# Regional Association Analysis of MetaQTLs Delineates Candidate Grain Size Genes in Rice

Anurag V. Daware[1], Rishi Srivastava[1], Ashok K. Singh[2], Swarup K. Parida[1]* and Akhilesh K. Tyagi[1,3]*

[1] National Institute of Plant Genome Research, New Delhi, India, [2] Division of Genetics, Indian Agricultural Research Institute, New Delhi, India, [3] Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi, India

Molecular mapping studies which aim to identify genetic basis of diverse agronomic traits are vital for marker-assisted crop improvement. Numerous Quantitative Trait Loci (QTLs) mapped in rice span long genomic intervals with hundreds to thousands of genes, which limits their utilization for marker-assisted genetic enhancement of rice. Although potent, fine mapping of QTLs is challenging task as it requires screening of large number of segregants to identify suitable recombination events. Association mapping offers much higher resolution as compared to QTL mapping, but detects considerable number of spurious QTLs. Therefore, combined use of QTL and association mapping strategies can provide advantages associated with both these methods. In the current study, we utilized meta-analysis approach to identify metaQTLs associated with grain size/weight in diverse Indian *indica* and aromatic rice accessions. Subsequently, attempt has been made to narrow-down identified grain size/weight metaQTLs through individual SNP- as well as haplotype-based regional association analysis. The study identified six different metaQTL regions, three of which were successfully revalidated, and substantially scaled-down along with *GS3* QTL interval (positive control) by regional association analysis. Consequently, two potential candidate genes within two reduced metaQTLs were identified based on their differential expression profiles in different tissues/stages of rice accessions during seed development. The developed strategy has broader practical utility for rapid delineation of candidate genes and natural alleles underlying QTLs associated with complex agronomic traits in rice as well as major crop plants enriched with useful genetic and genomic information.

Keywords: grain size, grain weight, metaQTL, QTL mapping, regional association mapping, rice

## INTRODUCTION

The major goal of contemporary molecular breeding is to understand complex inheritance pattern and genetic basis of agronomically important traits for rapid genomics-assisted crop improvement. Quantitative trait loci (QTLs) mapping is the most commonly adopted strategy for identifying QTLs (a genomic region with one or multiple genes) regulating different complex traits of agronomic importance in crop plants. Till date, genome scans for QTLs have been conducted for variety of agronomic traits in most of the major food crops, which

resulted in identification of numerous QTLs[1,2]. However, mere detection of QTLs is usually insufficient for their utilization in crop improvement programs. This is mainly because, confidence intervals for QTLs identified primarily by QTL mapping usually span tens of map units (cM) or hundreds of genes making it difficult to introgress these large genomic regions (QTL intervals) for genetic enhancement of elite crop varieties. Therefore, fine mapping and subsequent map-based cloning of these large QTL genomic intervals to one or few genes responsible for observed phenotypic trait variation are essential. However, fine mapping/positional cloning studies to scale-down a QTL region to candidate genes require development, maintenance and precise multi-location/year field phenotyping of large size (thousands of mapping individuals) advanced generation back-cross and/or RIL (recombinant inbred line) mapping populations. Further, it also requires screening of thousands of mapping individuals from developed mapping populations to identify suitable recombination events within trait-associated QTL intervals. Consequently, fine mapping of these QTLs into potential genes influencing phenotypic trait variation in crop plants requires considerable cost, labor, time and resources. This ultimately impedes fine mapping of majority of the identified QTLs into single gene and essentially affects efficient utilization of identified QTLs for crop improvement.

Association mapping is a proficient complementary approach to QTL mapping for rapid identification of genomic regions (genes) associated with complex quantitative traits. This strategy takes advantage of numerous historical recombinations accumulated within natural germplasm accessions during evolution/domestication and therefore, usually offers a much higher resolution as compared to QTL mapping (Mackay et al., 2009; Gupta et al., 2014). Although first introduced in humans, the utility of association mapping (including candidate gene-based, regional and genome-wide) is now well-demonstrated in multiple diploid as well as complex polyploid crop species to identify trait-associated genomic loci (genes/alleles). Since its inception, association mapping studies have contributed tremendously to our understanding on genetic architecture of complex agronomic traits (Aranzana et al., 2005; Ehrenreich et al., 2009; Meijón et al., 2014; Si et al., 2016; Song et al., 2016). Association mapping is found proficient to rapidly detect more number of natural allelic variations from germplasm accessions and thereby offers much higher resolution as compared to QTL mapping scanned on bi-parental mapping population. However, this strategy detects considerable number of false-positive marker trait associations due to presence of population structure among germplasm accessions used for association analysis (Breseghello and Sorrells, 2006; Sneller et al., 2009). This constraint of spurious marker-trait association can be addressed to certain extent by considering the significant effects of population structure in diverse statistical models as well-adopted by previous association mapping studies in crop plants (Kujur et al., 2015; Upadhyaya et al., 2015, 2016; Bajaj et al., 2016). Despite these efforts, association mapping result still suffers from significant degree of confounding due to strong population structure and cryptic relatedness among accessions used for association analysis. (Brachi et al., 2010). Recently, the process of complex trait dissection in crops has been accelerated due to advent of joint linkage-association analysis and/or QTL region-specific association analysis, which combines the advantages accompanied with both QTL mapping (efficiently detects highly authentic QTLs) and association mapping (higher resolution). The efficacy of this integrated strategy to scale-down and for subsequent fine mapping of the QTL regions into functionally relevant potential candidate genes and natural alleles explaining majority of the quantitative observed phenotypic trait variation in crop plants is well-documented (Lu et al., 2010; Saxena et al., 2014; Zhang et al., 2014; Kujur et al., 2015; Mammadov et al., 2015; Li et al., 2016).

Rice is one of the most consumed cereal food crop in the world and therefore, possesses huge socio-economic relevance. Grain size [grain length (GL) and grain width (GWi)] is one of the most important determinant of grain weight (GWt) and grain yield in rice, besides serving as a vital determinant of consumer preference especially for Basmati trade and commerce. Globally, more than 400 major QTLs regulating grain size/weight have been identified so far, predominantly by traditional linkage/QTL mapping approach and in few cases by genome-wide association studies (GWAS) using diverse bi-parental mapping populations and natural germplasm accessions, respectively (Fan et al., 2006; Zhao et al., 2011; Luo et al., 2013; Zuo and Li, 2014; Si et al., 2016). Genes underlying many of these major QTLs have been subsequently cloned and functionally characterized through fine-mapping/map-based cloning and various functional genomic strategies (Song et al., 2007; Shomura et al., 2008; Mao et al., 2010; Ishimaru et al., 2013). Although many vital QTLs/genes (for instance, *GS3*, *GS5*, *qGL3*, *GW2*, *GW5*, *GW8*, and *TGW6*, etc.) regulating grain size/weight are now well-known; however, majority of these QTLs/genes (except few like *GS3*) have failed to precisely explain grain size/weight trait variation existing within Indian *indica* and aromatic germplasm accessions of rice (Anand et al., 2013; Dixit et al., 2013; Lu et al., 2013). This can be attributed to disparities in genetic constitution of Indian *indica* and aromatic rice accessions from the Chinese *indica* and *japonica* accessions that have been used primarily for mapping of these known cloned QTLs/genes in rice. Meanwhile, many significant efforts have been made to elucidate the genetic basis of grain size/weight variability present within Indian *indica* and aromatic rice accessions by QTL mapping that led to identification of multiple long grain size/weight QTL intervals on different rice chromosomes (Amarawathi et al., 2008; Guleria et al., 2012; Marathi et al., 2012; Swamy et al., 2012; Shanmugavadivel et al., 2013; Vemireddy et al., 2015). Unlike other globally identified and well-dissected major grain size/weight QTLs, genes underlying these major QTLs have not been delineated yet. This has eventually hindered utilization of these QTLs in marker-assisted selection (MAS) for grain size/weight and grain yield improvement of elite Indian *indica* and aromatic rice accessions.

