



Usefulness of the cloned and fine-mapped genes/QTL for grain yield and related traits in *indica* rice breeding for irrigated ecosystems

Shanshan Liang^{a,b}, Chenxiao Sun^c, Guangjun Ren^d, Xiangqian Zhao^b, Meixue Zhou^a, David McNeil^a, Guoyou Ye^{b,*}

^a University of Tasmania, Private Bag 98, Hobart, Tasmania 7001, Australia

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

^c China National Rice Research Institute, Hangzhou 310006, China

^d Institute of Crop Science, Sichuan Academy of Agricultural Sciences, Chengdu 610066, Sichuan, China

ARTICLE INFO

Article history:

Received 14 September 2015

Received in revised form

27 November 2015

Accepted 28 November 2015

Available online 19 December 2015

Keywords:

Association analysis

Breeding

Grain yield

Marker-assisted selection

QTL

Rice (*Oryza sativa* L.)

ABSTRACT

Many genes/QTL for grain yield (GY) and yield related traits in rice have been cloned or fine-mapped in the last three decades. A collection of *indica* elite breeding lines and cultivars assembled in IRRI was used to test the usefulness of 39 well characterized yield related genes/QTL. The population of lines was phenotyped for GY and ten yield related traits under eight environments of three locations including Jiangxi, and Sichuan in China, and six season (2) and nitrogen rate (3) combinations in IRRI and genotyped using 46 markers tightly linked to the 39 target genes/QTL and 53 SSR markers evenly distributed on the genome. Using the 53 random SSR markers identified two major subpopulations. Association analyses were separately carried out for the whole population and the two subpopulations. All the 39 target genes/QTL were associated with two or more measured traits including traits not previously reported. *GW6* and *Gn1a* were associated with nine and eight traits, respectively. *Ghd7*, *qSPP7*, *SCM2*, and *SPP1* were associated with seven traits. *GIF1* and *Ltn* were associated with six traits. *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1*, and *qPH6-1* were associated with five traits. *D10*, *d27*, *DEP2*, *DWL1*, *Gnp4*, *Gw1-1*, *GW3*, *gw5*, *MOC1*, *PAP2*, *qGL7*, *qGL7-2* and *qGN4-1* were associated with four traits, *D88*, *Ghd8*, *GS5*, *Gw1-2*, *IPA1*, *qSH3* and *RPH* were associated with three traits. *ep3*, *gw8.1*, *gw9.1* and *qPDS3* were associated with two traits. A total of 16 genes/QTL were found to be associated with GY. *GS3*, *GW1-1* and *d27* were associated with GY in two testing environments and the others were only in one environment. Sixteen, six and ten genes/QTL were associated with panicle number per plant, grain number per panicle and thousand grain weight, respectively. Significant gene-by-environment interaction was present for all the studied genes/QTL. GY could not be well predicted using the markers significantly associated with the measured traits or all target markers based on stepwise multiple linear regression analysis. The adjusted coefficient of determination ranged from 0.024 to 0.191 for the final selected models considering the associated markers only and from 0.039 to 0.261 for the final selected models considering all target markers. Nevertheless the known genes might be explicitly utilized in developing more efficient selection criteria for enhancing selection accuracy.

© 2015 Elsevier B.V. All rights reserved.

Abbreviations: DTF, days to flowering; GIF, grain incomplete filling; GL, grain length; GN, grain number per panicle; GS, grain size; GW, grain width; GY, grain yield; HD, heading date; IPA, ideal plant architecture; PB, primary branches per panicle; PH, plant height; PN, panicle number per plant; SN, spikelet number per panicle; SR, seed setting rate; SB, secondary branches per panicle; TGW, thousand grain weight; TN, tiller number; WLT, withered leaf tip.

* Corresponding author. Fax: +63 2 580 5699.

E-mail address: g.ye@irri.org (G. Ye).

<http://dx.doi.org/10.1016/j.fcr.2015.11.014>

0378-4290/© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Rice (*Oryza sativa* L.) is one of the most important crops in the world. It has played a central role in human nutrition and been cultured for nearly 10,000 years (Molina et al., 2011). Great progress has occurred in rice production in the last decades due to the adoption of green revolution technology. Paddy rice production has increased from 257 million tons in 1966 to 744.4 million tons in 2014 (Khush, 2005; FAOSTAT, 2014). As world population grows, the global rice demand is estimated to rise from 439 million tons (milled rice) in 2010 to 496 million tons in 2020 and further

increase to 555 million tons in 2035 (GRiSP, 2010). At least 50% of the increase of rice production in the past has been due to the adoption of new cultivars. It is expected that breeding will still be one of the key approaches for further increasing in rice productivity. Rice breeders today face the challenge of how to increase productivity by effectively integrating well-established conventional breeding methods with new approaches offered by rapid advances in molecular marker technology and genomics.

For quantitative traits such as yield, many genes of small effects and environmental factors collectively determine the trait performance. The favorable alleles are likely to be spread across more than two lines, therefore requiring the assembly of alleles from different sources to a single inbred line in order to achieve significant improvement (Ye, 2010). Recurrent selection (RS) is a well established conventional breeding method which is designed to gradually increase the frequency of desirable alleles while maintaining genetic variability for future selections (Hallauer, 1985). RS is a cyclical breeding strategy involving three main steps: selection, evaluation and recombination of the best performing selections. Multiple genotypes are inter-mated to increase the chance of creating novel allelic combinations. Through multiple cycles of recombination between selected genotypes, linkage blocks are broken down and favorable genes are accumulated while the genetic diversity remains (Châtel et al., 2008).

The potential benefits of using molecular markers linked to the genes of interest in breeding programs, which have changed from phenotype-based toward a combination of phenotype- and genotype-based selection, have attracted much attention for more than three decades (Bernardo, 2008; Tester and Langridge, 2010). The success of marker-assisted selection (MAS) for traits of simple inheritance in many crops including rice has motivated rice breeders to search for QTL for complex traits, which account for a large proportion of phenotypic variation (major QTL) (Guo and Ye, 2014). Many yield-related genes/QTL have been identified and some of them are fine-mapped or cloned (Xing and Zhang, 2010). The effects of the well characterized genes/QTL were usually tested using specific populations. The use of these well-characterized genes/QTL in improving yield has started. However, significant improvement of GY in farm environments has yet been reported (Guo and Ye, 2014). A promising MAS method, known as genomic selection (GS) or genome-wide selection, has been recently introduced for using all trait-affecting genes to improve quantitative traits (Hayes et al., 2013; Meuwissen et al., 2001). GS uses genome-wide markers to predict the breeding (genotypic) values of the selected candidates. Once an accurate prediction model is developed using a reference population with genotyping and phenotypic observations, the model is then used to select genotypes within the selection population with genotyping data only. Simulation and empirical studies in self-pollinated crops, including wheat, barley and oat, have demonstrated the great potential of GS in improving quantitative traits (Jannink, 2010).

The irrigated rice breeding program at the International Rice Research Institute (IRRI) has a mission to develop new rice germplasm that produces high and stable yields across a broad range of environments. IRRI breeding lines have been directly released for production or used as crossing parents in many breeding programs in different countries. IRRI is adopting a new integrated breeding strategy to break yield barrier (GRiSP, 2010). The strategy utilizes RS to quickly pyramid the major QTL that have been proven useful in the breeding population in the first few selection cycles and maintain the genetic variation contributed by the many minor genes to be explored in later cycles and explores GS for reducing the length of breeding cycles and the cost of expensive phenotyping. A series of studies have been conducted to obtain essential information for designing more efficient mating and selection schemes of this general breeding strategy. As part of

this effort, 392 advanced lines and cultivars from many breeding programs in different countries representing the genetic diversity of breeding gene pools for irrigated ecosystems were collected to be used as part of the base population for future breeding at IRRI. This population was phenotyped via multi-environment trials (METs) in South-east Asia and genotyped using markers for well-characterized genes/QTL and genome-wide markers generated by genotyping by sequencing. Data collected so far has been used to analyse GEI for GY (Liang et al., 2015), investigate the factors affecting head rice yield and chalkiness (Zhou et al., 2015) and test the usefulness of important known genes/QTL for grain quality traits (Zhao et al., 2015). In this paper, the usefulness of 39 fine-mapped or cloned genes/QTL for GY and yield related traits was tested through association analysis. Prediction of GY with markers for these well characterized genes/QTL was investigated. The results can be used together with information from other studies to optimize the proposed integrated breeding strategy. The information gained will also be valuable to other rice breeding programs willing to exploit the well characterized genes/QTL for GY and related traits.

2. Materials and methods

2.1. Plant materials and phenotyping

Three hundred and ninety two rice lines developed for the irrigated lowland ecosystem were used in this study to achieve a large amount of genetic diversity. The majority of the lines were from IRRI (223). A good number of lines were also from PhilRice (31), CIAT (17), China (13) and Vietnam (11) (Supplementary Table 1). Phenotyping was conducted in Jiangxi (JX) and Sichuan (SC) in China and IRRI headquarters (Los Baños, Philippines). The experiments in JX and SC were for one crop season in 2012. At IRRI, the experiment was carried out in the dry season (DS) and wet season (WS) of 2012. Three nitrogen fertilizer application rates, no nitrogen, low (90 kg ha⁻¹), and high (180 kg ha⁻¹), were used to create three artificial environments in the DS, designated as DS1, DS2 and DS3, respectively. Similarly, three nitrogen fertilizer application rates, no nitrogen, low (45 kg ha⁻¹), and high (90 kg ha⁻¹), were used to create three artificial environments in the WS, designated as WS1, WS2 and WS3, respectively. Grain yield (GY) and ten related traits including grain number per panicle (GN), panicle number (PN), thousand grain weight (TGW), spikelets number per panicle (SN), seed setting rate (SR), number of primary branches per panicle (PB), number of secondary branches per panicle (SB), tiller number (TN), days to flowering (DTF) and plant height (PH), were tested. All 11 traits were measured for the five experiments at IRRI except DS2. In DS2, DTF, PH, PN, GY, and TGW were tested. GY, DTF, GN, PH, PN, TGW and SR were measured in SC, while GY, DTF, GN, PH and SR were measured in JX. Detailed trial description and trait measurement were given in Liang et al. (2015).

2.2. Markers and genotyping

Two sets of markers were used. The first set consisted of 53 random SSR markers evenly distributed on the whole genome (Supplementary Table 2). The second set consisted of 46 target markers, including SSR, STS and Indel markers which were tightly linked to 39 cloned or fine-mapped genes/QTL for GY or yield related traits (Table 1). Those target markers were initially chosen from the original publications, in which the known genes/QTL were reported. If there was no polymorphism in our population for any marker, another marker close to the initial marker was chosen. A total of 46 polymorphic markers for the 39 target genes/QTL were developed and screened for the population.

Table 1

Linked marker, chromosome location, known affected traits and primers of target genes/QTLs for GY and yield related traits in rice studied.

