three QTL were identified in the DC1, 8-way, DC12 and RMPRIL populations, respectively. *qPH1.1*, peaked at 37.0 Mb on chromosome 1 where the known gene *Psd1* locates (Li et al., 2014), was identified in all populations except DC2. The strongest association was found ($q=3.3E-05$) in RMPRIL in the DS. The percentage of phenotypic variance explained ranged from 2.7% (RMPRIL in the WS) to 9.4% (DC1 in the WS). The allele from parental line C increased PH with the average effect being from 12.0 mm (RMPRIL in the DS) to 18.2 mm (DC12 in the WS) (Table 2). *qPH1.2*, peaked at 38.7 Mb on chromosome 1 where the known gene *sd1/OsGA20ox2* for PH locates (Sasaki et al., 2002), was also identified in all populations except DC2. The percentage of variance explained ranged from 4.6% (RMPRIL in the WS) to 12.1% (DC1 in the DS) (Table 2). The allele from parental line C increased PH with the average effect being from 13.8 mm (8-way in the DS) to 21.5 mm (DC12 in the WS). *qPH2*, peaked at 29.6 Mb on chromosome 2, was identified in the DC12 in the WS. It explained 2.3% of the phenotypic variance. The allele from parental line H increased PH by an average of 10.6 mm (Table 2). *qPH6*, peaked at 26.3 Mb on chromosome 6, was identified in the DC12 and RMPRIL populations in the DS and WS. The percentage of phenotypic variance explained ranged from 1.2% (RMPRIL in the DS) to 3.3%

(DC12 in the DS). The allele from parental lines C, E, F, G and H increased PH with the average effect being from 6.4 mm (RMPRIL in the DS) to 9.9 mm (DC12 in the DS) (Table 2). *qPH12*, peaked at 22.5 Mb on chromosome 12, was identified in the 8-way population in the DS and WS. It explained 3.2% of the phenotypic variance in both seasons. The allele from parental line C increased PH by an average of 5.7 mm and 7.0 mm, respectively (Table 2).

3.4.10. Heading date

In the DC1 population, only two SNPs were significantly associated with HD, which was detected in the WS. In the DC2 population, only one SNP was significantly associated with HD, which was found in the DS. Two (three) SNPs, four (two), three (seven) were found to be significantly associated with HD in the DS (WS) in the 8-way, DC12 and RMPRIL populations, respectively (Fig 14). By delineating the significant markers within a physical distance of 2.5 Mb into a single QTL, a total of three QTL (*qHD3*, *qHD6* and *qHD10*) on chromosomes 3, 6 and 10 were detected for HD (Table 2). Among of them, one, one, two, two and three QTL were identified in the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. *qHD3*, peaked at 1.3 Mb on chromosome 3 where the known gene *DTH3/Hd9* locates (Bian et al., 2011), was identified in all populations except DC1. The highest peak ($q=7.4E-11$) was obtained in the RMPRIL population in the DS. The phenotypic variance accounted for ranged from 3.1% (8-way in the DS) to 7.5% (DC2 in the DS). The allele from parental lines A, B, C, G and H increased HD with the average effect ranging from 2.8 days (RMPRIL in the WS) to 4.7 days (RMPRIL in the DS) (Table 2). *qHD6*, peaked at 2.9 Mb on chromosome 6 where the known gene *Hd3a* for HD locates (Taoka et al., 2011), was identified in all populations except DC2. The highest peak ($1.1E-07$) was observed in the RMPRIL population in the WS. The phenotypic variance accounted for ranged from 1.8% (RMPRIL in DS) to 7.6% (8-way in WS). The allele from parental lines D and H increased HD with the average effect ranging from 3.1 days (RMPRIL in the DS) to 5.1 days (DC12 in the WS) (Table 2). *qHD10*, peaked at 9.3 Mb on chromosome 10, was identified in the RMPRIL population in the WS. It explained 1.4% of the phenotypic variance and the allele from parental lines A and E decreased HD by an average of 2.4 days (Table 2).

4. Discussion

In the present study GWAS was conducted using three MAGIC populations derived from eight elite *indica* lines for a range of agronomic traits. A total of 26 QTL on all chromosomes except 7, 9 and 11 were identified for 10 traits in the DS or WS using five populations. Six, two, 12, 10 and 20 out of the 26 QTL were identified in the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. Nine QTL including *qFLW4*, *qTGW3*, *qTGW5*, *qFGN4*, *qSBN4*, *qPH1.1*, *qPH1.2*, *qHD3* and *qHD6* corresponded to known QTL/genes recorded in the ricedata (<http://www.ricedata.cn>) database. Among them, six, one, eight, seven, and nine QTL were identified in the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. A total of 16 and 19 QTL were identified in the DS and WS, respectively. Nine QTL including *qPH1.1*, *qPH1.2*, *qPH6*, *qPH12*, *qHD6*, *qFLW4*, *qTGW3*, *qTGW5* and *qPBN3* were commonly detected in both of DS and WS. *qPH6*, *qPH12* and *qPBN3* were novel. These potentially novel QTL could be of great value for further elucidating the molecular genetics control of the traits investigated. These nine QTL plus *qHD3* ($q=7.4E-11$, RMPRIL in the DS) had large effects and should be further tested in larger populations to better define their positions and enable comparisons with underlying candidate genes. QTL with relatively small (large) effects and relatively small (big) q values will need to be validated in independent populations and environments to confirm their effects in further study.