In recent years, substantial progress has been made toward cataloging and characterizing vast number of available rice

---

[1]http://www.gramene.org

[2]http://qtaro.abr.affrc.go.jp

germplasm accessions. This has constituted phenotypically rich and genotypically diverse genetic resources comprising core/mini-core germplasm accessions and user-specific association panels (Ebana et al., 2008; Li et al., 2010; Ali et al., 2011; Zhang et al., 2011; Tiwari et al., 2015; International Rice Gene Bank[3]). Exploitation of these diverse genetic resources for various high-throughput genetic analysis and genomics-assisted crop improvement of rice is now conceivable with advent of next-generation sequencing (NGS)-led structural, functional and comparative genomic strategies. Numerous (Millions) sequence variants including SNPs (single nucleotide polymorphisms) and InDels (insertions–deletions) for thousands of resequenced rice accessions belonging to different *indica*, *japonica*, *aus*, aromatic and wild populations are available at a genome-wide scale in public-domain (Youens-Clark et al., 2010; Zhao et al., 2014; Alexandrov et al., 2015; McCouch et al., 2016). Accessibility to whole genome resequencing- and array-based SNP genotyping data for large number of diverse accessions provides unique opportunity in utilizing this information for regional association analysis of known QTL intervals to expedite their validation and delineation of underlying genes in rice (Chen et al., 2014; Zhao et al., 2014; Singh et al., 2015; McCouch et al., 2016). This strategy not only provides a way to narrow-down known QTLs to candidate genes, but also delivers information regarding favorable natural SNP allelic variants of genes regulating complex quantitative traits in rice accessions that got

---

[3]http://irri.org/our-work/research/genetic-diversity/international-rice-genebank
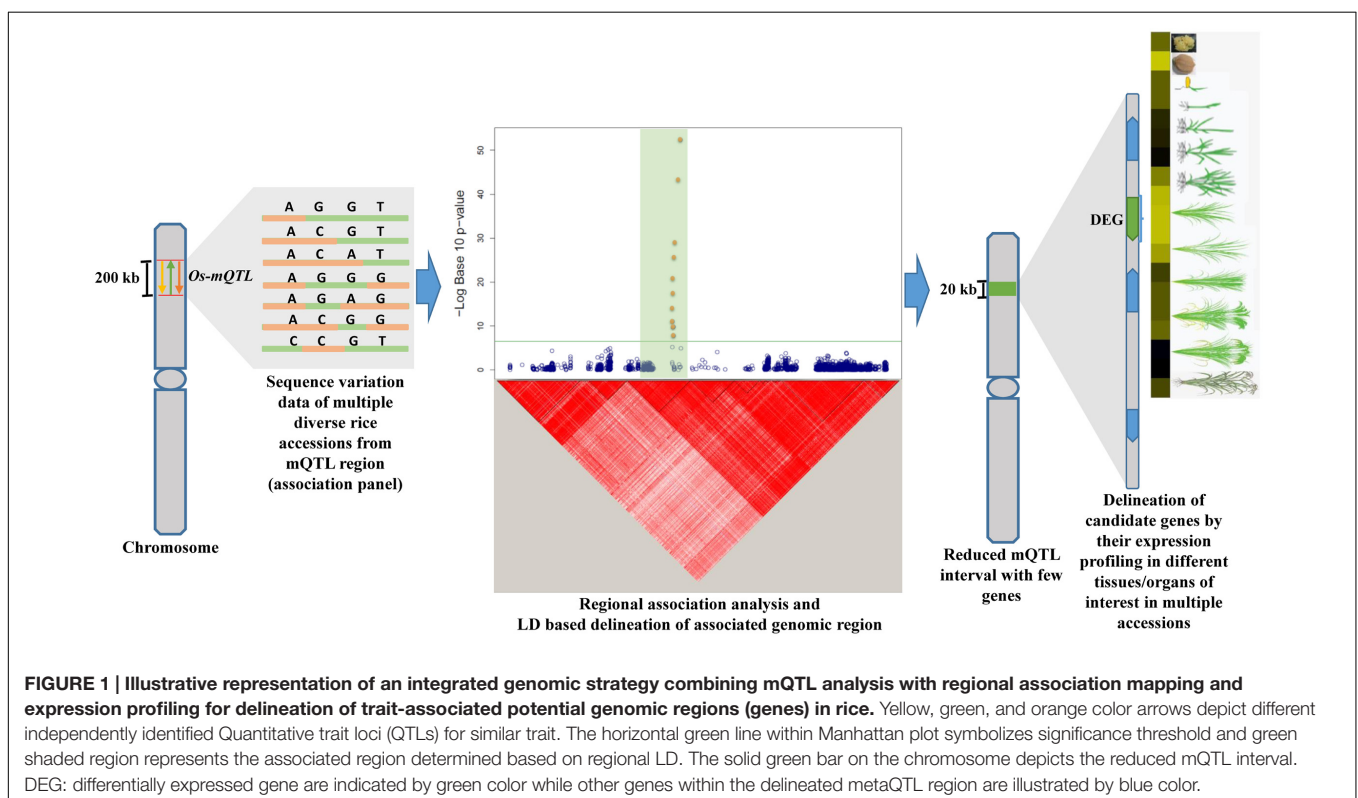
domesticated and fixed by natural selection, and subsequently adapted in diverse agro-climatic conditions. Therefore, this combinatorial strategy can be instrumental for dissection of complex quantitative traits and marker-aided genetic enhancement to develop cultivars with high GWt and yield in rice (Ruggieri et al., 2014; Zhang et al., 2014; Kujur et al., 2015).

Keeping above prospects in view, the current study employed an integrated strategy combining traditional metaQTL analysis and regional association mapping (**Figure 1**) to validate and narrow-down long confidence intervals of six important metaQTL regions governing grain size/weight in Indian *indica* and aromatic rice accessions. This study will take us one step closer toward identifying functionally relevant potential genes/alleles underlying known grain size/weight QTLs mapped on Indian germplasm accessions-derived low-density genetic linkage maps. This will essentially assist us to select superior natural allelic variants from the potential QTLs/genes detected for grain size/weight and their effective deployment in yield improvement especially of Indian *indica* and aromatic rice accessions.

## META-ANALYSIS IDENTIFIES ROBUST GENOMIC REGIONS ASSOCIATED WITH RICE GRAIN SIZE/WEIGHT

To identify genetic basis of grain size/weight variability in Indian *indica* and aromatic rice accessions, primarily the meta-analysis