Gene/QTL	Marker	Chr ^a	Pos (Mb) ^b	Known trait ^c	Forward primer	Backward primer	Remark	Type
SCM2	RM20547	6	27.0	Culm strength	CTCTTCTTCTGTCCTCTTGG	CCATCTTCATTACCGACCTCTGC	Cloned	SSR
DEP2	RM21964	7	25.3	Dense, erect panicle	AAGACAGCCTTCAAGGGATTGG	GTACGTGCACCGAGCAGAGC	Cloned	SSR
D88	RM6742	3	5.4	Dwarfness	CCAAGTTATCCAAGCTTCGTTTCG	AGAACGACCTTTCGAGGGAGAGC	Fine mapped	SSR
D10	RM3411	1	31.3	Dwarfness and TN	CGTCTCCAGATGGTCCAC	ATGGGACTCCCGTACTCTCTC	Cloned	SSR
htd1	RM17307	4	27.3	Dwarfness, TN	AGAGCTTGGAGGCACCAATACCG	AGAAAGAACTCCGGCCACCTTCG	Cloned	SSR
DWL1	HL921	3	34.9	Dwarfness, WLT	ATGGCTTCAGACTTCAGAGT	CAAATTAACCTTCAGGCAAG	Fine mapped	STS
DWL1	HL944	3	35.0	Dwarfness, WLT	TGCCAGCTAGCGAGCCTAA	ATTGCAGCGAGCTACACG	Fine mapped	STS
ep3	STS5803.7	2	9.0	Erect panicle, SB	GAATGGATGGATGGATCGAG	GTAGGGTCCGGCGAGATATG	Cloned	STS
GIF1	RM16942	4	20.4	GIF, TGW	CCAGTACTCTCGTCCACTCTCC	ATCGCTTTCACGTCACCAAGG	Cloned	SSR
qGL7	RID711	7	28.5	GL, TGW, SN	GCACATGCATGCTAGGACAT	AGCCGGTAAATTTCTTGACAC	Fine mapped	InDel
qGL7-2	Indel1	7	24.6	GL, TGW, SN	CCATAGTAAGACGACCTT	GATATTCTGTGACGAGTT	Fine mapped	InDel
qGL7-2	RM21945	7	24.9	GL, TGW, SN	CTACACAAGTGAAGGCCATCAGG	GTTCTAGGGTGTCTTTCATGAGC	Fine mapped	SSR
Gn1a	RM3604	1	5.1	GN, GY	CAGGAACCAACCTTCTCTTGACC	GTCAGACTCCGATCTGGGATGG	Cloned	SSR
Gn1a	RM10316	1	5.3	GN, GY	AAGATCGCTGGGAGATCTGTAGG	GCATGCTAATTAGTCAGCCTTGG	Cloned	SSR
Gnp4	Y48	4	19.6	GN	TCACCATATGGAAGCATCAAG	TATGTGTTTGTTCATGTGCAC	Cloned	SSR
qGN4-1	nkssr04.19	4	31.3	GN, PB, SB, PN	CTGGAATCACAAACCACGAC	GCTACCTCAAGCTCCACGAC	Fine mapped	SSR
qGN4-1	RM3276	4	30.7	GN, PB, SB, PN	TCCGTCTCGACTCTTCCATC	GATGAGACACACCGGACATG	Fine mapped	SSR
GS3	RGS1	3	16.5	GS	TCCACCTGCAGATTTCTTCC	GCTGGTCTTGACATCTCTCT	Cloned	SSR
GS5	C62	5	3.4	GS	GATTGACTGATAAATGACAGC	CTAAGTCCCATGGAATTAC	Cloned	SSR
GS5	RM574	5	3.4	GS	AAACTAGCCACGGTTTGGTAGGG	AGGGTGGCAGGGATGTAATTTCC	Cloned	SSR
gw5	RMw513	5	5.3	GW, GL	GTATTTGTTTGTGCGATTTC	TAGGACCATAGATGTGAGTTA	Cloned	SSR
qGY2-1	RM279	2	2.9	GY	CCTCTCACTCAGTGGACTCTCC	CCTCACCTAGGCTTTGATATGC	Fine mapped	SSR
Ghd8	RM22483	8	4.3	GY, HD, PH	GACCATGGTGTGAGTGTGACAGG	CAAGTCTACCTCAACCGCTACC	Cloned	SSR
Ghd7	RM5436	7	9.1	HD, GY, PH, GN	CAAGGGGGTGTCTCTATG	GTTGCTCGTCTACATGTGC	Cloned	SSR
gw9.1	RM24718.CNR111	9	20.9	HD, GY, SN, GN, PH	GACCAACGTGCATGTGACTT	GCTTGCACTAGGGCTCCTT	Fine mapped	SSR
IPA1	RM23422	8	25.2	IPA, TN, GY	GTCGGTCACGAAGTTCAGATCC	TCAGGCAAAAGTTGAAGATGGTAGC	Cloned	SSR
Nop(t)	M9	6	24.0	Non-panicle	AATGAATAGATTACCACATGCTA	TGCCTCTTATTTTACTTTCTTT	Fine mapped	STS
qPH6-1	RM3414	6	2.9	PH	TAGGGCAATTGTGCAAGTGG	TTGGGAATTGGGTAGGACAG	Fine mapped	SSR
qPH6-1	RM19417	6	3.0	PH	CGGTCCTTGAATCTTCTGC	GGTGACGCTTATGAATGTACCG	Fine mapped	SSR
RPH	RM481	7	NA	PH	TAGTAGCCGATTGAATGGC	CTCCACCTCTATGTTGTG	Fine mapped	SSR
d27	RM26985	11	21.4	PH, TN	CACAAGACAACCTTCAATGG	GGCTTAGGAGCGTTTATAGG	Cloned	SSR
qSH3	RM16	3	22.9	Seed shattering	GTGGCCAGGAGTAGTTGTCTCC	GACGTGTACACATAGCCAAATCATCC	Fine mapped	SSR
PAP2	RM15937	3	30.9	SN	GGAAGAACCTGCGTATCAAGACC	CCACACGGAAGCAGAATTAGCC	Cloned	SSR
PAP2	RM15948	3	31.2	SN	GAGAGTGGGTGAGAAGGATCAGC	ACAGCAGATTGTGTGCATCTGG	Cloned	SSR
qPDS3	RM14820	3	10.8	SN	AGGTCGTGATGTCCCTCTCC	AAACCATCGTGGCATCATCTCC	Fine mapped	SSR
qSPP7	RM5499	7	10.0	SN	GGACGAAAGGGTATTGTGATTGG	CCTCAAGGTGGTCTCTCTCTCC	Fine mapped	SSR
SPP1	YN27	1	6.5	SN	TACCACTGAACCCACGTGTC	GCTGCCCTTTGTCTCACGTT	Fine mapped	InDel
Gw1-1	RM10376	1	6.3	TGW	TTAGTTTAACCCGACCGTACACC	GGTCGTGAATTGGTGTCAAGC	Fine mapped	SSR
Gw1-2	RM1344	1	7.0	TGW	CTGCAATCCGAGTAGGAAGC	TGAGCATTCACCTCCGATCTG	Fine mapped	SSR
GW3	RM15206	3	16.8	TGW	CATTCTTTTGGCTCGATCTTTC	AAGCGCCATAATCCAGGAACC	Fine mapped	SSR
gw3.1	JL109	3	16.6	TGW	TGGAGCTGTGGAAGTACTGGA;	TCCTGTAGCTACCTGTCTAT	Fine mapped	SSR
GW6	RM20201	6	20.2	TGW	TTAGAGGTAACGGAGGCACAACC	GATGGCTTGAGAGCGTTTGTAGG	Fine mapped	SSR
gw8.1	RM23201.CNR151	8	21.5	TGW, GL	GTTCTTTCCGGTGACGAGATC	CGCTGCAGATGACGAGATAC	Cloned	SSR
GW2	RM12827	2	8.1	TGW, GW	GCTCTGGCACCGAGATTATTATAGC	GAGAGACTGCGACCTCTGTAGCC	Cloned	SSR
Ltn	ssr0649.23	8	25.3	TN	TCATCCTTAAGACGGTATCACA	CTCCCTCTCCGTTTCATATTC	Fine mapped	SSR
MOC1	RM20373	6	23.9	TN	GACGACGTGTGTTGACTTCTGC	CCAGTTCACCAACACAATGAGC	Cloned	SSR

^a chr, chromosome.^b Pos, position, the physical position in Mb.^c Known trait, traits which are controlled by the target genes/QTLs. GIF, grain incomplete filling; GL, grain length; GN, grain number per panicle; GS, grain size; GW, grain width; GY, grain yield; HD, heading date; IPA, ideal plant architecture; PB, primary branch number per panicle; PH, plant height; PN, panicle number; SB, secondary branch number per panicle; SN, spikelet number per panicle; SR, seed setting rate; TGW, thousand grain weight; TN, tiller number; WLT, withered leaf tip.

For DNA extraction the young leaves from a single plant were collected, frozen in liquid nitrogen, and stored at -80°C . DNA isolation was conducted using a cetyltrimethyl ammonium bromide method described by Murray and Thompson (1980). DNA concentration was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Rockland, DE, USA). For PCA we followed the protocol described by Santos et al. (1993) with minor modification. DNA was diluted to $15\text{ ng }\mu\text{L}^{-1}$. Amplification reactions were carried out in $10\text{ }\mu\text{L}$ containing 10 mM Tris-HCl pH 8.3, 50 mM KCl , 1.5 mM MgCl_2 , $50\text{ }\mu\text{M}$ each of dNTPs, $0.2\text{ }\mu\text{M}$ each of primers, $0.5\text{ U Taq polymerase}$ (Qiagen or homemade) and $10\text{--}30\text{ ng}$ of DNA template. PCR reactions were performed using PTC-200 thermal cycler (MJ Research Inc.) or G-Storm thermal cycler. The PCR conditions were as follows: 94°C for 5 min, followed by 35 cycles of 30 s at 94°C , 30 s at 55°C , and 45 s at 72°C with a final extension of 7 min at 72°C . The PCR amplification products were separated by running on 8% non-denaturing polyacrylamide gels

and observed under the Alphamager (Proteinsimple Inc., California USA) by SYBR® Safe staining method.

2.3. Statistical analysis

2.3.1. Phenotypic analysis

All trials were separately analyzed by fitting an appropriate spatial model with rows and columns using PBTools (bbi.irri.org) and R (R Core Team, 2015). The best linear unbiased estimations (BLUE) from the best-fit model were used as input data for association analysis. A detailed analysis of phenotypic data has been reported by Liang et al. (2015).

2.3.2. Diversity statistics, population structure and kinship

Polymorphism information content (PIC), gene diversity, number of alleles per locus and heterozygosity were computed using PowerMarker version 3.25 (Liu and Muse, 2005). STRUCTURE 2.3.4

(Falush et al., 2003; Pritchard et al., 2000) was used to analyze population structure using 53 random markers. The program was run with different values of the number of clusters (k) from one to ten using the admixture model with correlated allele frequency. Four separate groups of 20 runs were carried out for each K with a burning period of 5000 followed by 100,000 Markov Chain Monte Carlo (MCMC) repeats. The most likely number of subpopulations was determined using the Delta K method proposed by Evanno et al. (2005). For the optimal k value, the Q matrix (population membership estimates) was extracted from STRUCTURE run. To display the familial relationship, a heatmap of pairwise kinship of the 360 lines was built up in R (R Core Team, 2015).

2.3.3. Association analysis

Rare marker alleles, occurring at a frequency less than 5% among all the genotypes for any marker were excluded. BLUE obtained from the phenotypic analysis was used as phenotypes for each trait and line. Kinship matrix (K) calculation and association analysis were conducted using TASSEL v.2.0.1 (Bradbury et al., 2007). To choose the most appropriate model for association analysis four models implemented in TASSEL were compared using the 53 random markers for selected trait-environment combinations. The four models were (i) naïve model, the general linear model (GLM) without controlling population structure and relatedness; (ii) Q model, GLM including Q population structure coefficients as covariates; (iii) K model, the mixed linear model (MLM) including the K matrix to account for the genetic relatedness between genotypes; (iv) QK model, the MLM including both Q matrix and K matrix as correction for population structure and genetic relatedness between genotypes. The selected trait-environment combinations were GY in four testing environments including DS2, JX, SC and WS2, and four traits including DTF, GN, GY and TGW in WS2. Association analysis was also carried out within each of the subpopulations. Separate analysis was carried out for all measured traits in each of the eight testing environments and the average environment defined as the average across the testing environments.

To declare significant associations between markers and traits, positive false discovery rate (pFDR; q value) was calculated using the QVALUE in R (Storey, 2002; Storey and Tibshirani, 2003). Since the target markers had been reported to be closely linked to GY or yield related traits and some of these genes have been cloned and validated, a q value cut off of 0.05 was used.

2.3.4. Prediction of GY

Stepwise multiple linear regression (MLR) analysis was used to select an optimal model for predicting GY in each of the testing environments and the average environment. Two sets of markers were used. They were all the 46 target markers and the markers significantly associated with GY or yield related traits identified in each environment using the K model in the present study. To evaluate the relative contribution of each marker and develop the prediction model for GY (Y), the following formula was adopted.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$$

where Y is the GY, α is the intercept, β_n is the coefficient, X_n is the marker effect.

MLR were conducted using the MASS package in the R (R Core Team, 2015). The stepwise program computed a sequence of MLRs in a stepwise manner. The marker with least AIC value was removed from the regression at each step. The adjusted R^2 of the final model was estimated accordingly.

3. Results

3.1. Population structure, genetic diversity and genetic relationship

Population structure analysis indicated that the most likely number of subpopulations was two based on the change of K . Subpopulations 1 and 2 consisted of 155 and 205 lines, respectively (Fig. 1). Subpopulation 1 was dominated by IRRI materials with one line from AfricaRice and Pakistan, respectively, three lines from Indonesia, and four lines from Vietnam. Subpopulation 2 consisted of one line from Pakistan, three lines from AfricaRice, India, Indonesia, respectively, six lines from Bangladesh, and Vietnam, respectively, 14 lines from China, 19 lines from CIAT, 25 lines from PhilRice and the remaining 125 lines from IRRI. In summary, subpopulations 1 and 2 had 85.8% and 61.0% lines from IRRI breeding programs, respectively.

