



# Genome-wide association study of grain yield and related traits using a collection of advanced *indica* rice breeding lines for irrigated ecosystems



Shanshan Liang<sup>a,b</sup>, Liwen Wu<sup>c,d</sup>, Guangjun Ren<sup>e</sup>, Xiangqian Zhao<sup>b</sup>, Meixue Zhou<sup>a</sup>, David McNeil<sup>a</sup>, Guoyou Ye<sup>b,\*</sup>

<sup>a</sup> University of Tasmania, Private Bag 98, Hobart, Tasmania 7001, Australia

<sup>b</sup> Genetics and Biotechnology Division, International Rice Research Institute (IRRI), Los Baños, Laguna, Philippines

<sup>c</sup> State Key Laboratory of Rice Biology/Chinese National Center for Rice Improvement, China National Rice Research Institute, Hangzhou 310006, China

<sup>d</sup> Agricultural Genomics Institute in Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

<sup>e</sup> Institute of Crop Science, Sichuan Academy of Agricultural Sciences, Chengdu 610066, Sichuan, China

## ARTICLE INFO

### Article history:

Received 3 February 2016

Received in revised form 17 March 2016

Accepted 19 March 2016

Available online 24 March 2016

### Keywords:

Rice

Yield-related traits

QTL

Genome-wide association studies (GWAS)

Genotyping-by-sequencing

## ABSTRACT

Genome-wide association study (GWAS) is an effective approach for the identification of marker-trait associations (MTA) using more diverse germplasm not constrained by specific crossing. This study used a collection of 327 advanced *indica* breeding lines and varieties genotyped using a genotyping-by-sequencing method and phenotyped in eight environments to identify MTA for grain yield (GY) and 10 related traits. The current panel of *indica* breeding lines were divided into two subpopulations by three analytical methods including STRUCTURE, principal component analysis (PCA) and neighbour joining (NJ) tree. STRUCTURE and PCA gave the same classification of genotypes while the NJ tree results are slight difference in assigning genotypes to subpopulation from STRCUTURE and PCA. Wide variation was observed for all the 11 traits in the whole panel and the two subpopulations inferred by STRUCTURE analysis. Linkage disequilibrium (LD) analysis showed that LD decay varied across the chromosomes and average pairwise squared correlation coefficient ( $r^2$ ) dropped to half of its maximum value ( $r^2 = 0.25$ ) within a physical distance of 200 kb in the whole population. A MLM model controlling both population structure and cryptic relatedness was chosen to identify MTA for all the 11 tested traits. A total of 452 MTA that were delineated into 43 QTL were identified for all traits but PB, SB, and SR with 39 QTL being not reported before. Three QTL on chromosome 6, 9 and 12 were identified for GY but only in DS2. The numbers of QTL identified for the remaining traits varied from two to 26. Most of the detected QTL were found in only one environment. Four QTL were located in the regions containing genes/QTL previously identified for other related traits. The effects of identified QTL were relatively small with the highest percentage of phenotypic variance explained by a single QTL being 9.6%. The identified QTL are directly relevant to and can be more effectively used in breeding programs.

© 2016 Elsevier B.V. All rights reserved.

**Abbreviations:** AM, association mapping; DTF, days to flowering; GN, grain number per panicle; GBS, genotyping-by-sequencing; GEI, genotype-by-environment interaction; GLM, general liner model; GS, genomic selection; GWAS, genome-wide association studies; GY, grain yield; LD, linkage disequilibrium; MAF, minor allele frequency; MAS, marker-assisted selection; MET, multi-environment trials; MLM, mixed linear model; MSD, mean squared difference; MTA, marker-trait associations; NJ, neighbour joining; PB, primary branches per panicle; pH, plant height; PN, panicle number per plant; QEI, QTL-by-environment interaction; SN, spikelet number per panicle; SR, seed setting rate; SB, secondary branches per panicle; TGW, thousand grain weight; TN, tiller number per plant.

\* Corresponding author.

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

## 1. Introduction

Rice (*Oryza sativa* L.) is extensively cultivated on every continent in more than 100 countries and consumed by more than half of the world's population (Juliano, 1993). As world population grows, the global rice demand is estimated to rise from 439 million tons (milled rice) in 2010–496 million tons in 2020 and further increase to 555 million tons in 2035 (GRISP, 2010). Thus a significant increase in rice production is needed for the future of food security.

The Green Revolution in the 1960s greatly increased rice production with at least 50% of the increase being due to adoption of new cultivars. However, increased production potential of modern

rice cultivars has stagnated (Nguyen and Ferrero, 2006) partially due to its narrow genetic base resulting from narrow breeding populations and the population bottleneck that occurred during domestication. How to break the yield barrier is still the major challenge for rice breeders (GRiSP, 2010; Ye et al., 2013). On the other hand, as a model crop species for plant molecular biology and genomics, rice has more accumulated molecular and genomic information than most other crops. The utilization of this genetic information offers the rice breeding community a range of modern tools and methods for addressing this challenge. An integrated strategy was proposed by the International Rice Research Institute (IRRI) to increase breeding efficiency by effectively and efficiently utilizing well proven conventional breeding methods, new techniques and methods enabled by modern molecular biology and genomics and advanced methods in experimental design and data analysis (GRiSP, 2010; Ye et al., 2013). This strategy utilizes (marker-assisted) recurrent selection to quickly pyramid the major genes/QTL in the first few selection cycles and maintain genetic variation contributed by many minor genes to be explored in later cycles, explores genomic selection (GS) for reducing the breeding cycles and the costs of phenotyping and adopts advanced experimental design and data analysis methods to improve heritability (Ye et al., 2013).