**FIGURE 1 | Illustrative representation of an integrated genomic strategy combining mQTL analysis with regional association mapping and expression profiling for delineation of trait-associated potential genomic regions (genes) in rice.** Yellow, green, and orange color arrows depict different independently identified Quantitative trait loci (QTLs) for similar trait. The horizontal green line within Manhattan plot symbolizes significance threshold and green shaded region represents the associated region determined based on regional LD. The solid green bar on the chromosome depicts the reduced mQTL interval. DEG: differentially expressed gene are indicated by green color while other genes within the delineated metaQTL region are illustrated by blue color.
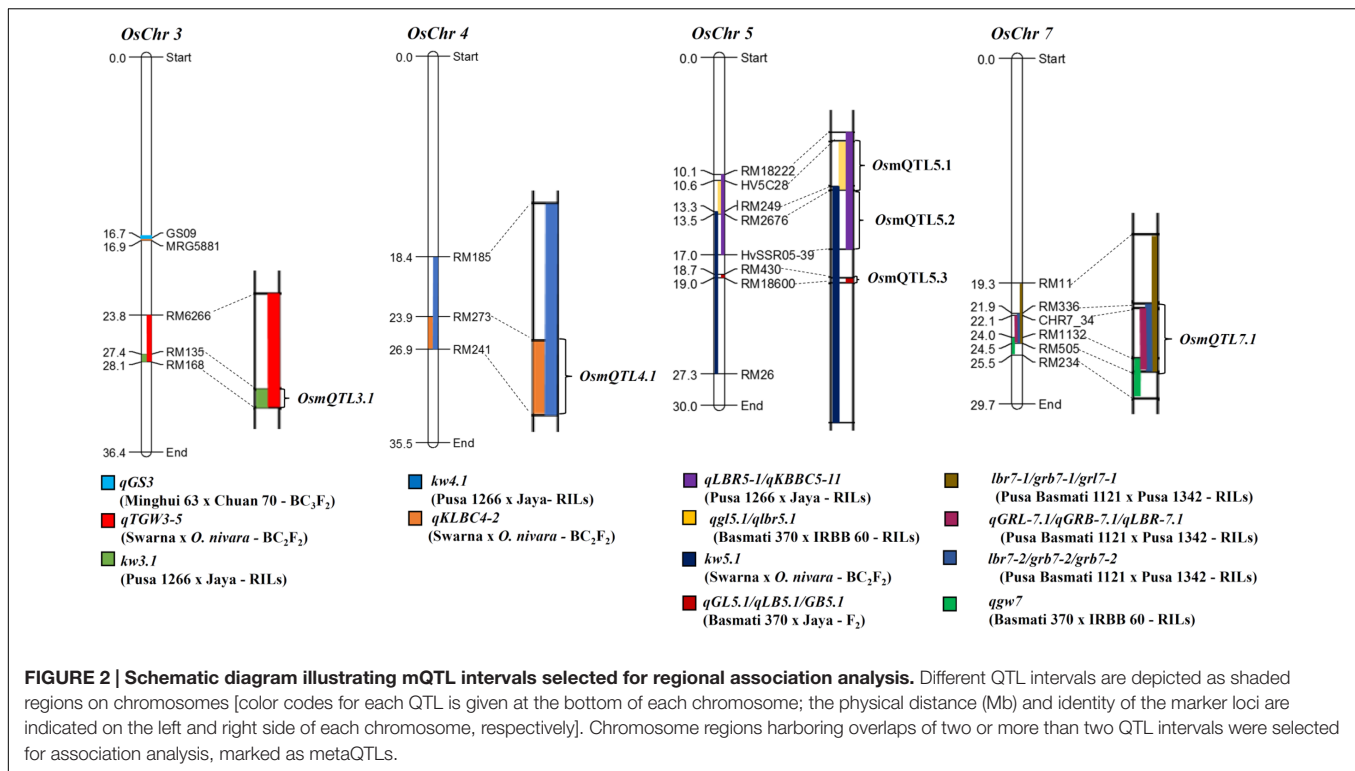
involving different QTL mapping studies was conducted. For this, extensive literature survey was conducted to prepare the comprehensive list of major grain size (GL, GWi, and GL/GWi) and GWt QTLs, which have been identified and mapped on different rice chromosomes using multiple preliminary and advanced generation mapping populations generated by inter-crossing among diverse Indian *indica* and aromatic rice accessions. The search resulted in identification of seven studies with 64 major/minor QTLs governing grain size and GWt traits mapped genetically on 12 rice chromosomes (Supplementary Table S1). The physical coordinates for the selected QTL intervals were defined by determining the physical positions (bp) of SSR (simple sequence repeat) markers flanking these QTLs as documented in RGAP (Rice Genome Annotation Project) pseudomolecule version 6.1[4]. Meta-analysis was further conducted using 64 identified grain size/weight QTLs in rice. For this, all these 64 QTLs were first projected on the reference rice genetic map and further meta-analysis was performed to detect robust metaQTLs on each chromosome using BioMercator V3 (Sosnowski et al., 2012) as per Deshmukh et al. (2012). Consequently, only those chromosomal metaQTL regions detected based on QTLs identified from at least two independent studies as well as with at least one of the QTL exhibiting LOD score ≥ 4 and phenotypic variance ≥ 10% were selected for further analysis. *GS3,* a well-known major grain size QTL interval served as a positive control and was also included as one of the metaQTL region (Fan et al., 2006).

Meta-analysis of 64 grain size/weight QTLs (by integrating their genetic and corresponding physical positions) detected six different mQTL (metaQTLs) regions (with at least one of the overlapping QTL exhibited a LOD score ≥ 4 and phenotypic variance ≥ 10) that were mapped on four different rice chromosomes. The confidence marker intervals for these six mQTLs ranged from 0.33 to 3.48 Mb. The structural annotation of the identified mQTLs revealed the presence of various protein-coding genes ranging from 42 to 526. A mQTL mapped on chromosome 5 (*OsmQTL5.3*) was found to be the shortest, spanning 0.33 Mb with 42 genes whereas another mQTL located on same chromosome (*OsmQTL5.2*) was longest spanning 3.48 Mb containing 526 genes (**Table 1** and **Figure 2**). The efficacy of meta-analysis in refining the confidence intervals for previously known QTLs as well as for validating their effects across different genetic backgrounds and environments is well-demonstrated (Goffinet and Gerber, 2000; Wu and Hu, 2012). Summarily, meta-analysis of major grain size/GWt QTLs in our study effectively identified six promising mQTL regions that are validated in different genetic backgrounds and environments. Therefore, these mQTLs have potential to serve as most efficient candidate targets either for fine mapping or their direct utilization in MAS for rice genetic enhancement as suggested previously (Ballini et al., 2008; Courtois et al., 2009; Khowaja et al., 2009; Wang et al., 2010).

---

[4]http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/

**TABLE 1 | Summary of metaQTL regions selected for association analysis.**

| metaQTL identities | Overlapping grain size/weight quantitative trait loci (QTLs) | Chromosomes | Start physical position (bp) | End physical position (bp) | Genomic intervals (bp) | Markers flanking metaQTLs | Number of genes within QTL intervals | References on QTL mapping studies |
|---|---|---|---|---|---|---|---|---|
| OsmQTL3.1 | kw3.1 and qTGW3-5 | OsChr3 | 27411818 | 28091534 | 679716 | RM135-RM168 | 108 | Marathi et al., 2012; Swamy et al., 2012 |
| OsmQTL4.1 | Kw4.1 and qKBLC4-2 | OsChr4 | 23863414 | 26857374 | 2993960 | RM273-RM241 | 500 | Guleria et al., 2012; Swamy et al., 2012 |
| OsmQTL5.1 | qLBR5-1, qKBBC5-1, qgl5.1, qlbr5.1, and kw5.1 | OsChr5 | 10600000 | 13480036 | 2880036 | HV5C28-RM2676 | 416 | Guleria et al., 2012; Swamy et al., 2012; Shanmugavadivel et al., 2013 |
| OsmQTL5.2 | Kw5.1, qLBR5-1, and qKBBC5-1 | OsChr5 | 13480036 | 16962564 | 3482528 | RM2676-HvSSR05-39 | 526 | Guleria et al., 2012; Swamy et al., 2012 |
| OsmQTL5.3 | Kw5.1, qGL5.1, qLB5.1, and GB5.1 | OsChr5 | 18691417 | 19018217 | 326800 | RM430-RM18600 | 42 | Swamy et al., 2012; Vemireddy et al., 2015 |
| OsmQTL7.1 | Lbr7, grb7-1, grl7-1, qGRL-7.1, qgw7, lbr7-2, grb7-2, and grl7-2 | OsChr7 | 22127494 | 24527013 | 2399519 | RM336-RM505 | 394 | Amarawathi et al., 2008; Singh et al., 2012; Shanmugavadivel et al., 2013 |
| GS3 | GS3 QTL (positive control) | OsChr3 | 16690372 | 16878706 | 188334 | GS09-MRG5881 | 23 | Fan et al., 2006 |

**FIGURE 2 | Schematic diagram illustrating mQTL intervals selected for regional association analysis.** Different QTL intervals are depicted as shaded regions on chromosomes [color codes for each QTL is given at the bottom of each chromosome; the physical distance (Mb) and identity of the marker loci are indicated on the left and right side of each chromosome, respectively]. Chromosome regions harboring overlaps of two or more than two QTL intervals were selected for association analysis, marked as metaQTLs.

## AN OPTIMIZED COMBINATORIAL mQTL AND REGIONAL ASSOCIATION MAPPING STRATEGY EFFICIENTLY SCALES-DOWN A KNOWN GS3 MAJOR QTL REGION

To optimize regional association mapping, association analysis for each of six mQTL genomic regions was performed independently using the individual SNP- and haplotype-based methods implemented in GAPIT and hapQTL, respectively. For this, sequence variation (SNPs) data of 505 accessions belonging to three different rice populations (*indica*, *japonica*, and *aus*) for selected mQTL genomic regions were retrieved from RiceVarMap database[5]. SNPs as well as accessions with greater than 1% missing data or minor allele frequency (MAF) < 5% were discarded. Filtered high-quality SNPs and accessions were further considered for association analysis (Supplementary Table S1). The phenotyping data of three major grain size/weight-related traits (GL, GWi, and GWt) were measured in a panel of 505 diverse rice accessions belonging to *indica*, *japonica,* and *aus* populations. Wider phenotypic variation (GL: varying from 6.71 to 11.23 mm with a mean 8.56 mm, GWi: 1.98 to 3.44 mm with a mean 2.87 mm and GWt: 15.38 to 37.47 g with a mean 24.27 g) with normal frequency distribution for all three gain size/weight traits among selected rice accessions were observed. To begin with the regional association analysis

[5]http://ricevarmap.ncpgr.cn/

was performed in a known mQTL region, which harbors *GS3*, a major gene regulating grain size especially in *indica* population, has been cloned and characterized earlier by fine mapping/map-based isolation, association mapping and functional genomic approaches (Fan et al., 2006, 2009; Mao et al., 2010). These previous studies demonstrated the significant role of a SNP allelic variant (C/A) causing non-synonymous amino acid substitution (serine to threonine) in a G-protein γ subunit gene delineated at a *GS3* QTL region in controlling long and short grain size differentiation in rice.