4.1. GWAS model selection is necessary for mapping using MAGIC population

In the present study, we adopted association analysis to identify marker-trait association. Many models have been proposed for association analysis and it is well-known that the results of association analysis using different models would be very different. MLM procedures have been developed to account for both population structure and unequal relatedness (Zhang et al., 2010). In the MLM, Q/P and K matrices were employed to reduce false positives derived from population structure and/or genetic relatedness. However, correcting for population structure not only reduces the frequency of false positives but also maybe entail false negatives in situations where a character state is strongly correlated with population structure (Cockram et al., 2008; Pasam et al., 2012). If alleles segregate in only one subpopulation, or totally different alleles are present in different subpopulations, models without accounting for population structure detect strong signals in the cloud of other, false signals, while the mixed model approach misses them

entirely. Therefore, the best model differs among populations and traits. The most appropriate model should be chosen by comparing different models. We compared six models and found that the K-model, QK model and PK model performed better than the naïve, Q and P models (Fig. 4). This was similar to the previous studies of comparing GWAS models in allogamous and autogamous species (Stich and Melchinger, 2009; Kang et al., 2008). We also found that the K-model, QK model and PK performed very similarly. In previous GWAS studies in rice, the MLM model took genome-wide patterns of genetic relatedness into account, greatly reduced false positives (Zhao et al., 2011; Huang et al., 2010, Zhao et al., 2015; Cui et al., 2015). Since we detected two and three subpopulations in the DC12 and RMPRIL we used MLM (PK) models to control both population structure and kinship to reduce the false positive associations.

4.2. Eight-way population is more powerful than the four-way populations

In the present study, the significant associations identified for 10 traits were more using the 8-way population than using the DC1, DC2 or DC12 population. Considering that the DC12 and 8-way populations had the same eight founders and very similar number of lines, our results suggested that the multiparent population derived from the 8-way cross is better than those derived from the two 4-way crosses and the use of two populations derived from the 4-way crosses that cover all the eight parents (i.e. the DC12 population) may fail to detect the QTL detectable using the 8-way population. This is consistent with the expectation that generating progeny from multiple parents offers an advantage in that more QTL can be detected. However, there were QTL that were not detected using the 8-way population were detected using the DC1 (i.e. *qPGN4*), DC2 (i.e. *qSBN12*) or DC12 population (i.e. *qPH2*, *qPH6* and *qPN2*). When only the QTL corresponding to known genes were considered, the QTL undetected by the 8-way population is the *qGN4* for the DC1 population. The more parents involved in developing a population usually implies that the frequencies of different alleles are lower and uneven across alleles. This results in reduced power for detecting QTL with small effects. If the objective is to identify QTL with relatively large effects to be directly used in breeding, the use of more parental lines is advantageous.

4.3. Combined analysis of multiple populations is beneficial

A strategy proposed to overcome the limitation of single biparental population in allelic diversity covered, mapping power and resolution is the joint analysis using multiple-line crosses (MC). QTL analyses across populations revealed a substantially higher number of QTL compared to the analyses of single biparental populations (Blanc et al., 2006; Negeri et al., 2011). The joint analyses also allow estimating QTL positions with higher precision (Steinhoff et al., 2011; Coles et al., 2010). Moreover, MC-QTL mapping allows the investigation of variation in allele substitution effects across different genetic backgrounds (Steinhoff et al., 2012). In the present study, more QTL were identified using the DC12 population (10 QTL) than both of the DC1 (6 QTL) and DC2 (2 QTL) populations, suggesting that the joint analysis had substantially higher QTL detection power. Although the number of lines of the DC12 population was much more than that of either DC1 or DC2, the gain in QTL identification was not simply obtained by the increased population size. Combining the DC1 and DC2 populations allows detecting QTL that are segregating only in one of the populations and as a result are not detectable in the other population. Clearly if all the resources were used to develop a single population, much higher power for detecting the QTL segregating in the population would be achieved. However, QTL that are not segregating in the population will be not be detected. It is also worthy pointing out that we used the unified mixed model of association mapping (Yu et al., 2008) in analyzing the DC12 and RMPRIIL populations, which assumes that the QTL is bi-allelic and identity-by-state alleles at the QTL position are identical over all the populations. This is the most parsimonious model but not always the most powerful model (Bardol et al., 2013). The so-called disconnected models, for which QTL effects are effectively fitted as nested effects within populations (Rebaï et al., 1997; Jannink and Jansen, 2001) might be more powerful.

The RMPRIIL population that contains 1071 lines of the DC1, DC2 and 8-way populations had higher power than other populations in the present study. This was expected, since RMPRIIL had the largest population size. We noted that most of the loci identified in the RMPRIIL population that were also detected in the DC1, DC2 or 8-way population were major loci with very small *q* values. If only QTL corresponding to the known genes were considered all the nine QTL were detected using the RMPRIIL while eight, seven, three and one were detected using the 8-way, DC12, DC1, and DC2, respectively. *qGN4* was detected using the RMPRIIL population but not the 8-way population. Combining multiple populations implies that population structure is more likely to be a problem and methods for controlling population

structure must be adopted. We used the mixed model based association analysis taking into consideration both of population structure and the genetic relatedness between genotypes. Such a model may overcompensate for population structure and relatedness and lead to false negatives. This is particularly true for traits associated with subpopulations (Stich et al., 2008). This could partially explain why some QTL were undetected by the joint analysis.

