New insights into QTNs and potential candidate genes governing rice yield via a multi-model genome-wide association study

Supriya Sachdeva[1], Rakesh Singh[1*], Avantika Maurya[1], Vikas K Singh[2], Uma Maheshwar Singh[3], Arvind Kumar[4] and Gyanendra Pratap Singh[5]

Abstract

Background Rice (*Oryza sativa* L.) is one of the globally important staple food crops, and yield-related traits are pre-requisites for improved breeding efficiency in rice. Here, we used six different genome-wide association study (GWAS) models for 198 accessions, with 553,229 single nucleotide markers (SNPs) to identify the quantitative trait nucleotides (QTNs) and candidate genes (CGs) governing rice yield.

Results Amongst the 73 different QTNs in total, 24 were co-localized with already reported QTLs or loci in previous mapping studies. We obtained fifteen significant QTNs, pathway analysis revealed 10 potential candidates within 100kb of these QTNs that are predicted to govern plant height, days to flowering, and plot yield in rice. Based on their superior allelic information in 20 elite and 6 inferior genotypes, we found a higher percentage of superior alleles in the elite genotypes in comparison to inferior genotypes. Further, we implemented expression analysis and enrichment analysis enabling the identification of 73 candidate genes and 25 homologues of *Arabidopsis*, 19 of which might regulate rice yield traits. Of these candidate genes, 40 CGs were found to be enriched in 60 GO terms of the studied traits for instance, positive regulator metabolic process (GO:0010929), intracellular part (GO:0031090), and nucleic acid binding (GO:0090079). Haplotype and phenotypic variation analysis confirmed that LOC_OS09G15770, LOC_OS02G36710 and LOC_OS02G17520 are key candidates associated with rice yield.

Conclusions Overall, we foresee that the QTNs, putative candidates elucidated in the study could summarize the polygenic regulatory networks controlling rice yield and be useful for breeding high-yielding varieties.

Keywords Multi-model GWAS, Quantitative trait nucleotides, Superior alleles, Yield, Ontology, Enrichment analysis, Candidate genes, *Oryza sativa* L, Networks

*Correspondence:
Rakesh Singh
rakesh.singh2@icar.gov.in
Full list of author information is available at the end of the article

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 2 of 18

Background

Rice is one of the major cereal crops that feeds over half of the human population and has been grown mainly in Asian countries for more than hundred decades [1–3]. Increasing yield potential has been a long-term objective in rice breeding and is imperative to overcome the global food crisis [4–9]. Rice yield is the most targeted complex trait that is a direct function of multiple factors including number and size of grains, productive tillers per plant, plant count per unit area, size of panicles and plant height [2, 7, 10–12] thus, the aforesaid component traits are prerequisite for achieving the desired yield in rice. This warrants the dissection of yield-related quantitative traits, discovery of the novel genetic factors and elucidate their molecular basis to meet the increased global rice demand.

New developments in genomic technologies, molecular markers, wealth of germplasm collections and genome information available on public portals have greatly facilitated deciphering of causative loci of complex traits for use in rice breeding [4, 13–19]. Essentially, Genome-wide association study (GWAS) has been evidenced to be a reliable approach for localization of genetic factors conferring quantitative traits to a narrow region [20–30] and has been successfully employed in different crop species such as rice, wheat, barley, chickpea, field pea, mungbean, linseed, flax and soybean [31–38]. Earlier genome studies reveal making use of dense coverage single nucleotide polymorphism markers in association analysis facilitates identification of functional variations governing yield related traits [39–45] connecting the functional implications of high throughput GWAS results and the known grain yield genes has become a standard approach for boosting rice productivity that may also be appropriate for other crop plants [46].

Extensive researches on identification of genetic variants for economically important traits have been performed using single-locus (SL) methodologies including MLM [47] and CMLM [48] that have limited quantitative trait nucleotides (QTNs) detection ability due to their polygenic nature and conservative Bonferroni correction factor involved [49, 50]. Multiple multi-dimensional/locus (ML) methods of genome analysis *viz.,* mrMLM [46], MLMM [51], FASTmrEMMA [50], FASTmrMLMM [52] were, therefore, proposed to estimate all the marker effects simultaneously with lesser running time, computational load and increased power accurately and identify the QTNs related to quantitative traits precisely yield related [12, 53–55]. Recent resequencing efforts at IRRI, Philippines also suggested harnessing germplasm diversity and novel alleles related to grain yield traits might play a vital role in rice breeding and improvement [6, 56, 57]. Furthermore, advances in functional genomics,

proteomics and metabolome studies, and their integration into a systems approach might prove instrumental in designing new higher yielding rice varieties for the future and accomplish the goals of rice biologists [46, 58]. CG-based association strategy has been recommended as a promising approach to develop 'tailored rice' suiting the ever-increasing rice demands [6].

The overall objective of our multi-model GWAS study was to identify the QTNs and CGs associated with yield traits *viz.,* plant height, days to flowering and plot yield across the selected subset of 3000 re-sequenced genome panel which were phenotyped during kharif 2020 and kharif 2021. We discovered common QTNs for the three aforesaid yield traits identified by both single and multi-locus models. We then conducted gene mining for the CGs in close vicinity of these significant QTNs and identified elite accessions with superior alleles in the selected subset which might accelerate selection breeding for higher yielding rice varieties. Additionally, we combined the expression data online available and pathway analysis results to better understand the molecular basis of these quantitative traits and improve grain yield in rice.

Material and methods

A simplified flow chart has been illustrated and presented (Fig. 1) to better understand the methodology adopted in execution of our multi-model GWAS study, and the steps implemented are as ensued:

Material, field trial and measurement of grain yield traits

The rice subset of 196 accessions from 3000 genomes re-sequenced panel (https://doi.org/https://doi.org/10.1186/2047-217X-3-7) collected from IRRI South Asia Regional Center (IRRI-SARC), Varanasi, India and 2 check varieties *viz.,* Pusa Basmati 1121 and Lalat were used as the materials in our study. These genotypes originating from 32 countries are presented in Table S1. The complete subset was planted in the ICAR-IARI farm, New Delhi, India during Kharif season of 2020 and 2021 with augmented RCB design in 7 blocks. The seeds were sown in a nursery bed and 21 days old seedlings were then replanted in puddled experimental plots in two rows with a distance of 20x15cms and four replications. Recommended farmer practices were followed for management of transplanted rice. Plant height(PlHt, cms), days to 50% flowering (DTF) and plot yield (PYD, kg/ha) was then recorded from five randomly selected healthy looking plants/plot as per the Rice Standard Evaluation System (http://www.knowledgebank.irri.org/images/docs/rice-standard-evaluation-system.pdf). All the three yield traits were visualized in R Studio, including the range, mean, standard deviation, coefficient of variation (CV), kurtosis and skewness (Table S2). Phenotypic correlations were also

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 3 of 18
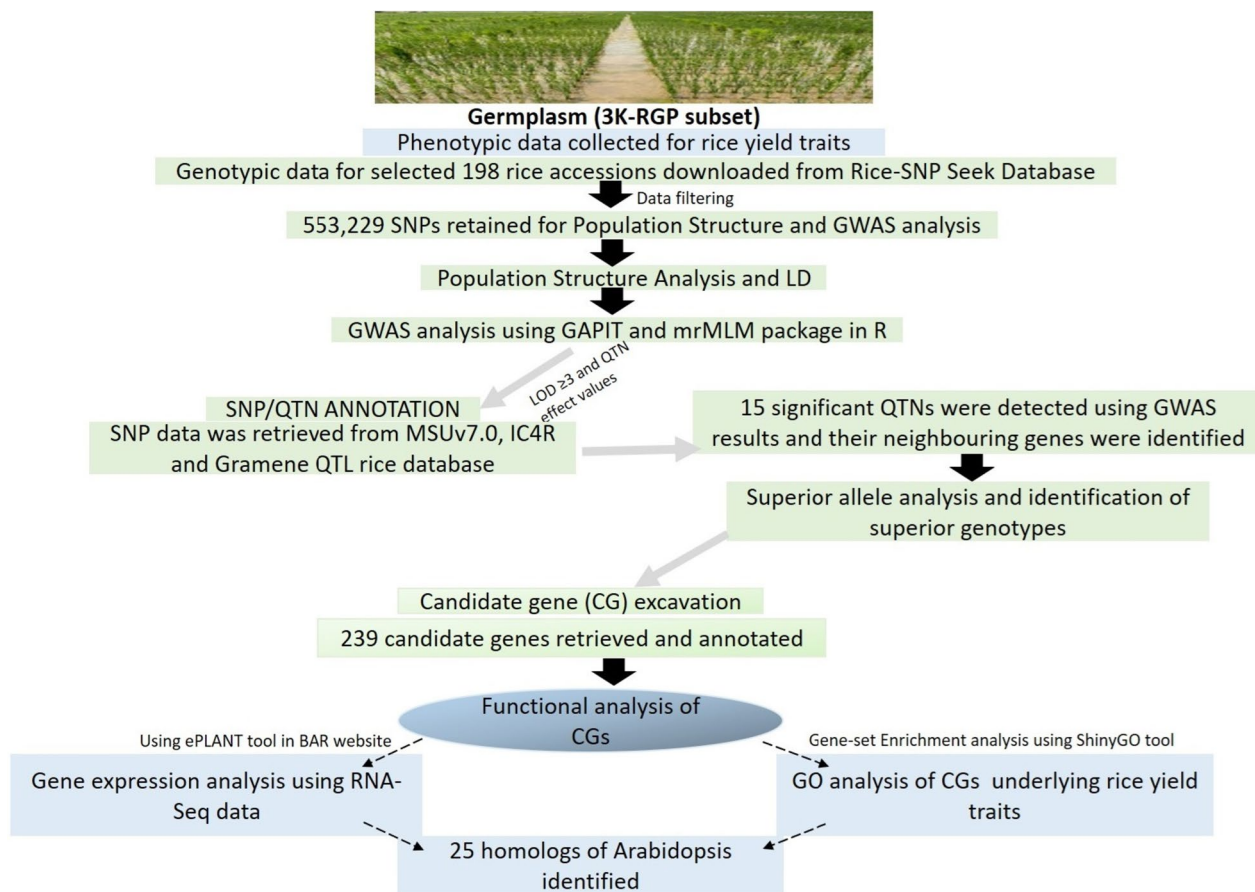
Fig. 1 Schematic representation of the methodology followed in our study
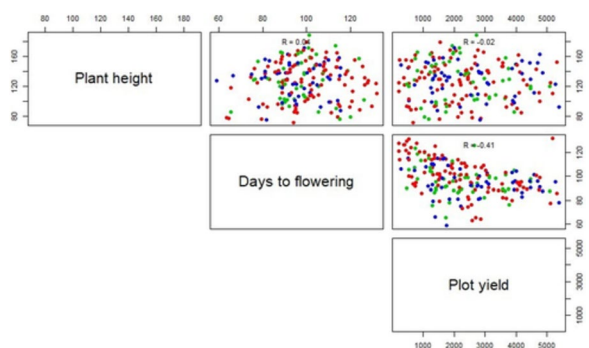

Fig. 2 Distribution of three yield traits in rice and Pearson's coefficients. Color gradients depict range of trait values. Blue, green and red dots indicate lower, moderate and high values of grain yield traits. Pl Ht, Plant height; DTF, Days to flowering; PYD, Plot yield

computed using the R package corrr (https://www.r-project.org/) (Fig. 2).