For regional association analysis, the SNP genotyping information and grain size/weight (GL, GWi, and GWt) phenotyping data of all 505 accessions as well as *indica* subset (365) of total 505 accessions were integrated using a CMLM interface of GAPIT independently. This analysis successfully revealed significant association (FDR corrected *P*-value of 0.002) of a known causal non-synonymous SNP (C/A at 16732087 bp on chromosome 3) reported previously in a G-protein γ subunit gene underlying a major *GS3* mQTL with grain size trait in *indica* population (but not in analysis with all 505 accessions) (**Figure 3A** and **Table 2**). The concerned SNP explained about 10% variation in GWi phenotype with A-allele positively affecting the grain size in rice. Like-wise, genetic association analysis was conducted for a *GS3* mQTL region for all three grain size/weight traits (GL, GWi, and GWt) using hapQTL (Xu and Guan, 2014), which employs haplotype-based strategy for establishing marker-trait association. This analysis identical to GAPIT detected significant association (Log$_{10}$ BF of 6.51 with *P*-value 3.18)

**FIGURE 3 | Individual SNP- and haplotype-based regional association analysis of a known *GS3* QTL region. (A)** Individual SNP-based regional association analysis. The blue color circle depicts different SNPs whereas the green line represents significance threshold. **(B)** Haplotype-based association analysis. The orange circles represent single SNPs; the blue circles represent core SNPs within haplotypes (HAP) used for association. The horizontal dotted yellow line represents the significance threshold. **(C)** LD block (based on mLHS) around the most significant SNP (marked by "+") is denoted with shaded lines. The top two ancestral haplotypes most significantly associated with grain length phenotype are designated as circles, with darker circles representing larger ancestral allele frequencies.
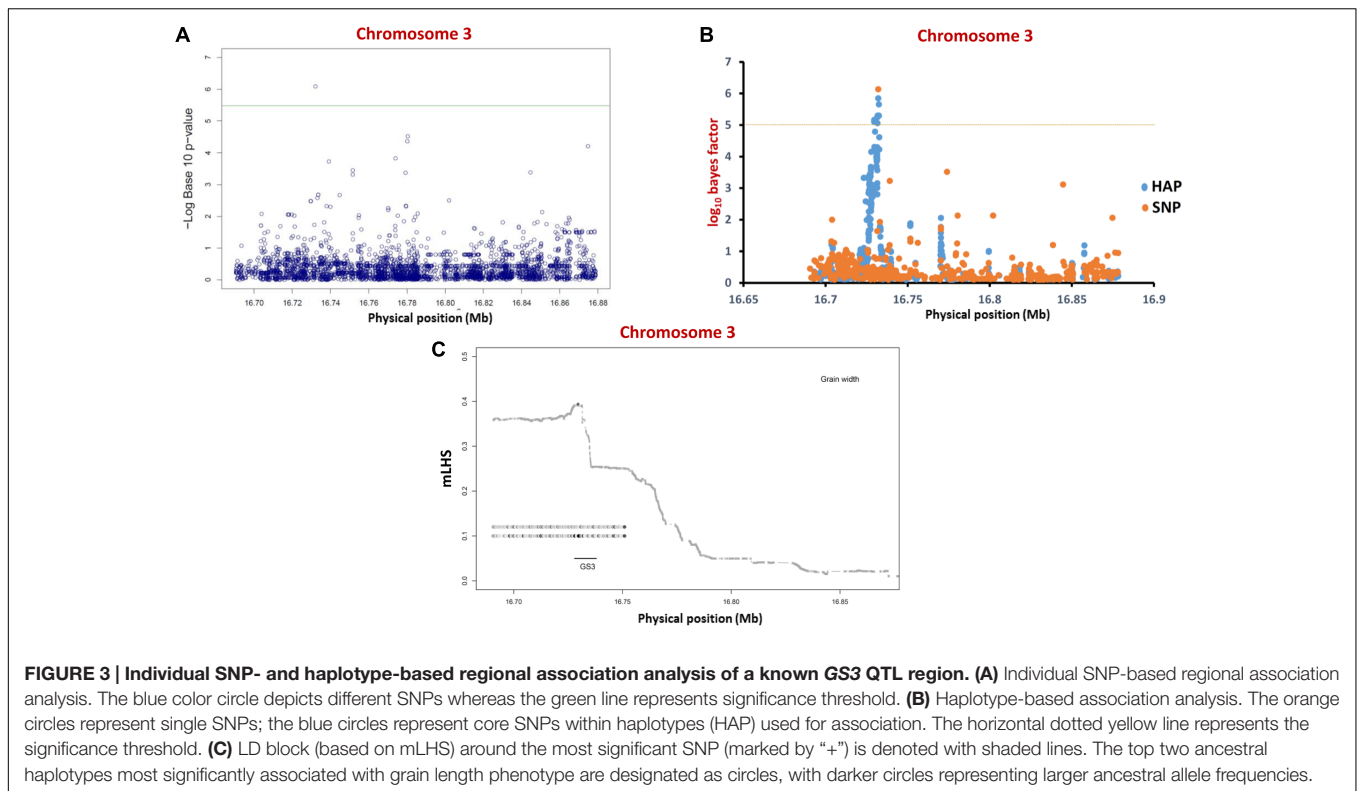
**TABLE 2 | Summary of individual-SNP based regional association analysis in metaQTL intervals.**

| metaQTL identities | Chromosomes | Population group (number of accessions) | Traits associated | Most significant SNPs* | Annotation | FDR corrected *p*-value | $R^2$ |
|---|---|---|---|---|---|---|---|
| *GS3* | *OsChr3* | *Indica* (365) | Grain width | sf0316732087 | Intergenic | 0.0025 | 0.101 |
| *OsmQTL5.3* | *OsChr5* | *Indica, japonica,* and *aus* (505) | Grain length | sf0518872800 | Intergenic | 0.00093 | 0.059 |

*SNP IDs are same as used in RiceVarMap database.*

between G-protein γ subunit gene-derived known causal non-synonymous SNP (C/A at 16732087 bp on chromosome 3) underlying a *GS3* QTL with grain size in *indica* subpopulation (but not in analysis with all 505 accessions) (**Figure 3B** and **Table 3**). Subsequently, haplotype block was constructed around this significant trait-associated SNP that harbors five protein coding genes including the one coding for known G-protein γ subunit gene (**Figure 3C** and **Table 3**). These outcomes are in accordance with earlier rice grain size trait association mapping studies (Takano-Kai et al., 2009; Zhao et al., 2011; McCouch et al., 2016). Henceforth, individual and haplotype-based regional association mapping strategies adopted in our study have proficiency to detect more accurate and non-spurious marker-trait association for grain size/weight in rice. Consequently, this approach has potential in rapid delineation of functionally relevant genes and natural alleles underlying not only for known mQTL intervals (like *GS3* QTL interval) but also targeting diverse major/minor grain size/weight mQTLs for which major effect genes are unknown in rice.

