



OPEN

Understanding complex genetic architecture of rice grain weight through QTL-meta analysis and candidate gene identification

C. Anilkumar^{1✉}, Rameswar Prasad Sah¹, T. P. Muhammed Azharudheen¹, Sasmita Behera¹, Namita Singh², Nitish Ranjan Prakash³, N. C. Sunitha⁴, B. N. Devanna¹, B. C. Marndi¹, B. C. Patra¹ & Sunil Kumar Nair²

Quantitative trait loci (QTL) for rice grain weight identified using bi-parental populations in various environments were found inconsistent and have a modest role in marker assisted breeding and map-based cloning programs. Thus, the identification of a consistent consensus QTL region across populations is critical to deploy in marker aided breeding programs. Using the QTL meta-analysis technique, we collated rice grain weight QTL information from numerous studies done across populations and in diverse environments to find constitutive QTL for grain weight. Using information from 114 original QTL in meta-analysis, we discovered three significant Meta-QTL (MQTL) for grain weight on chromosome 3. According to gene ontology, these three MQTL have 179 genes, 25 of which have roles in developmental functions. Amino acid sequence BLAST of these genes indicated their orthologue conservation among core cereals with similar functions. MQTL3.1 includes the *OsAPX1*, *PDIL*, *SAUR*, and *OsASN1* genes, which are involved in grain development and have been discovered to play a key role in asparagine biosynthesis and metabolism, which is crucial for source-sink regulation. Five potential candidate genes were identified and their expression analysis indicated a significant role in early grain development. The gene sequence information retrieved from the 3 K rice genome project revealed the deletion of six bases coding for serine and alanine in the last exon of *OsASN1* led to an interruption in the synthesis of α -helix of the protein, which negatively affected the asparagine biosynthesis pathway in the low grain weight genotypes. Further, the MQTL3.1 was validated using linked marker RM7197 on a set of genotypes with extreme phenotypes. MQTL that have been identified and validated in our study have significant scope in MAS breeding and map-based cloning programs for improving rice grain weight.

Grain weight in rice is conventionally measured with one 'thousand grains weight (TGW)' and considered as a critical grain yield component trait. The TGW is influenced by associated grain traits like grain length, grain width, and thickness¹. These traits are genetically controlled by many minor/major genes and are quantitatively inherited through generations. Over decades, researchers focused their investigations on grain weight to improve the grain yield in rice^{2,3}. With the advent of molecular breeding technologies, plant breeding and botanical views have gained a deeper grasp of genetic control of TGW^{4,5}. Application of molecular markers and linkage mapping techniques paved the path for the identification of genetic loci controlling TGW. In the last two decades, more than a hundred major effect quantitative trait loci (QTL) exclusively controlling TGW have been mapped using different bi-parental populations over various environmental conditions. However, only a few of them have been validated and cloned using isogenic lines⁶. Generally, QTL detected in different populations evaluated under different environments exhibit differential effects of phenotypic variance that create ambiguity in exploiting them in marker-assisted breeding programs. Further, bi-parental population based QTL identification is strongly influenced by contrasting parents, type of mapping population, size of population, growing environment, and choice of marker system that hinders application of these QTLs in breeding programs^{7,8}.

¹ICAR-National Rice Research Institute, Cuttack, India. ²Indira Gandhi Krishi Vishwavidyalaya, Raipur, India. ³ICAR-Central Soil Salinity Research Institute, Regional Research Station, Canning Town, India. ⁴University of Agricultural Sciences, Bangalore, India. ✉email: anilcgp@gmail.com

QTL identified through bi-parental populations play a very modest role in rice breeding due to inconsistencies in growing environments, leading to discontinuities in their use⁹. Thus, identification of constitutive, robust and large-effect QTL over environments and across populations is critical for deployment in genomics assisted breeding programs¹⁰. Impact of $G \times E$ interaction, may bring change in level of correlation between phenotype and underlying QTL across environments¹¹. This is also augmented with effects from differed genetic backgrounds^{12,13}, thereby limiting their utilization. The efficacy of identified QTL in genomics-assisted breeding is further hampered by undesirable epistatic and modifier effects of distinct genetic backgrounds¹⁴. The possible reasons behind the inability to use such identified QTL are their implausible genetic predictions (with low LOD scores, predicted in small populations, faulty genotyping, limited environmental evaluation, etc.) and ignorance of interaction effects of QTLs, which are important in determining QTL stability. Hence, it is important to identify a trait-related consensus genomic region that includes more than one QTL and is surrogated by a single marker¹⁵. Such genomic regions are more useful to incorporate at least one or a few alleles from any of the constituent QTL to improve the target trait. Although, several QTL for grain weight characters in rice have been identified; leveraging the available information to enhance the insights of the underlying genetic mechanisms is more appropriate.

A computational technique, Meta-QTL (MQTL) analysis, which combines the reported QTL information precisely, has been developed and deployed to refine QTL positions by constructing a consensus map^{16,17}. The consensus QTL predicted using meta-analysis of a group of QTL with a confidence interval of 95% is denoted as meta-QTL. These MQTL show high consistency, a small confidence interval, and a major effect on traits that can be effectively utilized in MAS breeding programs upon their validation in a set of germplasm accessions. To date, only a few MQTL studies have been reported in rice for grain yield^{18,19}, plant characters^{20,21} and yield under drought stress¹⁴. Further, these studies were done by considering more than one trait (meta-trait), which is generally contributed by several component traits. Such analysis cannot conclusively describe the genetic and molecular mechanisms of component traits²². Hence, no meta-analysis of genetic factors associated with exclusive evidence on grain weight alone is available.

In this milieu, the present study was hypothesized the presence of consistent large effect QTL controlling grain weight on rice genome and tested the hypothesis by genome-wide meta-analysis with objectives, (1) prediction of MQTL exclusively for grain weight considering reported QTL, (2) identification of peak (tightly linked) molecular markers associated with MQTL, (3) validation of identified MQTL using linked peak markers in a set of genotypes comprising of two extreme phenotypes and (4) identification of candidate genes in the MQTL interval and predict their possible biological, molecular and cellular functions in rice grain weight development.

Results

Identification of MQTL for grain weight. The information from 114 QTL found in 22 distinct independent research undertaken over the previous two decades was used to create the consensus genetic map (Supplementary file 1). The consensus map consisted of 1272 SSR markers evenly distributed over all 12 rice chromosomes (Fig. 1). While gathering the information, every precaution was made to ensure that only thousand grain weight specific QTL were extracted and no other seed shape related features were taken into account. Among all the QTL mapping studies included in the analysis, the mapping populations included a minimum of 39 and a maximum of 353 individuals tested under various environments. The number of QTL found in each investigation ranged from 3 to 22 and were spread across all 12 rice chromosomes (Fig. 2). The most QTL were found on chromosome 3 (22 QTL), whereas the fewest were found on chromosome 12 (4 QTL) (Fig. 3A). Further, the 114 QTL analyzed were grouped based on PVE (%) and LOD score. Almost 49% of the initial QTL were found to have a LOD score less than 5 (Fig. 3B). The PVE (%) of all 114 QTL ranges from 0.4 to 60% with a mean of 10.8%, with nearly 50% of them having a PVE (%) between 5 and 10% (Fig. 3C). After combining individual maps from 22 reported experiments, a consensus genetic map was created.

According to the meta-analysis description, chromosomes with less than nine original grain weight QTL were analysed using the Goffinet and Gerber²³ algorithm, which identified 36 MQTL across 11 chromosomes (except chromosome 3) with a 95% confidence interval and the lowest akaike information content AIC value (Table 1). The model that had a low AIC value while analysing was thought to indicate the importance of the MQTL discovered in this method. The chromosome 3 with 22 initial QTL was analysed using the Goffinet and Gerber algorithm as well as the Veyrieras approach resulted in identification of three MQTL with a 95 percent confidence interval. These three MQTL (MQTL3.1, MQTL3.2 and MQTL3.3) were declared true MQTL and considered for further analysis and validation, since they were identified using both methods with low AIC, corrected AIC (AICc and AIC3), Bayesian information criterion (BIC), and average weight of evidence (AWE) criteria. These three significant MQTL were located on chromosome 3 at 39.5 cM, 92.82 cM, and 116.66 cM, respectively. These MQTL had incredibly small intervals, with MQTL3.1 having 0.49 Mb, MQTL3.2 having 0.63 Mb, and MQTL3.3 having just 0.08 Mb intervals (Fig. 4). The weights of these MQTL, on the other hand, were highly significant, with values 0.19, 0.33 and 0.37, respectively (Table 2). The fact that MQTLs physical lengths were so modest in comparison to their genetic lengths demonstrates the significance of these three MQTL (Supplementary Figure S1).