SNP data genotyping, genetic diversity, structure analysis and LD

The genotypic data for the selected accessions was downloaded from the online repository of variants of rice, the Rice SNP-seek database (http://snp-seek.irri.org/). Rice dataset comprising of 553,229 SNPs was obtained across the rice genome with a minor allele frequency (MAF) >5% and major allele frequency <95% for structure analysis and GWAS. Plots showing MAF and uniform SNP density distribution across the chromosomes illustrate suitability of the SNP dataset for genetic dissection of yield related traits in rice (Fig. 3a-b). Genetic locus-based diversity estimates, such as gene diversity (PiPerBP), minor allele frequency (MAF), expected number of polymorphic sites per nucleotide (ThetaPerBP), observed heterozygosity (H_o) were calculated using TASSELv5.2.82 software (Table S3). The genetic distance matrix was then calculated, following the neighbour-joining method, and a phylogenetic tree was constructed, visualized using interactive tree of life (iTOL) software. Principal component analysis (PCA) was performed using R package prcomp to further validate the NJ results. Bayesian

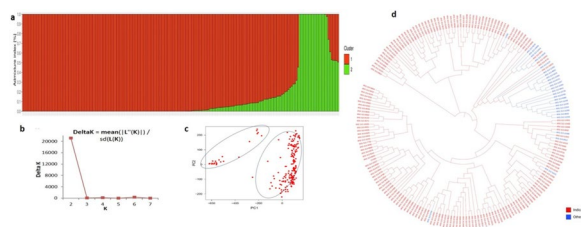Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 4 of 18
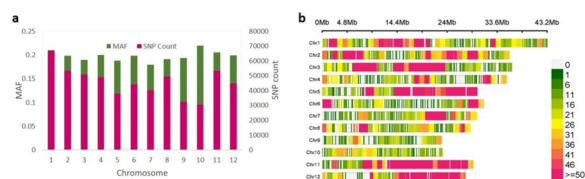
Fig. 3 The distribution of SNP markers on 12 rice chromosomes. **a** Distribution of average MAF. **b** SNP markers density across different chromosomes

model-based structure analysis was executed in STRUCTUREv2.3.4 [59] wherein the supposed number of subpopulations ranged from 1 to 7 and with each K repeated three times. The burnin period of 100,000 iterations and 100,000 MCMC (Markov Chain Monte Carlo) period were implemented for each run and the number of subgroups were identified using the Evanno criterion [60]. The optimum K value was then determined using the STRUCTURE HARVERSTER [61] (Fig. 4). Tidyverse and ggplot2 package were used to determine the LD between each pair of SNPs and the trend of Linkage disequilibrium (LD) decay was assessed using the squared coefficient of correlation (r^2) values of alleles and physical distance in Mb (Fig. S1).

GWAS and favourable allele analysis

GAPIT (Genome Association and Prediction Integrated) tool [62] and R package mrMLM [63] were used to perform GWAS in 198 rice accessions to identify the candidate QTNs including MLM [64], CMLM [47], mrMLM [49], FASTmrMLM [52] and FASTmrEMMA [50] and confirm the true associations with rice yield. Kinship matrix utilized in the study was calculated using TASSELv5.2.82 software. In single-locus models, the critical threshold of significant association was set similar

to recent GWAS studies [65–74]. In multi-locus GWAS, QTNs with Logarithm of Odds, LOD score of ≥ 3 were adopted as the significant SNPs associated with grain yield [8, 19, 75–77]. QTNs detected by both ML-GWAS models and SL-GWAS models were projected as candidates for rice yield related traits. Manhattan and QQ plots for all the three traits are presented in Fig. 5. Favourable allele of each of the commonly identified QTN was then identified using the QTN effect values and the genotype for code 1. In case, the QTN has a positive effect value, genotype for code 1 is considered the favourable SNP allele; in case the QTN has a negative effect, another genotype is considered favourable. Percentage of superior alleles was calculated for each of the common QTN. For every rice accession, proportion of favourable alleles in these QTNs was estimated as the count of favourable alleles divided by total count of QTNs and visualized as a heatmap.

SNP annotations, prediction and enrichment analysis of putative candidates

The online rice databases *viz.*, Genome Annotation database for rice (MSUv7.0, http://rice.uga.edu/); Information Commons for Rice (IC4R, http://ic4r.org/) and Gramene QTL (https://www.gramene.org/) were used to annotate the genes around the common QTNs discovered by different GWAS models. For all the identified gene loci, regions within LD decay distance were used to search for probable rice yield associated genes. With a view to further comprehend the molecular basis, we conducted gene-set enrichment analysis using ShinyGOv0.77 graphical tool wherein their biological functions, localization
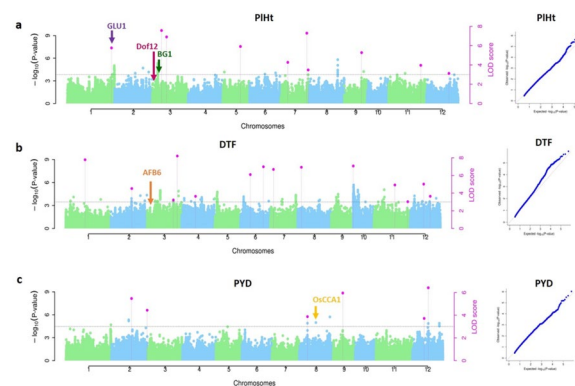

Fig. 4 Population Structure analysis, principal components and neighbour joining tree analysis representing 198 rice accessions. **a** Population structure (K=2); the areas of the two colors (red and green) depict the two subpopulations. **b** Plot of ΔK calculated for K =2. **c** Distribution of selected accessions along the three PCs. PC1 and PC2 refer to the first and second principal components respectively. Dots denote each accession of the rice subset. **d** Neighbour joining tree with different colours represent the subpopulations identified in our study i.e. red and blue represent the indica subspecies and other subpopulations, respectively

Fig. 5 Manhattan plots and quantile-quantile plots for PlHt, DTF and PYD using three multi-locus models. The horizontal dotted line indicates the threshold LOD score ≥3. The dots above the threshold value represent the significant QTNs at different rice chromosomes, the dots in pink colour represent QTNs detected by ≥2 models. Pl Ht, Plant height; DTF, Days to flowering; PYD, Plot yield

within plant cells and signalling pathways are disclosed [78].

Expression profile of candidate genes

The database BAR (The Bio-Analytical Resource for Plant Biology, https://bar.utoronto.ca/) was adopted to investigate all the candidates in different tissues to further illustrate the associations between genes and phenotypic differences. Subsequently, genes showing high expression in particular tissues were identified. Their homologs were also searched using the *Arabidopsis* Information Resource (TAIR) database (https://www.Arabidopsis.org/) and hypothesized on their possible functions using QuickGO tool (https://www.ebi.ac.uk/QuickGO/). The *Arabiodopsis* homologs expression analysis was then performed using the Affymetrix ATH1 whole genome genechip arrays normalized using the FPKM and Robust Multi-array average (RMA) method [79] at *Arabidopsis* eFP browser tool (https://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). The python package bioinfokit was used to create a heatmap illustrating the FPKM values of the homologs identified.

Analysis of haplotypes and phenotypic differences

To verify the associated locus between candidate genes and rice yield traits, SNP Seek software (https://snp-seek.irri.org/) was used to perform haplotype analysis considering the non-synonymous coding SNPs, then student's t test was performed to test the significant variations among the haplotypes. The haplotypes revealed and phenotypic distribution of each yield trait were then depicted as boxplots using ggplot2 package in R Studio.

Results

Phenotypic variance and correlations among yield traits

Three grain yield related traits, PlHt, DTF, and PYD were measured to study the variation in the subset panel of 198 selected accessions of rice. The scatter plot generated revealed that rice accessions show large variation for all the studied yield traits (Fig. 3). The mean value values recorded for PlHt, DTF and PYD were 127.25cm, 98.38 cm, and 2333.19 kg/ha, correspondingly. Furthermore, PlHt showed a wider variation in comparison to DTF ranging from 72.10 to 187.90 cms. Measures of variance for all these traits derived from descriptive analysis are presented in Table S2. Meanwhile, the PYD with maximum CV at 56% indicated that PYD had the highest extent of variation. Correlations among the three yield related traits were observed. PlHt and DTF were positively correlated with each other and negatively but significantly correlated with PYD. The Pearson coefficients, in overall were ranging from low to moderate.

Genotyping of selected subset of 3K panel

Genotypic data for 198 accessions of rice was retrieved from the 1M GWAS SNP dataset available at the Rice SNP Seek Database (http://snp-seek.irri.org/_download.zul). Followed by imputation in Beaglev5.4 software, 5,53,229 SNPs were filtered with a read depth of 10 and minor allele frequency >5% and mapped on the 12 rice chromosomes. Maximum SNPs were mapped on chromosome 1(66716) and minimum SNPs were mapped on chromosome 10 (30527) (Fig. 3a). SNP distribution of different loci was moderately uniform across the rice genome indicating suitability of filtered SNPs for molecular dissection of traits governing in rice (Fig. 3b).

Genetic divergence, population structure and LD

Genetic diversity analysis was conducted to to infer the extent of genetic variablity available in selected subset of rice accessions. The summary of the results of diversity statistics presented in Table S3 point towards considerable variability in the 198 accessions selected for our GWAS study. The ThetaPerBP values ranged from 0.09 to 0.38 (maximum value for the SNP marker) with a mean value of 0.23. The PiperBP for all genetic loci in the rice subset ranged from 0.05 to 0.5, with an average value of 0.28551. The average minor allele frequency, observed heterozygosity, and major allele frequency considering all the 198 accessions were depicted. The Bayesian model-based pattern of population structure was defined using high quality 5,53,229 SNPs in STRUCTURE software. Structure runs from K = 3 to K = 7 using 5,53,229 SNP markers are shown in Figure S3. Applying Evanno et al. (2005) method, maximum ΔK/K value of 2 was selected (Fig. 4b) suggesting existence of two subpopulations in the selected subset of re-sequenced 3000 rice genomes (Fig. 4a). The two subpopulations identified comprised of 173 (87.37%) and 25 (12.63%) accessions, correspondingly. The larger cluster, cluster 1 consisting of *indica* rice accessions belonged to *indica* subpopulations while the smaller and diverse cluster 2 included accessions from *japonica, aus/boro* and *intermediate* type subpopulations. In the case of K = 3, rice accessions were grouped into three populations in conformity with NJ clustering, of which each population had all admixed individuals (Figure S3). Above K = 3, improving model fit with incremental K values, propose that these values fail to display any significant phylogenetic structure (Falush et al., 2003, Prichard et al., 2007). To explore the genetic differentiation due to population structure among the 198 rice accessions as reflected by these 5,53,229 filtered genome-wide SNPs, we employed Neighbor Joining (NJ) analysis. The clustering pattern of NJ analysis corroborated with the results of population structure of all

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 6 of 18

the selected accessions though the accessions into three major subgroups (Fig. 4d). A total of 26.26% of rice accessions belonging to *ind1a, ind1b, ind2* and *ind3* were included in cluster 2. Whereas, cluster 1 was identified as the smallest cluster consisting of only 4.04% of *indica* rice accessions. It was predominated in accessions belonging to *indx* and *ind1b* subpopulations. However, cluster 3 was the largest cluster that constituted 69.69% of the total accessions. It was recognized as the most diverse with smaller statured to medium height accessions belonging to *indica, japonica, aus/boro* and *intermediate* type subpopulations. According to principal component analysis, there were three subpopulations in the selected subset of rice accessions as represented by population structure analysis (Fig. 4c)

The genome-wide LD decay for filtered SNPs was plotted as a scatterplot between the average r^2 values against the physical map distance in mega basepairs. As shown in graphical display (Fig. S1), r^2 value decreased with the increase in distance resulting in fastest decay within 100 kb, and then progressively slowed down, therefore we searched genomic ranges of common QTNs (50kb on either side of the QTN identified).