4.4. QTL-by-season interaction is important

In the present study, more QTL were detected in the WS (16) than in the DS (19). Since in many rice growing regions including Philippines the two growing seasons are very different in solar radiation and other climatic factors, disease and pest pressures and agronomic practices, QTL identification studies commonly reported that different QTL were identified for different seasons. For instance, in a recent GWAS study conducted using 369 advanced breeding lines from IRRI's irrigated rice breeding program only four out of the 52 identified QTL were significant for both of the DS and WS in 2012 at IRRI (Begum et al., 2015). In previous studies, it was common that more QTL was identified for the DS, since the accuracy of trials conducted at the DS was usually higher (Begum et al., 2015). It is difficult to figure out exactly why more QTL were detected in the WS in the present study. Nevertheless, nine QTL were commonly identified across the two seasons. Six out of the nine QTL were in chromosomal regions containing known QTL/genes. They were *qPH1.1*, *qPH1.2*, *qHD6*, *Hd3a*, *qFLW4*, *qTGW3* and *qTGW5* correspondent to *Psd1*, *sd1*, *Nal1/qFLW4/LSCHL4*, *GS3* and *qSW5*, respectively.

4.5. Clusters of QTL are present

Four clustered QTL regions were detected in the present study. The most interesting one is the cluster containing *qFLW4*, *qGN4* and *qSBN4* detected on chromosome 4, which is 33 kb away from the genes *SPIKELET NUMBER (SPIKE)* and *Nal1/qFLW4/LSCHL4*. Many QTL for morphological traits, including leaf size, spikelet number, panicle number, and root volume, have been identified in this chromosome region. The *NAL1* gene that is very close to our *qFLW4* was found to affect leaf width, GPS (green for photosynthesis) related to photosynthetic efficiency and SPIKE related to the number of branches per panicle. All previous studies reported that the desirable SPIKE and *Nal1/qFLW4/LSCHL4* alleles were from *japonica* rice cultivars. In the present study, the desirable allele of QTL *qFLW4* was from the parental line A (SAGC-08), an

indica variety. The allele increased FLW by an average of 1.76 mm and 2.59 mm in the DS and WS in the 8-way population, 1.86 mm and 1.65 mm in the DC12 population, 1.92 mm and 1.52 mm in the RMPRIL population, respectively. This QTL cluster could be useful for increasing yield by optimizing plant type and the source-sink relationship. The cluster containing *qTGW3*, *qGN3* and *qSF3.1* on chromosome 3 was 342 kb away from the *GS3*. The *GS3* locus has been frequently identified as a major QTL for both grain weight and grain length in the literature. The cluster containing *qHD3* and *qPBN3* on chromosome 3 was 623 bp away from the gene *DTH3*. *DTH3* affected flowering time and had no significant effect on other main agronomic traits (Bian et al., 2011). The cluster containing *qHD6* and *qPL6* on chromosome 6 was 1.9 kb away from the gene *Hd3a*. Recent studies showed that *Hd3a* functions as a mobile signal for branching in rice (Tsuji et al., 2015).

4.6. Phenotypic variation and usefulness in breeding

The three MAGIC populations were developed as part of the recurrent selection for population improvement and variety development program at IRRI. All the eight founders were elite lines and chosen based on their desirable characters for practical breeding. For the development of the MAGIC populations no artificial selection was made. However transgressive segregation was seen in the three MAGIC populations for all traits, particularly for GY, SN, SBN and GN (Table 1). From the three MAGIC populations 34 high yielding lines with satisfactory agronomic performance were selected by IRRI breeders. The 8-way, DC1 and DC2 contributed 21, seven and six lines, respectively. The 34 lines had about two times higher yield (46.9 g/plant) than the population average (24.7 g/plant) (data not shown). These selected lines are currently being tested together with lines developed by other IRRI breeding programs in the preliminary yield trials in IRRI. We also developed four single cross RILs populations using the same eight parental lines (AxB, CxD, ExF and GxH). The single cross populations were much lower in trait variations and the number of transgressive segregates than the MAGIC populations and not used in this study. Nevertheless, at least for yield related traits multiparental crosses seemed to be better than single crosses. Bults derived from outstanding plants of the same crosses (4-way and 8-way) were developed and progeny-tested to capture the chance to extract best recombinants for variety development and advance the population to the next cycle. More

than 2000 advanced lines other than the MAGIC lines have been developed and 500 were selected by breeders for irrigated ecosystems in South-Asia, South-east Asia and Africa.

5. Conclusions

Overall, our results suggested that MAGIC populations derived from diverse elite lines could be successfully used in identifying QTL through GWAS. A total of 26 QTL on all chromosomes except seven, nine and 11 were identified for 10 of the 14 measured traits. Nine of the QTL corresponded to known QTL/genes. Six, two, 12, 10 and 20 out of the 26 QTL were identified in the DC1, DC2, 8-way, DC12 and RMPRIL populations, respectively. The 8-way population was found to be more powerful than the two 4-way populations (DC1, DC2). The joint analysis, which combines different populations (e.g. DC12 and RMPRIL), increased the number of QTL identified and mapping resolution.

Abbreviations

RMPRIL: Rice multi-parent recombinant inbred line population; FLL: Flag leaf length; FLW: Flag leaf width; FLR: Flag length and width ratio; GN: Grain number per panicle; GY: Grain yield; HD: Heading date; PBN: Primary branch number; PH: Plant height; PL: Panicle length; PN: Panicle number; SBN: Secondary branch number; SF: Spikelet fertility; SN: Spikelet number per panicle; TGW: Thousand grain weight.

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

Acknowledgments

Thanks go to Ms Alaine Gulles for helping with data analysis. LM was supported by the Green Super Rice Project funded by the Bill & Melinda Gates Foundation. The research of GY was partially supported by the National Natural Science Foundation of China (grant 31461143014).