# INDIVIDUAL SNP-BASED REGIONAL ASSOCIATION ANALYSIS IDENTIFIES CANDIDATE GENES UNDERLYING GRAIN SIZE/WEIGHT mQTLs

For CMLM-based regional association analysis, high-quality SNP genotyping (of each selected mQTL genomic regions) and grain size/weight phenotyping (GL, GWi, and GWt) data for all 505 accessions (representing *indica*, *japonica,* and *aus* populations) were combined with their relative kinship matrix (K) and PCA [principal component (PC) analysis] information using P3D/compressed mixed linear model (CMLM) in GAPIT (Lipka et al., 2012; Kumar et al., 2015; Upadhyaya et al., 2015). The quantile-quantile plot generated in GAPIT output was employed to visualize relative distribution of observed and expected $-\log_{10}$ (*P*)-value for each SNP marker-trait association. The *P*-value threshold of significance was corrected for multiple comparisons based on false discovery rate (FDR cut-off $\leq 0.05$) (Benjamini and Hochberg, 1995). Further, association analysis was performed

**TABLE 3 | Summary of haplotype-based regional association analysis in metaQTL intervals.**

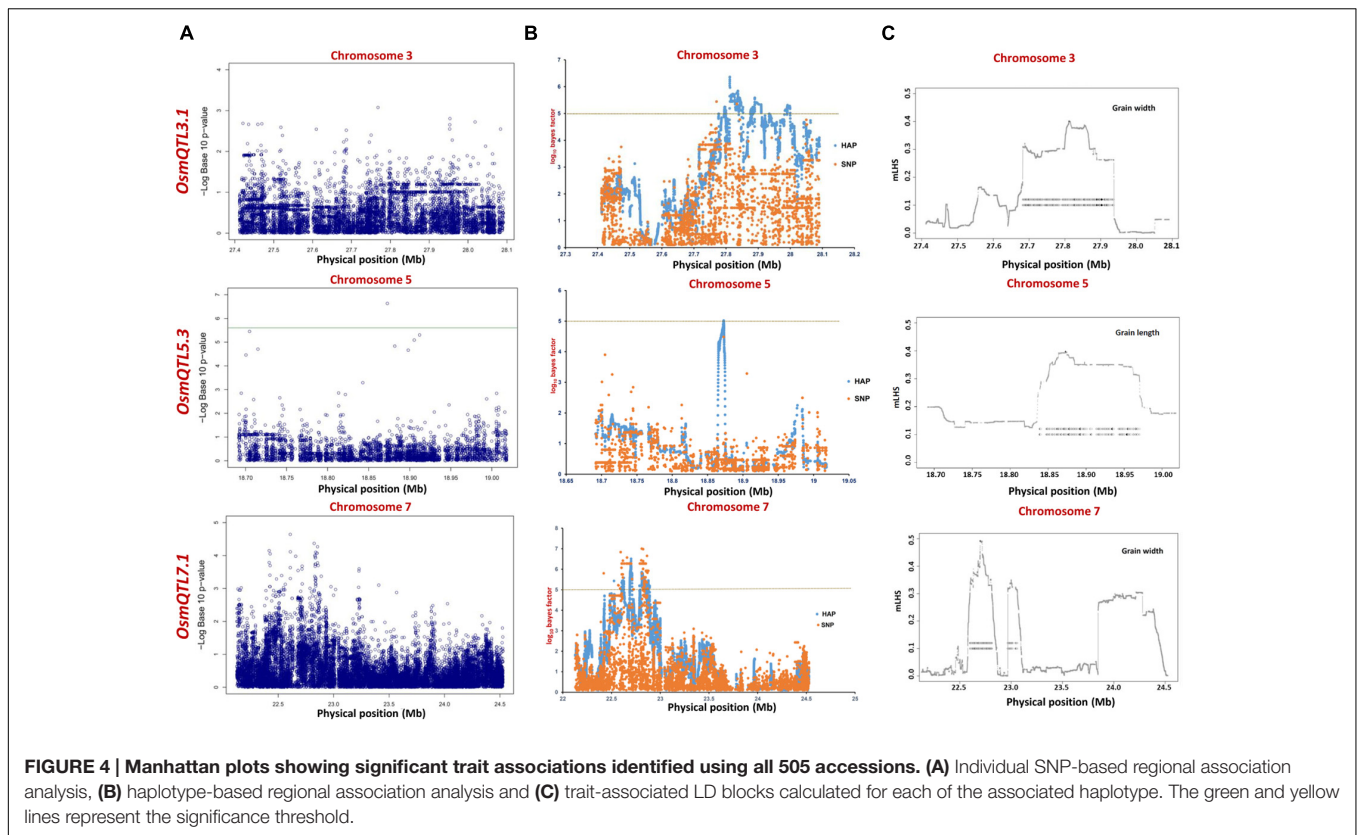| metaQTL identities | Chromo somes | metaQTL intervals (Mb) | Number of genes within metaQTL intervals | Population group (number of accessions) | Trait associated | Most significant SNPs* | Structural annotation | BF2 | P2 | BF1 | P1 | Delineated metaQTL intervals (Mb) | Number of genes within delineated intervals |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GS3 | OsChr3 | 16.69–16.88 (0.19) | 23 | Indica (365) | Grain width | sf0316732087 | Intergenic | 5.85 | 3.18 | 6.14 | 3.89 | 16.66–16.75 (0.09) | 5 |
| OsmQTL3.1 | OsChr3 | 27.41–28.09 (0.68) | 108 | Indica, japonica, and aus (505) | Grain width | sf0327811770 | Intergenic | 6.36 | 2.30 | −0.19 | 0.66 | 27.68–27.93 (0.25) | 34 |
| OsmQTL5.3 | OsChr5 | 18.69–19.01 (0.32) | 42 | Indica, japonica and aus (505) | Grain length | sf0518872800 | Intergenic | 5.01 | 2.08 | 4.482 | 4.76 | 18.84–18.97 (0.13) | 12 |
| OsmQTL7.1 | OsChr7 | 22.13–22.53 (2.4) | 394 | Indica, japonica, and aus (505) | Grain length | sf0722700042 | Intergenic | 6.51 | 2.78 | 0.796 | 0.41 | 22.60–23.06 (0.46) | 58 |
| OsmQTL3.1 | OsChr3 | 27.41–28.09 (0.68) | 108 | Indica (365) | Grain width | sf0327686607 | Intergenic | 5.81 | 2.03 | 0.138 | 0.50 | 27.67–27.70 (0.03) | 4 |
| OsmQTL5.3 | OsChr5 | 18.69–19.01 (0.32) | 42 | Indica (365) | Grain weight | sf0518859830 | Intergenic | 3.0 | 0.98 | 5.17 | 2.40 | 18.72–18.95 (0.23) | 27 |

*SNP IDs are same as used in RiceVarMap database.

using the genotyping and grain size/weight phenotyping data and diversity-related information (K and PCA with no PC covariates) of 365 accessions belonging to *indica* population following the aforesaid methods of GAPIT.

The integration of SNP genotyping data from six selected *OsmQTL* regions with grain size/weight-related phenotyping information (GL, GWi, and GWt) of all 505 accessions successfully detected significant association of a genomic locus underlying *OsmQTL5.3* with GL in rice (**Figure 4A** and **Table 2**). However, significant association could not be detected for any other *OsmQTL* regions with any of grain size/weight phenotypes when analysis was performed for *indica* subpopulation. A most significant SNP (G/T) (SNP ID: sf0518872800) within *OsmQTL5.3* was found to be located in the intergenic region between LOC_Os05g32430 (encoding NLI interacting factor-like phosphatase protein) and LOC_Os05g32440 (unknown expressed protein). Considering the known estimated average LD decay rate in rice (Huang et al., 2010), 20 protein-coding genes which lie within 100 kb up- and down-stream regions of a most significant SNP (G/T) (at 18872800 bp on a chromosome 5) were considered as putative candidates in regulating rice GL. None of the genomic loci underlying the rest five *OsmQTL* regions exhibited significant association with any of the three grain size/weight traits, when analyzed in all 505 accessions or 365 *indica* subpopulation (**Figures 4A**, **5A** and **Table 2**).