QTNs identified for targeted yield traits using GWAS models

We analyzed all the three yield traits by employing six GWAS models to identify a total of 73 QTNs. Of these, 23,25 and 25 QTNs were associated with PlHt, DTF

and PYD correspondingly. Manhattan plots portraying significant QTNs and the respective QQ plots for the abovementioned traits are shown (Fig. 5). Additionally, we identified the common QTNs across GWAS methodologies that were considered robust QTNs strongly associated with the three targeted yield traits. Accordingly, 15 common QTNs were detected concurrently by both SL- and ML approaches located on chromosome no 1,2,3,7,8,9,10,11 and 12 (Table 1). Their LOD scores ranged from 3.11 to 7.93 with consistent QTN effect values across different GWAS methods. Amongst the mrMLM GWAS methodologies, maximum number of common QTNs were identified by FASTmrMLM model (25) (Fig. 6a) and amongst the different combinations of GWAS models, MLM, CMLM, MLMM and
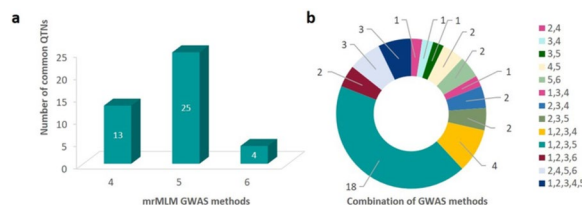
Fig. 6 Total number of QTNs identified in the GWAS study. **a** QTNs identified simultaneously across the mrMLM methods; **b** QTNs identified by different combinations of GWAS methods. GWAS methods: 1, MLM; 2, CMLM; 3, MLMM; 4, mrMLM; 5, FASTmrMLM; 6, FASTmrEMMA

Table 1 Common QTNs for three grain quantity traits in rice across different multi-methods

Trait	Method[a]	Marker	Chr	Position (bp)	QTN effect	LOD score	r² (%)[b]
Pl Ht	1,2,3,4	**41184096**	**1**	**41184096**	**7.82**	**5.76**	**5.64**
Pl Ht	5,6	**53308499**	**2**	10037576	-3.06, -8.86	3.56,3.15	1.45,2.92
PYD	1,2,3,5	**65474785**	**2**	22203862	**284.86**	**4.39**	**2.47**
PYD	1,2,3,5	77317239	2	34046316	257.91	3.91	2.24
PYD	1,2,3,5	**83271479**	**3**	4063306	**169.78**	**3.80**	**1.52**
Pl Ht	2,4,5,6	**238555432**	**7**	26223525	**7.43, 6.39, 8.97**	**7.29, 7.03, 3.19**	**7.34,6.29,3.03**
DTF	3,4	245045948	8	3016420	3.85	6.93	4.94
PYD	1,2,3,4,5	**280074957**	**9**	9602407	**-544.06**	**5.95**	**7.81**
DTF	1,2,3,4	293545796	10	60526	-4.74	7.07	7.43
Pl Ht	1,2,3,4	341803503	11	25110946	5.81	3.96	4.44
Pl Ht	3,5	**325332817**	**11**	8640260	**-10.84**	**4.82**	**6.28**
PYD	1,2,3,5	339876697	11	23184140	239.78	3.11	1.50
DTF	4,5	**359450698**	**12**	13737035	**-5.86, -3.57**	**5.02**	**7.85,3.5**
DTF	1,2,3,5	**368582572**	**12**	22868909	**-4.22**	**7.93**	**6.78**
PYD	1,3,4	**357969622**	**12**	12255959	**-587.86**	**3.72**	**8.61**

Pl Ht Plant height; *DTF* Days to flowering, *PYD* Plot yield

[a] MLM, CMLM,MLMM,mrMLM,FASTmrMLM,FASTmrEMMA were indicated by , respectively

[b] r²(%), proportion of total phenotypic variation explained by each QTN. Bold text indicates the QTNs appeared to be in vicinity of genes or QTLs associated with Pl Ht, DTF and PYD

FASTmrMLM identified the maximum number of QTNs (18) (Fig. 6b). We mapped one robust QTN on chromosome which was detected by all the three mrMLM GWAS models, 238555432 with variable LOD score values of 3.19 to 7.29 and percentage of variance explained (r^2) ranging from 3.03 to 7.34 (Table 1).

Based on gene annotations and prior studies, 10 out of 15 common QTNs have been confirmed to be located within or in proximity of the known yield related genes *viz.*, 41184096, 53308499, 325332817, 359450698, 357969622, 368582572, 65474785, 83271479, 238555432, 280074957 and the remaining 5 QTNs were newly discovered with phenotypic variance of 1.5% to 7.43% (Table 1). These results established the experimental accuracy of our study on QTN detection and also a rapid way to select genotypes with superior allelic compositions for breeding high yielding rice.

Superior allele analysis in selected rice 3K subset

On the basis of average trait values of PlHt, DTF and PYD, we identified 20 elite genotypes having higher phenotypic values and 6 inferior genotypes with lower phenotypic values (Table 2, Fig. 7). The percentage of superior alleles across the 15 QTNs in each of the elite genotypes varied from 33.33% to 73.33% of which 80% (16 of 20 genotypes) displayed >50% PSA and the remaining 4 genotypes (20%) exhibited PSA <50%. On the other hand, the percentages for superior alleles in inferior genotypes fluctuated from 13.33% to 40% wherein all the 6 genotypes displayed PSA <50% indicating that 20 genotypes carry all the desired alleles for the targeted yield traits and can be employed in molecular breeding in rice.

Additionally, we discovered few common superior alleles in multiple elite genotypes. For instance, seven genotypes IRIS_313-9966, IRIS_313-9262, IRIS_313-9139, IRIS_313-9551, IRIS_313-8305, IRIS_313-9066, IRIS_313-8641 and IRIS_313-10401 carry all the superior alleles 41184096, 25110946, 100375576, 8640260, 23184140 and 22203862 and the superior alleles 8640260 and 22203862 occurred in all the 20 genotypes (Fig. 8). We predict that these common superior alleles might strongly influence the grain yield traits and generate valuable information for breeding improved rice in future.

### Excavation of candidate loci around robust QTNs identified

We searched the probable CGs for each of the 15 QTNs. Consequently, we identified 239 genes and 73 genes were highly expressed in specific tissues *viz.,* shoot, leaves and rice seed grains and inflorescence as per the BAR tool (Table 3). Data on gene annotations revealed that 39 of 73 genes (53.42%) were annotated previously in 16 pathways

Table 2 Phenotypic averages of plant height, days to flowering and plot yield and proportion of superior alleles in 26 rice genotypes across 15 common QTNs

Line	PSA (%)	Pl Ht(cms)	DTF	PYD (kg/ha)
IRIS_313-8305	60.00	120.10	97.75	4461.25
IRIS_313-8586	66.67	108.90	98.50	3352.50
IRIS_313-8641	60.00	103.60	100.12	3129.17
IRIS_313-8721	66.67	121.80	89.62	4363.33
IRIS_313-8900	66.67	117.50	95.50	4465.42
IRIS_313-9066	60.00	119.80	87.75	5098.33
IRIS_313-9139	60.00	111.50	88.75	4741.66
IRIS_313-9262	66.67	111.20	93.16	4105.42
IRIS_313-9400	46.67	108.70	92.50	2539.58
IRIS_313-9551	60.00	95.90	95.25	3033.33
IRIS_313-9917	60.00	109.70	92.25	2770.00
IRIS_313-9966	73.33	100.00	99.12	4777.08
IRIS_313-10020	66.67	111.40	101.50	4587.92
IRIS_313-10294	66.67	98.10	88.75	3949.58
IRIS_313-10307	73.33	121.90	101.50	2215.00
IRIS_313-10333	40.00	103.30	89.75	2242.92
IRIS_313-10337	33.33	107.00	96.25	3974.17
IRIS_313-10361	46.67	85.40	94.25	2563.75
IRIS_313-10401	53.33	107.60	92.00	2592.50
IRIS_313-11048	53.33	104.20	88.00	3645.42
IRIS_313-8437	**33.33**	**134.10**	**115.00**	**1680.83**
IRIS_313-11010	**26.67**	**154.90**	**122.00**	**453.75**
IRIS_313-11059	**40.00**	**162.91**	**104.50**	**1222.92**
IRIS_313-11138	**13.33**	**155.40**	**114.00**	**239.17**
IRIS_313-11588	**40.00**	**156.70**	**123.75**	**737.08**
IRIS_313-11902	**26.67**	**145.80**	**113.50**	**1061.67**

Bold font indicates the 6 lines with lower average values of Pl Ht, DTF, PYD and non-bold font indicates the 20 lines with higher average phenotypic values

Pl Ht Plant height; *DTF* Days to flowering, *PYD* Plot yield
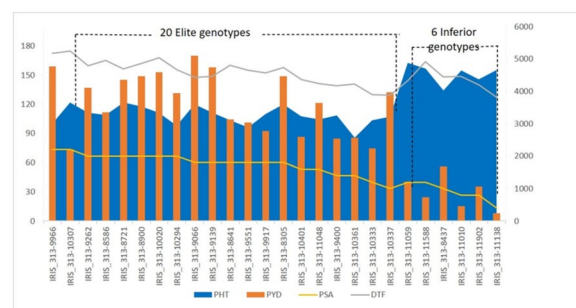
Fig. 7 Distribution of elite alleles and yield traits in 26 higher and lower average phenotypic value genotypes

using ShinyGOv0.77 tool (Fig. 9). Homologs of the 73 candidates were identified in model plant *Arabidopsis*

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 8 of 18
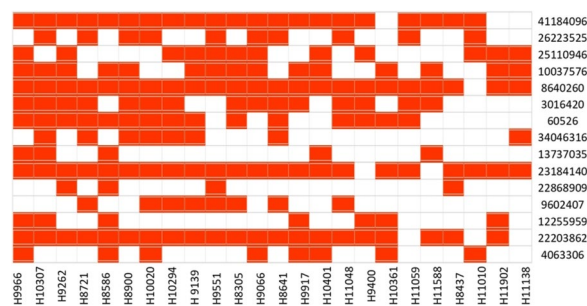
Fig. 8 Heatmap presenting the distribution of superior allelic combinations for the 15 QTNs identified in 20-elite and 6-inferior genotypes. Red and white coloured blocks correspond to the superior and inferior alleles

thaliana using TAIR database and anticipated their putative roles (Table 4).