PIC, gene diversity, the number of alleles per locus and heterozygosity in the whole population and within subpopulations are summarized in Table 2. For the whole population, a total of 388 alleles were obtained from the 99 loci including the target and random markers scored for the 360 genotypes, with an average 3.919 alleles per locus varying from two to 11. The average PIC value was 0.3787, ranging from 0.0273 to 0.7826. The average diversity index was 0.4181, ranging from 0.0275 to 0.8079. The average heterozygosity was 0.0058, ranging from 0 to 0.0611, while 65 loci were homozygous. Within subpopulation 1, a total of 325 alleles were detected from the 99 loci, ranging from two to seven alleles per genetic marker, with an average of 3.283 alleles per locus. The corresponding PIC values ranged from 0.0127 to 0.7754, with an average of 0.3399. The genetic diversity at each marker ranged from 0.0128 to 0.8024 with an average of 0.3778. The average heterozygosity was 0.0060, ranging from 0 to 0.0710, while 71 loci were homozygous. For subpopulation 2, a total of 378 alleles were obtained, with an average 3.818 alleles per locus varying from 2 to 8. The corresponding PIC values was 0.3873, varying from 0.0284 to 0.7824. The genetic diversity at each marker ranged from 0.0288 to 0.8077 with an average of 0.4262. The average heterozygosity was 0.0056, ranging from 0 to 0.0732, while 70 loci were homozygous. The genetic diversity within the two subpopulations was consistent with the population structure results. The subpopulation 2 consisted of lines with more diverse origins.

The kinship coefficient among the 360 lines ranged from 0 to 1.80, with a mean of 0.49. There were about 2.33% unrelated genotype pairs (kinship = 0), 6.96% distantly related genotype pairs with kinship being lower than 0.10 and 0.03% highly related genotype pairs with kinship being higher than 1.50 (Fig. 2a). The majority of the genotype pairs had kinship similar to half-sibs (kinship = 0.5). Generally, the relationship within a breeding population was greater than among the breeding populations. The unequal familial relatedness among genotypes could be easily seen in the heatmap (Fig. 2b).

3.2. Association analysis

3.2.1. Comparison between models

To choose an appropriate model for association analysis, the associations between the 53 random markers and the four traits including DTF, GN, GY, and TGW were analysed using four analytical models including the naïve, Q , K and QK models. Model comparison was made using the quantile–quantile (QQ) plot, which is based on the observed and expected P values. The ideal distribution of P values should follow a uniform distribution with less deviation from the expected P values (Kang et al., 2008). Fig. 3 gives the QQ plots for the associations between markers and GY measured in DS2, JX, SC and WS2 using the four models. It showed that P values of both

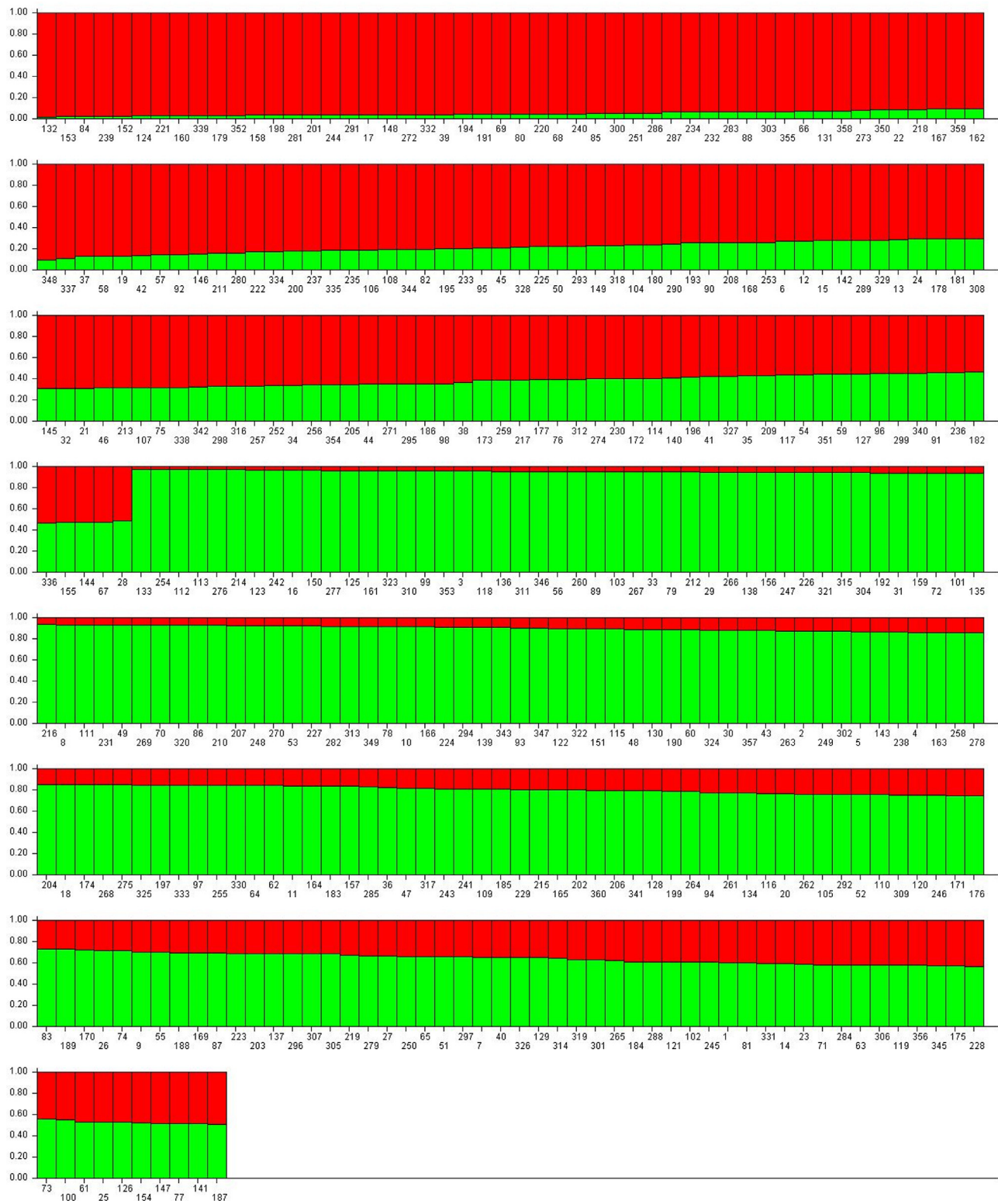


Fig. 1. Population structure of 360 *indica* rice genotypes analysed with 53 unlinked SSR markers using the admixture model implemented in STRUCTURE. Two different sub-populations are indicated by different colors. Each genotype is represented by a vertical bar broken into colored segments with lengths proportional to each of the *K* inferred ancestral populations. The number at the bottom of each vertical bar corresponds to the code of rice genotype (1–360).

Table 2
Diversity and polymorphism information for the whole population and the two subpopulations.

Group	No. of lines	Gene diversity	Heterozygosity	PIC	Average allele/locus
Whole population	360	0.4181	0.0058	0.3787	3.919
Subpopulation1	155	0.3778	0.0060	0.3399	3.283
Subpopulation2	205	0.4262	0.0056	0.3873	3.818

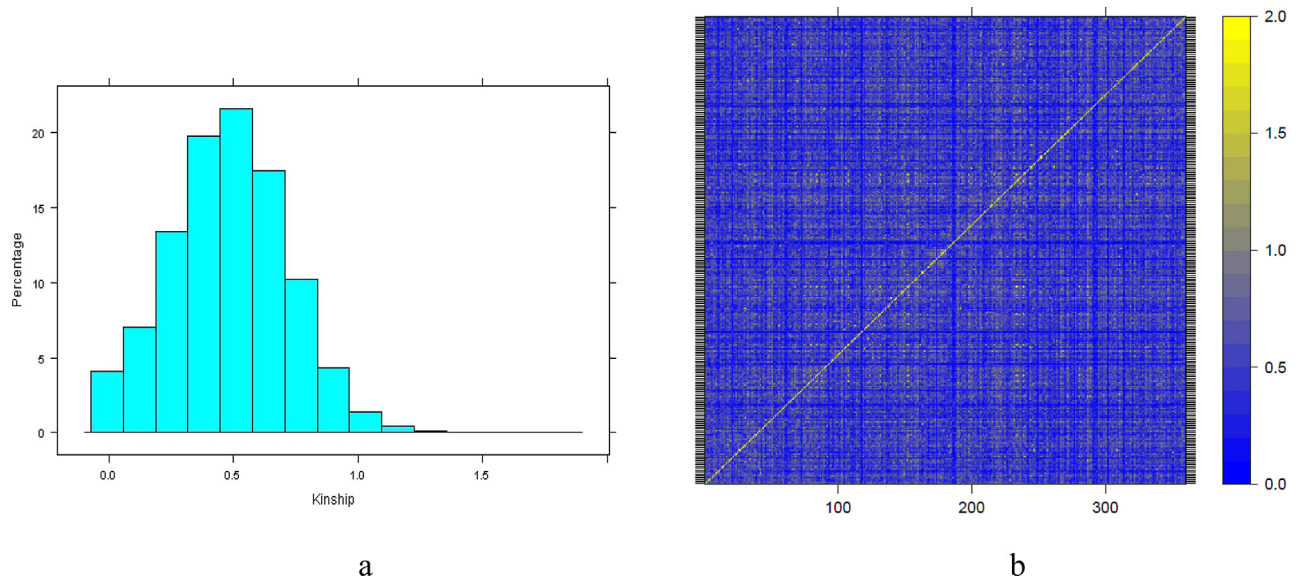


Fig. 2. Genetic relatedness between 360 *indica* lines analyzed using 53 random markers. (a) Distribution of the pairwise kinship coefficients estimated using 53 random markers. (b) Heatmap of the estimated pairwise kinship coefficients.

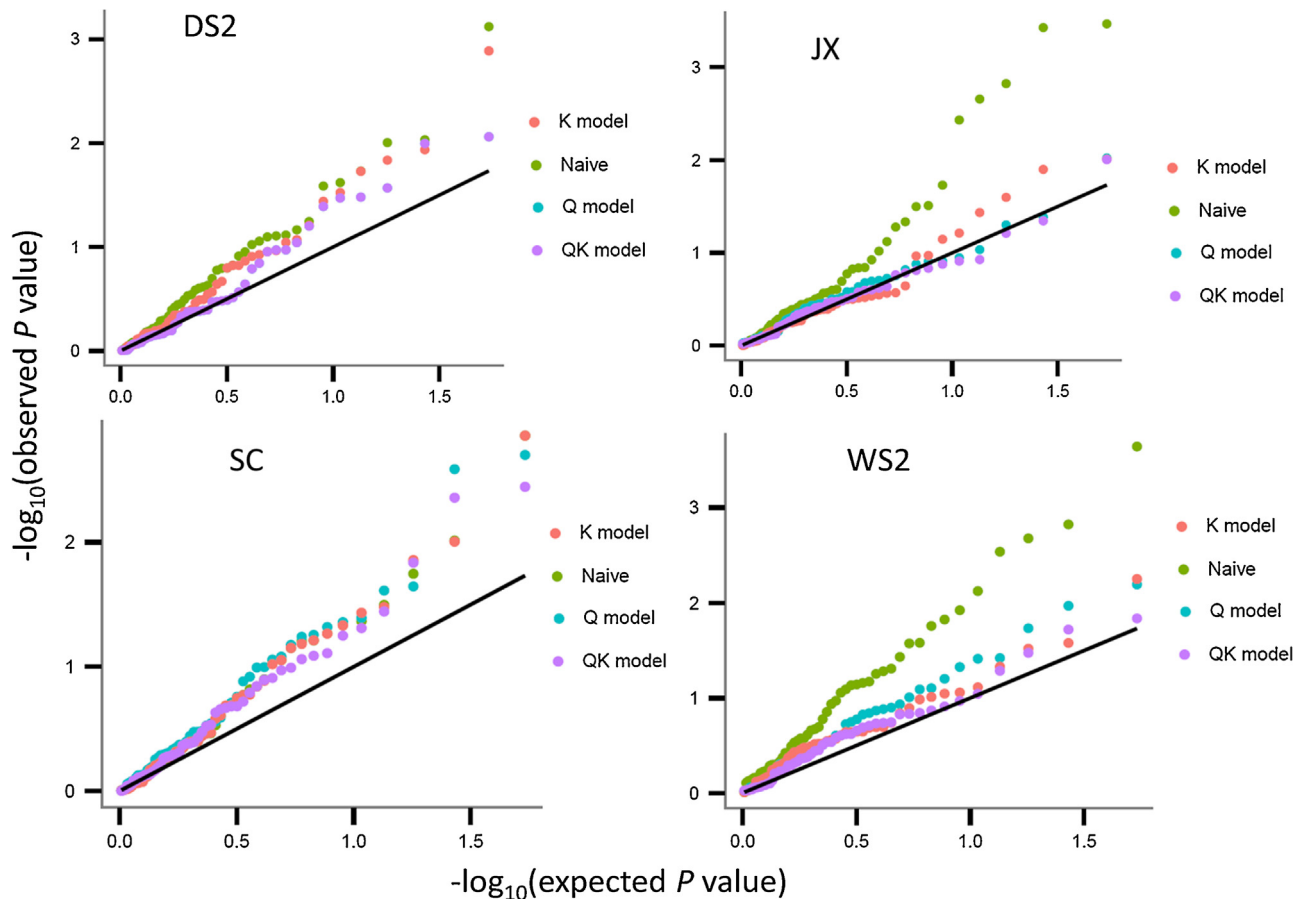


Fig. 3. Log quantile–quantile (QQ) P value plot for naïve, Q, K and QK association models for GY measured in DS2, JX, SC and WS3.