Mining genes underlying MQTL regions. For discovering the genes underlying MQTL regions, the relative physical positions of the closest flanking markers of MQTL were sought from the Gramene database. Genes within intervals of three MQTL on chromosome 3 were obtained from Rice MSU (<http://rice.plantbiology.msu.edu/>). Information on a total of 179 genes was extracted from intervals of identified MQTLs (Supplementary file 2). The MQTL3.1, having an interval of 0.49 Mb, accommodates a total of 73 genes, while MQTL3.2, with an interval of 0.63 Mb, accommodates 93 genes. The MQTL3.3 was found to have a very small interval of just 0.08 Mb containing 13 genes with a higher weight factor than the other two MQTL (Table 2). Even though



Figure 1. Distribution of the markers on the rice consensus map constructed and utilized for meta-QTL analysis in the present study (Color intensity from white to red indicate low to high density of marker, respectively).

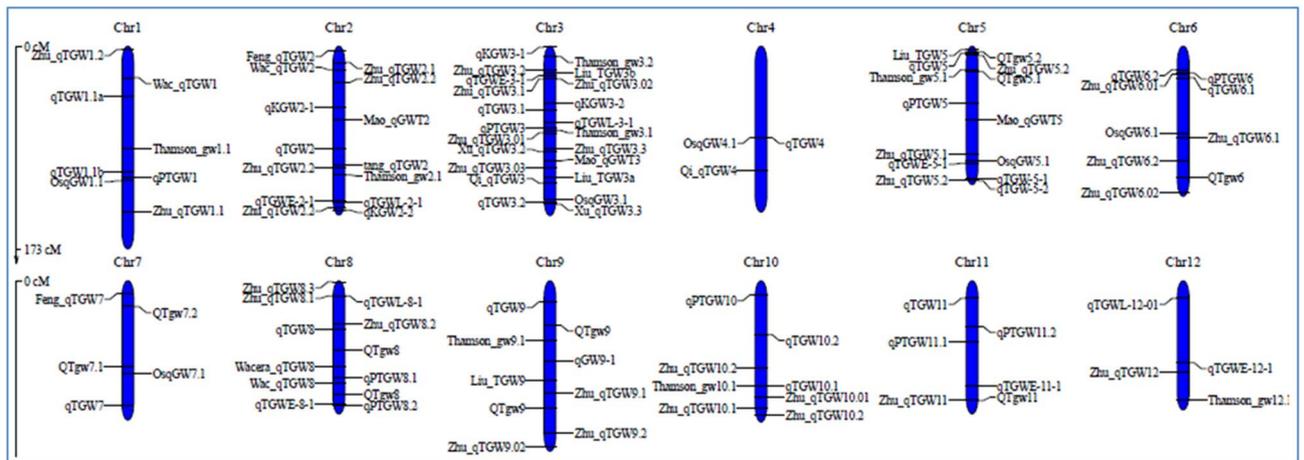


Figure 2. Distribution of reported grain weight QTL on all 12 rice chromosomes (name of the first author of original report is used as prefix while naming the QTL in this study).

179 genes were found within the intervals of identified MQTL, only 25 genes were characterized for different biological functions (Table 3), including five genes controlling biological processes associated with grain weight and shape. Even though MQTL3.2 had the largest number of genes among the three MQTL, a greater number of characterized genes were found in MQTL3.1, followed by MQTL3.2, and a very small number were found in MQTL3.3. Among the characterized genes, four loci in the MQTL3.1 interval (*LOC_Os03g17690*, *LOC_Os03g17860*, *LOC_Os03g18050* and *LOC_Os03g18130*) and one locus in MQTL3.3 interval (*LOC_Os03g51330*) were found associated with grain development related functions. Conserved domain information for all the

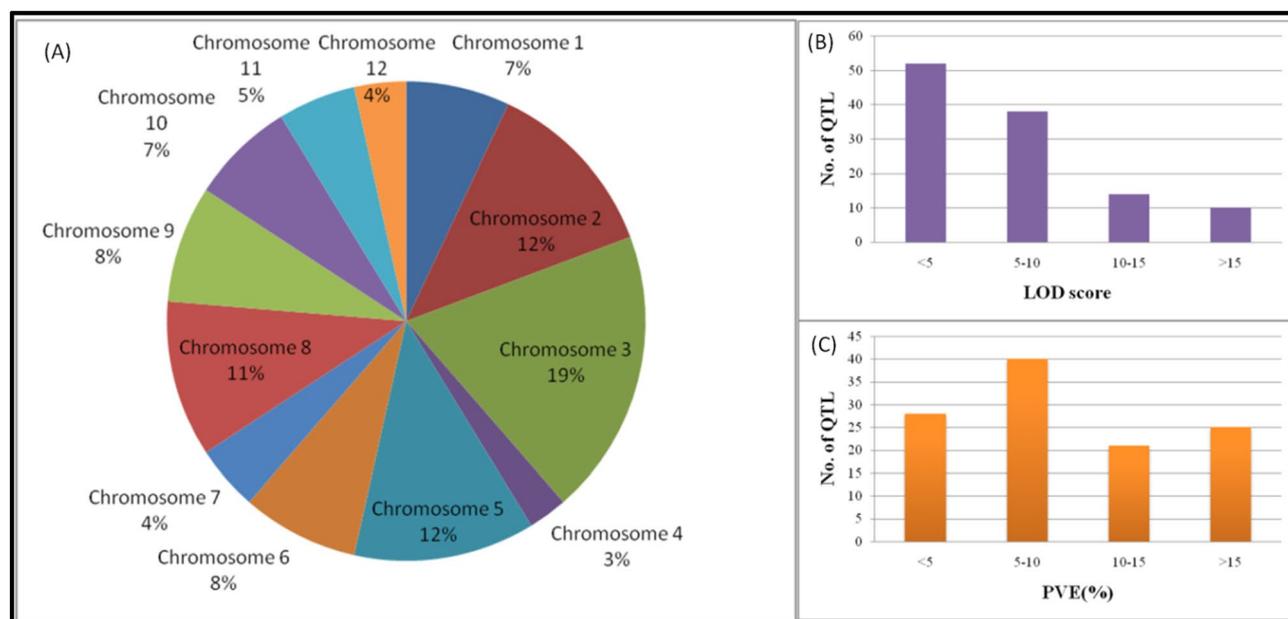


Figure 3. Detailed information on initial QTL utilized for QTL-meta analysis (A) chromosome distribution of initial QTL (B) LOD score information of initial QTL and (C) information on phenotypic variation explained by initial QTL considered for the study.

Chromosome	Number of QTL	Number of MQTL	Model	Method
1	8	4	Model 4	Gerber and Goffinet
2	14	4	Model 4	Gerber and Goffinet
3	22	3	AIC, AICc, AIC3, BIC, AWE	Veyrieras
4	3	3	Model 3	Gerber and Goffinet
5	14	4	Model 4	Gerber and Goffinet
6	9	2	Model 2	Gerber and Goffinet
7	5	3	Model 3	Gerber and Goffinet
8	12	4	Model 4	Gerber and Goffinet
9	9	3	Model 3	Gerber and Goffinet
10	8	3	Model 3	Gerber and Goffinet
11	6	4	Model 4	Gerber and Goffinet
12	4	2	Model 2	Gerber and Goffinet
Total	114	39		

Table 1. Chromosome-wise list of initial QTL and identified MQTL.

genes under three MQTL were used to understand the gene functions and found many ubiquitination pathway related genes, auxin biosynthesis genes, pectin methyl biosynthesis genes, and sucrose metabolism regulating genes in these regions (Supplementary file 3). Amino acid sequence BLAST of 25 characterized genes underlying three MQTL regions identified orthologue protein sequences in the other four core cereals. Complete syntenic relationship between rice and sorghum for all 25 amino acid sequences followed by maize and barley lacking synteny to amino acid sequences of *LOC_Os03g42290* and *LOC_Os03g18120*, respectively. However, wheat genome displayed lower synteny than the other three cereals (Supplementary file 4).