Expression profile of the homologous genes

The *Arabidopsis* eFP browser tool at BAR database demonstrates the normalized FPKM expression data obtained for specific tissues, including 1st node, young leaves, mature leaf, shoots, cotyledons, seeds and rosette vegetative and after flowering. Heatmap of the homologous genes depicting the FPKM expression of the homologs in different tissues are shown in Fig. 10 (Table S4). As shown, AT1G07920 and AT01G18700 had the highest FPKM values in 1st node, young leaf, mature leaves, shoots, cotyledons, seeds and rosette. For flowering time QTNs, AT04G38120, AT01G63940, AT01G33540, AT02G38230, AT02G32810 and AT01G55860 had higher expression in vegetative rosette and rosette after flowering indicating a potential association between these candidates and flowering time. For PlHt QTNs, AT04G28660, AT01G67250, AT01G18700, AT01G60900, AT04G28990 and AT01G15220 had higher expression in cotyledons, seeds and rosette suggesting these genes may have specifically a large role on yield. Furthermore, AT01G07920, AT01G54150, AT01G36730, AT03G49880, AT05G39850, AT01G26940 and AT01G10830 found in QTNs for plot yield showed higher expressions in cotyledons, seeds and inflorescense, and were identified to associate with rice yield.

Gene enrichment of identified candidates

ShinyGO tool was used to conduct GO functional enrichment analysis of candidates underlying the targeted yield traits. According to enrichment analysis, 40 of 73 candidates were annotated and significantly enriched for 60 GO terms ($p < 0.05$). Out of 60 GO terms, 34,13 and 13 GO terms were enriched in the category of biological, cellular and metabolic component. The bar plot depicting the most significant pathways of these CGs are shown in red colour (Fig. 9). Genes enriched in biological processes comprised of GO terms associated with biogenesis, cell development, reproduction, anatomical structure development, regulation of metabolic processes, signalling and stress responses (Fig. 11). Interestingly, one GO term, GO:0010929, a positive regulator of auxin mediated signalling pathway also governs several other pathways mediating growth, reproduction and architecture of rice plants. The 25 genes involved in cellular processes included organelle membrane, protein complexes, envelope, photosynthetic and endomembrane system. Another enriched term, GO:0031090 was associated with different membrane-bound organelles within the cell. Among the molecular components, GO:0090079 with translation regulator and nucleic acids binding activity has a central role in initiation, activation and termination of polypeptide synthesis at the site of ribosomal assembly. These findings demonstrate the influence of these CGs on rice yield and provide valuable insight into their genetic basis and discover new yield related genes in rice.

Haplotype and phenotypic difference analysis of candidate genes

To further confirm the association of the candidate loci and rice yield traits using SNPs within the candidate genes. LOC_OS09G15770 (9634451-9638121) was analyzed to reveal the intragenic variation affecting the yield in rice and to identify superior haplotype. Fig. 12(a) depicts the linkage disequilibrium and haplotype block with five SNPs at 9636961 bp, 9637028 bp, 9637406 bp, 9637426 bp and 9637435 bp. The 198 accessions were categorized into 3 haplotypes based on these 5 SNPs *viz.,* HapA (GTTCG), HapB (TCTCG), HapC (GCCCG). Amongst these haplotypes, HapA recorded the highest average PYD (3118.67) whereas the HapB presented the lowest average PYD (1932.77) (Fig. 12b). A student t test showed that significant differences existed between the haplotypes (P-value = 0.014). There was also a significant variation in PYD between haplotypes of LOC_OS02G36710 (P-value = 0.0017). Therefore, we deduce the candidate genes LOC_OS09G15770 and LOC_OS02G36710 to be associated with grain yield in rice. Supplementary Figure S2 depicts the results of haplotype block and phenotypic variation in LOC_OS02G17520, which was identified for plant height. We may presume that candidate loci LOC_OS09G15770, LOC_OS02G36710 and LOC_OS02G17520 might be associated with rice yield.

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 9 of 18

Table 3 Showing detailed information on pathway annotations of 73 candidate loci

QTN name	Gene name[a]	Chromosome	Position	GO number	Annotation
41184096	**LOC_Os01G71190**	**1**	**41186974..41188839**	**GO:0031090**	**Photosystem II reaction center PSB28 protein**
41184096	LOC_Os01G71106	1	41160726..41155813	-	Disease resistance protein (TIR-NBS-LRR class) family
41184096	LOC_OS03G07960	3	4069589..4072442	GO:0031090	expressed protein
41184096	LOC_Os03G08020	3	4506887-4507439	GO:0003746	elongation factor Tu, putative, expressed
10037576	**LOC_Os02G17460**	**2**	**10049191..10042478**	**-**	**tesmin/TSO1-like CXC domain containing protein, expressed**
10037576	**LOC_OS02G17520**	**2**	**10084266..10087922**	**GO:0034622**	**cytochrome c-type biogenesis protein ccmH precursor, putative, expressed**
10037576	LOC_OS02G17470	2	10050837..10057784	GO:0003723	RNA-binding protein-related, putative, expressed
22203862	LOC_OS02G36710	2	22174407..22181019	GO:0034622	galactosyltransferase family protein, putative, expressed
4063306	LOC_OS03G08050	3	4111409..4113361	GO:0006414	elongation factor Tu, putative, expressed
4063306	LOC_Os03G07994	3	4089784..4086997	-	expressed protein
4063306	LOC_Os03G08000	3	4090003..4094134	-	expressed protein
4063306	LOC_Os03G08010	3	4095194..4097239	-	elongation factor Tu, putative, expressed
26223525	LOC_OS07G43890	7	26242366..26245311	GO:0031090	emp24/gp25L/p24 family protein, putative, expressed
26223525	**LOC_Os07G43870**	**7**	**26230679..26240915**	**-**	**heat shock protein DnaJ, putative, expressed**
3016420	LOC_Os08G05670	8	3038465..3031282	-	HEAT repeat family protein, putative, expressed
3016420	LOC_Os08G05530	8	2965358..2968015	-	LSM domain containing protein, expressed
3016420	**LOC_Os08G05570**	**8**	**2977391..2982700**	**-**	**monodehydroascorbate reductase, putative, expressed**
9602407	LOC_Os09G15790	9	9643529..9646172	GO:0003924	ras-related protein, putative, expressed
9602407	LOC_Os09G15775	9	9636409..9637761	-	expressed protein
9602407	LOC_Os09G15760	9	9629854..9631896	-	expressed protein
9602407	LOC_OS09G15770	9	9634451..9638121	-	CPuORF13 - conserved peptide uORF-containing transcript, expressed
9602407	LOC_Os09G15720	9	9595103..9593570	-	expressed protein
23184140	LOC_OS11G38959	11	23184447..23187634	GO:0003723	40S ribosomal protein S9-2, putative, expressed
23184140	LOC_Os11g38970	11	23189532..23190792	-	expressed protein
23184140	LOC_Os11g38990	11	23212467..23217461	-	peptidyl-prolyl cis-trans isomerase, putative, expressed
23184140	LOC_Os11g38880	11	23142279..23140918	-	hypothetical protein
23184140	LOC_Os11g38910	11	23166977..23167787	-	pyruvate decarboxylase isozyme 1, putative, expressed
23184140	LOC_Os11g38930	11	23176604..23178856	-	tRNA-splicing endonuclease subunit Sen2, putative, expressed
60526	LOC_Os10g01030	10	40060..38648	-	hypothetical protein
60526	LOC_Os10g01010	10	22576..22214	-	transposon protein, putative, Ac/Ds sub-class, expressed
60526	**LOC_Os10g01110**	**10**	**103441..100327**	**-**	**OsSCP44 - Putative Serine Carboxypeptidase homologue, expressed**
60526	LOC_Os10g01080	10	82250..80964	-	SOR/SNZ family protein, putative, expressed
60526	LOC_Os10g01060	10	67119..72971	-	protein kinase family protein, putative, expressed
60526	LOC_Os10g01044	10	58337..45941	-	isoflavone reductase, putative, expressed
25110946	LOC_Os11g41820	11	25153456..25148703	GO:0003723	RNA recognition motif containing protein, expressed
8640260	LOC_Os11g15230	11	8598973..8596981	-	expressed protein
8640260	LOC_Os11g15280	11	8635205..8631629	-	TNP1, putative, expressed
13737035	**LOC_Os12g24170**	**12**	**13782378..13761534**	**-**	**beta-galactosidase precursor, putative, expressed**
13737035	**LOC_Os12g24080**	**12**	**13697068..13702579**	**-**	**HECT-domain domain containing protein, expressed**
13737035	LOC_Os12g24130	12	13735026..13738628	-	expressed protein
22868909	LOC_Os12g37210	12	22827572..22824580	-	expressed protein
12255959	LOC_Os12g21789	12	12267520..12264653	-	expressed protein
12255959	**LOC_Os12g21710**	**12**	**12204428..12208058**	**-**	**nnrU, putative, expressed**

Bold font signifies genes that we propose correlate with PlHt, DTF and PYD in rice

[a] Indicates the gene locus that correlates with the QTN (before the gene name in the same row)
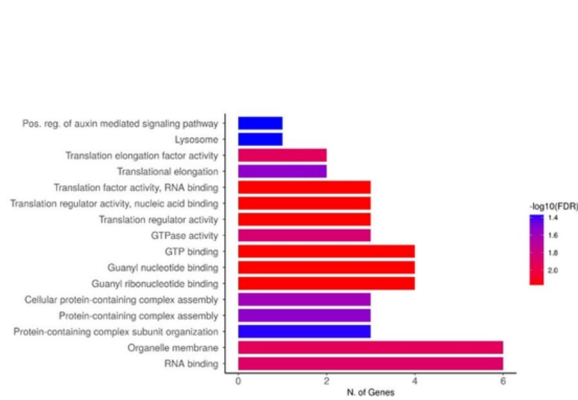
Fig. 9 Barplot showing gene enrichment pathways for 73 candidate genes

Discussion

Improving the productivity has been a major bottleneck for rice breeders ensuring food supply to an ever-growing global population. Mostly, yield traits are complex and modulated by various polygenes. GWAS predictions on the gene loci governing rice yield traits based on different regression models and identification of associated SNP markers might reveal new insights into the molecular mechanism underlying rice yield. Classic models of GWAS have been employed widely to discover variations in a gene particularly in several crops [16, 80] although, such models suffer from critical test corrections and ignore the effect of multiple gene loci in overall [8, 81, 82] therefore, ML-models with higher statistical power were developed. In our investigation, we employed six different GWAS models including classic and ML-models both in a 3K-RGP subset of 198 accessions to detect QTNs associated with three grain yield traits (PlHt, DTF and PYD). Fifteen QTNs were identified by two or more GWAS models instantaneously (Table 1). Remarkably, we dissected multiple gene clusters for the targeted traits; for example, the SNP marker 65474785 on chromosomal number 3 at 22203862 bp, was located in close vicinity of genes controlling number of rice grains/plant, plant height and heading date, LOC_Os03g07920 (OsBG1) [83], LOC_Os03g07360 (dof12) [84], LOC_Os3g08850(AFB6) [85]. Moreover, 41184096 located on chromosome 1 showed overlap

with LOC_Os01g71340 (OsGLU1) [86] encoding membrane bound endo-1,4-β-D-glucanase that