## HAPLOTYPE-BASED REGIONAL ASSOCIATION ANALYSIS DETECTS NOVEL ASSOCIATIONS AT GRAIN SIZE/WEIGHT mQTLs

Haplotype-based regional association analysis was performed by integrating high-quality SNP (for each of the selected mQTL genomic regions) genotyping data with grain size/weight phenotyping information (GL, GWi, and GWt) of all 505 rice accessions belonging to three different rice populations (*indica, japonica,* and *aus*) using hapQTL. To determine population structure, PCA was performed with filtered high-quality SNPs using TASSEL 5.0 and top three PCs were used as covariates in regression model during analysis. Threshold for significance BF (Bayesian factor) was selected as maximum $\log_{10}$ BF (bayes factor) under null for all three grain size/weight phenotypes combined plus one (Xu and Guan, 2014). In entire analysis, number of upper level clusters (S), which represent source population were set to 3 and lower level clusters (K) signifying the ancestral haplotypes, were set to 10. Multiple (10) EM (expectation-maximization) runs were used to avoid uncertainty in inferring LD. LD blocks around the most significant associations were quantified by computing mLHS (local haplotype sharing between SNP markers) between the most significant SNP marker loci and the all remaining SNPs. Largest genomic regions, which have an mLHS value 2.5 times greater than background mLHS (i.e., 0.25 for $K = 10$) value, were defined as trait-associated LD blocks. Like-wise,
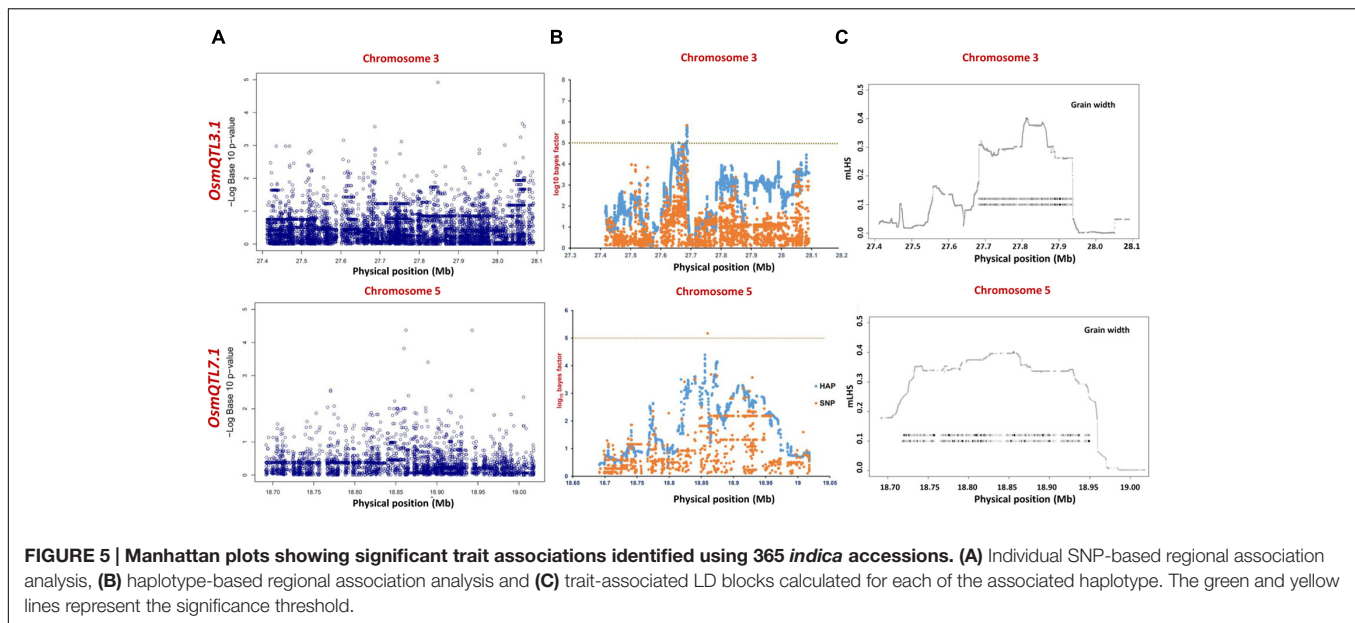
**FIGURE 4 | Manhattan plots showing significant trait associations identified using all 505 accessions. (A)** Individual SNP-based regional association analysis, **(B)** haplotype-based regional association analysis and **(C)** trait-associated LD blocks calculated for each of the associated haplotype. The green and yellow lines represent the significance threshold.

haplotype-based association analysis (with no PC covariates) was performed independently by integrating genotyping data with grain size/weight phenotyping information of 365 *indica* accessions as per aforementioned methods.

The haplotype-based association analysis of six selected *OsmQTL* regions using hapQTL in all 505 accessions detected significant association of *OsmQTL5.3* (SNP T/G at 18872800 bp on chromosome 5 with $\log_{10}$ BF of 5.02 for the haplotype) with GL (**Figure 4B** and **Table 3**). The outcome is agreed-well with our aforementioned CMLM-based association analysis with all 505 accessions, thereby further augmenting the authenticity of detected marker-trait association in rice. In addition, hapQTL analysis detected association of *OsmQTL3.1* (SNP G/A at 327811770 bp on chromosome 3 with $\log_{10}$ BF of 6.36 for haplotype) and *OsmQTL7.1* (SNP A/G at 22700042 bp on chromosome 7 with $\log_{10}$ BF of 6.51 for haplotype) with GWi and GL, respectively. Subsequent analysis with *indica* subpopulation could only detect significant association of *OsmQTL3.1* (SNP T/C at 27686607 bp on chromosome 3 with $\log_{10}$ BF of 5.81 for the haplotype) with GWi and *OsmQTL5.3* (SNP G/A at 18859830 bp on chromosome 5 with $\log_{10}$ BF of 5.17 for the haplotype) with GWt (**Figure 5B** and **Table 3**). Interestingly, two additional novel marker-trait associations (detected both in *indica* subpopulations and in all 505 accessions) aside those identified using GAPIT were detected using hapQTL (**Figures 4A,B**). This highlights the power of haplotype-led association strategies in detecting novel associations, which are difficult to detect with individual SNP-based association methods due to its inability to deal

with allelic as well as locus heterozygosity (Xu and Guan, 2014). The mLHS was further quantified around core SNPs (of most significant haplotypes) to determine the LD blocks (trait-associated regions) harboring the candidate genes associated with the analyzed grain size/weight trait. The LD block quantified using all 505 accessions around the most significant SNP (MSP) within *OsmQTL3.1* (G/A at 327811770 bp on chromosome 3) spanned from 27.68 to 27.94 Mb containing 34 genes whereas those calculated around the MSP within *OsmQTL5.3* (T/G at 18872800 bp on chromosome 5) and *OsmQTL7.1* extended from 18.84 to 18.97 Mb and 22.60 to 23.06 Mb harboring 12 and 58 rice genes, respectively (**Figure 4C** and **Table 3**). Like-wise, LD bocks were also estimated around MSPs detected with *indica* subpopulation using 365 *indica* accessions. The LD block around MSP detected within *OsmQTL3.1* (T/C at 27686607 bp on chromosome 3) extended from 27.67 to 27.71 Mb with only four genes, whereas those around MSP detected within *OsmQTL5.3* (T/G at 18872800 bp on chromosome 5) spanned from 18.72 to 18.95 Mb with 37 genes (**Figure 5C** and **Table 3**).

Summarily, haplotype-based association analysis and subsequent LD block quantification using genotyping data of all 505 accessions significantly narrowed-down original confidence marker intervals for *OsmQTL3.1, OsmQTL5.3,* and *OsmQTL7.1* from 0.68 (108 genes), 0.33 (42 genes), and 2.40 (394 genes) Mb to 0.25 (34 genes), 0.13 (12 genes), and 0.46 (58 genes) Mb, respectively (**Figure 4** and **Table 3**). In case of similar analysis with *indica* subpopulation, the intervals for *OsmQTL3.1*

**FIGURE 5 | Manhattan plots showing significant trait associations identified using 365 *indica* accessions. (A)** Individual SNP-based regional association analysis, **(B)** haplotype-based regional association analysis and **(C)** trait-associated LD blocks calculated for each of the associated haplotype. The green and yellow lines represent the significance threshold.

and *OsmQTL5.3* were scaled-down from 2.4 (108 genes) and 0.33 (42 genes) Mb to 0.03 (4 genes) and 0.23 (27 genes) Mb, respectively (**Figure 4** and **Table 3**). In case of *OsmQTL5.3*, mLHS calculation with all 505 accessions produced smaller LD block as compared to those calculated with 365 *indica* accessions, showing evidence of inverse correlation between number of accessions used for mLHS calculation and size of LD block as suggested previously (Xu and Guan, 2014). However, in case of *OsmQTL3.1*, the LD block was found much smaller in *indica* analysis, which can be explained by lower LD in *indica* subset (365 accessions) (McCouch et al., 2016). This provides further opportunity to narrow-down these mQTL regions by increasing the number of germplasm accessions for haplotype-based association analysis and subsequent quantification of LD blocks as well as through subpopulation-specific association analysis. Moreover, the mLHS-based quantification of LD blocks has a greater potential to scale-down the longer mQTL intervals into shorter genomic intervals as compared to *R* square-based quantification used traditionally for association analysis. Therefore, haplotype-based regional association analysis has profound efficacy for rapid delineation of candidate genes and natural alleles underlying major QTLs in rice (Xu and Guan, 2014).