naïve and Q models deviated from the $Y=X$ line, especially in JX and WS2. In contrast, the K and QK models showed an improved fit for P values and the two models behaved similarly. Fig. 4 gives the QQ plots for the associations between markers and DTF, GN, GY and TGW measured in WS2. The P values were greatly inflated

for the naïve and Q models. For DTF, GY and TGW, the K and QK models behaved similarly, while for GN, the QK model had a better fit than the K model. The result suggested that mixed linear models using either K alone or both of Q and K sufficiently accounted for structure and familial relatedness among the genotypes in our

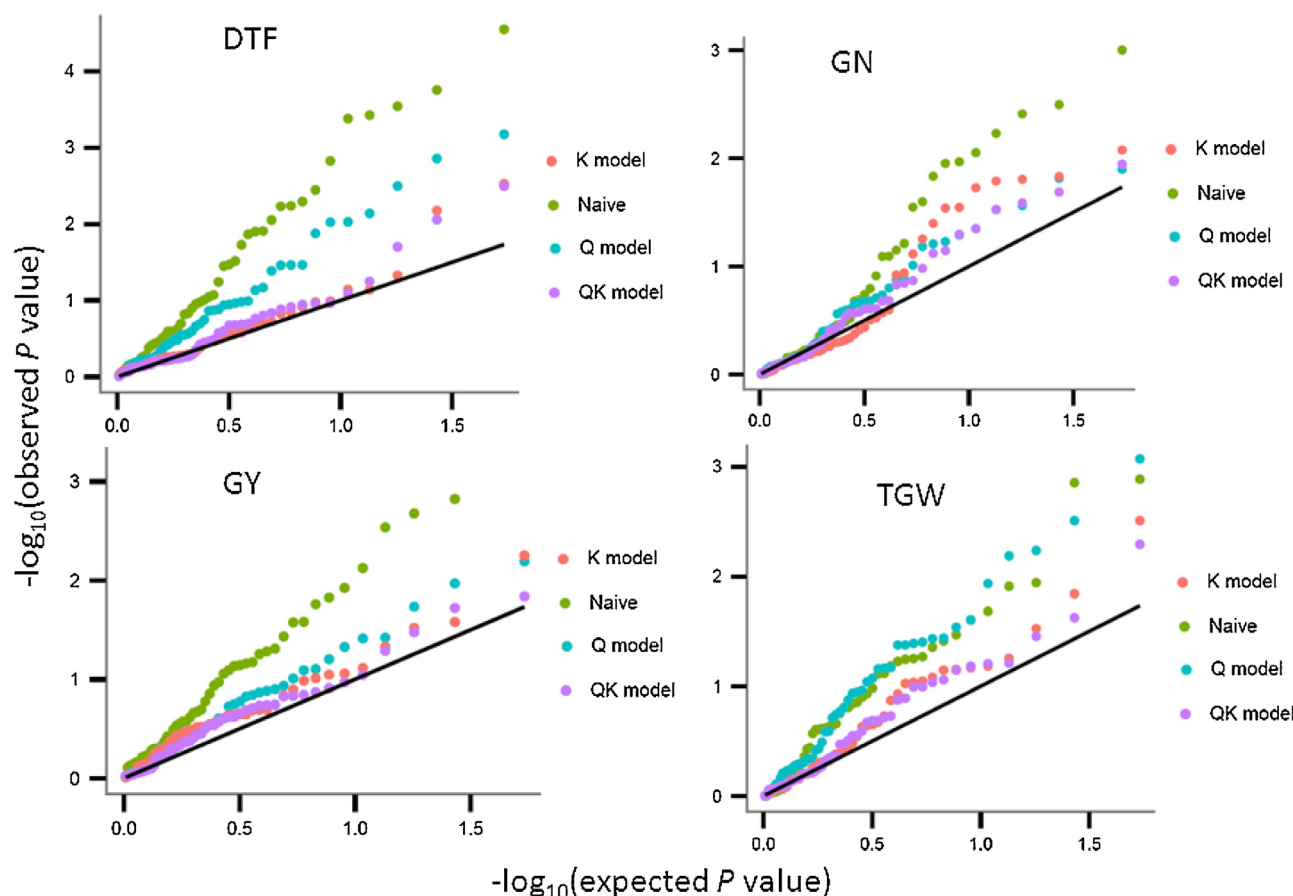


Fig. 4. Log quantile–quantile (QQ) P value plots for naïve, Q, K and QK association models for DTF, GN, GY and TGW measured in WS2.

population. Considering the insignificant difference between the K and QK model and the likelihood that the QK model might be too stringent, only the results from the K model were presented.

3.2.2. Association between genes/QTL and traits

Based on the K model applied to the whole population and the two subpopulations, all the 39 genes/QTL were found to be associated with two or more traits in one or more testing environments. A brief summary of significant associations between the genes/QTL and the 11 traits is presented in Table 3. GW6 was associated nine traits while *ep3*, *gw8.1*, *gw9.1*, and *qPDS3* were associated with only two traits (Table 3). All the 39 target genes/QTL were associated with one or more traits in the average environment (Table 3). The number of significant gene/QTL trait associations (GTA) identified varied greatly across the eight testing environments ranging from seven (DS2) to 34 (DS3). The two seasons were similar in the total number of significant GTA identified. Nitrogen fertilizer rate had different effects on the number of GTA in the DS and WS. In the DS, DS3 (high nitrogen rate) had the highest number of GTA (34), followed by DS1 (no nitrogen) (eight) and DS2 (medium nitrogen level) (seven). In the WS, the highest number of GTA (20) was found under the WS1 (no nitrogen), followed by WS2 (low nitrogen rate) (14) and WS3 (high nitrogen rate) (ten).

A total of 16 genes/QTL were found to be associated with GY (Table 4). Ten genes/QTL were found significant using the whole population (Table 4). GS3 was significant in two testing environments as well as the average environment. The remaining genes/QTL were significant in only one testing environment. Five and seven genes/QTL were significant in the subpopulation 1 and subpopulation 2, respectively (Table 4). These included *GIF1*, GS3, *Gw1-1*, *gw3.1*, *qPH6-1*, and *qSPP7* that were also found significant

using the whole population. *d27* was associated with GY in the subpopulation 1 in two testing environments, while the remaining genes/QTL were in only one testing environment.

Twenty five genes/QTL were found to be associated with DTF (Table 5). Seventeen genes/QTL showed significant associations with DTF in the whole population (Table 5). GS3 was associated with DTF in five out of the eight testing environments as well as the average environment. *Ltn* and *Ghd7* were associated with DTF in four and three testing environments, respectively. GS5 was associated with DTF in three testing environments and the average environment. *GIF1* and GW2 were associated with DTF in two testing environments. *gw3.1* and *qGY2-1* were associated with DTF in one testing environment and the average environment. The remaining genes/QTL were significant only in one testing environment or the average environment. A total of 23 genes/QTL showed significant association with DTF in the subpopulation 2 (Table 5). Eight of the genes/QTL including *DEP2*, *DWL1*, *Gn1a*, GW3, *gw5*, GW6, *htd1* and *qPDS3* was not associated with DTF in the whole population. No significant associations were found using the subpopulation 2. GS3 was significant in five testing environments and the average environment. *GIF1* and *qSPP7* were significant in four testing environments and the average environment. *Ghd7* was significant in three testing environments and the average environment. *Gn1a* was significant in four testing environments. GS5 and *Ltn* were significant in three testing environments, while *D88* and *qPH6-1* were in two testing environments. *RPH* was significant in one testing environment and the average environment. The remaining genes/QTL were significant in only one testing environment.

All the 39 target genes/QTL were associated with PH (Table 6). The associations between six genes/QTL and PH were found in one environment using the whole population, which were *Ghd8*, *Gnp4*,

Table 3

Associations between 39 target genes/QTLs and GY and related traits assessed in eight testing environments and the average environment using *K* model in the whole population and subpopulations.

Gene/QTL	DTF	GN	GY	PB	PH	PN	SB	SN	SR	TGW	TN	No. of traits ^a
<i>D10</i>	0 ^b	0	1	1	2	0	0	0	1	0	0	4
<i>d27</i>	1	0	2	1	1	0	0	0	0	0	0	4
<i>D88</i>	2	0	0	1	1	0	0	0	0	0	0	3
<i>DEP2</i>	1	0	0	0	1	0	0	1	1	0	0	4
<i>DWL1</i>	1	0	0	1	1	2	0	0	0	0	0	4
<i>ep3</i>	0	0	0	0	1	0	0	0	0	0	1	2
<i>Ghd7</i>	4	4	0	0	1	0	3	5	2	3	0	7
<i>Ghd8</i>	1	0	0	1	2	0	0	0	0	0	0	3
<i>GIF1</i>	5	3	1	1	1	0	0	0	2	0	0	6
<i>Gn1a</i>	4	4	1	1	1	1	5	5	0	0	0	8
<i>Gnp4</i>	0	1	1	1	2	0	0	0	0	0	0	4
<i>GS3</i>	7	0	3	1	1	0	0	0	0	5	0	5
<i>GS5</i>	3	0	0	1	1	0	0	0	0	0	0	3
<i>Gw1-1</i>	0	3	2	1	2	0	0	0	0	0	0	4
<i>Gw1-2</i>	0	1	0	0	1	1	0	0	0	0	0	3
<i>GW2</i>	2	2	0	1	1	0	0	1	0	0	0	5
<i>GW3</i>	1	0	0	1	1	0	0	0	0	3	0	4
<i>gw3.1</i>	2	0	1	1	2	0	0	0	1	0	0	5
<i>gw5</i>	1	0	0	1	2	0	0	0	0	2	0	4
<i>GW6</i>	1	1	0	1	1	1	0	1	1	6	2	9
<i>gw8.1</i>	0	0	0	1	1	0	0	0	0	0	0	2
<i>gw9.1</i>	0	0	0	1	1	0	0	0	0	0	0	2
<i>htd1</i>	1	0	1	1	2	0	0	0	0	1	0	5
<i>IPA1</i>	1	0	0	1	1	0	0	0	0	0	0	3
<i>Ltn</i>	5	3	0	3	2	0	0	2	1	0	0	6
<i>MOC1</i>	0	0	0	1	1	0	0	0	1	0	1	4
<i>Nop(t)</i>	0	1	1	1	1	0	0	2	0	0	0	5
<i>PAP2</i>	0	0	1	1	1	0	0	0	2	0	0	4
<i>qGL7</i>	0	1	1	0	1	0	0	0	1	0	0	4
<i>qGL7-2</i>	0	1	1	0	1	0	0	0	0	1	0	4
<i>qGN4-1</i>	0	0	1	2	1	0	0	0	0	1	0	4
<i>qGY2-1</i>	2	0	0	2	2	0	0	1	0	0	2	5
<i>qPDS3</i>	1	0	0	0	1	0	0	0	0	0	0	2
<i>qPH6-1</i>	2	1	1	0	2	0	0	0	0	6	0	5
<i>qSH3</i>	1	0	0	0	1	0	1	0	0	0	0	3
<i>qSPP7</i>	5	4	1	0	1	0	3	5	2	0	0	7
<i>RPH</i>	3	0	0	1	1	0	0	0	0	0	0	3
<i>SCM2</i>	2	2	0	1	1	0	1	4	0	1	0	7
<i>SPP1</i>	0	2	0	1	2	1	2	3	0	0	1	7
Whole population ^c	17	13	10	28	6	4	4	8	7	10	1	37
Subpopulation ^d	23	11	12	4	39	1	6	10	8	3	5	39
Total ^e	25	16	16	29	39	5	6	11	11	10	5	39

^a Number of traits which the target gene/QTL was associated with.

^b Number of environments where the association between the target gene/QTL and trait were detected.

^c Number of target genes/QTLs associated with a trait in the whole population.

^d Number of target genes/QTLs associated with a trait in the subpopulations.

^e Total number of target genes/QTLs associated with a trait.