Table 4 Details of 25 homologous genes in *Arabidopsis thaliana*

| Gene name	GO number	Homologous gene	Annotation
LOC_OS01G71190	**GO:0031090**	**AT4G28660**	**Photosystem II reaction center PSB28 protein**
LOC_OS02G55640	**GO:0034622**	**AT1G67250**	**Proteasome maturation factor UMP1**
LOC_OS02G17520	**GO:0031090**	**AT1G15220**	**Cytochrome c biogenesis protein family**
LOC_OS02G17470	**GO:0003723**	**AT4G28990**	**RNA-binding protein-related**
LOC_OS03G08020	**GO:0003723**	**AT1G07920**	**GTP binding Elongation factor Tu family protein**
LOC_OS03G08050	**GO:0005525**	**AT1G07920**	**GTP binding Elongation factor Tu family protein**
LOC_OS03G08000	**-**	**AT1G54150**	**E3 Ubiquitin ligase family protein**
LOC_OS03G08010	-	AT1G07920	GTP binding Elongation factor Tu family protein
LOC_OS07G43890	GO:0031090	AT1G14010	Emp24/gp25L/p24 family/GOLD family protein
LOC_OS07G43870	**-**	**AT1G18700**	**DNAJ heat shock N-terminal domain-containing protein**
LOC_OS08G05670	**-**	**AT4G38120**	**ARM repeat superfamily protein**
LOC_OS08G05570	**-**	**AT1G63940**	**Monodehydroascorbate reductase 6**
LOC_OS09G15775	**-**	**AT1G36730**	**Translation initiation factor IF2/IF5**
LOC_OS09G15770	**GO:0005525**	**AT1G36730**	**Translation initiation factor IF2/IF5**
LOC_OS11G41820	**GO:0003723**	**AT1G60900**	**"U2 snRNP auxilliary factor, large subunit, splicing factor"**
LOC_OS11G38959	**GO:0003723**	**AT5G39850**	**Ribosomal protein S4**
LOC_OS11G38990	**-**	**AT1G26940**	**Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein**
LOC_OS11G38930	-	AT3G45590	Splicing endonuclease 1
LOC_OS10G01110	**-**	**AT1G33540**	**Serine carboxypeptidase-like 18**
LOC_OS10G01080	**-**	**AT2G38230**	**Pyridoxine biosynthesis 1.1**
LOC_OS10G01044	-	AT3G18890	NAD(P)-binding Rossmann-fold superfamily protein
LOC_OS12G24170	**-**	**AT2G32810**	**Beta galactosidase 9**
LOC_OS12G24080	**-**	**AT1G55860**	**Ubiquitin-protein ligase 1**
LOC_OS12G21789	**-**	**AT3G49880**	**Glycosyl hydrolase family protein 43**
LOC_OS12G21710	**-**	**AT1G10830**	**15-cis-zeta-carotene isomerase**

Bold font shows genes that we suggest correlate with plant height, days to flowering and plot yield in rice
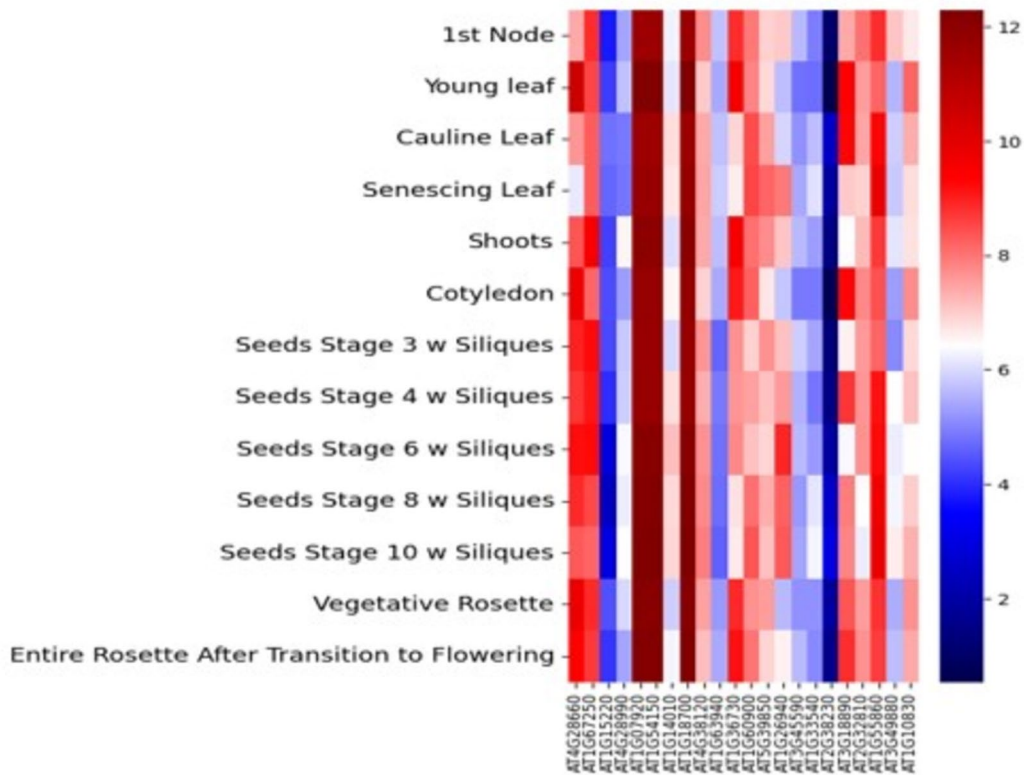
Fig. 10 Heatmap depicting FPKM expression values of the 25 homologs identified for the three grain yield traits. The red color indicates high expression, and the blue color shows low expression. y-axis is log2 (FPKM +1)
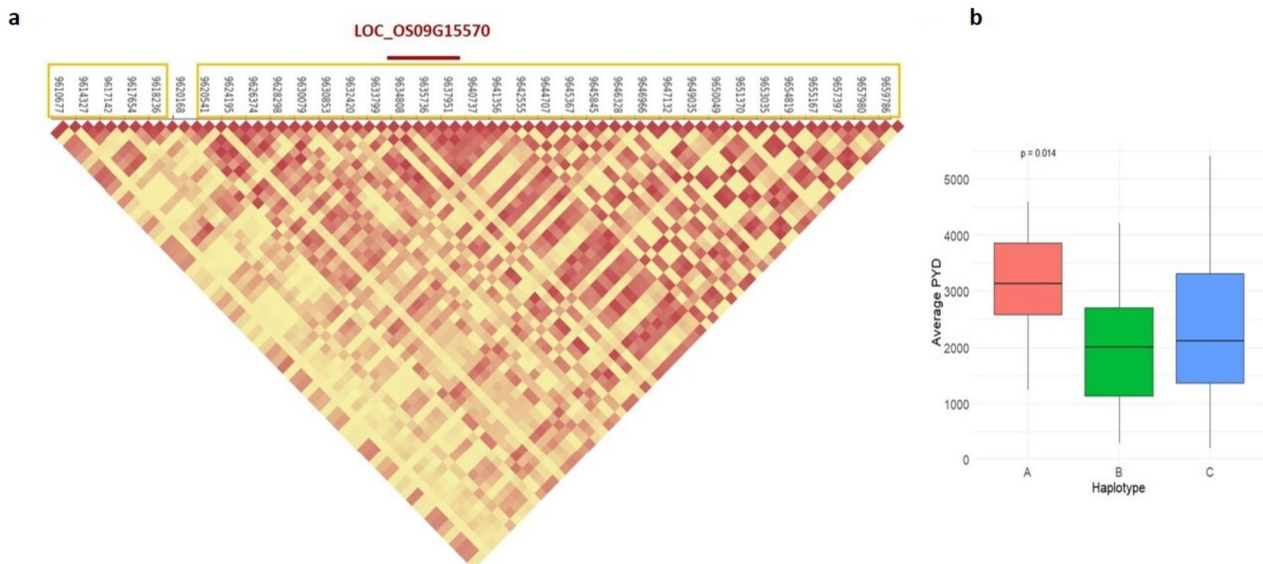
Fig. 11 Gene ontology tree depicting annotations of putative SNP markers. Boxes in the diagram represent the GO terms corresponding to a seven-digit ID number preceded by GO and their functional description. Yellow coloured nodes at the bottom represent the significantly enriched GO terms in enrichment analysis. Arrowheads to the left and right indicate the direction of relationships between different GO terms/ nodes at upper and lower levels

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 12 of 18
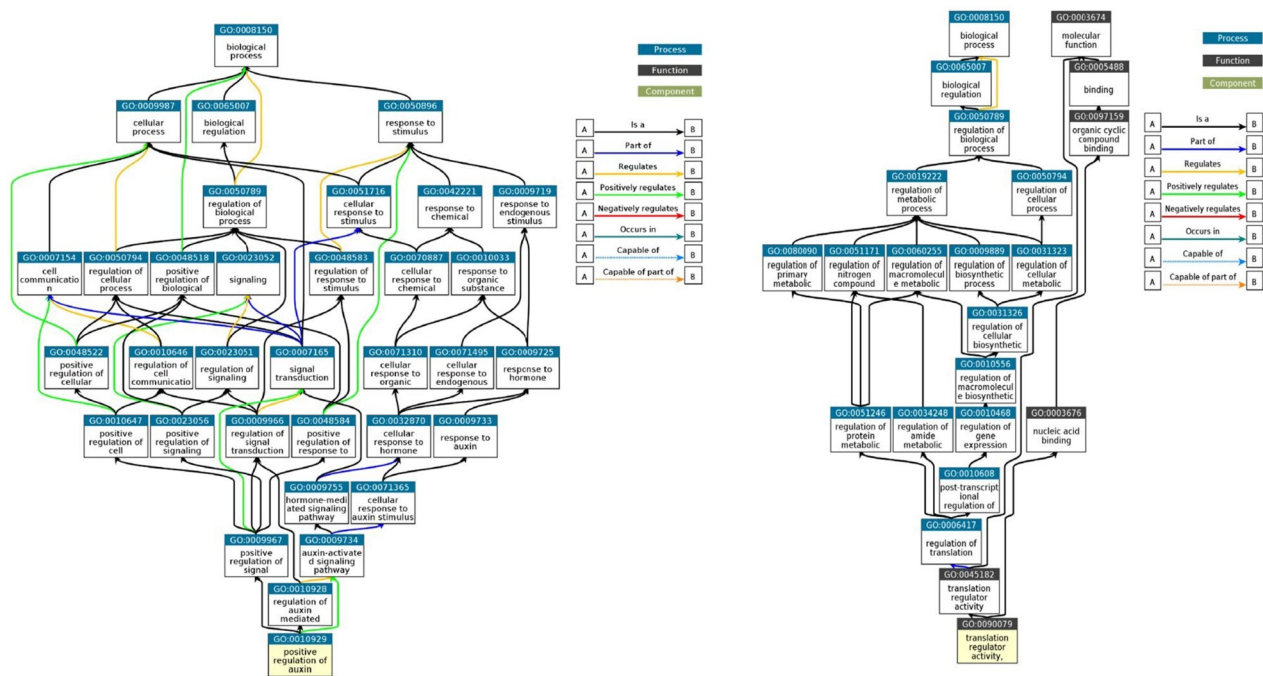
Fig.12 Results of haplotype and phenotypic difference analysis for the candidate gene LOC_OS09G15770. **a** LD and haplotype block with five SNPs within LOC_OS09G15770. **b** boxplot of PYD trait among the three haplotypes of LOC_OS09G15770. The SNP markers in LD region around the significant QTN (10037576) are shown in yellow color

participates in signalling pathways induced by plant hormones gibberellins (GA) and brassinosteroids (BR) and facilitates cell development. Glu1 mutants exhibit dwarfism due to defects in elongation of internodes in rice [5]. Furthermore, marker 245045948 at chromosome 8 was found to be closely associated with LOC_08g06110 (OsCCA1) regulating tiller production and panicle development in rice [7, 19]. These results demonstrated that the QTNs unravelled in our study possibly will be beneficial in modulating grain yield related traits in rice.