## IN SILICO GLOBAL EXPRESSION PROFILING DELINEATES CANDIDATE GENES AT mQTL INTERVALS NARROWED-DOWN BY REGIONAL ASSOCIATION ANALYSIS

Integration of expression profiling-based functional evidences with association analysis is very popular approach for further delineating candidate genes within associated LD blocks and

has been successfully employed in variety of crops (Das et al., 2016). In order to further prioritize candidate genes delineated by regional association mapping, global *in silico* expression profiling of these genes were performed utilizing the whole genome microarray profiling data for diverse vegetative and reproductive stages of Nipponbare cultivar available at RiceXPro database[6]. This includes microarray expression profiles of different tissues such as total inflorescence, floral organs (anther, pistil, lemma, and palea) at early reproductive development stages, embryo and endosperm during early grain development stages as well as seed maturation stages.

Expression profiles of all candidate genes within delineated mQTL regions (validated and narrowed-down by regional association mapping) were compared with aforesaid various vegetative and reproductive tissues using global *in silico* expression analysis. Most of the genes displayed differential expression when compared between various tissues at different stages of reproductive development. Some of the genes within the delineated genomic intervals were found to express at very high levels in spikelet hull (lemma and palea) during early reproductive development stages and therefore, can be involved in spikelet hull development and essentially in seed development. For instance, one gene (LOC_Os05g32460) delineated within reduced *OsmQTL5.3* region (0.1 Mb), which codes for VQ domain-containing protein and another gene (LOC_Os03g48660) located within reduced *OsmQTL3.1* interval (0.13 Mb) encoding DUF966 protein were found to be highly expressed in both lemma and palea during early stages of seed development and in endosperm during seed maturation (Supplementary Figure S1). Interestingly, the genes encoding VQ-domain containing proteins have been previously reported to be involved in seed development and maturation in plants (Wang et al., 2010; Jing and Lin, 2015).

---

The orthologs of DUF966 protein coding gene (LOC_Os03g48660) are known to be regulated by auxin, which have well established role in seed development along with other hormones (Locascio et al., 2014). Therefore, both of these genes can be considered as most promising candidates for further functional validation.

Collectively, the outcomes of the present study indicate the potential utility of high-resolution association mapping in reduction of grain size/weight mQTL intervals derived from diverse QTL mapping studies in rice. The reduced grain size/weight major mQTLs can be directly utilized in MAS for rice genetic improvement given their consistent effects in different genetic backgrounds and environments. Moreover, two potential candidate genes exhibiting similarity with earlier characterized *Arabidopsis* known grain size/weight orthologs (Locascio et al., 2014), delineated by us employing an integrated strategy (mQTL region-specific association analysis and expression profiling) can further be used for functional validation studies and translational genomic analyses for rice genetic enhancement. With the availability of huge sequencing- and array-based SNP genotyping information for numerous phenotypically well-characterized germplasm accessions of major food crops such as maize, wheat, soybean etc., the developed strategy can be readily extended to these crop plants for identification of valid candidate genes and natural alleles associated with important agronomic traits. This essential information will be of immense use in quick molecular genetic dissection of complex quantitative traits to expedite genomics-assisted crop improvement.

## AUTHOR CONTRIBUTIONS

AD written the manuscript and performed all analysis. RS helped in data analysis. AS, SP, and AT conceptualized and guided the study.

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00807/full#supplementary-material

## REFERENCES

Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., et al. (2015). SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 43, 1023–11027. doi: 10.1093/nar/gku1039

Ali, M. L., McClung, A. M., Jia, M. H., Kimball, J. A., McCouch, S. R., and Eizenga, G. C. A. (2011). A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Sci.* 51, 2021–2035. doi: 10.2135/cropsci2010.11.0641

Amarawathi, Y., Singh, R., Singh, A. K., Singh, V. P., Mohapatra, T., Sharma, T. R., et al. (2008). Mapping of quantitative trait loci for Basmati quality traits in rice (*Oryza sativa* L.). *Mol. Breed.* 21, 49–65. doi: 10.1007/s11032-007-9108-8

Anand, D., Baunthiyal, M., Singh, A., Gopala Krishnan, S., Singh, N. K., Prabhu, K. V., et al. (2013). Validation of gene based marker-QTL association for grain dimension traits in rice. *J. Plant Biochem. Biotechnol.* 22, 467–473. doi: 10.1007/s13562-012-0176-4

Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., et al. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* 1:e60. doi: 10.1371/journal.pgen.0010060

Bajaj, D., Upadhyaya, H. D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Identification of candidate genes for dissecting complex branch number trait in chickpea. *Plant Sci.* 245, 61–70. doi: 10.1016/j.plantsci.2016.01.004

Ballini, E., Morel, J. B., Droc, G., Price, A., Courtois, B., Notteghem, J. L., et al. (2008). Genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Mol. Plant Microbe Interact.* 21, 859–868. doi: 10.1094/MPMI-21-7-0859

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., et al. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 6:e1000940. doi: 10.1371/journal.pgen.1000940

Breseghello, F., and Sorrells, M. E. (2006). Association analysis as a strategy for improvement of quantitative traits in plants. *Crop Sci.* 46, 1323–1330. doi: 10.2135/cropsci2005.09-0305

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., et al. (2014). A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant* 7, 541–553. doi: 10.1093/mp/sst135

Courtois, B., Ahmadi, N., Khowaja, F. S., Price, A. H., Rami, J. F., Frouin, J., et al. (2009). Rice root genetic architecture: meta-analysis from a drought QTL database. *Rice* 2, 115–128. doi: 10.1007/s12284-009-9028-9

Das, S., Singh, M., Srivastava, R., Bajaj, D., Saxena, M. S., Rana, J. C., et al. (2016). mQTL-seq delineates functionally relevant candidate gene harbouring a major QTL regulating pod number in chickpea. *DNA Res.* 23, 53–65.

Deshmukh, R., Sonah, H., Kondawar, V., Tomar, R. S. S., and Deshmukh, N. (2012). Identification of meta quantitative trait loci for agronomical traits in rice (*Oryza sativa*). *Indian J. Genet. Plant Breed.* 72, 264–270.

Dixit, N., Dokku, P., Amitha Mithra, S. V., Parida, S. K., Singh, A. K., Singh, N. K., et al. (2013). Haplotype structure in grain weight gene *GW2* and its association with grain characteristics in rice. *Euphytica* 192, 55–61. doi: 10.1007/s10681-012-0852-4

Ebana, K., Kojima, Y., Fukuoka, S., Nagamine, T., and Kawase, M. (2008). Development of mini core collection of Japanese rice landrace. *Breed. Sci.* 58, 281–291. doi: 10.1270/jsbbs.58.281

Ehrenreich, I. M., Hanzawa, Y., Chou, L., Roe, J. L., Kover, P. X., and Purugganan, M. D. (2009). Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics* 183, 325–335. doi: 10.1534/genetics.109.105189

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

Fan, C., Yu, S., Wang, C., and Xing, Y. A. (2009). Causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. *Theor. Appl. Genet.* 118, 465–472. doi: 10.1007/s00122-008-0913-1

Goffinet, B., and Gerber, S. (2000). Quantitative trait loci: a meta-analysis. *Genetics* 155, 463–473.