Appendix A. Supplementary data

Suppl. Fig.1. Distribution of 14 traits in the DC1, DC2 and 8-way populations measured in the DS and WS of 2014.

Suppl. Fig. 2. Correlation coefficients among different traits estimated in the DS and WS for the DC1, DC2 and 8-way populations.

Suppl. Fig. 3. Kinship between lines in the DC1 (A), DC2 (B), 8-way (C), DC12 (D), RMPRIL (E) *indica* rice populations.

Suppl. Fig. 4. LD patterns of the DC1 (A), DC2 (B), 8-way (C), DC12 (D), RMPRIL (E) *indica* rice populations.

Suppl. Table 1 Correlation coefficients among different traits estimated in the DS and WS for the DC1, DC2 and 8-way populations.

Author Contributions

Conceived and designed the experiments: GY and HL. Performed the experiments: LM, XZ, and KP. Analyzed the data: LM and GY. Contributed reagents/materials/analysis tools: GY. Wrote the paper: LM and GY.

Acknowledgments

Thanks go to Ms Violeta Bartolome for helping with data analysis.

Author details:

¹ Plant Breeding, Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Los Baños, Laguna, Philippines

²Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

References

Bandillo, N., Raghavan, C., Muyco, P.A., Sevilla, M.A.L., Lobina, I.T., Dilla-Ermita, C.J., Tung, C.-W., McCouch, S., Thomson, M., Mauleon, R., Singh, R.K., Gregorio, G., Redoña, E.,

- Leung, H., 2013. Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice* (N. Y). 6, 6–11. doi:10.1186/1939-8433-6-11
- Bardol, N., Ventelon, M., Mangin, B., Jasson, S., Loywick, V., Couton, F., Derue, C., Blanchard, P., Charcosset, A., Moreau, L., 2013. Combined linkage and linkage disequilibrium QTL mapping in multiple families of maize (*Zea mays* L.) line crosses highlights complementarities between models based on parental haplotype and single locus polymorphism. *Theor. Appl. Genet.* 126, 2717–2736. doi:10.1007/s00122-013-2167-9
- Begum, H., Spindel, J.E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B., McCouch, S.R., 2015. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS One* 10, e0119873. doi:10.1371/journal.pone.0119873
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* doi:10.2307/2346101
- Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Sci.* 48, 1649–1664. doi:10.2135/cropsci2008.03.0131
- Bian, X.F., Liu, X., Zhao, Z.G., Jiang, L., Gao, H., Zhang, Y.H., Zheng, M., Chen, L.M., Liu, S.J., Zhai, H.Q., Wan, J.M., 2011. Heading date gene, *dth3* controlled late flowering in *O. Glaberrima* Steud. by down-regulating *Ehd1*. *Plant Cell Rep.* 30, 2243–2254. doi:10.1007/s00299-011-1129-4
- Blanc, G., Charcosset, A., Mangin, B., Gallais, A., Moreau, L., 2006. Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theor. Appl. Genet.* 113, 206–224. doi:10.1007/s00122-006-0287-1
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi:10.1093/bioinformatics/btm308
- Breseghele, F., Sorrells, M.E., 2006. Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323–1330. doi:10.2135/cropsci2005.09-0305
- Cavanagh, C., Morell, M., Mackay, I., Powell, W., 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr. Opin. Plant Biol.* 11, 215–221. doi:10.1016/j.pbi.2008.01.002

687 Chen, M., Luo, J., Shao, G., Wei, X., Tang, S., Sheng, Z., Song, J., Hu, P., 2012. Fine mapping
688 of a major QTL for flag leaf width in rice, qFLW4, which might be caused by alternative
689 splicing of NAL1. *Plant Cell Rep.* 31, 863–872. doi:10.1007/s00299-011-1207-7

690 Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. *J. Am.*
691 *Stat. Assoc.* 74, 829–836. doi:10.2307/2286407

692 Cockram, J., White, J., Leigh, F.J., Lea, V.J., Chiapparino, E., Laurie, D.A., Mackay, I.J., Powell,
693 W., O’Sullivan, D.M., 2008. Association mapping of partitioning loci in barley. *BMC*
694 *Genet.* 9, 16. doi:10.1186/1471-2156-9-16

695 Coles, N.D., McMullen, M.D., Balint-Kurti, P.J., Pratt, R.C., Holland, J.B., 2010. Genetic
696 control of photoperiod sensitivity in maize revealed by joint multiple population analysis.
697 *Genetics* 184, 799–812. doi:10.1534/genetics.109.110304

698 Deshmukh, R., Singh, A., Jain, N., Anand, S., Gacche, R., Singh, A., Gaikwad, K., Sharma, T.,
699 Mohapatra, T., Singh, N., 2010. Identification of candidate genes for grain number in rice
700 (*Oryza sativa* L.). *Funct. Integr. Genomics* 10, 339–347. doi:10.1007/s10142-010-0167-2

701 Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X., Zhang, Q., 2006. GS3, a major QTL
702 for grain length and weight and minor QTL for grain width and thickness in rice, encodes a
703 putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi:10.1007/s00122-
704 006-0218-1

705 Fitzgerald1, M.A., McCouch, S.R., Hall, R.D., 2009. Not just a grain of rice: the quest for
706 quality. *Trends Plant Sci.* 14, 133–139. doi:10.1016/j.tplants.2008.12.004