We studied pathway annotations and conducted GO analysis to identify the putative genes associated with rice yield (Table 3, text highlighted in bold). LOC_Os01G71190 (photosystem II reaction center protein, Psb28) belonging to the family of light induced genes in rice has regulatory role in stabilizing PSII activity. Knockouts of LOC_Os01G71190 caused a pale green phenotype and dwarfism [87]; might serve as a key gene for cell growth and development located near the OsGLU1 gene (157kb) that controls cell wall biosynthesis in coordination with GA and BR [5]. LOC_02G17460 is a member of polycomb group of proteins that encode TSO1-like cysteine rich (CXC) domain containing proteins highly expressed in the shoot apex, carpels, pollen, and seeds; works as a regulator of gene subsets involved in cell division and proliferation [88]. TSO1 insertional mutants also show phenotypic aberrations in pollen grains and

carpels which indirectly affects the grain yield [89, 90]. LOC_Os02G17520 is a cytochrome c biogenesis ccmH precursor that has been confirmed as an essential housekeeping gene [91]. CCMH, a thiol-disulfide oxidoreductase within the mitochondrial membrane which is involved intricately in cytochrome c maturation and electron transfers between enzymes involved in photosynthesis and respiration [91]. Insertional mutation studies demonstrated lethal defects of knockout CCMH gene by arresting the embryo development at torpedo stage [92] and therefore, may play a crucial role in growth and embryo development. LOC_Os07G43870 codes for heat shock protein DnaJ that improves rice architecture by modulating gibberellin homeostasis [93] thereby, enhancing rice yield. Notably, a loss of function mutation ($NAL11^{-923\text{del}-1552}$) in Narrow Leaf 11 gene, HSP with DnaJ domain results in GA defects and inhibits chloroplast development suggesting the role of LOC_Os07G43870 and gibberellins in determining the productivity in rice [94]. LOC_08G05570 is a monohydrate reductase gene (MDHAR) that participates in the ascorbate-glutathione cycle and mutants generated by RNAi exhibited decreased chlorophyll, relative ascorbate-dehydroascorbate ratios which in turn reduce biomass and rice yield [95]. LOC_Os10G01110 (OsSCP44) is annotated as a putative serine carboxypeptidase homologue that has a strong association with grain weight,

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 13 of 18

therefore, likely to regulate rice yield. Serine peptidases with SP10 domain are thought to be involved in protein degradation and activate flowering genes by cleaving floral repressors [96]. LOC_Os10G01060 (a type of kinase family protein), LOC_Os02G55640 (proteasome maturation factor, UMP1), LOC_OsO3G08000(expressed protein) and LOC_Os03G08010 (elongation factor Tu) maintain a balance between growth, development and stress responses in plants by acting as sensors for different plant hormones and influencing grain yield components [97]. Recent studies revealed that plants with enhanced expression of receptor like-kinases (RLKs) displayed an increase in seed yield and induced earlier flowering [2, 55, 98–104]. LOC_Os12G21789 encodes a protein belonging to glycosyl hydrolase family that primarily works in strigolactone (SL) signalling pathway and is predicted to be closely related to rice yield traits. RNAi knockdown mutant OsD14L showed a reduction in size of panicles, seed grains leading to high tillering and dwarf phenotype in rice [3, 7, 105–107]. LOC_Os12G24170 is a precursor of beta-galactosidase enzyme (BGAL) that mediates the physiological process of seed germination through cross talks among GA and Abscisic acid (ABA) to determine grain yield in rice [3]. Knockout studies revealed that BGAL9 regulates different physiological functions and stress responses in rice [108]. Transcriptomics studies very well implied the upregulation of OsBAGL1, OsBAGL4, OsBAGL8 and OsBAGL11 genes during seed germination and their influence on plant growth and development [109]. LOC_Os12G24080 annotated as HECT-domain containing protein belonging to the family of Ubiquitin ligases (E3s) that play a crucial role in multiple biological processes which includes flowering [110, 111]. HECT ligases associated with proteasomes orchestrate plant growth and development severely [112]. HAF1, E3 ubiquitin ligase interacts with Heading date1(Hd1) gene and alters the flowering times in short- and long-day rice plants [113] signifying that HAF1 is critical to precisely modulate timing of Hd1 accumulation for the photoperiod induction. Another study revealed that Flowering Related RING Protein 1 (*FRRP1*), a E3 ubiquitin ligase possibly regulates flowering time and rice yield by reducing monoubiquitination of H2B histone proteins and resultant changes in length of grains, panicles and plant height [114]. LOC_Os12G21710 annotated as the high-tillering and dwarf 12 gene (htd12) encodes a 15-cis-carotene isomerase (Z-ISO) protein, a member of nitrite and nitric oxide reductase U (NnrU) family pf proteins that regulates carotenoids and SL biosynthetic pathways and modifies plant architecture [115]. T20, a carotene isomerase localized in chloroplasts suppresses outgrowth of rice tillers by cross talk among carotenoids, SL and ABA [116].

MIT3 gene coding for a carotenoid isomerase, mutants revealed high tillering, leaf variegations and a dwarfed phenotype implicating the role of carotene isomerases in governing rice yield by attuning the signalling pathways [11]. These putative genes appear to be closely associated to rice yield traits and could be utilized to develop functional markers for use in rice breeding programs.

We identified 25 *Arabidopsis* homologues to the 73 rice CGs, nineteen of which with higher expression in distinct tissues, such as leaves, shoots, cotyledons, seed stages, vegetative rosette and rosette after flowering were identified to affect grain yield in rice (Table 4). LOC_Os01G71190 seems to involve in GO:0031090. The homologous gene in *Arabidopsis* is AT4G28660, which governs different biological processes affecting the plant growth [117, 118]. LOC_Os02G55640 participates in the GO:0034622 pathway and its homologue in *Arabidopsis*, AT1G67250 have roles in protein degradation pathway (ERAD) and assembly of proteasomes with documented associations with grain yield [82, 119]. Ortholog of LOC_Os02G17520(Cytochrome c biogenesis family protein) with QTN 53308499 in its genic region is characterized by ABC-type transporter and ATP hydrolysis activity. Its involvement in ABA mediated pathways [120] and ATP binding, suggests potential association with rice development [46, 121]. Interestingly, there are two candidates, LOC_Os02G17470 (RNA binding protein-related) and LOC_Os07G43870(DNA J heat shock protein) gene which act as a positive regulator of pollen and pollen tube growth which is essentially important for reproduction in plants [122, 123]. Also, LOC_Os03G08000 (E3 ubiquitin ligase family protein), involved in proteolysis and regulation of developmental processes, has been found to be located near to the QTN, 83271479 associated with PYD and PlHt indicating a possible role in grain yield [9, 124, 125]. Interesting, yield related gene monohydrate reductase gene LOC_Os08G05770, was observed in neighbouring region of the QTN associated with DTF, signifying a possible pleotropic effect on all the grain yield related traits [38, 126–128]. Extensive studies revealed another CG, LOC_Os12G21710 encodes a 15-cis-zeta-carotene isomerase ortholog in *Arabidopsis* and it participates in unique functions in photosynthesis, signalling pathways via carotenoid derivatives and govern plant physiology, growth and stress responses in plants [1, 129, 130]. In *Arabidopsis*, serine-carboxypeptidase like-18 gene, LOC_Os10G01110, belonging to SCPL family is assumed to regulate seed germination and yield traits in rice *viz.,* grain weight and grain size due to SNP polymorphisms in allele *GS5-1* and *GS5-2* alleles in response to ABA [75, 131, 132]. Amongst the other notable yield related genes near the locus include LOC_Os12G24170 (BGAL9) and LOC_Os12G21789 (Glycosyl

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 14 of 18

hydrolase family protein 43) which has a key role in plant cell growth and development, explicitly for microspore and pollen germination by modifying the cell wall components [133–135].

In the current study, six GWAS models were employed for yield related traits on selected subset of rice accessions using 5,53,229 SNPs and the average phenotypic data recorded in the IARI field. For plant height, 23 QTNs were identified, whereas 25 QTNs were identified for plot yield and days to flowering each. Amongst the 73 different QTNs in total, 24 were co-localized with already reported QTLs or loci in previous mapping studies. We obtained 15 significant QTNs and 10 candidates underlying the three aforementioned yield traits. Meanwhile, we selected 20 elite genotypes for breeding high yielding rice and identified 25 homologues of *Arabidopsis*, 19 of which might regulate yield traits in rice. Subsequently, gene annotation, gene ontology and enrichment analysis showed 40 CGs were found to be enriched in GO terms of the studied traits. LOC_OS09G15770, LOC_OS02G36710 and LOC_OS02G17520 were confirmed as key candidates by tissue-specific expression analysis, and haplotype and phenotypic variation analysis. Selecting the identified elite genotypes with increased frequency of desirable alleles for grain yield traits might speed up the rhythm of rice improvement and address the challenges relating to food security and sustainable rice production. Furthermore, the identification of QTNs and CGs in our study provided insights into the regulatory mechanisms and genetic associations of these yield related traits in rice.

Abbreviations

GWAS	Genome wide-association study
SNPs	Single nucleotide polymorphisms
QTN	Quantitative trait nucleotide
SL	Single-locus
ML	Multi-locus
PlHt	Plant height
DTF	Days to 50% flowering
PYD	Plot yield
CV	Coefficient of variation
PiPerBP	Nucleotide diversity per base pair/ Genetic diversity
ThetaPerBP	The expected number of polymorphic sites per nucleotide
LD	Linkage Disequilibrium
H_o	Observed heterozygosity
NJ	Neighbour joining
PCA	Principal component analysis
GAPIT	Genome Association and Prediction Integrated Tool
LOD	Logarithm of Odds
MSUv7.0	Genome Annotation database for Rice
IC4R	Information Commons for Rice
BAR	Bio-Analytical Resource for Plant Biology
RMA	Robust Multi-array average
TAIR	*Arabidopsis* Information Resource
GA	Gibberellins
BR	Brassinosteroids
SL	Strigolactone
BA	Abscisic acid
CG	Candidate gene
GO	Gene ontology

Supplementary Information

Supplementary material 1.

Supplementary material 2.

Supplementary material 3.

Supplementary material 4.

Supplementary material 5.

Supplementary material 6.

Supplementary material 7.