Table 4

Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and GY identified in the whole population of and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulations	
		Env ^a	<i>q</i>	Env (Sub) ^b	<i>q</i>
<i>D10</i>	RM3411	– ^c	–	JX (1)	0.001
<i>d27</i>	RM26985	–	–	DS3 (1), JX (1)	0.005–0.027
<i>GIF1</i>	RM16942	SC	0.004	SC (2)	0.033
<i>Gn1a</i>	RM10316	–	–	DS2 (2)	0.037
<i>Gnp4</i>	Y48	SC	0.026	–	–
<i>GS3</i>	RGS1	Ave, DS1, DS3	0.0002–0.044	Ave (1), DS1 (2), WS2 (2), DS1 (4)	0.014–0.047
<i>Gw1-1</i>	RM10376	SC	0.004	JX (1)	0.006
<i>gw3.1</i>	JL109	DS1	0.021	DS1 (2)	0.014
<i>htd1</i>	RM17307	SC	0.008	–	–
<i>Nop(t)</i>	M9	SC	0.029	–	–
<i>PAP2</i>	RM15937	–	–	JX (1)	0.013
<i>qGL7</i>	RID711	SC	0.013	–	–
<i>qGL7-2</i>	Indel1	–	–	JX (1)	0.003
<i>qGN4-1</i>	nkssr04.19	–	–	DS1 (2)	0.031
<i>qPH6-1</i>	RM19417	SC	0.012	–	–
<i>qPH6-1</i>	RM3414	SC	0.010	SC (2)	0.033
<i>qSPP7</i>	RM5499	SC	0.010	SC (2)	0.033

^a Env indicates the environments where significant associations were found.

^b Sub indicates significant association was found in the subpopulation.

^c Indicates association was not significant and thus *q* was not given.

Table 5Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and DTF identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulations	
		Env ^a	q	Env (Sub) ^b	q
<i>d27</i>	RM26985	DS2	0.035	– ^c	–
<i>D88</i>	RM6742	JX	0.032	JX (2), WS1 (2)	0.047–0.048
<i>DEP2</i>	RM21964	–	–	WS1 (2)	0.037
<i>DWL1</i>	HL944	–	–	WS1 (2)	0.038
<i>Ghd7</i>	RM5436	Ave, WS1, WS2, WS3	0.001–0.040	Ave (2), WS1 (2), WS2 (2), WS3 (2)	0.005–0.035
<i>Ghd8</i>	RM22483	DS3	0.010	DS3(2)	0.043
<i>GIF1</i>	RM16942	DS2, DS3	0.015–0.017	Ave (2), DS1 (2), DS2 (2), DS3 (2), WS1 (2)	0.006–0.036
<i>Gn1a</i>	RM3604	–	–	SC (2), WS1 (2), WS2 (2), WS3 (2)	0.018–0.045
<i>GS3</i>	RGS1	DS1, DS2, DS3, WS2, WS3	0.004–0.024	Ave (2), DS1 (2), DS2 (2), DS3 (2), WS1 (2), WS2 (2)	0.011–0.045
<i>GS5</i>	RM574	DS1, DS2, DS3	0.008–0.041	DS1 (2), DS2 (2), DS3 (2)	0.010–0.024
<i>GW2</i>	RM12827	DS2, DS3	0.010–0.041	DS3 (2)	0.043
<i>GW3</i>	RM15206	–	–	JX (2)	0.047
<i>gw3.1</i>	JL109	Ave, WS1	0.033–0.040	WS1 (2)	0.027
<i>gw5</i>	RMw513	–	–	WS1 (2)	0.027
<i>GW6</i>	RM20201	–	–	WS1 (2)	0.048
<i>htd1</i>	RM17307	–	–	WS1 (2)	0.049
<i>IPA1</i>	RM23422	DS3	0.030	–	–
<i>Ltn</i>	ssr0649.23	Ave, SC, WS1, WS2, WS3	0.033–0.050	WS1 (2), WS (2), WS3 (2)	0.027–0.045
<i>qGY2-1</i>	RM279	Ave, WS1	0.033–0.040	WS1 (2)	0.027
<i>qPDS3</i>	RM14820	–	–	WS1 (2)	0.027
<i>qPH6-1</i>	RM19417	JX	0.045	WS1 (2), JX (2)	0.027–0.041
<i>qSH3</i>	RM16	JX	0.008	JX(2)	0.042
<i>qSPP7</i>	RM5499	Ave	0.033	Ave (2), JX (2), WS1 (2), WS2 (2), WS3 (2)	0.035–0.048
<i>RPH</i>	RM481	WS3	0.033	Ave (2), WS1 (2)	0.027–0.035
<i>SCM2</i>	RM20547	JX	0.045	WS1 (2)	0.027

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.

gw3.1, *htd1*, *Ltn* and *qPH6-1* (Table 6). All the 39 target genes/QTL were significant in the subpopulation 2 while only *D10* was found in the subpopulation 1. *Ghd8*, *Gw1-1*, *gw5*, *qGY2-1* and *qPH6-1* were significant in one testing environment and the average environment, while the remaining genes/QTL were associated with PH only in the average environment.

A total of 16 genes/QTL were associated with GN (Table 7). Thirteen genes/QTL were found significant using the whole population (Table 7). *Gn1a*, *Ghd7* and *qSPP7* were significant in four out of the eight testing environments. *GIF1* and *Ltn* were significant in two testing environments and the average environment. *SCM2* and *SPP1* were significant in two testing environments. *Gw1-1* and *GW2* were significant in one testing environment and the average environment. The remaining genes/QTL were associated with GN in one testing environment or the average environment. Five genes/QTL including *Gnp4*, *Gw1-1*, *qGL7*, *qGL7-2* and *qSPP7* were found to be associated with GN in the subpopulation 1 and eight genes/QTL including *Ghd7*, *GIF1*, *Gw1-1*, *GW2*, *Ltn*, *Nop(t)*, *qPH6-1* and *qSPP7* in the subpopulation 2. Eight of the 11 genes/QTL except *Gnp4*, *qGL7*, and *qGL7-2* were among those found using the whole population. *Gw1-1* and *GIF1* were significant in two testing environments as well as the average environment. *GW2* and *Ltn* were significant in one testing environment and the average environment in the subpopulation 2. The remaining genes/QTL were significant only in one testing environment.

Eleven genes/QTL were associated with SN (Table 8). Eight genes/QTL were significant in the whole population (Table 8). *Gn1a* was significant in five out of the eight testing environments as well as the average environment. *Ghd7* and *qSPP7* were significant in four testing environments and the average environment. *SCM2* and *SPP1* were associated with SN in the average environment and three and two testing environments, respectively. *Ltn* was significant in one testing environment and the average environment, while *GW2* and *Nop(t)* were only in one testing environment. A total of ten genes/QTL were found to be associated with SN in the subpopulation 2. Except *DEP2*, *GW6* and *qGY2-1* the other genes showed

significant association in the whole population. *SCM2* was significant in two testing environments and the average environment while *Ghd7* was in two testing environments. *Gn1a*, *Ltn* and *qSPP7* were significant in one testing environment and the average environment. The remaining genes/QTL were significant in only one testing environment.

Only five genes/QTL were associated with PN (Table 9). Four were found using the whole population and one using the subpopulation 1. Using the whole population *DWL1* was associated with PN in DS2 and the average environment, *Gn1a* and *Gw1-2* in DS3 and *GW6* in DS2. *SPP1* was associated with PN in WS2 using the subpopulation 1.

Only *GW6* was found to be associated with TN using the whole population (Table 10). Analysis in each of the two subpopulations found four additional associated genes/QTL (Table 10). *qGY2-1* was associated with TN in WS1 and the average environment in the subpopulation 2, while *ep3*, *GW6*, *MOC1* and *SPP1* were associated with TN in only one testing environment in the subpopulation 1.

Ten genes/QTL were found to be associated with TGW using the whole population (Table 11). *qPH6-1* was significant in five out of the eight testing environments as well as the average environment. *GS3* and *GW6* were significant in four testing environments and the average environment. *Ghd7* and *GW3* were significant in two testing environments and the average environment. *gw5* was associated in WS1 and the average environment. *htd1*, *qGL7-2*, *GN4-1* and *SCM2* were associated in only one testing environment. Three genes/QTL were found to be associated with TGW in the subpopulation 2 (Table 11). *GW6* was significant in three testing environments. *qPH6-1* was significant in SC and WS3. *GS3* showed significant association only in WS3.

A total of 29 genes/QTL were found to be associated with PB (Table 12). Twenty-eight, and four genes/QTL were found in the whole population and the subpopulation 2, respectively, while no significant association was found in the subpopulation 1. *Ltn* was significant in DS3, WS2 and the average environment in the whole population, while the remaining genes/QTL were significant only in

Table 6Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and PH identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>D10</i>	RM3411	– ^c	–	Ave (2), SC (2)	0.02–0.036
<i>d27</i>	RM26985	–	–	Ave (2)	0.027
<i>D88</i>	RM6742	–	–	Ave (2)	0.017
<i>DEP2</i>	RM21964	–	–	Ave (2)	0.017
<i>DWL1</i>	HL921	–	–	Ave (2)	0.022
<i>DWL1</i>	HL944	–	–	Ave (2)	0.033
<i>ep3</i>	STS5803.7	–	–	Ave (2)	0.029
<i>Ghd7</i>	RM5436	–	–	Ave (2)	0.033
<i>Ghd8</i>	RM22483	WS2	0.004	Ave (2), WS2 (2)	0.0127–0.029
<i>GIF1</i>	RM16942	–	–	Ave (2)	0.021
<i>Gn1a</i>	RM10316	–	–	Ave (2)	0.022
<i>Gn1a</i>	RM3604	–	–	Ave (2)	0.022
<i>Gnp4</i>	Y48	SC	0.047	Ave (2)	0.021
<i>GS3</i>	RGS1	–	–	Ave (2)	0.033
<i>GS5</i>	C62	–	–	Ave (2)	0.029
<i>GS5</i>	RM574	–	–	Ave (2)	0.022
<i>Gw1-1</i>	RM10376	–	–	Ave (2), JX (2)	0.0004–0.013
<i>Gw1-2</i>	RM1344	–	–	Ave (2)	0.021
<i>GW2</i>	RM12827	–	–	Ave (2)	0.029
<i>GW3</i>	RM15206	–	–	Ave (2)	0.022
<i>gw3.1</i>	JL109	WS2	0.046	Ave (2)	0.013
<i>gw5</i>	RMw513	–	–	Ave (2), JX (2)	0.022–0.044
<i>GW6</i>	RM20201	–	–	Ave (2)	0.021
<i>gw8.1</i>	RM23201.CNR151	–	–	Ave (2)	0.022
<i>gw9.1</i>	RM24718.CNR111	–	–	Ave (2)	0.029
<i>htd1</i>	RM17307	SC	0.009	Ave (2)	0.021
<i>IPA1</i>	RM23422	–	–	Ave (2)	0.021
<i>Ltn</i>	ssr0649.23	WS1	0.006	Ave (2)	0.021
<i>MOC1</i>	RM20373	–	–	Ave (2)	0.029
<i>Nop(t)</i>	M9	–	–	Ave (2)	0.021
<i>PAP2</i>	RM15937	–	–	Ave (2)	0.027
<i>PAP2</i>	RM15948	–	–	Ave (2)	0.029
<i>qGL7</i>	RID711	–	–	Ave (2)	0.027
<i>qGL7-2</i>	RM21945	–	–	Ave (2)	0.027
<i>qGL7-2</i>	Indel1	–	–	Ave (2)	0.029
<i>qGN4-1</i>	RM3276	–	–	Ave (2)	0.022
<i>qGN4-1</i>	nkssr04.19	–	–	Ave (2)	0.033
<i>qGY2-1</i>	RM279	–	–	Ave (2), JX (2)	0.012–0.04
<i>qPDS3</i>	RM14820	–	–	Ave (2)	0.021
<i>qPH6-1</i>	RM19417	SC	0.004	Ave (2), SC (2)	0.020–0.029
<i>qPH6-1</i>	RM3414	–	–	Ave (2)	0.027
<i>qSH3</i>	RM16	–	–	Ave (2)	0.017
<i>qSPP7</i>	RM5499	–	–	Ave (2)	0.029
<i>RPH</i>	RM481	–	–	Ave (2)	0.017
<i>SCM2</i>	RM20547	–	–	Ave (2)	0.013
<i>SPP1</i>	YN27	–	–	Ave (2)	0.027

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.

DS3. *Gn1a*, *Ltn*, *qGN4-1* and *qGY2-1* were significant in the average environment using the subpopulation 2.

A total of six genes/QTL were associated with SB. Four genes/QTL were significant in the whole population (Table 13). *Gn1a* was significant in five testing environments as well as the average environment. *Ghd7* and *qSPP7* were in two testing environments and the average environment. *SPP1* was only in the average environment. Four (*Ghd7*, *Gn1a*, *qSPP7* and *SCM2*) and two (*qSH3* and *SPP1*) genes/QTL were associated with SB in the subpopulation 1 and subpopulation 2, respectively (Table 13). *Gn1a* was significant in DS1 and the average environment, while the remaining five genes/QTL were associated with SB in only one testing environment.