707 Fujita, D., Trijatmiko, K.R., Tagle, A.G., Sapasap, M.V., Koide, Y., Sasaki, K., Tsakirpaloglou,
708 N., Gannaban, R.B., Nishimura, T., Yanagihara, S., Fukuta, Y., Koshiha, T., Slamet-Loedin,
709 I.H., Ishimaru, T., Kobayashi, N., 2013. NAL1 allele from a rice landrace greatly increases
710 yield in modern indica cultivars. *Proc. Natl. Acad. Sci. U. S. A.* 110, 2–7.
711 doi:10.1073/pnas.1310790110

712 Guo, L.B., Ye, G.Y., 2014. Use of major quantitative trait loci to improve grain yield of rice.
713 *Rice Sci.* 21, 65–82. doi:10.1016/S1672-6308(13)60174-2

714 Holland, J., 2007. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* 10,
715 156–161. doi:10.1016/j.pbi.2007.01.003

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H.,
 Varshney, R.K., Cavanagh, C.R., 2015. MAGIC populations in crops: current status and
 future prospects. *Theor. Appl. Genet.* 128, 999–1017. doi:10.1007/s00122-015-2506-0

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li,
 M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K.,
 Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q. F., Li, J.,
 Han, B., 2010. Genome-wide association studies of 14 agronomic traits in rice landraces.
Nat. Genet. 42, 961–967. doi:10.1038/ng.695

Jannink, J.L., Jansen, R., 2001. Mapping epistatic quantitative trait loci with one-dimensional
 genome searches. *Genetics* 157, 445–454.

Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., Eskin, E., 2008.
 Efficient control of population structure in model organism association mapping. *Genetics*
 178, 1709–1723. doi:10.1534/genetics.107.080101

Khush, G.S., 2005. What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol. Biol.*
 59, 1–6. doi:10.1007/s11103-005-2159-5

Li, R., Xia, J., Xu, Y., Zhao, X., Liu, Y.G., Chen, Y., 2014. Characterization and genetic
 mapping of a Photoperiod-sensitive dwarf 1 locus in rice (*Oryza sativa* L.). *Theor. Appl.*
Genet. 127, 241–250. doi:10.1007/s00122-013-2213-7

Li, Z., Ye, G., Yang, M.E.I., Liu, Z., Lu, D., Mao, X., Wu, Q., Li, X., 2014. Genetic
 characterization of a multiparent recombinant inbred line of rice population. *Res. Crop.* 15,
 28–37. doi:10.5958/j.2348-7542.15.1.004

Liang, S., Ren, G., Liu, J., Zhao, X., Zhou, M., McNeil, D., Ye, G. 2015. Genotype-by-
 environment interaction is important for grain yield in irrigated lowland rice. *Field Crops*
Res. 180, 90-99

Liang, S., Sun, C., Ren, G., Zhao, X., Zhou, Z., McNeil, D., and Ye G, 2016. Usefulness of the
 cloned and fine-mapped genes/QTL for grain yield and related traits in *indica* rice breeding
 for irrigated ecosystems. *Field Crops Res.*, <http://dx.doi.org/10.1016/j.fcr.2015.11.014>

Lipka, A.E., Kandianis, C.B., Hudson, M.E., Yu, J., Drnevich, J., Bradbury, P.J., Gore, M. A,
 2015. From association to prediction: statistical methods for the dissection and selection of
 complex traits in plants. *Curr. Opin. Plant Biol.* 24, 110–118. doi:10.1016/j.pbi.2015.02.010

- Meng, L., Guo, L., Ponce, K., Zhao, X., Ye, G. 2016 Characterization of three indica rice multiparent advanced generation intercross (MAGIC) populations for QTL identification. *The Plant Genome (Accepted)*
- Miura, K., Ashikari, M., Matsuoka, M., 2011. The role of QTLs in the breeding of high-yielding rice. *Trends Plant Sci.* 16, 319–326. doi:10.1016/j.tplants.2011.02.009
- Negeri, A.T., Coles, N.D., Holland, J.B., Balint-Kurti, P.J., 2011. Mapping QTL controlling southern leaf blight resistance by joint analysis of three related recombinant inbred line populations. *Crop Sci.* 51, 1571–1579. doi:10.2135/cropsci2010.12.0672
- Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5269–5273. doi:10.1073/pnas.76.10.5269
- Pasam, R.K., Sharma, R., Malosetti, M., Eeuwijk, F.A. van, Haseneyer, G., Kilian, B., Graner, A., 2012. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 12. doi:10.1186/1471-2229-12-16
- Pauli, F., Torelli, N., 2015. Relabelling in bayesian mixture models by pivotal units. *arXiv Prepr. arXiv 1501, 05478.*
- Peng, S., Huang, J., Sheehy, J.E., Laza, R.C., Visperas, R.M., Zhong, X., Centeno, G.S., Khush, G.S., Cassman, K.G., 2004. Rice yields decline with higher night temperature from global warming. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9971–9975. doi:10.1073/pnas.0403720101
- Qiu, X., Pang Y., Yuan, Z. Xing, D., Xu, J., Dingkuhn, M., Li, Z., **Ye, G.** 2015. Genome-wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of *Indica* Rice Germplasm. ***PloS One* 10:e0145577. Doi:10.1371/journal.pone.0145577**
- R Development Core Team, 2014. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria, 2012.
- Rebaï, A., Blanchard, P., Perret, D., Vincourt, P., 1997. Mapping quantitative trait loci controlling silking date in a diallel cross among four lines of maize. *Theor. Appl. Genet.* 95, 451–459. doi:10.1007/s001220050582
- Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Matsuoka, M., 2002. A mutant gibberellin-synthesis gene in rice. *Nature* 416, 701–702.