Availability of data and materials
The genotype datasets analysed in our GWAS study are available at the Rice SNP-Seek Database (https://snp-seek.irri.org/) and the details of the genotypes used in the study can be found in supplementary information file.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for application
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Division of Genomic Resources, ICAR-NBPGR, Pusa, New Delhi, India. [2]International Rice Research Institute (IRRI), South Asia Hub, ICRISAT, Hyderabad, India. [3]International Rice Research Institute (IRRI), South Asia Regional Centre (ISARC), Varanasi, India. [4]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Telangana, India. [5]ICAR-National Bureau of Plant Genetic Resources, Pusa, New Delhi, India.

References

1. Jia K, Baz L, Al-Babili S. From carotenoids to strigolactones. J Exp Bot. 2017;69(9):2189–204.
2. Aloryi KD, Okpala NE, Amo A, Bello SF, Akaba S, Tian X. A meta-quantitative trait loci analysis identified consensus genomic regions and candidate genes associated with grain yield in rice. Front Plant Sci. 2022;13:1035851.
3. Li H, Li X, Wang G, Zhang J, Wang G. Analysis of gene expression in early seed germination of rice: landscape and genetic regulation. BMC Plant Biol. 2022;22(1):1–14.
4. Qiu X, Pang Y, Yuan Z, Xing D, Xu J, Dingkuhn M, Li Z, Ye G. Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. PLoS One. 2015;10(12):e0145577.
5. Liu F, Wang P, Zhang X, Li X, Yan X, Fu D, Wu G. The genetic and molecular basis of crop height based on a rice model. Planta. 2018;247:1–26.
6. Abbai R, Singh VK, Nachimuthum VV, Sinha P, Selvaraj R, Vipparla AK, Singh AK, Singh UM, Varshney RK, Kumar A. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. Plant Biotechnol J. 2019;17(8):1612–22.
7. Deveshwar P, Prusty A, Sharma S, Tyagi AK. Phytohormone-mediated molecular mechanisms involving multiple genes and QTL govern grain number in rice. Front Genet. 2020;11:586462.
8. Zhong H, Liu S, Sun T, Kong W, Deng X, Peng Z, Li Y. Multi-locus genome-wide association studies for five yield-related traits in rice. BMC Plant Biol. 2021;21:364.
9. Yan P, Zhu Y, Wang Y, Ma F, Lan D, Niu F, Dong S, Zhang X, Hu J, Liu S, Guo T. A new RING finger protein, PLANT ARCHITECTURE and GRAIN NUMBER 1, affects plant architecture and grain yield in rice. Int J Mol Sci. 2022;23(2):824.
10. Xing Y, Zhang Q. Genetic and molecular bases of rice yield. Annu Rev Plant Biol. 2010;61:421–42.
11. Liu L, Xie T, Peng P, Qiu H, Zhao J, Fang J, Patil SB, Wang Y, Fang S, Chu J, Yuan S. Mutations in the MIT3 gene encoding a carotenoid isomerase lead to increased tiller number in rice. Plant Sci. 2018;267:1–10.
12. Li L, Wu X, Chen J, Wang S, Wan Y, Ji H, et al. Genetic dissection of epistatic interactions contributing yield-related agronomic traits in rice using the compressed mixed model. Plants. 2022;11(19):2504.
13. Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, Numa H, et al. The rice annotation project database (RAP-DB): 2008 update. Nucleic Acids Res. 2008;36:D1028–33. https://doi.org/10.1093/nar/gkm978.
14. McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, et al. Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci USA. 2009;106(30):12273–8.
15. Agrama HA, Yan W, Jia M, Fjellstrom R, McClung AM. Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. Nat Sci. 2010;2(04):247.
16. Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 2010;42(11):961-U976.
17. Wickland DP, Battu G, Hudson KA, Diers BW, Hudson ME. A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow GB-eaSy. BMC Bioinformatics. 2017;18:1–12.
18. Selvaraj R, Singh AK, Singh VK, Abbai R, Habde SV, Singh UM, Kumar A. Superior haplotypes towards development of low glycemic index rice with preferred grain and cooking quality. Sci Rep. 2021;11(1):1–15.
19. Zhang J, Wang S, Wu X, Han L, Wang Y, Wen Y. Identification of QTNs, QTN-by-environment interactions and genes for yield-related traits in rice using 3VmrMLM. Front Plant Sci. 2022;13:995609.
20. Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He J, et al. Improving the accuracy of whole genome prediction for complex traits using the results of genome wide association studies. PloS One. 2014;9:e93017.
21. Begum H, Spindel JE, Lalusin A, Borromeo T, Gregorio G, Hernandez J, Virk P, Collard B, McCouch SR. Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). PLoS One. 2015;10(3):e0119873.
22. Reig-Valiente JL, Marqués L, Talón M, Domingo C. Genome-wide association study of agronomic traits in rice cultivated in temperate regions. BMC Genomics. 2018;19:1–11.
23. Bollinedi H, Yadav AK, Vinod KK, Gopala Krishnan S, Bhowmick PK, Nagarajan M, Neeraja CN, Ellur RK, Singh AK. Genome-wide association study reveals novel marker-trait associations (MTAs) governing the localization of Fe and Zn in the rice grain. Front Genet. 2020;11:213.
24. Zhou X, Huang X. Genome-wide association studies in rice: How to solve the low power problems? Mol Plant. 2019;12:10–2.
25. Zhang P, Zhong K, Zhong Z, Tong H. Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.). BMC Plant Biol. 2019;19:1–12.
26. Alqudah AM, Sallam A, Baenziger PS, Börner A. GWAS: Fast-Forwarding gene identification and characterization in temperate cereals: Lessons from barley–A review. J Adv Res. 2020;1(22):119–35.
27. Verma RK, Chetia SK, Dey PC, Rahman A, Saikia S, Sharma V, Sharma H, Sen P, Modi MK. Genome-wide association studies for agronomical traits in winter rice accessions of Assam. Genomics. 2021;113(3):1037–47.
28. Ravikiran KT, Gopala Krishnan S, Abhijith KP, Bollinedi H, Nagarajan M, Vinod KK, et al. Genome-wide association mapping reveals novel putative gene candidates governing reproductive stage heat stress tolerance via two staggered sowings in rice. Front Genet. 2022;13:876522. https://doi.org/10.3389/fgene.2022.876522.
29. Abhijith KP, Gopala Krishnan S, Ravikiran KT, Dhawan G, Kumar P, Vinod KK, Bhowmick PK, Nagarajan M, Seth R, Sharma R, Badhran SK. Genome-wide association study reveals novel genomic regions governing agronomic and grain quality traits and superior allelic combinations for Basmati rice improvement. Front Plant Sci. 2022;13:994447.
30. Sundaramoorthy M, Ramasamy SP, Rajagopalan VR, Ramalingam AP, Ayyenar B, Mohanavel V, et al. Pilot scale genome wide association mapping identified novel loci for grain yield traits in rice. Plant Physiol Rep. 2022;27(1):11–21.
31. Kumar S, Kumari J, Bhusal N, Pradhan AK, Budhlakoti N, Mishra DC, et al. Genome-Wide Association Study Reveals Genomic Regions Associated with Ten Agronomical Traits in Wheat under Late-Sown Conditions. Front Plant Sci. 2020;11:549743.
32. Xie D, Dai Z, Yang Z, Tang Q, Sun J, Yang X, Song X, Lu Y, Zhao D, Zhang L, Su J. Genomic variations and association study of agronomic traits in flax. BMC Genomics. 2018;19(1):1–12.
33. Singh N, Agarwal N, Yadav HK. Genome-wide SNP-Based Diversity Analysis and Association Mapping in Linseed (*Linum usitatissimum* L.). Euphytica. 2019;215:139. https://doi.org/10.1007/s10681-019-2462-x.
34. Jaiswal V, Gupta S, Gahlaut V, Muthamilarasan M, Bandyopadhyay T, Ramchiary N, et al. Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (*Setaria italica* L.) Using ddRAD Sequencing. Sci Rep. 2019;9:5020.
35. Fang Y, Liu S, Dong Q, Zhang K, Tian Z, Li X, Li W, Qi Z, Wang Y, Tian X, Song J. Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. Front Plant Sci. 2020;9(11):9.
36. Reddy VRP, Das S, Dikshit HK, Mishra GP, Aski M, Meena SK, et al. Genome-Wide Association Analysis for Phosphorus Use Efficiency Traits in Mungbean (*Vigna radiata* L Wilczek) Using Genotyping by Sequencing Approach. Front Plant Sci. 2020;11:537766.
37. Chaurasia S, Singh A, Kumar A, Songachan LS, Yadav MC, Kumar S, et al. Genome-wide Association Mapping Reveals Key Genomic Regions for Physiological and Yield-Related Traits under Salinity Stress in Wheat (*Triticum aestivum* L.). Genomics. 2021;113(5):3198–215.
38. Saroha A, Pal D, Gomashe SS, Kaur V, Ujjainwal S, Rajkumar S, et al. Identification of QTNs associated with flowering time, maturity, and plant height traits in *Linum usitatissimum* L. using genome-wide association study. Front Genet. 2022;13:811924.
39. You ZH, Huang ZA, Zhu Z, Yan GY, Li ZW, Wen Z, et al. PBMDA: a novel and effective path-based computational model for miRNA-disease association prediction. PLoS Comput Biol. 2017;13:e1005455.
40. Dehury B, Behera SK, Negi S. Overcoming the limitation of GWAS platforms using systems biology approach. Curr Bioinformatics. 2017;12:156–70.

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 16 of 18

41. Torres RO, Henry A. Yield stability of selected rice breeding lines and donors across conditions of mild to moderately severe drought stress. Field Crops Res. 2018;220:37–45.

42. Wu TY, Gruissem W, Bhullar NK. Targeting intra-cellular transport combined with efficient uptake and storage significantly increases grain iron and zinc levels in rice. Plant Biotechnol J. 2019;17:9–20.

43. Subedi SR, Sandhu N, Singh VK, Sinha P, Kumar S, Singh S, et al. Genome-wide association study reveals significant genomic regions for improving yield, adaptability of rice under dry direct seeded cultivation condition. BMC Genomics. 2019;20(1):1–20.

44. Bhandari A, Sandhu N, Bartholome J, Cao-Hamadoun T-V, Ahmadi N, Kumari N, Kumar A. Genome-wide association study for yield and yield related traits under reproductive stage drought in a diverse indica-aus rice panel. Rice. 2020;13(1):1–22.

45. Ren M, Huang M, Qiu H, Chun Y, Li L, Kumar A, Fang J, Zhao J, He H, Li X. Genome-wide association study of the genetic basis of effective tiller number in rice. Rice. 2021;14(1):1–13.

46. Jiang J, Xing F, Wang C, Zeng X. Identification and analysis of rice yield-related candidate genes by walking on the functional network. Front Plant Sci. 2018;9:1685.

47. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60.

48. Li M, Liu X, Bradbury P, Yu J, Zhang Y, Todhunter RJ, Buckler ES, Zhang Z. Enrichment of statistical power for genome-wide association studies. BMC Biol. 2014;12:1–10.

49. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6:19444.

50. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, Wang SB, Dunwell JM, Zhang YM, Wu R. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform. 2018;19(4):700–12.

51. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44:825–30.

52. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies. biorxiv. 2018;341784. https://doi.org/10.1101/341784.

53. Cui Y, Zhang F, Zhou Y. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. Front Plant Sci. 2018;5(9):1464.

54. Tibbs Cortes L, Zhang Z, Yu J. Status and prospects of genome-wide association studies in plants. Plant Genome. 2021;14(1):e20077.