Ontology of genes within the MQTL regions and expression analysis of candidate genes. The GO analysis of genes within MQTL regions was performed by taking genes in individual MQTL. Gene ontology analysis within the MQTL3.1 and MQTL3.3 regions was unsuccessful due to low gene numbers and gene ontology terms. However, MQTL3.2, with a 0.63 Mb interval containing 93 genes, was found to have sufficient GO terms to map. Ontology annotation for these genes predicted the involvement of genes in many biological processes such as biological regulation, metabolic process, macromolecule synthetic pathways, transcription, gene expression, etc., indicating direct or indirect association of these genes in grain development (Supplementary Figure S2). Further, the five genes identified as directly related to grain development and considered as candidate genes (Table 4) were analyzed separately for functional annotation using the ShinyGO v0.741 (<http://bioinformatics.sdsstate.edu/go/>) online tool. These genes were found to be associated with many asparagine metabolism

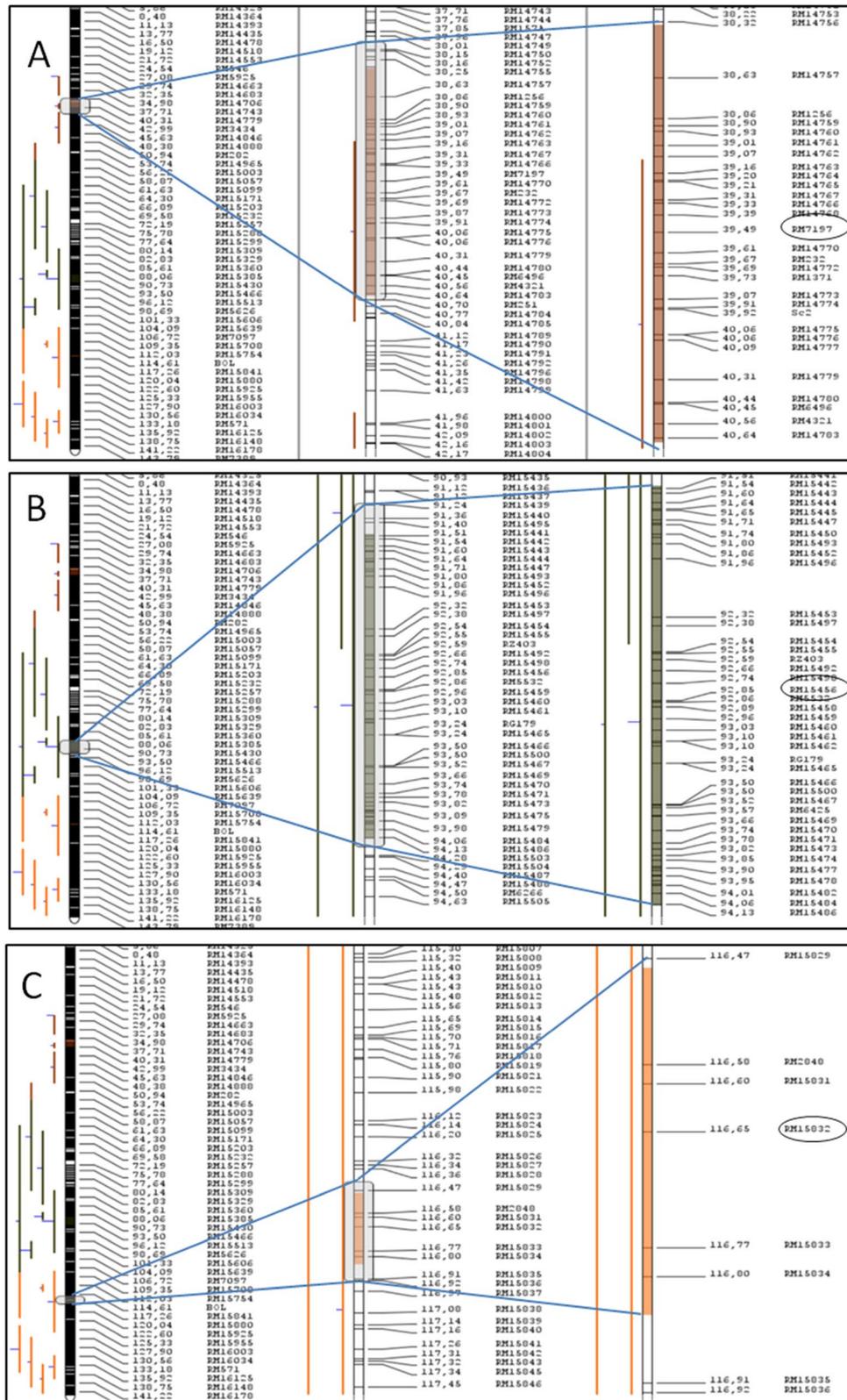


Figure 4. Magnified view of significant meta-QTL identified in the study (A) MQL3.1 having interval of 0.49 Mb with peak marker RM7197, (B) MQL3.2 having interval of 0.63 Mb with peak marker RM15456 and (C) MQL3.3 having interval of 0.08 Mb with peak linked marker RM15832.

Sl No	Meta-QTL	Chromosome	Positions	Start (cM)	End (cM)	Peak marker	Weight	CI (95%)	Interval (Mb)	Number of genes in the interval	Flanking markers
1	MQTL3.1	3	39.5	38.32	40.64	RM7197	0.19	2.33	0.49	73	RM14757-RM4321
2	MQTL3.2	3	92.82	91.54	94.13	RM15456	0.33	2.64	0.63	93	RM15442-RM15484
3	MQTL3.3	3	116.66	116.47	116.8	RM15832	0.37	0.36	0.08	13	RM15829-RM15834

Table 2. List and information on identified significant MQTL in the study.

Sl. No	MQTL	Locu ID	Position
1	MQTL3.1	LOC_Os03g17460	9,710,092—9,708,206
2	MQTL3.1	LOC_Os03g17470	9,714,748—9,712,370
3	MQTL3.1	LOC_Os03g17480	9,719,907—9,717,252
4	MQTL3.1	LOC_Os03g17570	9,768,656—9,759,479
5	MQTL3.1	LOC_Os03g17610	9,797,352—9,799,915
6	MQTL3.1	LOC_Os03g17690*	9,843,327—9,846,747
7	MQTL3.1	LOC_Os03g17700	9,850,473—9,847,700
8	MQTL3.1	LOC_Os03g17780	9,893,421—9,896,633
9	MQTL3.1	LOC_Os03g17790	9,900,400—9,899,758
10	MQTL3.1	LOC_Os03g17860*	9,955,138—9,952,788
11	MQTL3.1	LOC_Os03g17870	9,957,429—9,958,013
12	MQTL3.1	LOC_Os03g17980	9,961,503—9,960,708
13	MQTL3.1	LOC_Os03g18050*	10,057,658—10,058,281
14	MQTL3.1	LOC_Os03g18110	10,105,981—10,108,026
15	MQTL3.1	LOC_Os03g18120	10,116,219—10,120,924
16	MQTL3.1	LOC_Os03g18130*	10,124,384—10,119,873
17	MQTL3.1	LOC_Os03g18140	10,147,977—10,153,118
18	MQTL3.1	LOC_Os03g18150	10,163,441—10,165,477
19	MQTL3.2	LOC_Os03g41600	23,127,983—23,128,906
20	MQTL3.2	LOC_Os03g42020	23,341,805—23,347,331
21	MQTL3.2	LOC_Os03g42100	23,434,712—23,433,316
22	MQTL3.2	LOC_Os03g42110	23,442,337—23,438,157
23	MQTL3.2	LOC_Os03g42200	23,475,827—23,472,278
24	MQTL3.2	LOC_Os03g42290	23,533,463—23,538,865
25	MQTL3.3	LOC_Os03g51330*	29,370,719—29,373,265

Table 3. List of characterized genes underlying identified significant MQTL. *genes directly involved in grain weight or grain shape related biological processes.