Eleven genes/QTL were associated with SR (Table 14). Seven genes/QTL associated with SR in the whole population. *Ghd7* was associated with SR in SC and the average environment. The remaining six genes/QTL were associated with SR in only one testing environment or the average environment. Four and five significant associations were found in the subpopulation 1 and subpopulation 2, respectively (Table 14). *D10*, *GW6*, *MOC1* and *PAP2* were found

in the subpopulation 1 while *GIF1*, *qSPP7*, *Ghd7*, *Ltn* and *PAP2* in the subpopulation 2. *GIF1* and *qSPP7* were associated with SR in one testing environment and the average environment. *PAP2* was associated with SR in JX and SC. The remaining genes/QTL were associated with SR in one testing environment or the average environment.

3.3. Prediction of GY using markers for known genes

Two prediction models were developed to predict GY using stepwise MLR for each of the eight testing environments and the average environment. One was developed using significant markers associated with any of the measured traits in any of the environments (associated marker model) and the other using all the markers for the 39 target genes/QTL (all marker model). Table 15 lists the selected markers and adjusted R^2 in the final models for each environment. In the average environment, all 39 genes/QTL were associated with at least one of the 11 tested traits, thus there was only one model. Sixteen markers were included in the model

Table 7Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and GN identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>Ghd7</i>	RM5436	DS3, SC, WS1, WS2	9.9E-06-0.014	SC (2)	0.0007
<i>GIF1</i>	RM16942	Ave, JX, SC	2.2-E05-0.039	Ave (2), JX (2), SC (2)	4.5-E08-0.049
<i>Gn1a</i>	RM10316	DS3, SC, WS1, WS2	0.009–0.044	– ^c	–
<i>Gn1a</i>	RM3604	WS2	0.004	–	–
<i>Gnp4</i>	Y48	–	–	WS3 (1)	0.008
<i>Gw1-1</i>	RM10376	Ave, JX	3.6E-10-4.0E-13	Ave (2), JX (2), SC (2)	1.9E-10-3.7E-10
<i>Gw1-2</i>	RM1344	SC	0.044	–	–
<i>GW2</i>	RM12827	Ave, JX	0.0004–0.001	Ave (2), JX (2)	0.014–0.047
<i>GW6</i>	RM20201	SC	0.040	–	–
<i>Ltn</i>	ssr0649.23	Ave, DS3, JX	1.9E-19-0.017	Ave (2), JX (2)	1.9E-10-3.7E-10
<i>Nop(t)</i>	M9	SC	0.040	SC (2)	0.015
<i>qGL7</i>	RID711	–	–	SC (1)	0.008
<i>qGL7-2</i>	RM21945	–	–	SC (1)	0.042
<i>qPH6-1</i>	RM3414	SC	0.002	SC (2)	0.005
<i>qSPP7</i>	RM5499	DS3, SC, WS1, WS2	0.0001–0.017	SC (1), SC (2)	0.0007–0.042
<i>SCM2</i>	RM20547	DS1, SC	0.031–0.045	–	–
<i>SPP1</i>	YN27	DS3, WS2	0.006–0.033	–	–

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.**Table 8**Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and SN identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole panel		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>DEP2</i>	RM21964	– ^c	–	DS3 (2)	0.028
<i>Ghd7</i>	RM5436	Ave, DS3, SC, WS1, WS2	0.008–0.034	DS2 (2), SC (2)	0.028–0.049
<i>Gn1a</i>	RM10316	Ave, DS3, SC, WS1, WS2	0.001–0.031	Ave (2), DS2 (2)	0.028–0.045
<i>Gn1a</i>	RM3604	WS2	0.038	–	–
<i>GW2</i>	RM12827	WS2	0.038	–	–
<i>GW6</i>	RM20201	–	–	DS23 (2)	0.030
<i>Ltn</i>	ssr0649.23	Ave, DS3	0.022–0.047	Ave (2), DS3 (2)	0.028–0.046
<i>Nop(t)</i>	M9	WS1	0.036	Ave (2)	0.046
<i>qGY2-1</i>	RM279	–	–	Ave (2)	0.046
<i>qSPP7</i>	RM5499	Ave, DS3, SC, WS1, WS2	0.08–0.034	Ave (2), DS3 (2)	0.028–0.046
<i>SCM2</i>	RM20547	Ave, DS1, SC, WS2	0.002–0.045	Ave (2), DS1 (2), SC (2)	0.044–0.049
<i>SPP1</i>	YN27	Ave, DS3, WS2	0.009–0.047	DS3 (2)	0.0278

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.**Table 9**Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and PN identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>DWL1</i>	HL921	Ave, DS2	0.045–0.047	– ^c	–
<i>Gn1a</i>	RM3604	DS3	0.028	–	–
<i>Gw1-2</i>	RM1344	DS3	0.019	NE	–
<i>GW6</i>	RM20201	WS2	0.007	–	–
<i>SPP1</i>	YN27	–	–	WS2 (1)	0.033

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.**Table 10**Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and TN identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>ep3</i>	STS5803.7	– ^c	–	WS2 (1)	0.033
<i>GW6</i>	RM20201	WS2	0.03	WS1 (1)	0.004
<i>MOC1</i>	RM20373	–	–	WS1 (1)	0.042
<i>qGY2-1</i>	RM279	–	–	Ave (2), WS1 (2)	0.001–0.011
<i>SPP1</i>	YN27	–	–	WS2 (1)	0.014

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.

Table 11Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and TGW identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>Ghd7</i>	RM5436	Ave, DS3, SC	0.030–0.047	– ^c	–
<i>GS3</i>	RGS1	Ave, DS1, DS2, DS3, WS3	0.011–0.047	WS3 (2)	0.039
<i>GW3</i>	RM15206	Ave, SC, WS2	0.03–0.038	–	–
<i>gw5</i>	RMw513	Ave, WS1	0.042–0.048	–	–
<i>GW6</i>	RM20201	Ave, DS3, WS1, WS2, WS3	0.003–0.048	SC (2), WS2 (2), WS3 (2)	0.018–0.026
<i>htd1</i>	RM17307	SC	0.021	–	–
<i>qGL7-2</i>	RM21945	SC	0.004	–	–
<i>qGN4-1</i>	nkssr04.19	SC	0.05	–	–
<i>qPH6-1</i>	RM19417	Ave, DS3, WS1, SC, WS2, WS3	0.00003–0.048	SC (2), WS3 (2)	0.007–0.039
<i>qPH6-1</i>	RM3414	Ave, SC	0.001–0.037	SC (2)	0.017
<i>SCM2</i>	RM20547	JX	0.002	–	–

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.**Table 12**Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and PB identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>D10</i>	RM3411	DS3	0.025	– ^c	–
<i>d27</i>	RM26985	DS3	0.025	–	–
<i>D88</i>	RM6742	DS3	0.024	–	–
<i>DWL1</i>	HL921	DS3	0.039	–	–
<i>DWL1</i>	HL944	DS3	0.027	–	–
<i>Ghd8</i>	RM22483	DS3	0.045	–	–
<i>GIF1</i>	RM16942	DS3	0.041	–	–
<i>Gn1a</i>	RM3604	–	–	Ave (2)	0.014
<i>Gnp4</i>	Y48	DS3	0.024	–	–
<i>GS3</i>	RGS1	DS3	0.024	–	–
<i>GS5</i>	C62	DS3	0.025	–	–
<i>Gw1-1</i>	RM10376	DS3	0.032	–	–
<i>GW2</i>	RM12827	DS3	0.024	–	–
<i>GW3</i>	RM15206	DS3	0.024	–	–
<i>gw3.1</i>	JL109	DS3	0.024	–	–
<i>gw5</i>	RMw513	DS3	0.050	–	–
<i>GW6</i>	RM20201	DS3	0.025	–	–
<i>gw8.1</i>	RM23201.CNR151	DS3	0.024	–	–
<i>gw9.1</i>	RM24718.CNR111	DS3	0.030	–	–
<i>htd1</i>	RM17307	DS3	0.025	–	–
<i>IPA1</i>	RM23422	DS3	0.024	–	–
<i>Ltn</i>	ssr0649.23	Ave, DS3, WS2	0.0001–0.036	Ave (2)	0.008
<i>MOC1</i>	RM20373	DS3	0.024	–	–
<i>Nop(t)</i>	M9	DS3	0.024	–	–
<i>PAP2</i>	RM15937	DS3	0.024	–	–
<i>qGN4-1</i>	nkssr04.19	DS3	0.045	–	–
<i>qGN4-1</i>	RM3276	DS3	0.025	Ave (2)	0.027
<i>qGY2-1</i>	RM279	DS3	0.024	Ave (2)	0.027
<i>RPH</i>	RM481	DS3	0.039	–	–
<i>SCM2</i>	RM20547	DS3	0.025	–	–
<i>SPP1</i>	YN27	DS3	0.025	–	–

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.**Table 13**Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and SB identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>Ghd7</i>	RM5436	Ave, DS3, WS1	0.008–0.02	WS1 (2)	0.037
<i>Gn1a</i>	RM10316	Ave, DS1, DS3, WS1, WS2	0.0003–0.04	Ave (2), DS1 (2)	0.006–0.010
<i>Gn1a</i>	RM3604	WS1	0.049412	– ^c	–
<i>qSH3</i>	RM16	–	–	WS3 (1)	0.013
<i>qSPP7</i>	RM5499	Ave, DS3, WS1	0.01–0.03	WS1 (2)	0.037
<i>SCM2</i>	RM20547	–	–	DS1 (2)	0.046
<i>SPP1</i>	YN27	Ave	0.047085	WS2 (1)	0.013

^a Env indicates the environments where significant associations were found.^b Sub indicates significant association was found in the subpopulation.^c Indicates association was not significant and thus q was not given.

Table 14Significant associations ($q \leq 0.05$) between the 39 target genes/QTLs and SR identified in the whole population and two subpopulations.

Gene/QTL	Marker	Whole population		Subpopulation	
		Env ^a	q	Env (Sub) ^b	q
<i>D10</i>	RM3411	– ^c	–	JX (1)	0.043235
<i>DEP2</i>	RM21964	Ave	0.018	–	–
<i>Ghd7</i>	RM5436	Ave, SC	0.018–0.048	Ave (2)	0.043095
<i>GIF1</i>	RM16942	Ave	0.045	Ave (2), SC (2)	0.012–0.043
<i>gw3.1</i>	JL109	Ave	0.039	–	–
<i>GW6</i>	RM20201	–	–	Ave (1)	0.010028
<i>Ltn</i>	ssr0649.23	–	–	DS1 (2)	0.046
<i>MOC1</i>	RM20373	–	–	Ave (1)	0.032938
<i>PAP2</i>	RM15937	SC	0.030	JX (1), SC (2)	0.012–0.036
<i>qGL7</i>	RID711	SC	0.030	–	–
<i>qSPP7</i>	RM5499	Ave	0.009	Ave (2), SC (2)	0.041–0.043

^a Env indicates the environments where significant associations were found.^b Sub indicates the subpopulation where significant association was found.^c Indicates association was not significant and thus q was not given.**Table 15**The final model for predicting GY in each environment resulted from the multiple linear regressions and adjusted R^2 .