775 Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., Yano, M., 2008.
 776 Deletion in a gene associated with grain size increased yields during rice domestication. *Nat.*
 777 *Genet.* 40, 1023–1028. doi:10.1038/ng.169
 778 Steinhoff, J., Liu, W., Maurer, H.P., Würschum, T., Longin, C.F.H., Ranc, N., Reif, J.C., 2011.
 779 Multiple-line cross quantitative trait locus mapping in european elite maize. *Crop Sci.* 51,
 780 2505–2516. doi:10.2135/cropsci2011.03.0181
 781 Steinhoff, J., Liu, W., Reif, J.C., Porta, G. Della, Ranc, N., Würschum, T., 2012. Detection of
 782 QTL for flowering time in multiple families of elite maize. *Theor. Appl. Genet.* 125, 1539–
 783 1551. doi:10.1007/s00122-012-1933-4
 784 Stich, B., Melchinger, A.E., 2009. Comparison of mixed-model approaches for association
 785 mapping in rapeseed, potato, sugar beet, maize, and Arabidopsis. *BMC Genomics* 10, 94.
 786 doi:10.1186/1471-2164-10-94
 787 Stich, B., Möhring, J., Piepho, H.P., Heckenberger, M., Buckler, E.S., Melchinger, A.E., 2008.
 788 Comparison of mixed-model approaches for association mapping. *Genetics* 178, 1745–1754.
 789 doi:10.1534/genetics.107.079707
 790 Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. *Proc. Natl.*
 791 *Acad. Sci. U. S. A.* 100, 9440–9445. doi:10.1073/pnas.1530509100
 792 Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., Yamaguchi, M., Nakashima,
 793 C., Purwestri, Y.A., Tamaki, S., Ogaki, Y., Shimada, C., Nakagawa, A., Kojima, C.,
 794 Shimamoto, K., 2011. 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen.
 795 *Nature* 476, 332–335.
 796 Tester, M., Langridge, P., 2010. Breeding technologies to increase crop production in a changing
 797 world. *Science* (80-.). 327, 818–822. doi:10.1126/science.1183700
 798 Tsuji, H., Tachibana, C., Tamaki, S., Taoka, K., Kyojuka, J., Shimamoto, K., 2015. Hd3a
 799 promotes lateral branching in rice. *Plant J.* 82, 256–266. doi:10.1111/tpj.12811
 800 Van Inghelandt, D., Reif, J.C., Dhillon, B.S., Flament, P., Melchinger, A.E., 2011. Extent and
 801 genome-wide distribution of linkage disequilibrium in commercial maize germplasm. *Theor.*
 802 *Appl. Genet.* 123, 11–20. doi:10.1007/s00122-011-1562-3
 803 Wickham, H., Chang, W., 2013. An implementation of the grammar of graphics. *R Packag.*
 804 version.

- Xing, Y., Zhang, Q., 2010. Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442. doi:10.1146/annurev-arplant-042809-112209
- Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi:10.1038/ng1702
- Zhang, G., Li, S., Wang, L., Ye, W., Zeng, D., 2014. LSCHL4 from japonica cultivar, which is allelic to NAL1, increases yield of indica super rice 93-11. *Mol. Plant* 7, 1350–1364. doi:10.1093/mp/ssu055
- Zhang, Z., Ersoz, E., Lai, C., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., Buckler, E.S., 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546
- Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D., McCouch, S.R., 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2, 467. doi:10.1038/ncomms1467
- Zhao, X., Zhou L., Ponce, K., Ye, G., 2015. The Usefulness of Known Genes/Qtls for Grain Quality Traits in an Indica Population of Diverse Breeding Lines Tested using Association Analysis. *Rice* 8, 1-13

Table 1 Performance of 14 traits of the parental lines and MAGIC populations measured in IRRI headquarter (Los Banos, Philippines) in the dry season (DS) and wet season (WS) of 2014.