MQTL	Locus ID	MSU ID	GO term	Gene description
MQTL3.1	Os03g0285700	LOC_Os03g17690	Metabolic process	Similar to L-ascorbate peroxidase
MQTL3.1	Os03g0287900	LOC_Os03g17860	Cellular homeostasis	Similar to Protein disulfide isomerase
MQTL3.1	Os03g0290300	LOC_Os03g18050	Molecular function	Auxin responsive protein
MQTL3.1	Os03g0291500	LOC_Os03g18130	Biosynthetic process	Asparagine synthase domain containing protein
MQTL3.3	Os03g0723000	LOC_Os03g51330	Cellular component	GRAS transcription factor domain containing protein

Table 4. Details of candidate genes identified under MQTL regions associated with grain development.

related pathways (Fig. 5). These genes play a significant role in understanding the source-sink relationship in rice²⁴. Micro-array based expression analysis datasets from RiceXPro revealed changes in the expression patterns of five candidate genes studied. Among five candidate genes, locus *LOC_Os03g17860* (*PDIL* gene) was found to have continuous expression from early stages of ovary and embryo development to later stages of maturation (Fig. 6). Locus *LOC_Os03g18050* (*SAUR* gene) was found to up regulate early ovary development followed by down regulation during embryo development. Other three candidate genes were found to have significant up regulation during early stages of ovary and embryo development.

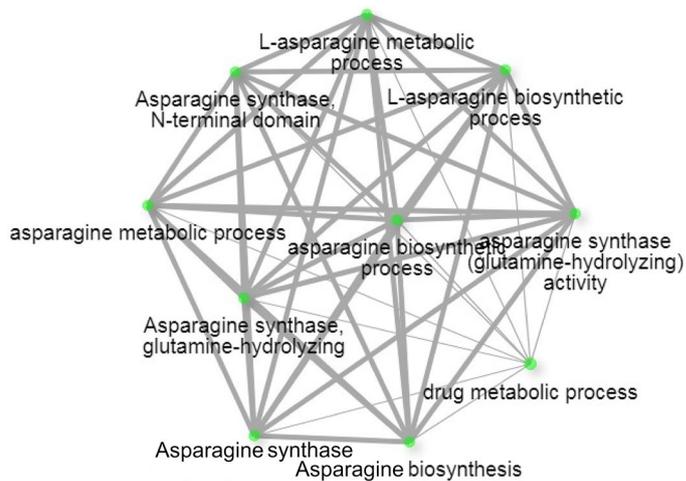


Figure 5. Network display of five grain related genes underlying identified MQTL involved in asparagine biosynthesis and metabolism pathways having significant role in regulating source to sink relation in rice. The green nodes in the network indicate different asparagine biosynthesis and metabolic pathways and grey connecting lines indicate interaction between the processes of pathways to contribute towards establishing source to sink relationship.

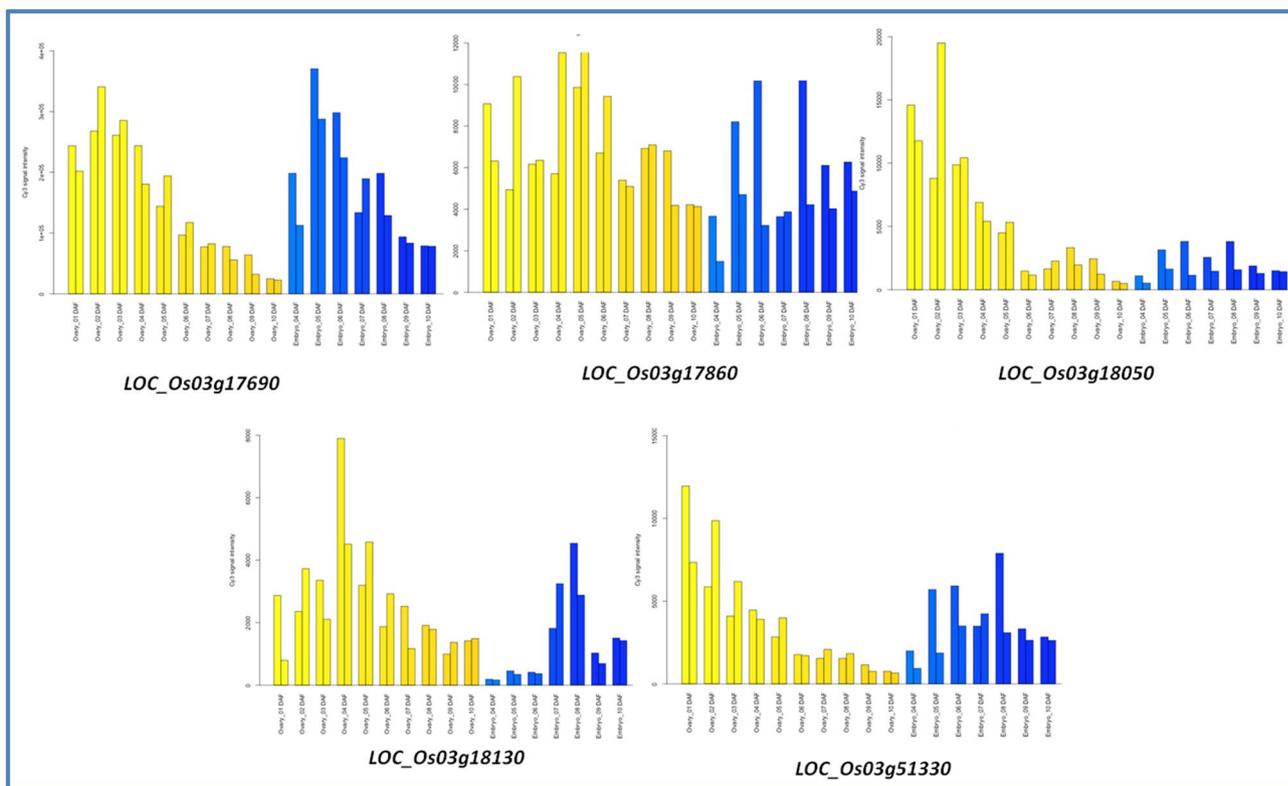


Figure 6. Graphical representation of expression of candidate genes at early grain development stage. The yellow bars taken in two replicates indicate expression of candidate genes in ovary at different days after flowering, whereas the blue bars indicate expression of genes in embryo at different days after flowering, and represented in two replications.

Validation of MQTL. Considering the consensus map as reference, three peak markers present in the middle of the identified MQTL region were opted for validation of MQTL. Markers RM7197, RM15456, and RM15832 from MQTL3.1, MQTL3.2, and MQTL3.3, respectively, were used for validation of these MQTL. Out of three selected markers, only RM7197 could clearly differentiate the genotypes into two phenotypic classes, i.e., high and low grain weight. Hence, MQTL3.1 was validated beyond doubt by RM7197. The allele with an amplicon size of 100 bp was found specifically associated with low grain weight genotypes while the alternate allele

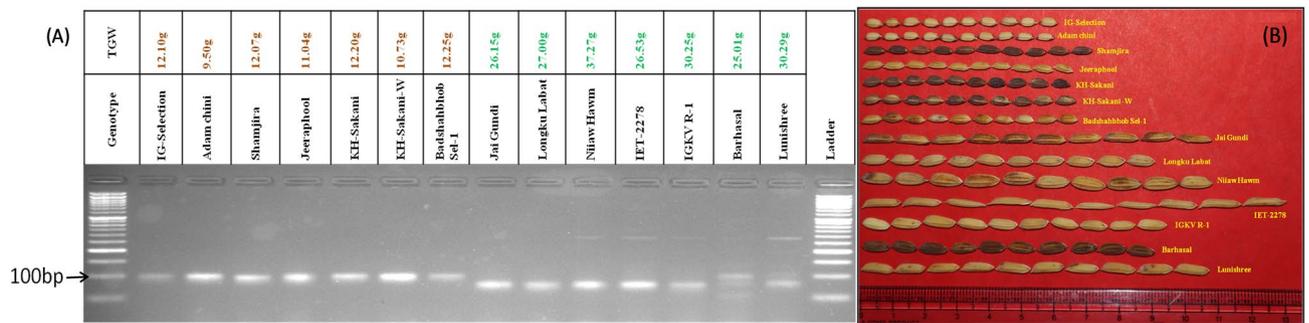


Figure 7. (A) Validation of marker RM7197, peak marker of MQTL3.1 on set of seven low grain weight and seven high grain weight genotypes. The allele with 100 bp linked to low grain weight and 85 bp allele linked to high grain weight. (B) Representation of grain size and shape of genotypes used for validation of identified MQTL.