Env ^a	Marker ^b	Selected markers ^c	Gene/QTL ^d	Adjusted R^2
Com	Assoc	HL944, M9, RGS1, RM12827, RM14820, RM16, RM20373, RM20547, RM21964, RM279, RM3604, RMW513, ssr0649.23, STS5803.7, Y48, YN27	<i>DWL1</i> , <i>Nop(t)</i> , <i>GS3</i> , <i>GW2</i> , <i>qPDS3</i> , <i>qSH3</i> , <i>MOC1</i> , <i>SCM2</i> , <i>DEP2</i> , <i>qGY2-1</i> , <i>Gn1a</i> , <i>gw5</i> , <i>Ltn</i> , <i>ep3</i> , <i>Gnp4</i> , <i>SPP1</i>	0.158
	All	HL944, M9, RGS1, RM12827, RM14820, RM16, RM20373, RM20547, RM21964, RM279, RM3604, RMW513, ssr0649.23, STS5803.7, Y48, YN27	<i>DWL1</i> , <i>Nop(t)</i> , <i>GS3</i> , <i>GW2</i> , <i>qPDS3</i> , <i>qSH3</i> , <i>MOC1</i> , <i>SCM2</i> , <i>DEP2</i> , <i>qGY2-1</i> , <i>Gn1a</i> , <i>gw5</i> , <i>Ltn</i> , <i>ep3</i> , <i>Gnp4</i> , <i>SPP1</i>	0.158
DS1	Assoc	nkssr04.19, RGS1	<i>qGN4-1</i> , <i>GS3</i>	0.070
	All	C62, HL944, M9, nkssr04.19, RGS1, RM14820, RM19417, RM20373, RM21964, RM26985, RM3411, RM574	<i>GS5</i> , <i>DWL1</i> , <i>Nop(t)</i> , <i>qGN4-1</i> , <i>GS3</i> , <i>qPDS3</i> , <i>qPH6-1</i> , <i>MOC1</i> , <i>DEP2</i> , <i>d27</i> , <i>D10</i>	0.151
DS2	Assoc	HL921, RM10316, RM12827	<i>DWL1</i> , <i>Gn1a</i> , <i>GW2</i>	0.024
	All	JL109, RGS1, RM15937, RM22483, RM23422, Y48	<i>gw3.1</i> , <i>GS3</i> , <i>PAP2</i> , <i>Ghd8</i> , <i>IPA1</i> , <i>Gnp4</i>	0.039
DS3	Assoc	C62, HL944, JL109, M9, RM10316, RM20201, RM20373, RM20547, RM279, RM3411, RM5436, RM574, RMW513, Y48	<i>GS5</i> , <i>DWL1</i> , <i>gw3.1</i> , <i>Nop(t)</i> , <i>Gn1a</i> , <i>GW6</i> , <i>MOC1</i> , <i>SCM2</i> , <i>qGY2-1</i> , <i>D10</i> , <i>Ghd7</i> , <i>gw5</i> , <i>Gnp4</i>	0.191
	All	HL944, Indel1, JL109, RM10316, RM14820, RM15206, RM19417, RM20201, RM20373, RM20547, RM21945, RM279, RM3411, RM3414, RM5436, STS5803.7, Y48	<i>DWL1</i> , <i>qGL7-2</i> , <i>gw3.1</i> , <i>Gn1a</i> , <i>qPDS3</i> , <i>GW3</i> , <i>qPH6-1</i> , <i>GW6</i> , <i>MOC1</i> , <i>SCM2</i> , <i>qGY2-1</i> , <i>D10</i> , <i>Ghd7</i> , <i>ep3</i> , <i>Gnp4</i>	0.230
JX	Assoc	RM20547, RM279, RM3411	<i>SCM2</i> , <i>qGY2-1</i> , <i>D10</i>	0.063
	All	RM20373, RM20547, RM21945, RM3414	<i>MOC1</i> , <i>SCM2</i> , <i>qGL7-2</i> , <i>qPH6-1</i>	0.117
SC	Assoc	Indel1, M9, RID711, RM10316, RM10376, RM15937, RM16942, RM17307, RM19417, RM20547, RM21945, RM3411	<i>qGL7-2</i> , <i>Nop(t)</i> , <i>qGL7</i> , <i>Gn1a</i> , <i>Gw1-1</i> , <i>PAP2</i> , <i>GIF1</i> , <i>htd1</i> , <i>qPH6-1</i> , <i>SCM2</i> , <i>D10</i>	0.149
	All	M9, RGS1, RID711, RM10376, RM14820, RM15937, RM15948, RM19417, RM20547, RM21945, RM21964, RM24718.CNR111, RM26985, RM3276, RM3604, RM5499, YN27	<i>Nop(t)</i> , <i>GS3</i> , <i>qGL7</i> , <i>Gw1-1</i> , <i>qPDS3</i> , <i>PAP2</i> , <i>qPH6-1</i> , <i>SCM2</i> , <i>qGL7-2</i> , <i>DEP2</i> , <i>gw9.1</i> , <i>d27</i> , <i>qGN4-1</i> , <i>Gn1a</i> , <i>qSPP7</i> , <i>SPP1</i>	0.193
WS1	Asso	M9, RGS1, RM5436, RM6742, RMW513, ssr0649.23	<i>Nop(t)</i> , <i>GS3</i> , <i>Ghd7</i> , <i>D88</i> , <i>gw5</i> , <i>Ltn</i>	0.106
	All	HL921, M9, nkssr04.19, RM1344, RM15206, RM19417, RM21964, RM26985, RM279, RM3276, RM3411, RM574, RM6742, RMW513, ssr0649.23, Y48, YN27	<i>DWL1</i> , <i>Nop(t)</i> , <i>qGN4-1</i> , <i>Gw1-2</i> , <i>GW3</i> , <i>qPH6-1</i> , <i>DEP2</i> , <i>d27</i> , <i>qGY2-1</i> , <i>D10</i> , <i>GS5</i> , <i>D88</i> , <i>gw5</i> , <i>Ltn</i> , <i>Gnp4</i> , <i>SPP1</i>	0.229
WS2	Assoc	RGS1, RM10316, RM15206, RM19417, RM20201, RM3604, RM5436, ssr0649.23	<i>GS3</i> , <i>Gn1a</i> , <i>GW3</i> , <i>qPH6-1</i> , <i>GW6</i> , <i>Ghd7</i> , <i>Ltn</i>	0.116
	All	HL921, RGS1, RM10316, RM10376, RM15206, RM15937, RM17307, RM20201, RM21964, RM23201.CNR151, RM24718.CNR111, RM279, RM3276, RM481, RM5436, RM574, RM6742, RMW513, ssr0649.23, Y48, YN27	<i>DWL1</i> , <i>GS3</i> , <i>Gn1a</i> , <i>Gw1-1</i> , <i>GW3</i> , <i>PAP2</i> , <i>htd1</i> , <i>GW6</i> , <i>DEP2</i> , <i>gw8.1</i> , <i>gw9.1</i> , <i>qGY2-1</i> , <i>qGN4-1</i> , <i>RPH</i> , <i>Ghd7</i> , <i>GS5</i> , <i>D88</i> , <i>gw5</i> , <i>Ltn</i> , <i>Gnp4</i> , <i>SPP1</i>	0.261
WS3	Assoc	RGS1, RM16, RM19417, RM20201, RM3604	<i>GS3</i> , <i>qSH3</i> , <i>qPH6-1</i> , <i>GW6</i> , <i>Gn1a</i>	0.097
	All	C62, HL921, nkssr04.19, RM10316, RM14820, RM15948, RM17307, RM19417, RM20201, RM20373, RM20547, RM23201.CNR151, RM279, RM3276, RM3411, RM3414, RM481, RM6742	<i>GS5</i> , <i>DWL1</i> , <i>qGN4-1</i> , <i>Gn1a</i> , <i>qPDS3</i> , <i>PAP2</i> , <i>htd1</i> , <i>qPH6-1</i> , <i>GW6</i> , <i>MOC1</i> , <i>SCM2</i> , <i>gw8.1</i> , <i>qGY2-1</i> , <i>D10</i> , <i>RPH</i> , <i>D88</i>	0.225

^a Env, Environment where the GY was evaluated.^b All indicates all the 46 markers for the 39 known genes were used. Assoc indicates only markers identified to be associated with GY or yield related traits were used.^c Markers included in the final model.^d Known genes/QTL with which the markers in the model were associated.

and captured 15.8% of the phenotypic variation. For the three environments in the DS, the final associated markers models for DS1, DS2 and DS3 included two out of eight, three out of seven and 14 out of 38 markers and explained 7.0%, 2.4% and 19.1% of the total phenotypic variation, respectively. The final all marker models for DS1, DS2 and DS3 consisted of 12, six and 17 markers and captured 15.1%, 3.9% and 23.0% of the phenotypic variation, respectively. There were two and ten common markers between the two final models for DS1 and DS3, respectively. For DS2, there was no common marker between the two final models. For JX and SC, three out of 17 and 12 out of 22 markers were included in the associated marker models, which explained 6.3% and 14.9% of the total phenotypic variation, respectively. Four and 17 markers were selected in the all marker models and explained 11.7% and 19.3% of the total variation, respectively. There were two common markers between the two models for JX and seven common markers between the two models for SC. The final associated marker models for WS1, WS2, and WS3 included six out of 21, eight out of 15 and five out of ten markers and explained 10.6%, 11.6% and 9.7% of the total variation, respectively. The final all marker models for WS1, WS2, and WS3 had 17, 21 and 18 markers and captured 22.9%, 26.1% and 22.5% of the phenotypic variation, respectively. The number of common markers between the two final models for WS1, WS2 and WS3 were four, six and two, respectively.

4. Discussion

4.1. Genetic diversity can be further increased by introducing Chinese breeding lines

One of the objectives of the newly designed irrigated breeding program at IRRI is to increase the genetic diversity of its breeding populations by incorporating elite lines from other large breeding programs. The genetic diversity and PIC value of the present population were 0.4181 and 0.3787, respectively. The PIC value was close to and comparable with Huang et al.'s (2015) results in terms of the diversity level and the population size. They found that the PIC value of 299 inbred *indica* rice varieties cultivated in Guangdong province of China assessed using 34 SSR markers was 0.38. The genetic diversity and PIC value in our study were lower than those of the populations of 83 global *indica* lines (0.694 and 0.665) and 495 Chinese *indica* lines (0.623 and 0.595) assessed by Wang et al. (2014) with 84 SSR.

The average allele per locus was 3.92 in the present population. This was slightly higher than that of Huang et al.'s study (2015), in which the average allele per locus was 3.3 for 34 SSR markers. The average allele number was also higher than that of the population comprised of 328 *indica* accessions mostly collected in China studied using 100 SSR (Jin et al., 2010), in which the average allele per locus ranged from 2.60 to 3.36 among the 6 subpopulations. The average number of alleles per locus in our study was much smaller than Wang et al.'s (2014) results, which reported that the average allele per locus for the 84 SSR in the populations of 83 global lines and 495 Chinese *indica* lines were 10.5 and 13.0, respectively. Considering that our population consisted of lines mainly from IRRI it seemed that genetic diversity can be greatly increased by introducing Chinese breeding lines. The few Chinese breeding lines included in the present study could not represent the genetic diversity of Chinese breeding lines.

4.2. Careful association analysis is important

The usefulness of 39 well-characterized genes/QTL was tested using association analysis in this study. It is well-known that population structure and unequal relatedness among individuals greatly

affect the results of association analysis and must be carefully taken into account (Pritchard and Rosenberg, 1999; Yu et al., 2006). The 360 *indica* lines of the present study were divided into two subpopulations using a model based analysis implemented in STRUCTURE (Falush et al., 2003; Pritchard et al., 2000). The kinship among the 360 lines was unequal. The estimated kinship coefficient ranged from 0 to 1.80 with an average 0.49. Distantly related genotype pairs accounted for 6.96% and highly related genotype pairs accounted for only 0.03%. The majority of the genotype pairs had kinship similar to half-sibs. The association results from models with and without population structure and relationship among individuals were significantly different (Figs. 3 and 4). The naïve model identified the highest number of marker trait associations, in which there must be a lot of false positives as many of the observed *P* values significantly deviated from the expected *P* values in the QQ plot. The *Q* model fitted the expected *P* values better than the naïve model but much worse than the *QK* and *K* models, indicating that controlling only for the population structure was useful but not enough to reduce the spurious associations. The *K* model and *QK* models performed similarly for most traits and in most testing environments, suggesting that controlling the unequal familial relatedness could sufficiently control the spurious associations in the present population. We also found that some genes/QTL were only detectable in one of the two subpopulations using the *K* model, although the subpopulation was much smaller than the whole population.

4.3. Usefulness of the tested genes/QTL

All the studied genes/QTL were associated with at least two of the 11 measured traits in one of the eight testing environments or the average environment. The numbers of genes/QTL associated with GY, DTF, PH, GN, SN, PN, TN, TGW, PB, SB and SR were 16, 25, 39, 16, 11, five, five, ten, 29, six and 11, respectively (Tables 4–14). These genes/QTL could thus be used in improving the corresponding target traits in breeding for different objectives. A few genes/QTL were associated with trait(s) in more than three testing environments. *GS3* was associated with DTF in six testing environments and the average environment. *GIF1*, *qSPP7* and *Ltn* were associated with DTF in four testing environments as well as the average environment. *Gn1a* associated with DTF in four testing environments. *Ghd7* was associated with DTF in three testing environment and the average environment. *GS5* was associated with DTF in three testing environments. *Ghd7*, *Gn1a*, and *qSPP7* were associated with GN in four testing environment. *qPH6-1* and *GW6* were associated with TGW in five testing environments and the average environment. *Gn1a* was associated with SB in four testing environment as well as the average environment. *Ghd7*, *Gn1a*, and *qSPP7* were associated with SN in four testing environments and the average environment, while *SCM2* in three testing environments and the average environment. These genes/QTL could be regarded as the important ones for the corresponding traits. The following observed characteristics need to be carefully considered when these genes/QTL are explored in breeding.