Trait	Env.	Parental Lines								MAGIC population					
		A	B	C	D	E	F	G	H	DC1		DC2		8-way	
		Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean±SD	Range	Mean±SD	Range	Mean±SD	Range
GY (g/plant)	DS	24.1	14.4	15.3	24.9	30.3	27.2	21.5	28.1	25.2±7.3	7.0-49.9	25.8±7.3	9.3-49.1	25.9±8.1	7.7-67.1
	WS	21.8	22.0	21.7	24.8	18.6	19.8	22.3	25.4	20.3±5.7	5.3-38.8	20.4±5.7	5.3-40.0	20.0±5.9	6.0-42.1
PN	DS	9.9	14.1	8.7	11.9	10.9	13.3	9.9	15.7	11.2±2.5	6.0-20.3	12.3±2.5	6.7-23.3	11.9±2.6	5.0-22.3
	WS	10.0	11.8	8.8	9.8	11.0	11.7	11.7	11.7	10.0±2.4	3.7-21.7	10.5±2.7	3.7-22.2	10.6±2.3	4.2-22.2
GN	DS	139.9	70.0	99.3	103.9	130.2	93.2	83.3	90.6	115.3±29.4	64.1-222.4	102.0±29.4	35.6-217.9	109.9±27.0	44.1-227.6
	WS	129.4	181.5	148.9	221.2	113	166.9	134.7	215.0	146.1±62.8	20.8-394.9	146.3±59.4	20.8-394.9	141.2±58.8	28.5-348.4
	WS	16.8	13.8	20.3	17.0	15.7	10.8	15.7	18.8	16.1±2.5	8.5-23.9	16.0±2.6	8.5-23.9	16.4±2.7	8.4-23.9
TGW (g)	DS	26.8	27.3	25.2	22.8	30.3	25.5	32.1	26.9	26.6±3.0	20.3-34.9	28.1±3.0	20.3-36.2	27.9±3.3	20.3-39.5
	WS	27.4	28.1	26.0	24.6	29.6	27.4	32.9	24.8	27.2±3.0	20.9-35.6	27.9±3.2	20.5-37.5	28.3±3.4	19.9-39.5
PL (mm)	DS	226.1	238.2	238.1	252.0	259.5	242.7	278.8	248.0	248.2±18.8	185.6-296.6	244.4±18.8	183.0-296.3	247.3±18.5	192.4-309.7
	WS	246.2	263.7	291.3	279.5	246.5	260.3	298.8	277.2	250.7±36	119.2-341.0	248.6±35.4	90.3-341.0	249.1±32.6	105.3-345.3
PBN	DS	11.5	10.8	10.5	10.0	8.3	10.0	12.0	9.3	9.6±1.4	6.8-14.8	9.3±1.4	6.0-13.2	10.2±1.4	5.5-16.5
	WS	10.8	9.8	8.9	10.1	10.8	9.7	12.1	10.1	9.5±1.9	4.0-16.0	9.3±1.9	2.5-16.0	9.5±2.0	3.0-15.5
SBN	DS	40.8	24.0	42.7	27.7	25.8	23.7	33.3	34.8	26.8±7.5	12.1-59.5	23.2±7.5	8.5-51.9	26.1±7.4	5.4-56.8
	WS	37.9	20.2	23.2	22.0	29.9	22.7	24.5	22.3	27.3±9.2	8.0-54.5	24.8±8.9	4.7-56.7	25.4±9.5	4.3-61.2
SN	DS	183.6	97.4	132.4	138.4	154.5	104.7	114.8	128.0	146.7±34.3	86.1-263.0	126.0±34.3	59.8-246.7	142.7±34.5	74.2-258.3
	WS	203.4	223.0	199.0	244.3	138.9	189.2	168.3	251.0	179.5±67.8	42.0-423.6	175.6±64.1	40.5-423.6	172.5±62.5	33.5-365.4
SF (%)	DS	76.7	73.5	74.2	75.2	85.2	89.0	72.0	70.5	78.8±9.1	40.3-95.4	81.4±9.1	52.7-96.7	77.8±11.2	26.8-97.3
	WS	63.6	81.0	74.6	90.4	81.4	88.3	76.1	85.4	80.5±11.5	41.8-99.2	82.7±10.9	41.8-99.2	81.0±11.6	35.6-98.7
FLL (mm)	DS	322	355.2	297.2	366.4	344.6	325.0	301.5	349.6	345.9±62.0	223.3-548.9	323.0±62.0	194.7-489	330.3±61.2	176.0-526.0
	WS	308.3	331.2	356.8	370.0	336.7	300.5	339.0	345.8	327.1±50.8	182.7-458.6	322.2±50.1	182.7-471.1	326.1±60.3	143.7-681.2
FLR	DS	17.5	23.1	15.1	21.2	19.1	24.0	17.4	20.2	19.8±4.0	12.3-39.4	18.9±4.0	9.6-29.8	18.4±3.7	10.0-39.1
	WS	19.1	24.6	17.4	22.0	21.3	27.9	21.9	18.6	20.6±3.9	12.9-45.4	20.5±3.9	12.2-45.4	20.3±4.7	10.5-47.3
FLW (mm)	DS	18.7	15.9	19.7	17.2	18.1	13.5	17.6	17.6	17.7±2.4	11.9-24.2	17.4±2.4	11.4-27.4	18.2±2.3	12.1-25.9
PH (cm)	DS	106.7	115.0	133.0	109.3	114.0	107.3	112.7	116.0	109.9±18.2	76.7-163.7	107.1±18.2	59.7-163.5	109.7±12.7	76.7-151.5
	WS	114.3	119.3	140.8	106.7	106.2	116.3	130.0	125.2	110.6±18.8	59.6-172.1	110.3±17.1	59.6-172.1	112.3±14.9	74.2-167.7
HD (days)	DS	97.3	99.0	92.3	96.3	87.0	90.7	99.7	96.0	92.0±6.7	73.7-115.7	91.1±6.7	78.4-111.4	94.8±6.7	78.0-116.4
	WS	86.3	100.7	89.3	90.7	86.0	92.0	94.3	89.3	85.1±6.0	68.8-117.8	85.5±5.7	68.8-117.8	87.0±6.2	70.8-116.8

Grain yield (GY, g/plant), panicle number (PN), grain number per panicle (GN), Thousand grain weight (TGW, g), flag leaf length (FLL, mm), panicle length (PL, mm), primary branch number (PBN), secondary branch number (SBN), spikelet fertility (SF, %), spikelet number per panicle (SN), flag length and width ratio (FLR), flag leaf width (FLW, mm), plant height (PH, cm), heading date (HD, days).

Table 2 QTL results for 10 traits in five populations.