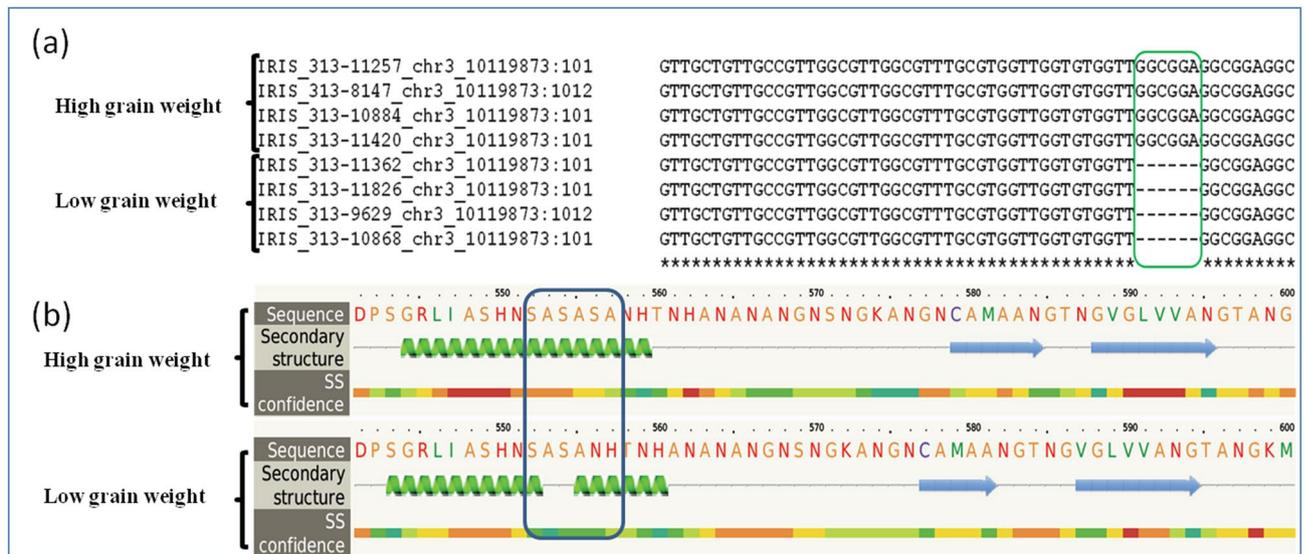


Figure 8. Sequence variation between low and high grain weight genotypes for *LOC_Os03g18130* under MQTL3.1 retrieved from 3 k rice genome project. (a) A six base pair deletion in the last exon of the gene in low grain weight genotypes and (b) two amino acids, alanine and serine deleted in amino acid coding sequence which led to interruption of secondary structure (α -helix) of asparagine synthetase b protein in low grain weight genotypes.

with an amplicon size of 85 bp was specific to higher grain weight (Fig. 7). The validated SSR marker RM7197 is located at 9,888,524 bp on rice chromosome three between identified candidate genes *LOC_Os03g17690* and *LOC_Os03g17860*. Even though the marker sequence was found within the inter-genic region, it was found to surrogate all four candidates in the MQTL3.1 region. However, the other two MQTL were not clearly validated in the study, indicating the need for the identification of more appropriate markers for further validation of those MQTL.

The gene sequence differences between eight genotypes from the 3 K rice genome project (Supplementary file 6) revealed sequence variations between genotypes for five candidate genes identified under two MQTL. However, differences in sequences for loci *LOC_Os03g17690*, *LOC_Os03g17860*, and *LOC_Os03g18050* under MQTL3.1, and locus *LOC_Os03g51330* under MQTL3.3 did not distinguish low grain weight and high grain weight genotype classes. Interestingly, the gene sequence of eight genotypes at locus *LOC_Os03g18130* under MQTL3.1 exhibited prominent sequence differences for genotypes of different grain weight classes. Three single nucleotide changes distinguished genotype classes at 317 bp, 349 bp, and 395 bp, followed by two nucleotide deletions in high grain weight genotypes, were noticed at 2037 bp from 5' end. Unfortunately, these SNPs and two nucleotide differences were found in the intronic regions of the gene and their contribution to the grain weight difference was ruled out. A six nucleotide deletion at 3646 bp in low grain weight genotypes was observed, which is located on the last exon of the gene (Fig. 8a). The deleted nucleotides code for alanine and serine amino acids in high grain weight genotypes. The deletion of two amino acids interrupted the α -helix in the secondary structure of the asparagine synthetase B protein (Fig. 8b), which also made minimal changes to the 3D structure (Supplementary Figure S3). Since the deletion was found in the coding region, this sequence change is expected to have an impact on gene expression.

Discussion

Understanding the genetic architecture of complex quantitative traits is a prelude to crop improvement. Grain weight in rice, being a metric trait, has a significant impact on grain yield, and thus needs a thorough dissection and deployment in the rice yield improvement breeding programme. Several researchers over the last couple of decades have characterised and mapped several genetic loci responsible for inheritance of rice grain weight^{25–28}. However, many of these findings done on different bi-parental populations often resulted in higher confidence interval that limited their validation and utilization in practical plant breeding^{29,30}. The information on QTL identified in independent experiments over various populations across diverse environments can be integrated to identify true and consistent QTL through a meta-analysis approach²³. Few studies have located MQTL^{14,19,31} to mine candidate genes.

In the present study, meta-QTL analysis was conducted specifically for thousand grain weight in rice. For comprehensive identification of consistent QTL responsible for grain weight in rice, information from 22 independent investigations identifying 114 QTL was used (Supplementary file 1). The initial QTL identified were spread across 12 rice chromosomes, indicating metric inheritance of the trait with several causative alleles over all the chromosomes (Fig. 2). A similar trend of QTL distribution on all chromosomes was reported by Islam et al.³² for salt tolerance in rice. Even though 39 MQTL were identified from 114 original QTL, only three MQTL on chromosome 3 were considered significant based on AIC information (Table 1). The presence of a greater number of initial QTL for grain weight on chromosome 3 hinted at the chances of identification of a consistent MQTL (Fig. 3A). Further, earlier researchers like Xu et al.³³ and Zhu et al.³⁴ identified a few major QTL on chromosome 3 for grain weight. Three MQTL identified on chromosome 3 were found to have very low interval and comparatively high weight (Table 2), which is advantageous for utilization in strategic breeding programs³⁵. The genome-wide association studies carried out on two diverse rice panels did not identify any QTL either for grain weight^{36,37} or for grain yield³⁸ on chromosome three owing to population size and marker density used in these studies. Consensus projection of all the minor and major QTL identified in different studies into a single or few MQTL, which explains the variation of all included QTL, signifies the power of meta-QTL identification. Hence, identified consistent MQTL with short interval have potential use in marker assisted backcross breeding programs for rapid elimination of linkage drag. The identified MQTL regions were found rich in gene content, harboring a total of 179 gene loci within their intervals. Several of these genes were not annotated, and only 25 of them, were to be found characterized and annotated. These included important genes like *GSTL3* (*LOC_Os03g17460*) and *GSTL1* (*LOC_Os03g17480*) responsible for stress tolerance^{39,40}, while *PRR73* (*LOC_Os03g17570*) regulates salt tolerance⁴¹ in rice. However, these characterized genes are not related to grain weight and are instead associated with some important biological processes. The Gene Ontology survey helped understand the functions of genes within MQTL. The gene ontology plot from the MQTL3.2 region depicted the roles of loci in different biological and metabolic processes (Supplementary Figure S2). These genes are associated with many physiological and molecular pathways related to grain development⁴². The group of genes significantly associated with GO terms from MQTL regions might enrich phenotype expression of grain weight and aid in selection. Further, a sequence blast to identify conserved domains for all 179 genes unraveled the information on important genes and their functions, which contribute towards the biological process of grain development. Genes responsible for the ubiquitination pathway, auxin biosynthesis pathway, pectin methyl transferase coding genes and sucrose metabolism pathway regulating genes were found within the identified MQTL regions. These pathways have significant roles to contribute for early grain development, including cell division in endosperm⁴³ and grain maturation by sink organ sucrose metabolism^{44,45}. The comparison of orthologue amino acid sequences of other closely related cereals revealed that these proteins are conserved across species and are responsible for important biological functions, leading to plant development as similar to rice. The availability of orthologue regions in closely related cereals increases the significance of identified MQTL and also verifies their stability^{46,47}. The orthologue sequence search also included the candidate genes identified in this study, the protein expressions of amino acid sequences in other cereals were found to be similar to those in the rice plant. Thus, that verifies the stability of identified candidate genes in this study.