To implement this strategy, 392 advanced breeding 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 breeding population. This population has been phenotyped via multi-environment trials (METs) in South-east Asia and genotyped using markers for well-characterized genes/QTL and genome-wide markers to obtain essential information for designing more efficient mating and selection schemes of this general breeding strategy. Phenotypic analysis and basic genotypic analysis were used to remove erroneous lines and lines unadapted to the targeted tropical and subtropical environments and to investigate genotype-by-environment interaction (GEI) for GY (Liang et al., 2015). Thirty-nine cloned or fine-mapped genes/QTL for GY and related traits were tested for their usefulness through association analysis (Liang et al., 2016). Significant gene-by-environment interaction was found for all of the genes/QTL and GY could not be well predicted using the markers significantly associated with the measured traits. More marker-trait associations (MTA) need to be identified for grain yield improvement.

Considering that the majority of the well characterized genes/QTL were identified using biparental populations derived from contrasting parents, it is not unexpected that their transferability to breeding populations is low. An alternative method for the detection of MTA is association mapping (AM). AM identifies MTA caused by linkage disequilibrium (LD), which is the non-random association of alleles at separate loci. AM utilizes ancestral recombination events to identify MTA and provides comparatively higher mapping resolution than the biparental linkage analysis (Zhu et al., 2008). AM has been successfully employed in rice to identify MTA for a range of traits, including yield and yield components (Agrama et al., 2007), cold tolerance (Cui et al., 2013), grain quality, flowering time (Ordóñez et al., 2010) and harvest index (Li et al., 2012). Many associated markers found in these studies were in regions where QTL have previously been identified. The development of several next generation sequencing platforms makes GWAS more and more popular. Using shallow sequencing, a total of 80 MTA for 14 agronomic traits were identified, explaining an average of about 36% of the phenotypic variance (Huang et al., 2010). Among these loci, six were close to known genes *OsC1* (chromosome 6), *ALK* (chromosome 6), *Rc* (chromosome 6), *qSW5* (chromosome 5) and *GS3* (chromosome 3), respectively. The same method was applied to a larger collection of 950 worldwide rice varieties and

identified a total of 32 new loci for flowering time and 10 grain-related traits (Huang et al., 2012). Zhao et al. (2011) performed GWAS using a global collection of 413 rice accessions genotyped using a high-quality custom-designed 44,100 SNP array and phenotyped for 34 traits including morphological, developmental and agronomic traits over two consecutive crop seasons. Dozens of common variants influencing numerous complex traits were identified.

AM is usually conducted using panels of diverse germplasm. While maximizing the genetic diversity, which is beneficial for identifying novel QTL and candidate genes that underlie traits of interest, these panels may not be adapted to identifying relevant variations directly useful in breeding programs (Bordes et al., 2014). Any identified QTL should be validated in a breeding population before they can be used in selection. AM in locally-adapted breeding material should be of more practical benefit to breeders (Bernardo, 2008), since identified beneficial alleles could be incorporated into the creation of new cultivars with limited deleterious effect (Pauli et al., 2014). Empirical studies in wheat and barley demonstrated that GWAS using elite breeding populations is an effective strategy for integrating new genomic technologies into the development of superior cultivars (Bordes et al., 2014; Mohammadi et al., 2015; Pauli et al., 2014; Pozniak et al., 2012). Recently, GWAS performed using a population of elite *indica* rice breeding lines from IRRI's irrigated breeding program genotyped with GBS identified 52 QTL for 11 agronomic traits, with some QTL being co-localized with QTL identified before (Begum et al., 2015). However, the genome-wide reduction of genetic diversity caused by intensive selection during breeding could reduce the efficiency of AM for some traits. A population mixing lines from several breeding programs, each led by independent breeders with their own germplasm, could increase the level of diversity for an efficient use of AM (Bordes et al., 2014).

The main objective of this study was to identify MTA/QTL for GY and 10 related traits using the base breeding population assembled for future *indica* rice breeding for irrigated ecosystem at IRRI. More than 321k markers were generated by the newly developed genotyping-by-sequencing (GBS) method and 76k high quality markers with good genome coverage were used. Population structure of the panel was studied using a model-based Bayesian cluster analysis implemented in STRUCTURE (Falush et al., 2003; Pritchard et al., 2000a), principal component analysis (PCA) and nearest neighbour joining (NJ) methods (Saitou and Nei, 1987). LD patterns of the whole populations and the inferred subpopulations were investigated. To reduce false positives caused by population structure and unequal relatedness among genotypes (lines) four statistical models were compared using the QQ plot and mean squared difference (MSD) between expected and observed *p* values from analysis proposed by Stich et al. (2008). The PK model was chosen to detect MTA for all the trait environment combinations. A total of 452 MTA that were delineated into 43 QTL were identified for GY and seven related traits.

## 2. Materials and methods

### 2.1. Plant materials and phenotyping

Three hundred and ninety two rice varieties or advanced 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 (225). A good number of lines were also from PhilRice (38), CIAT (19), China (14) and Vietnam (10) (Supplementary Table S1). Phenotyping was conducted in Jiangxi (JX) and Sichuan (SC) in China and IRRI headquarters (Los Baños, Philippines). The experiments in JX and SC were conducted in one crop season in 2012. At IRRI, the experiments were carried

Download English Version:

<https://daneshyari.com/en/article/6374448>

Download Persian Version:

<https://daneshyari.com/article/6374448>

[Daneshyari.com](https://daneshyari.com)