Trait	QTL	Position	Env.	DC1		DC2		8-way		DC12		RMPRIL		Gene Symbol	Reference
				R ²	Effect	R ²	Effect	R ²	Effect	R ²	Effect	R ²	Effect		
PN	<i>qPN2</i>	25103588	WS	-	-	-	-	-	-	3.3	1.9	-	-		
GN	<i>qGN3</i>	17019705	WS	-	-	-	-	-	-	-	-	1.8	25.1		
	<i>qGN4</i>	31052578	DS	6.3	26.3	-	-	-	-	-	-	1.4	12.2	<i>qGN4-1</i>	[27]
	<i>qGN8</i>	2175226	DS	-	-	-	-	-	-	-	-	1.4	16.0		
	<i>qTGW2</i>	16086722	DS	-	-	-	-	4.6	-1.9	-	-	-	-		
TGW	<i>qTGW3</i>	17019705	DS	-	-	-	-	6.5	-2.7	-	-	2.7	-1.9	<i>GS3</i>	[30]
			WS	8.0	-2.1	-	-	5.8	-2.5	4.2	-2.1	3.9	-2.3		
	<i>qTGW5</i>	5697111	DS	7.9	2.4	-	-	-	-	4.0	2.3	1.5	1.5	<i>qSW5</i>	[31]
			WS	7.5	2.3	-	-	3.6	2.1	5.4	2.6	3.2	2.1		
PL	<i>qPL6</i>	2924222	WS	-	-	-	-	4.3	18.7	-	-	2.2	15.5		
	<i>qPL8</i>	26038950	WS	-	-	-	-	3.3	22.1	-	-	-	-		
PBN	<i>qPBN3</i>	1270943	DS	-	-	-	-	-	-	-	-	2.6	0.6		
			WS	-	-	-	-	-	-	-	-	1.7	0.7		
SBN	<i>qPBN4</i>	15852364	DS	-	-	-	-	-	-	-	-	1.8	-0.8		
	<i>qSBN4</i>	33157662	WS	-	-	-	-	3.3	6.2	-	-	1.6	6.0	<i>SPIKELET NUMBER (SPIKE)</i>	[32]
	<i>qSBN12</i>	3144839	DS	-	-	6.4	-5.1	-	-	-	-	-	-		
SF	<i>qSF3.1</i>	17019705	WS	-	-	-	-	-	-	-	-	2.3	5.8		
	<i>qSF3.2</i>	31835485	WS	-	-	-	-	-	-	-	-	1.9	-7.0		
FLW	<i>qFLW4</i>	31052578	DS	-	-	-	-	5.8	1.8	6.1	1.9	6.2	1.9	<i>Nal1/qFLW4/LSCHL4</i>	[28, 29]
			WS	-	-	-	-	3.4	2.6	3.7	1.6	3.1	1.5		
	<i>qFLW6</i>	6399722	DS	-	-	-	-	-	-	-	-	1.5	0.8		
	<i>qFLW10</i>	146531	WS	-	-	-	-	-	-	-	-	1.3	-1.1		
PH	<i>qPH1.1</i>	37030733	DS	8.0	15.7	-	-	3.4	10.2	5.1	15.5	2.7	12.0	<i>Psd1</i>	[33]
			WS	9.4	17.4	-	-	4.6	14.6	5.9	18.2	2.7	13.1		
	<i>qPH1.2</i>	38652270	DS	12.1	15.6	-	-	5.9	13.8	8.2	18.7	5.6	16.2	<i>sd1/OsGA20ox2</i>	[34]
			WS	10.2	15.2	-	-	6.6	17.7	9.3	21.5	4.6	16.8		
	<i>qPH2</i>	29568762	WS	-	-	-	-	-	-	2.3	10.6	-	-		
	<i>qPH6</i>	26318220	DS	-	-	-	-	-	-	3.3	9.9	1.2	6.4		
			WS	-	-	-	-	-	-	-	-	2.0	7.7		
	<i>qPH12</i>	22544640	DS	-	-	-	-	3.2	5.7	-	-	-	-		
HD			WS	-	-	-	-	3.2	7.0	-	-	-	-		
	<i>qHD3</i>	1270943	DS	-	-	7.5	4.6	3.1	3.5	5.1	4.2	5.5	4.7	<i>DTH3</i>	[35]
			WS	-	-	-	-	-	-	-	-	2.5	2.8		
	<i>qHD6</i>	2924222	DS	-	-	-	-	3.1	3.5	-	-	1.8	3.1	<i>Hd3a</i>	[36]
			WS	7.5	4.7	-	-	7.6	4.6	5.7	5.1	4.0	4.4		
	<i>qHD10</i>	9260265	WS	-	-	-	-	-	-	-	-	1.4	2.4

Fig.1. Principal Component Analysis (PCA) of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations.

(A)

(B)

(C)

(D)

(E)

Fig. 2. Phylogenetic trees of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations.

Fig. 3. LD decay (r^2) of marker-pairs over all chromosomes as a function of physical distance (Mb) of five *indica* rice populations. The viewport figure is zoomed in 0 to 10 Mb is on the top right.

Fig. 4. Comparison of different statistical models. Quantile-Quantile plots for the trait plant height of DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations measured in the dry season of 2014 in IRRI.

Fig. 5. Genome-wide association studies for PN of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 6. Genome-wide association studies for GN of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$.

Fig. 7. Genome-wide association studies for TGW of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 8. Genome-wide association studies for PL of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 9. Genome-wide association studies for PBN of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 10. Genome-wide association studies for SBN of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 11. Genome-wide association studies for SF of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 12. Genome-wide association studies for FLW of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log₁₀ q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 13. Genome-wide association studies for PH of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$

Fig. 14. Genome-wide association studies for HD of the DC1 (A), DC2 (B), 8-way (C), DC12 (D) and RMPRIL (E) *indica* rice populations. Negative log10 q-value from a genome-wide scan are plotted against physical position on each of 12 chromosomes. Horizontal solid line indicates the genome-wide significance threshold $q < 0.05$