Four loci from MQTL3.1 and one from MQTL3.3 were reported to have associations with grain development related processes. Locus *LOC_Os03g17690* found in MQTL3.1 was associated with *OsAPX1* gene plays significant role in embryogenesis⁴⁸ and also in heat stress tolerance⁴⁹. The Gene *PDIL* (*LOC_Os03g17860*) was responsible for accumulation of seed storage⁵⁰, gene *SAUR* (*LOC_Os03g18050*) was influencing reproductive development⁵¹ while *OsASN1* (*LOC_Os03g18130*) regulating asparagine dependent rice development⁵². These functions of identified gene loci within MQTL3.1 define its importance in the improvement of rice grain weight. None of the genes identified within the MQTL3.2 region were found directly associated with grain weight. However, their importance in other biological processes leading to grain yield was found to be significant. Among the genes within MQTL3.3, gene *OsGRAS19* (*LOC_Os03g51330*) was reported to regulate grain shape in rice⁵³. Apart from directly influencing grain development processes, these genes were found to be associated with asparagine biosynthesis and asparagine metabolic processes (Fig. 5). It is reported that asparagine along with glutamine regulates source to sink relations in rice plant⁵². Glutamine and asparagine are the two important forms of N transport from root to shoot in rice plants. Further, reproductive organ specific expression of identified candidate genes under defined MQTL regions indicated the role of early embryo development stages for enhancing grain weight. Utilization of these MQTL in plant breeding enhances the chance of improving source to sink relations, thereby increasing grain weight and yield in rice.

Among three random rice markers used for validation of MQTL on a set of extreme genotypes consisting of seven low grain weights and seven high grain weights, only one marker (RM7197) clearly differentiated two extreme categories (Fig. 7). Segregation of marker amplicons with low and high seed weight classes of genotypes validates their linkage with the respective MQTL. The validated marker was the peak marker of MQTL3.1,

indicating the marker was tightly linked to MQTL3.1. The validated marker, RM7197, has scope in practical marker aided plant breeding to incorporate the MQTL3.1 region into an elite background to enhance grain weight. The association of the RM7197 marker with MQTL3.1 indicates its association with all other genes underlying MQTL3.1. Hence, it can be used as a surrogate to track them in marker-assisted rice breeding programs for improving grain weight. The remaining two markers related to the remaining two MQTL were not clearly validated owing to very few QTL/genes involved in the MQTL formation. In MQTL3.2, only one candidate gene and a few minor QTL were reported in the region; whereas, within MQTL3.3, no candidate genes were identified and only two minor QTL were present. Therefore, the marker linked to very small numbers of alleles responsible for grain weight might not be possible to associate with a complete phenotype of grain weight in genotypes with extreme phenotypes. Hence, there was a lack of association with phenotype in the validation.

The six nucleotide deletion on the last exon of the gene was noticed in the gene sequence of low grain weight genotypes. The deleted sequence 5'GGCGGA3' with the complementary sequence 5'TCCGCC3' in high grain weight genotypes codes for two amino acids, serine and alanine, which play a significant role in asparagine biosynthesis⁵⁴. The efficiency of serine: glyoxylate aminotransferase or alanine: glyoxylate aminotransferase (AGT1) activity in the asparagine biosynthetic pathway significantly increases when glycolate is an amino acceptor and serine or alanine as an amino donor⁵⁵. The deletion of coding sequences of these amino acids negatively impacts the asparagine biosynthetic pathway, which indirectly reduces the strength of the sink-to-source relationship. Hence, this change in gene sequence may be one of the genetic causes of the change in grain weight. This evidence from gene sequence information of genotypes from the 3 K rice genome project reiterated the importance of identified MQTL for grain weight.

Conclusion

By collating the results of independent investigations conducted over various environments and identification of true consensus genomic region for grain weight paves new avenue for utilization in practical rice breeding. In this study, three significant and consistent consensus genomic regions for grain weight were identified in rice through a QTL-meta analysis approach. The intervals of these MQTL contains many candidate genes responsible for asparagine metabolic pathway and GRAS transcription factor domains which regulates source to sink relationship and grain weight in rice. This provides insight into pathways underlying grain weight character and genes responsible for variable grain weight in rice. The marker RM7197 validated for identified meta-QTL has significant scope in incorporating all the genes underlying MQTL3.1 for improving grain weight in rice. The outcome of this investigation has significant application in practical rice breeding programs to incorporate these MQTL for grain weight improvement through marker-aided breeding programs.

Materials and methods

Bibliographic survey and data generation. A comprehensive survey of literature related to QTL mapping for grain weight in rice published over the last two decades was performed to aggregate the information (Supplementary file 1). The information on the number and type of molecular markers, genetic map, type and size of mapping populations, parents used, LOD scores, genetic distances in linkage groups and proportion of phenotypic variance explained (R^2 values) were collated from published literature. Information on a total of 114 QTL mapped exclusively for grain weight over environments across populations along with their genetic maps were collected, tabulated, and summarized. A few studies with incomplete information on genetic maps were excluded from the study. The population size of progenies in the mapping populations of various studies varied from 39 to 353 progenies evaluated over different locations encompassing different studies. Separate input text files for map and QTL information were created as per the requirements of Biomeqator v4.2¹⁷. This information was considered as primary information for the present investigation.

Consensus map development and QTL projection. Forehand to MQTL analysis, a consensus genetic map was developed based on collated information of genetic maps from published studies using Biomeqator v4.2^{16,17}. The genetic map well saturated with 19,180 SSR markers available in the Gramene database (<https://archive.gramene.org/markers/>) was used as the reference genetic map for developing the consensus map. The information on individual map files for 114 grain weight specific QTL (original/initial QTL) was integrated on to the consensus map for QTL projection. In order to incorporate information from the SNP markers linked to original QTL, the positions of SNP markers in the rice genome were determined and the closest SSR markers corresponding to them were chosen for projecting QTL on the consensus map as suggested by Khahani et al.⁵⁶. The QTL projection on the consensus map was based on R^2 explained by QTL, LOD score, QTL positions, and confidence intervals of initial QTL. For the original QTL, the confidence interval (CI) used to project them on the consensus map was calculated following the Darvasi and Soller⁵⁷ equation: $\frac{530}{NR^2}$, in which N is the size of the population and R^2 is the phenotypic variation explained by the QTL. The QTL position on the consensus map was projected by considering the marker interval of the QTL on the original map and the corresponding position on the consensus map. A Gaussian distribution rule was used to calculate the new CI on the consensus map as suggested by Veyrieras et al.¹⁶.

QTL meta-analysis. A consensus genetic map of each chromosome with projected QTL for grain weight was subjected to MQTL analysis following the default parameter setting in BioMercator v4.2 (<https://urgi.versailles.inra.fr/Tools/BioMercator-V4>)^{16,17,58}. The method of analysis was selected based on initial QTL numbers on each chromosome; if a chromosome contains more than nine QTL, the 'Veyrieras' method or otherwise the 'Gerber and Goffinet' method was followed. An appropriate model for the identification of MQTL was chosen based on Akaike Information Criterion (AIC), corrected Akaike Information Criterion (AICc and AIC3),

Bayesian Information Criterion (BIC) and/or Average Weight of Evidence (AWE) criteria¹⁶. The lowest number of models predicted by at least three of these criteria was considered the best. The MQTL identified by the best model based on the lowest value for the aforesaid criteria was reported²⁹.

Discovery of genes within Meta-QTL regions. Based on the results of MQTL analysis, the anchor positions of significant MQTL were identified by considering flanking marker positions obtained from the Gramene database. For the process called mapping, physical positions of flanking markers were used as inputs in the search field of Rice MSU (<http://rice.plantbiology.msu.edu/>) to discover and extract information on the number of genes, locus IDs, and annotated data related to MQTL regions. For each loci identified under MQTL region, the information on conserved domain and gene function was extracted from Rice MSU (<http://rice.plantbiology.msu.edu/>) and NCBI-conserved domain database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Further, amino acid sequences of characterized genes under identified MQTL regions were used to find orthologues in other core cereals using the NCBI orthologue search database (<https://www.ncbi.nlm.nih.gov/>). The locus IDs of genes from the MQTL regions were used as input for the gene ontology (GO) search.