Guleria, S., Sharma, V., Marathi, B., Anand, S., Singh, N. K., Mohapatra, T., et al. (2012). Molecular mapping of grain physico-chemical and cooking quality traits using recombinant inbred lines in rice (*Oryza sativa* L.). *J. Plant Biochem. Biotechnol.* 21, 1–10. doi: 10.1007/s13562-011-0064-3

Gupta, P. K., Kulwal, P. L., and Jaiswal, V. (2014). Association mapping in crop plants: opportunities and challenges. *Adv. Genet.* 85, 109–147. doi: 10.1016/b978-0-12-800271-1.00002-0

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* 45, 707–711. doi: 10.1038/ng.2612

Jing, Y., and Lin, R. (2015). The VQ motif-containing protein family of plant-specific transcriptional regulators. *Plant Physiol.* 169, 371–378. doi: 10.1104/pp.15.00788

Khowaja, F. S., Norton, G. J., Courtois, B., and Price, A. H. (2009). Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis. *BMC Genomics* 10:276. doi: 10.1186/1471-2164-10-276

Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M., et al. (2015). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* 5:9468. doi: 10.1038/srep09468

Kumar, V., Singh, A., Mithra, A. S. V., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046

Li, X., Yan, W., Agrama, H., Hu, B., Jia, L., Jia, M., et al. (2010). Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica* 138, 1221–1230. doi: 10.1007/s10709-010-9521-5

Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., et al. (2016). Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Front. Plant. Sci.* 7:833. doi: 10.3389/fpls.2016.00833

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Locascio, A., Roig-Villanova, I., Bernardi, J., and Varotto, S. (2014). Current perspectives on the hormonal control of seed development in *Arabidopsis* and maize: a focus on auxin. *Front. Plant Sci.* 5:412. doi: 10.3389/fpls.2014.00412

Lu, L., Shao, D., Qiu, X., Sun, L., Yan, W., Zhou, X., et al. (2013). Natural variation and artificial selection in four genes determine grain shape in rice. *New Phytol.* 200, 1269–1280. doi: 10.1111/nph.12430

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., et al. (2010). Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19585–19590. doi: 10.1073/pnas.1006105107

Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shangguan, Y., et al. (2013). An-1 encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell* 25, 3360–3376. doi: 10.1105/tpc.113.113589

Mackay, T. F., Stone, E. A., and Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* 10, 565–577. doi: 10.1038/nrg2612

Mammadov, J., Sun, X., Gao, Y., Ochsenfeld, C., Bakker, E., Ren, R., et al. (2015). Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (*Zea mays* L.). *BMC Genomics* 10:916. doi: 10.1186/s12864-015-2171-3

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

Marathi, B., Guleria, S., Mohapatra, T., Parsad, R., Mariappan, N., Kurungara, V. K., et al. (2012). QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (*Oryza sativa* L.). *BMC Plant Biol.* 12:137. doi: 10.1186/1471-2229-12-137

McCouch, S., Wright, M., Tung, C. W., Maron, L., McNally, K., Fitzgerald, M., et al. (2016). Open access resources for genome wide association mapping in rice. *Nat. Commun.* 7:10532. doi: 10.1038/ncomms10532

Meijón, M., Satbhai, S. B., Tsuchimatsu, T., and Busch, W. (2014). Genome-wide association study using cellular traits identifies a new regulator of root development in *Arabidopsis*. *Nat. Genet.* 46, 77–81. doi: 10.1038/ng.2824

Ruggieri, V., Francese, G., Sacco, A., D'Alessandro, A., Rigano, M. M., Parisi, M., et al. (2014). An association mapping approach to identify favourable alleles for tomato fruit quality breeding. *BMC Plant Biol.* 14:337. doi: 10.1186/s12870-014-0337-9

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031

Shanmugavadivel, P. S., Amitha Mithra, S. V., Dokku, P., Anand Raj Kumar, K., Rao, G. J. N., Singh, A., et al. (2013). Mapping quantitative trait loci (QTL) for grain size in rice using a RIL population from Basmati × indica cross showing high segregation distortion. *Euphytica* 194, 401–416. doi: 10.1007/s10681-013-0964-5

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* 40, 1023–1028. doi: 10.1038/ng.169

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., and Hou, Q. (2016). *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48, 447–456. doi: 10.1038/ng.3518

Singh, N., Jayaswal, P. K., Panda, K., Mandal, P., Kumar, V.,Singh, B., et al. (2015). Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* 5:11600. doi: 10.1038/srep11600

Singh, R., Singh, A. K., Sharma, T. R., Singh, A., and Singh, N. K. (2012). Fine mapping of grain length QTLs on chromosomes 1 and 7 in Basmati rice (*Oryza sativa* L.). *J. Plant Biochem. Biotechnol.* 21, 157–166. doi: 10.1007/s13562-011-0080-3

Sneller, C. H., Mather, D. E., and Crepieux, S. (2009). Analytical approaches and population types for finding and utilizing QTL in complex plant populations. *Crop Sci.* 49, 363–380. doi: 10.2135/cropsci2008.07.0420

Song, J., Liu, Z., Hong, H., Ma, Y., Tian, L., Li, X., et al. (2016). Identification and validation of loci governing seed coat color by combining association mapping and bulk segregation analysis in soybean. *PLoS ONE* 7:e0159064. doi: 10.1371/journal.pone.0159064

Song, X., Huang, W., Shi, M., Zhu, M., and Lin, H. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39, 623–630. doi: 10.1038/ng2014

Sosnowski, O., Charcosset, A., and Joets, J. (2012). BioMercator V3: an upgrade of genetic map compilation and QTL meta-analysis algorithms. *Bioinformatics* 28, 2082–2083. doi: 10.1093/bioinformatics/bts313

Swamy, B. P., Kaladhar, K., Shobha Rani, N., Prasad, G. S., Viraktamath, B. C., Reddy, G. A., et al. (2012). QTL analysis for grain quality traits in 2 BC₂F₂ populations derived from crosses between *Oryza sativa* L. cv Swarna and 2 accessions of *O. nivara*. *J. Hered.* 103, 442–452. doi: 10.1093/jhered/esr145

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., et al. (2009). Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* 182, 1323–1334. doi: 10.1534/genetics.109.103002

Tiwari, K. K., Singh, A., Pattnaik, S., Sandhu, M., Kaur, S., Jain, S., et al. (2015). Identification of a diverse mini-core panel of Indian rice germplasm based on genotyping using microsatellite markers. *Plant Breed.* 134, 164–171. doi: 10.1111/pbr.12252

Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci. Rep.* 6:24050. doi: 10.1038/srep24050

Upadhyaya, H. D., Bajaj, D., Das, S., Saxena, M. S., Badoni, S., Kumar, V., et al. (2015). A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol. Biol.* 89, 403–420. doi: 10.1007/s11103-015-0377-z

Vemireddy, L. R., Noor, S., Satyavathi, V. V., Srividhya, A., Kaliappan, A., Parimala, S., et al. (2015). Discovery and mapping of genomic regions governing

economically important traits of Basmati rice. *BMC Plant Biol.* 15:207. doi: 10.1186/s12870-015-0575-5

Wang, A., Garcia, D., Zhang, H., Feng, K., Chaudhury, A., Berger, F., et al. (2010). The VQ motif protein IKU1 regulates endosperm growth and seed size in *Arabidopsis*. *Plant J.* 63, 670–679. doi: 10.1111/j.1365-313X.2010.04271.x

Wu, X. L., and Hu, Z. L. (2012). Meta-analysis of QTL mapping experiments. *Methods Mol. Biol.* 871, 145–171. doi: 10.1007/978-1-61779-785-9_8

Xu, H., and Guan, Y. (2014). Detecting local haplotype sharing and haplotype association. *Genetics* 197, 823–838. doi: 10.1534/genetics.114.164814

Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., Declerck, G., Derwent, P., et al. (2010). Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39, D1085–D1094. doi: 10.1093/nar/gkq1148

Zhang, D., Song, H., Cheng, H., Hao, D., Wang, H., Kan, G., et al. (2014). The acid phosphatase-encoding gene *GmACP1* contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* 10:e1004061. doi: 10.1371/journal.pgen.1004061

Zhang, P., Li, J., Li, X., Liu, X., Zhao, X., and Lu, Y. (2011). Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS ONE* 6:e27565. doi: 10.1371/journal.pone.0027565

Zhao, H., Yao, W., Ouyang, Y., Yang, W., Wang, G., and Lian, X. (2014). RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res. D* 10, 18–22.

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467. doi: 10.1038/ncomms1467

Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 48, 99–118. doi: 10.1146/annurev-genet-120213-092138