Firstly, for all genes/QTL environment had large effects and significant gene-by-environment interaction (GEI) was present. The total number of genes/QTL associated with the measured traits varied greatly across the testing environments ranging from seven in DS2 to 34 in DS3. Both of season and nitrogen rates affected the number of GTA, although season had less effect than nitrogen level. The numbers of GTA found in the DS and WS were 49 and 44, respectively. Nitrogen rate had different effects on the GTA in the DS and WS. In the DS, the numbers of significant genes/QTL were eight, seven and 34 in DS1, DS2 and DS3, respectively while in the WS they were 20, 14 and ten in WS1, WS2 and WS3, respectively.

The importance of GEI was clear, since no gene/QTL was significantly associated with a trait in all the eight testing environments

and that the associated genes/QTL changed from environment to environment for all traits. Some of the genes/QTL even had effects in opposite directions in different environments. One allele of *Gn1a* had opposite effect on DTF in WS2 compared with in other three environments including SC, WS1 and WS3. Similarly, one allele of *GIF* and *Ltn* had opposite effect on DTF in DS3 and WS3. One allele of *qSPP7* had opposite effect on DTF in JX and WS3 compared to in WS1 and WS2 (data was not given).

Secondly, all the 39 genes/QTL were found to be associated with two or more traits in the present study. *GW6* was associated with nine out of the 11 traits excepting GY and SB. *Gn1a* was associated with eight traits excepting SR, TGW and TN. *Ghd7*, *qSPP7*, *SCM2* and *SPP1* associated with seven traits. *GIF1* and *Ltn* were associated with six traits. *GS3*, *GW2*, *gw3.1*, *htd1*, *Nop(t)*, *qGY2-1* and *qPH6-1* were associated with five traits. For the remaining 24 genes/QTL, 13, seven and four were found to be associated with four, three and two traits, respectively (Table 3). Seventeen of the 39 target genes/QTL were known to affect more than one trait previously (Table 1). The underlying biology of pleiotropy might be that a gene was involved in multiple pathways. For instance, *Ghd7* was involved in multiple pathways for traits contributable to yield, such as phytohormone synthesis pathway and photoperiod pathway (Wei et al., 2010; Xue et al., 2008) and biotic and abiotic stresses (Weng et al., 2014). On the other hand, it is possible that genetic strongly correlated agronomic traits are controlled by the same biological processes.

Many of the tested known genes/QTL were reported to be associated with only one trait in previous reports. This may simply due to the limited number of traits measured. In the present study, we tested 11 yield related traits and found that all the target genes/QTL were multiple-trait associated. Therefore, it is important for MAS to investigate the effects of the target genes on other important agronomic traits. This is particularly important when genes/QTL with limited traits information are used.

Thirdly, all the genes/QTL were associated with traits unreported previously (Table 3). For instance, the grain weight controlling gene, *GW6* (Guo et al., 2009), was associated with TGW and other eight traits except GY and SB. *Gn1a* regulating grain number and enhancing GY (Ashikari et al., 2005) was found to be associated with GN, GY and other six traits including DTF, PB, PH, PN, SB and SN. *Ghd7*, an important regulator of PH, HD, GN and GY (Xue et al., 2008; Lu et al., 2012), was found to be associated with PH, DTF and GN. It was also associated with SB, SN, SR and TGW in one or more environments. This could be partially explained by the fact that most of the gene/QTL fine mapping and cloning studies usually focused on only one or few traits and whether the target gene/QTL affects other traits was not studied. The present study evaluated GY and ten related traits and provided a chance to identify associations between genes/QTL with traits not measured previously, which furthered our understanding of agronomic importance of the known genes/QTL.

4.4. GY cannot be well predicted using markers for known genes

Using the markers for genes/QTL associated with yield and related traits to predict GY under each environment captured from 2.4% (DS2) to 19.1% (DS3) of the phenotypic variation. The number of selected markers in the final prediction model varied from three (DS2) to 14 (DS3). The percentage of variation captured by markers in the final models derived from all the target markers varied from 3.9% (DS2) to 26.1% (WS2). As expected, using significant markers identified by association analysis were worse than using all markers. Clearly, the predictions could not be used to replace phenotypic selection based on GY. Although it was often claimed that a single cloned gene could increase GY sustainably, none of the 39 tested genes/QTL has enough power to predict GY well in the present study. Nevertheless, it seems that it is

possible to develop more efficient selection criterion and/or schemes by explicitly incorporating the known genes with large effects.

GY prediction based on markers could be further improved by using high density and high throughput genotyping readily available in rice. One option is to identify more marker-trait associations through genome-wide association studies (GWAS). The other option is to exploit genomic selection that utilizes markers in linkage disequilibrium with all genes affecting GY and captures interactions between genes (Desta and Ortiz, 2014; Nakaya and Isobe, 2012). Genomic selection has been applied in rice to predict the breeding values of GY (Spindel et al., 2015). The prediction accuracies ranged from 0.31 to 0.34 for GY and outperformed prediction using MLR and pedigree records alone. Genomic selection could be an effective tool to increase the selection efficiency, especially for those traits with complex genetic architecture.

5. Conclusion

The usefulness of 39 genes/QTL were tested using association analysis in a population of wide range of varieties or elite breeding lines, which were chosen to represent the genetic diversity of some *indica* breeding gene pools for irrigated ecosystems (Supplementary Table 1). Most of the lines are used as parental lines in IRRI's irrigated breeding programs including pedigree breeding and RS. All the studied genes/QTL were associated with at least two of the 11 measured traits in one or more of the eight testing environments or the average environment. For GY, the number of significant genes/QTL varied from 0 to eight across testing environments and the average environment. It was found that significant GEI was present for all genes, most of the GTA were not reported previously and all of the genes/QTL were associated with at least two tested traits. GY could not be well predicted using markers for these known genes. New marker-trait associations need to be identified using GWAS with high density markers and advanced statistical approaches. GS models based on large number of markers for genes affecting GY and captures small QTL effects and interactions between genes should be exploited. The well characterized genes/QTL with large effects can be used explicitly in developing GS selection criteria to increase selection efficiency.

Conflict of interest

The authors declare they have no conflict of interest.

Author contribution

GY and DM conceived and designed the experiments, SL, GR and XZ performed the experiments, SL CS and GY analyzed the data and wrote the paper and DM and XZ commented on the manuscript before submission.

Acknowledgments

Thanks go to Ms Violeta Bartolome for helping with data analysis. SL was supported by a scholarship provided by the Monsanto's Beachell-Borlaug International Scholars Program. The research of GY was partially supported by the National Natural Science Foundation of China31461143014.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.fcr.2015.11.014>.

References

- Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H., Matsuoka, M., 2005. Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745, <http://dx.doi.org/10.1126/science.1113373>.
- Bernardo, R., 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48, 1649, <http://dx.doi.org/10.2135/cropsci2008.03.0131>.
- 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, <http://dx.doi.org/10.1093/bioinformatics/btm308>.
- Châtel, M., Ospina, Y., Rodriguez, F., Lozano, V.H., Delgado, H., 2008. Upland rice composite population breeding and selection of promising lines for Colombian savannah ecosystem. *Pesqui. Agropecu. Trop.* 38, 1–5.
- Desta, Z.A., Ortiz, R., 2014. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19, 592–601, <http://dx.doi.org/10.1016/j.tplants.2014.05.006>.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620, <http://dx.doi.org/10.1111/j.1365-294X.2005.02553.x>.
- Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587, <http://dx.doi.org/10.1111/j.1471-8286.2007.01758.x>.
- FAOSTAT, FAO-STAT Database—Agricultural Production (2014) <http://apps.fao.org>.
- GRiSP, Global Rice Science Partnership (GRiSP) Proposal, IRRI (2010).
- Guo, L., Ma, L., Jiang, H., Zeng, D., Hu, J., Wu, L., Gao, Z., Zhang, G., Qian, Q., 2009. Genetic analysis and fine mapping of two genes for grain shape and weight in rice. *J. Integr. Plant Biol.* 51, 45–51, <http://dx.doi.org/10.1111/j.1744-7909.2008.00793.x>.
- Guo, L.B., Ye, G.Y., 2014. Use of major quantitative trait loci to improve grain yield of rice. *Rice Sci.* 21, 65–82, [http://dx.doi.org/10.1016/s1672-6308\(13\)60174-2](http://dx.doi.org/10.1016/s1672-6308(13)60174-2).
- Hallauer, A.R., 1985. Compendium of recurrent selection methods and their application. *Crit. Rev. Plant Sci.* 3, 1–33.
- Hayes, B.J., Cogan, N.O.I., Pemberton, L.W., Goddard, M.E., Wang, J., Spangenberg, G.C., Forster, J.W., 2013. Prospects for genomic selection in forage plant species. *Plant Breed.* 132, 133–143, <http://dx.doi.org/10.1111/pbr.12037>.
- Huang, M., Wu, Y., Tao, X., Liu, Y., Yang, G., Chen, Z., 2015. Genetic diversity of main inbred indica rice varieties applied in Guangdong province as revealed by molecular marker. *Rice Sci.* 22, 1–8.
- Jannink, J.-L., 2010. Dynamics of long-term genomic selection. *Genet. Sel. Evol.* 42, 35, <http://dx.doi.org/10.1186/1297-9686-42-35>.
- Jin, L., Lu, Y., Xiao, P., Sun, M., Corke, H., Bao, J., 2010. Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor. Appl. Genet.* 121, 475–487, <http://dx.doi.org/10.1007/s00122-010-1324-7>.
- 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, <http://dx.doi.org/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, <http://dx.doi.org/10.1007/s11103-005-2159-5>.
- 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, <http://dx.doi.org/10.1016/j.fcr.2015.05.014>.
- Liu, K., Muse, V.S., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129, <http://dx.doi.org/10.1093/bioinformatics/bti282>.
- Lu, L., Yan, W., Xue, W., Shao, D., Xing, Y., 2012. Evolution and association analysis of Ghd7 in rice. *PLoS One* 7, e34021, <http://dx.doi.org/10.1371/journal.pone.0034021>.
- Meuwissen, T.H., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B.A., Bustamante, C.D., Boyko, A.R., Purugganan, M.D., 2011. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8351–8356, <http://dx.doi.org/10.1073/pnas.1104686108>.
- Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326, <http://dx.doi.org/10.1093/nar/8.19.4321>.
- Nakaya, A., Isobe, S.N., 2012. Will genomic selection be a practical method for plant breeding? *Ann. Bot.* 110, 1303–1316, <http://dx.doi.org/10.1093/aob/mcs109>.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>.
- Santos, F.R., Pena, S.D., Epplen, J.T., 1993. Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum. Genet.* 90, 655–656.
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., McCouch, S.R., 2015. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11, e1004982, <http://dx.doi.org/10.1371/journal.pgen.1004982>.
- Storey, J.D., 2002. A direct approach to false discovery rates. *J. R. Stat. Soc. B* 64, 479–498.
- Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445.
- Tester, M., Langridge, P., 2010. Breeding technologies to increase crop production in a changing world. *Science* 327, 818–822, <http://dx.doi.org/10.1126/science.1183700>.
- Wang, C.-H., Zheng, X.-M., Xu, Q., Yuan, X.-P., Huang, L., Zhou, H.-F., Wei, X.-H., Ge, S., 2014. Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity (Edinb)* 112, 489–496, <http://dx.doi.org/10.1038/hdy.2013.130>.
- 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 Physiol.* 153, 1747–1758, <http://dx.doi.org/10.1104/pp.110.156943>.
- Weng, X., Wang, L., Wang, J., Hu, Y., Du, H., Xu, C., Xing, Y., Li, X., Xiao, J., Zhang, Q., 2014. Grain number, plant height, and heading date7 is a central regulator of growth, development, and stress response. *Plant Physiol.* 164, 735–747, <http://dx.doi.org/10.1104/pp.113.231308>.
- Xing, Y., Zhang, Q., 2010. Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* 61, 421–442, <http://dx.doi.org/10.1146/annurev-arplant-042809-112209>.
- 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. *Nat. Genet.* 40, 761–767, <http://dx.doi.org/10.1038/ng.143>.
- Ye, G., 2010. New directions in breeding methodologies and breeding informatics. In: Board of Trustees Meeting Report. IRRI (April).
- Zhao, X., Zhou, L., Ponce, K., Ye, G., 2015. The usefulness of known genes/QTL for grain quality traits in an indica population of diverse breeding lines tested using association analysis. *Rice* 8, 29, <http://dx.doi.org/10.1186/s12284-015-0064-3>.
- Zhou, L., Liang, S., Ponce, K., Marundon, S., Ye, G., Zhao, X., 2015. Factors affecting head rice yield and chalkiness in indica rice. *Field Crops Res.* 172, 1–10, <http://dx.doi.org/10.1016/j.fcr.2014.12.004>.