Gene ontology analysis and expression analysis. Selecting appropriate gene ontology terms from the list of mapped genes in a specific genomic region is termed as GO annotation. The information on locus ID of genes discovered in the MQTL region was functionally classified following the single enrichment analysis (SEA) tool in the web-based AgriGOv2.0 (<http://systemsbiology.cau.edu.cn/agriGOv2/>) interface. The necessary parameters like Benjamini-Yekutieli (FDR under dependency) under multiple adjustment test and significance at 5% were set before analysis³². The GO terms include the involvement of genes in biological processes, cellular processes, metabolic processes and regulation, etc. Among these genes, grain development related genes were identified based on available literature and were considered as candidate genes for grain weight. The expression data of these genes were downloaded from RiceXPro database. RXP_0011 dataset, which contains micro-array based expression data of grain early stage development, was considered for further interpretation.

Validation of identified MQTL by linked markers. In traditional QTL validation, it is important to develop near isogenic or transgenic lines for assessing the effect of identified QTL. Validation of MQTL conceptually differs from that of QTL validation owing to its property of inclusive QTL³². According to the algorithmic definition of meta-analysis, MQTL identification in a condensed genomic region assures the validation of known QTL in the region²³. Hence, the identification of a linked surrogate marker and validating it on a set of genotypes with extreme phenotypes is practically significant. Significant association of marker alleles with phenotypes ensures the presence of at least a few alleles of QTL present in the MQTL. Several researchers validated identified MQTL following the aforesaid approach^{14,32,60}. In the present study, MQTL associated peak-SSR markers from each MQTL were opted for validation on a set of 14 genotypes, comprising seven genotypes each from extreme low and high grain weight categories available at ICAR-National Rice Research Institute Cuttack. The experimental genotypes were selected from a panel of 300 genotypes evaluated over the last five years, and a mean of thousand grain weight data was considered. The genotypes with less than 12.50 g were considered as low grain weight set and those with more than 25.00 g were considered as high grain weight set. Among selected genotypes, Niiaw Hawm (37.27 g) recorded the highest grain weight, while Adam Chini (9.50 g) recorded the lowest grain weight (Supplementary file 5). Seedlings were grown in pots at laboratory conditions and genomic DNA was extracted using the CTAB method⁶¹. The quality and quantity of genomic DNA were tested before thermal amplification of regions corresponding to MQTL-linked markers. Amplified fragments were visualized using gel electrophoresis and a documentation system.

Further, the gene sequences of four loci under MQTL3.1 and one locus under MQTL3.3 for a set of eight genotypes, four low grain weight and four high grain weight⁶² in the 3 K rice genome project were retrieved from the rice functional genomics and breeding (RFGB) database (<https://www.rmbreeding.cn/public/searchbak>). The indica accessions of Indian origin were selected to understand the gene sequence differences between low grain weight and high grain weight genotypes (Supplementary file 6). The gene sequences of eight genotypes were subjected to multiple sequence alignment using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) tool (<https://www.ebi.ac.uk/Tools/msa/muscle/>). Upon identification of sequence differences between genotypes, their position on the gene sequence and functional contributions to variation were explored. The protein structure variation due to sequence differences was analyzed using the Phyre2 webportal (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>)⁶³ and represented in for visualization.

Data availability

All data included in this study are available upon request by contact with the corresponding author. The data used for gene expression analysis was retrieved from RXP_0011 data set in RiceXPro database (<https://ricexpro.dna.affrc.go.jp/Zapping/>) and sequence information for eight genotypes was retrieved from rice functional genomics and breeding database (<https://www.rmbreeding.cn/>).

Received: 2 February 2022; Accepted: 25 July 2022

Published online: 16 August 2022

References

1. Weng, J. *et al.* Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18**(12), 1199–1209 (2008).
2. Yuan, L., Denning, G. & Mew, T. Hybrid rice breeding for super high yield. *Denning GL, Mew TW, editors*, pp.10–12 (1998).

3. Kumar, A. *et al.* Breeding high-yielding drought-tolerant rice: Genetic variations and conventional and molecular approaches. *J. Exp. Bot.* **65**(21), 6265–6278 (2014).
4. Ahmar, S. *et al.* Conventional and molecular techniques from simple breeding to speed breeding in crop plants: recent advances and future outlook. *Int. J. Mol. Sci.* **21**(7), 2590 (2020).
5. Katara, J. L. *et al.* Genomics assisted breeding: The need and current perspective for rice improvement in India. *ORYZA-An International Journal of Rice* **58**(1), 61–68 (2021).
6. Huang, R. *et al.* Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci.* **18**(4), 218–226 (2013).
7. Li, W. T. *et al.* Meta-analysis of QTL associated with tolerance to abiotic stresses in barley. *Euphytica* **189**(1 SPL), 31–49 (2013).
8. Zhao, X., Peng, Y., Zhang, J., Fang, P. & Wu, B. Identification of QTLs and Meta-QTLs for seven agronomic traits in multiple maize populations under well-watered and water-stressed conditions. *Crop Sci.* **58**(2), 507–520 (2018).
9. Lei, L. *et al.* Genetic dissection of rice (*Oryza sativa* L.) tiller, plant height, and grain yield based on QTL mapping and metaanalysis. *Euphytica* **214**(7), 1–17 (2018).
10. Bernier, J., Atlin, G. N., Serraj, R., Kumar, A. & Spaner, D. Breeding upland rice for drought resistance. *J. Sci. Food Agric.* **88**(6), 927–939 (2008).
11. Tuberosa, R. *et al.* Mapping QTLs regulating morpho-physiological traits and yield: Case studies, shortcomings and perspectives in drought-stressed maize. *Ann. Bot.* **89**(7), 941–963 (2002).
12. Rong, J. *et al.* Meta-analysis of polyploid cotton QTL shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. *Genet.* **176**(4), 2577–2588 (2007).
13. Gobu, R., Shiv, A., Anilkumar, C., Basavaraj, P. S., Harish, D., Adhikari, S., Vinita, R., Umesh, H. & Sujatha, M. Accelerated crop breeding towards development of climate resilient varieties. *Climate change and Indian Agriculture: Challenges and Adaptation Strategies*. ICAR-National Academy of Agricultural Research Management, Hyderabad, Telangana, India, pp. 49–70 (2020).
14. Swamy, B. M., Vikram, P., Dixit, S., Ahmed, H. U. & Kumar, A. Meta-analysis of grain yield QTL identified during agricultural drought in grasses showed consensus. *BMC Genom.* **12**(1), 1–18 (2011).
15. Izquierdo, P. *et al.* Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **131**(8), 1645–1658 (2018).
16. Veyrieras, J. B., Goffinet, B. & Charcosset, A. MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments. *BMC Bioinform.* **8**(1), 1–16 (2007).
17. Sosnowski, O., Charcosset, A. & Joets, J. BioMercator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics* **28**(15), 2082–2083 (2012).
18. Swamy, B. M. & Sarla, N. Meta-analysis of yield QTLs derived from inter-specific crosses of rice reveals consensus regions and candidate genes. *Plant Mol. Biol. Rep.* **29**(3), 663–680 (2011).
19. Wu, Y. *et al.* Quantitative trait loci identification and meta-analysis for rice panicle-related traits. *Mol. Genet. Genom.* **291**(5), 1927–1940 (2016).
20. Khawaja, F. S., Norton, G. J., Courtois, B. & Price, A. H. Improved resolution in the position of drought-related QTLs in a single mapping population of rice by meta-analysis. *BMC Genom.* **10**(1), 1–14 (2009).
21. Yang, L. *et al.* QTL mapping for heading date, leaf area and chlorophyll content under cold and drought stress in two related recombinant inbred line populations (Japonica rice) and meta-analysis. *Plant Breed.* **137**(4), 527–545 (2018).
22. Liu, H. *et al.* Major genomic regions responsible for wheat yield and its components as revealed by meta-QTL and genotype–phenotype association analyses. *Planta* **252**(4), 1–22 (2020).
23. Goffinet, B. & Gerber, S. Quantitative trait loci: A meta-analysis. *Genet.* **155**(1), 463–473 (2000).
24. Yabuki, Y. *et al.* A temporal and spatial contribution of asparaginase to asparagine catabolism during development of rice grains. *Rice* **10**(1), 1–10 (2017).
25. Fan, C. *et al.* 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**(6), 1164–1171 (2006).
26. Bian, J. M. *et al.* QTL mapping and correlation analysis for 1000-grain weight and percentage of grains with chalkiness in rice. *J. Genet.* **92**(2), 281–287 (2013).
27. Zhu, Y. *et al.* Fine mapping of qTGW10--20.8, a QTL having important contribution to grain weight variation in rice. *Crop J.* **7**(5), 587–597 (2019).
28. Feng, Y. *et al.* Validation of a QTL for grain size and weight using an introgression line from a cross between *Oryza sativa* and *Oryza minuta*. *Rice* **14**(1), 1–12 (2021).
29. Price, A. H. Believe it or not, QTLs are accurate!. *Trends Plant Sci.* **11**, 213–216 (2006).
30. Mirdar Mansuri, R. *et al.* Salt tolerance involved candidate genes in rice: an integrative meta-analysis approach. *BMC Plant Biol.* **20**(1), 1–14 (2020).
31. Courtois, B. *et al.* Rice root genetic architecture: Meta-analysis from a drought QTL database. *Rice* **2**(2), 115–128 (2009).
32. Islam, M., Ontoy, J. & Subudhi, P. K. Meta-analysis of quantitative trait loci associated with seedling-stage salt tolerance in rice (*Oryza sativa* L.). *Plants* **8**(2), 33 (2019).
33. Xu, F. *et al.* Rapid identification of major QTLs associated with rice grain weight and their utilization. *PLoS ONE* **10**(3), e0122206 (2015).
34. Zhu, Y. *et al.* Fine mapping of qTGW10–20.8, a QTL having important contribution to grain weight variation in rice. *Crop J.* **7**(5), 587–597 (2019).
35. Venske, E. *et al.* Meta-analysis of the QTLome of Fusarium head blight resistance in bread wheat: Refining the current puzzle. *Front. Plant Sci.* **10**, 727 (2019).
36. Yang, L. V. *et al.* Genome-wide association analysis and allelic mining of grain shape-related traits in rice. *Rice Sci.* **26**, 384–392 (2019).
37. Nayak, A. K. *et al.* Genetic dissection of grain size traits through genome-wide association study based on genic markers in rice. *Rice Sci.* **29**, 3 (2022).
38. Muhammed Azharudheen, T. P. *et al.* Genome-wide association analysis for plant type characters and yield using cgSSR markers in rice (*Oryza sativa* L.). *Euphytica* **218**, 69 (2022).
39. Song, Y. *et al.* Proteomic analysis of copper stress responses in the roots of two rice (*Oryza sativa* L.) varieties differing in Cu tolerance. *Plant Soil* **366**(1), 647–658 (2013).
40. Lian, L. *et al.* PEPC of sugarcane regulated glutathione S-transferase and altered carbon–nitrogen metabolism under different N source concentrations in *Oryza sativa*. *BMC Plant Biol.* **21**(1), 1–15 (2021).
41. Wei, H., Wang, X., He, Y., Xu, H. & Wang, L. Clock component OsPRR73 positively regulates rice salt tolerance by modulating OsHKT2; 1-mediated sodium homeostasis. *EMBO J.* **40**(3), e105086 (2021).
42. Guo, N., Zhang, S., Gu, M. & Xu, G. Function, transport, and regulation of amino acids: What is missing in rice?. *Crop J.* **9**(3), 530–542 (2021).
43. Liu, X. *et al.* Polycomb protein OsFIE2 affects plant height and grain yield in rice. *PLoS ONE* **11**(10), e0164748 (2016).
44. Deng, X. *et al.* OsINV3 and its homolog, OsINV2, control grain size in rice. *Int. J. Mol. Sci.* **21**(6), 2199 (2020).
45. Chen, K., Lyskowski, A., Jaremko, Ł & Jaremko, M. Genetic and molecular factors determining grain weight in rice. *Front. Plant Sci.* **12**, 605799 (2021).

46. Khahani, B., Tavakol, E., Shariati, V. & Rossini, L. Meta-QTL and ortho-MQTL analyses identified genomic regions controlling rice yield, yield-related traits and root architecture under water deficit conditions. *Sci. Rep.* **11**(1), 1–18 (2021).
47. Kumar, S. *et al.* Meta-QTLs, ortho-MQTLs, and candidate genes for thermotolerance in wheat (*Triticum aestivum* L.). *Mol. Breed.* **41**(11), 1–22 (2021).
48. Huang, L. *et al.* The ascorbate peroxidase APX1 is a direct target of a zinc finger transcription factor ZFP36 and a late embryogenesis abundant protein OsLEA5 interacts with ZFP36 to co-regulate OsAPX1 in seed germination in rice. *Biochem. Biophys. Res. Commun.* **495**(1), 339–345 (2018).
49. He, Y. *et al.* PREMATURE SENESCENCE LEAF 50 Promotes Heat Stress Tolerance in Rice (*Oryza sativa* L.). *Rice* **14**(1), 1–7 (2021).
50. Yasuda, H., Hirose, S., Kawakatsu, T., Wakasa, Y. & Takaiwa, F. Overexpression of BiP has inhibitory effects on the accumulation of seed storage proteins in endosperm cells of rice. *Plant Cell Physiol.* **50**(8), 1532–1543 (2009).
51. Jain, M. & Khurana, J. P. Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J.* **276**(11), 3148–3162 (2009).
52. Luo, L. *et al.* OsASN1 plays a critical role in asparagine-dependent rice development. *Int. J. Mol. Sci.* **20**(1), 130 (2019).
53. Lin, Z. *et al.* Novel OsGRAS19 mutant, D26, positively regulates grain shape in rice (*Oryza sativa*). *Funct. Plant Biol.* **46**(9), 857–868 (2019).
54. Gaufichon, L., Rothstein, S. J. & Suzuki, A. Asparagine metabolic pathways in arabidopsis. *Plant Cell Physiol.* **57**, 675–689 (2016).
55. Zhang, Q. *et al.* Characterization of Arabidopsis serine:glyoxylate aminotransferase, AGT1, as an asparagine aminotransferase. *Phytochemistry* **85**, 30–35 (2013).
56. Khahani, B., Tavakol, E., Shariati, V. & Fornara, F. Genome wide screening and comparative genome analysis for Meta-QTLs, ortho-MQTLs and candidate genes controlling yield and yield-related traits in rice. *BMC Genom.* **21**(1), 1–24 (2020).
57. Darvasi, A. & Soller, M. A simple method to calculate resolving power and confidence interval of QTL map location. *Behav. Genet.* **27**(2), 125–132 (1997).
58. Arcade, A. *et al.* BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* **20**(14), 2324–2326 (2004).
59. Chardon, F. *et al.* QTL meta-analysis in Arabidopsis reveals an interaction between leaf senescence and resource allocation to seeds. *J. Exp. Bot.* **65**(14), 3949–3962 (2014).
60. Prakash, N. R. *et al.* Meta-analysis and validation of genomic loci governing seedling and reproductive stage salinity tolerance in rice. *Physiol. Plant.* **174**, e13629 (2022).
61. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15 (1987).
62. Niu, Y. *et al.* Identification and allele mining of new candidate genes underlying rice grain weight and grain shape by genome-wide association study. *BMC Genomics* **22**, 602 (2021).
63. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

Acknowledgements

We thank ICAR-National Rice Research Institute, Cuttack for financial support for this work. We also thank ICAR-NBPGR for sharing high grain weight genotype for the study. All authors have declared that there are no conflicts of interests regarding this article.

Author contributions

A.C., R.P.S. and M.A.T.P. conceptualized the research; S.B. and N.S. performed lab experiments; N.R.P., S.N.C. and D.B.N. formal data analysis; B.C.M. and SKN provided genetic material; A.C., R.P.S. and D.B.N. prepared draft manuscript, B.C.P. and S.N.C. edited the draft manuscript. All authors contributed to improve the manuscript and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17402-w>.

Correspondence and requests for materials should be addressed to C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022