55. Wang J, Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. Genom Proteom Bioinformatics. 2021;19(4):629–40.

56. Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK. Accessing genetic diversity for crop improvement. 2010;13(2):167-73.

57. Varshney RK, Singh VK, Kumar A, Powell W, Sorrells ME. Can genomics deliver climate-change ready crops? Curr Opin Plant Biol. 2018;45:205–11.

58. Li Y, Gealy D. Weedy red rice has novel sources of resistance to biotic stress. Crop J. 2018;6(5):443–50.

59. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945–59.

60. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–20.

61. Earl DA, VonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;1(4):359–61.

62. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28(18):2397–9.

63. Zhang YW, Tamba CL, Wen YJ, Li P, Ren WL, Ni YL, Gao J, Zhang YM. mrMLM v4. 0.2: an R platform for multi-locus genome-wide association studies. Genom Proteom Bioinformatics. 2020;18(4):481–7.

64. Zhang YM, Mao Y, Xie C, Smith H, Luo L, Xu S. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L). Genetics. 2005;169(4):2267–75.

65. Sehgal D, Autrique E, Singh R, Ellis M, Singh S, Dreisigacker S. Identification of genomic regions for grain yield and yield stability and their epistatic interactions. Sci Rep. 2017;7(1):1–12.

66. Ta KN, Khong NG, Ha TL, Nguyen DT, Mai DC, Hoang TG, et al. A genome-wide association study using a Vietnamese landrace panel of rice (*Oryza sativa*) reveals new QTLs controlling panicle morphological traits. BMC Plant Biol. 2018;18:1–15.

67. Sui M, Wang Y, Bao Y, Wang X, Li R, Lv Y, et al. Genome-wide association analysis of sucrose concentration in soybean (Glycine max L) seed based on high-throughput sequencing. Plant Genome. 2020;13(3):e20059.

68. Li W, Yang B, Xu J, Peng L, Sun S, Huang Z, et al. A genome-wide association study reveals that the 2-oxoglutarate/malate translocator mediates seed vigor in rice. Plant J. 2021;108(2):478–91.

69. Bheemanahalli R, Knight M, Quinones C, et al. Genome-wide association study and gene network analyses reveal potential candidate genes for high night temperature tolerance in rice. Sci Rep. 2021;11(1):6747.

70. Islam AS, Mustahsan W, Tabien R, Awika JM, Septiningsih EM, Thomson MJ. Identifying the genetic basis of mineral elements in rice grain using genome-wide association mapping. Genes. 2022;13(12):2330.

71. Hu C, Jiang J, Li Y, Song S, Zou Y, Jing C, Zhang Y, Wang D, He Q, Dang X. QTL mapping and identification of candidate genes using a genome-wide association study for heat tolerance at anthesis in rice (*Oryza sativa* L.). Front Genet. 2022;13:983525.

72. Rathan ND, Krishna H, Ellur RK, Sehgal D, Govindan V, Ahlawat AK, Krishnappa G, Jaiswal JP, Singh JB, Sv S, Ambati D. Genome-wide association study identifies loci and candidate genes for grain micro-nutrients and quality traits in wheat (*Triticum aestivum* L.). Sci Rep. 2022;12(1):7037.

73. Haghi R, Ahmadikhah A, Fazeli A, Shariati V. Candidate genes for anthocyanin pigmentation in rice stem revealed by GWAS and whole-genome resequencing. Plant Genome. 2022;15(3):e20224.

74. Tadesse W, Gataa ZE, Rachdad FE, Baouchi AE, Kehel Z, Alemu A. Single- and multi-trait genomic prediction and genome-wide association analysis of grain yield and micronutrient-related traits in ICARDA wheat under drought environment. Mol Genet Genom. 2023;18:1–2.

75. Duan P, Xu J, Zeng D, Zhang B, Geng M, Zhang G, Huang K, Huang L, Xu R, Ge S, Qian Q. Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. Mol Plant. 2017;10(5):685–94.

76. Kumari J, Lakhwani D, Jakhar P, Sharma S, Tiwari S, Mittal S, et al. Association mapping reveals novel genes and genomic regions controlling grain size architecture in mini core accessions of Indian National Genebank wheat germplasm collection. Front Plant Sci. 2023;14:1148658. https://doi.org/10.3389/fpls.2023.1148658.

77. Zhao Q, Shi XS, Wang T, Chen Y, Yang R, Mi J, Zhang YW, Zhang YM. Identification of QTNs, QTN-by-environment interactions, and their candidate genes for grain size traits in main crop and ratoon rice. Front Plant Sci. 2023;14:1119218.

78. Akond Z, Ahsan MA, Alam M, Mollah MNH. Robustification of GWAS to explore effective SNPs addressing the challenges of hidden population stratification and polygenic effects. Sci Rep. 2021;11(1):13060.

79. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.

80. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet. 2011;43(2):159–62.

81. Ma X, Feng F, Wei H, Mei H, Xu K, Chen S, Li T, Liang X, Liu H, Luo L. Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. Front Plant Sci. 2016;29(7):1801.

82. He J, Zhuang Y, Li C, Sun X, Zhao S, Ma C, Lin H, Zhou H. SIMP1 modulates salt tolerance by elevating ERAD efficiency through UMP1A-mediated proteasome maturation in plants. New Phytol. 2021;232(2):625–41.

83. Liu L, Tong H, Xiao Y, Che R, Xu F, Hu B, Liang C, Chu J, Li J, Chu C. Activation of Big Grain1 significantly improves grain size by regulating auxin transport in rice. Proc Natl Acad Sci. 2015;112(35):11102–7.

84. Wu Q, Li D, Li D, Liu X, Zhao X, Li X, Li S, Zhu L. Overexpression of OsDof12 affects plant architecture in rice (*Oryza sativa* L.). Front Plant Sci. 2015;6:833.

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 17 of 18

85. He Q, Yang L, Hu W, Zhang J, Xing Y. Overexpression of an auxin receptor OsAFB6 significantly enhanced grain yield by increasing cytokinin and decreasing auxin concentrations in rice panicle. Sci Rep. 2018;8(1):14051.

86. Zhou HL, He SJ, Cao YR, Chen T, Du BX, Chu CC, Zhang JS, Chen SY. OsGLU1, a putative membrane-bound endo-1, 4-ß-d-glucanase from rice, affects plant internode elongation. Plant Mol Biol. 2006;60:137–51.

87. Jung KH, Lee J, Dardick C, Seo YS, Cao P, Canlas P, et al. Identification and functional analysis of light-responsive unique genes and gene family members in rice. PLoS Genet. 2008;4(8):e1000164.

88. Simmons AR, Davies KA, Wang W, Liu Z, Bergmann DC. SOL1 and SOL2 regulate fate transition and cell divisions in the *Arabidopsis* stomatal lineage. Development. 2019;146(3):dev171066.

89. Andersen SU, Algreen-Petersen RG, Hoedl M, Jurkiewicz A, Cvitanich C, Braunschweig U, Schauser L, Oh SA, Twell D, Jensen EØ. The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc in vitro and TSO1 is required for both male and female fertility in *Arabidopsis thaliana*. J Exp Bot. 2007;58(13):3657–70.

90. Sijacic P, Wang W, Liu Z. Recessive antimorphic alleles overcome functionally redundant loci to reveal TSO1 function in *Arabidopsis* flowers and meristems. PLoS Genet. 2011;7(11):e1002352.

91. Sanders C, Turkarslan S, Lee DW, Daldal F. Cytochrome c biogenesis: the Ccm system. Trends Microbiol. 2010;18(6):266–74 Meyer EH, Giegé P, Gelhaye E, Rayapuram N, Ahuja U, Thöny-Meyer L, Grienenberger JM, Bonnard G. AtCCMH, an essential component of the c-type cytochrome maturation pathway in Arabidopsis mitochondria, interacts with apocy-tochrome c. Proc Natl Acad Sci. 2005;102(44):16113-16118.

92. Meinke DW. Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in *Arabidopsis*. New Phytologist. 2020;226(2):306–25.

93. Luo L, Xie Y, Yu S, Yang J, Chen S, Yuan X, Guo T, Wang H, Liu Y, Chen C, Xiao W. The DnaJ Domain-containing Heat Shock Protein NAL11 Determines Plant Architecture by Mediating GA homeostasis in Rice (*Oryza sativa* L.). New Phytologist. 2023;237(6):2163–79.

94. Huang Y, Du L, Wang M, Ren M, Yu S, Yang Q. Multifaceted roles of zinc finger proteins in regulating various agronomic traits in rice. Front Plant Sci. 2022;13:974396.

95. Kim JJ, Kim YS, Park SI, Mok JE, Kim YH, Park HM, Kim IS, Yoon HS. Cytosolic monodehydroascorbate reductase gene affects stress adaptation and grain yield under paddy field conditions in *Oryza sativa* L. japonica. Mol Breed. 2017;37:1–13.

96. Tripathi LP, Sowdhamini R. Cross genome comparisons of serine proteases in Arabidopsis and rice. BMC Genomics. 2006;7(1):1–31.

97. Zhu Q, Feng Y, Xue J, Chen P, Zhang A, Yu Y. Advances in receptor-like protein kinases in balancing plant growth and stress responses. Plants. 2023;12(3):427.

98. Planas-Riverola A, Gupta A, Betegón-Putze I, Bosch N, Ibañes M, Caño-Delgado AI. Brassinosteroid signaling in plant development and adaptation to stress. Development. 2019;146(5):dev151894.

99. Lozano-Elena F, Caño-Delgado AI. Emerging roles of vascular brassinosteroid receptors of the BRI1-like family. Curr Opin Plant Biol. 2019;51:105–13.

100. Chen W, Lv M, Wang Y, Wang PA, Cui Y, Li M, Wang R, Gou X, Li J. BES1 is activated by EMS1-TPD1-SERK1/2-mediated signaling to control tapetum development in *Arabidopsis* thaliana. Nat Commun. 2019;10(1):4164.

101. Ackerman-Lavert M, Savaldi-Goldstein S. Growth models from a brassinosteroid perspective. Curr Opin Plant Biol. 2020;53:90–7.

102. Hohmann U, Ramakrishna P, Wang K, Lorenzo-Orts L, Nicolet J, Henschen A, Barberon M, Bayer M, Hothorn M. Constitutive activation of leucine-rich repeat receptor kinase signaling pathways by BAK1-INTERACTING RECEPTOR-LIKE KINASE3 chimera. Plant Cell. 2020;32(10):3311–23.

103. Ren D, Ding C, Qian Q. Molecular bases of rice grain size and quality for optimized productivity. Sci Bull. 2023;68(3):314–50.

104. Vicentini G, Biancucci M, Mineri L, Chirivì D, Giaume F, Miao Y, Kyozuka J, Brambilla V, Betti C, Fornara F. Environmental control of rice flowering time. Plant Commun. 2023;4(5):100610.

105. Ikeda M, Miura K, Aya K, Kitano H, Matsuoka M. Genes offering the potential for designing yield-related traits in rice. Curr Opin Plant Biol. 2013;16(2):213–20.

106. Wang YP, Tang SQ, Chen HZ, Wu ZF, Zhang H, Duan EC, Shi QH, Wu ZM. Identification and molecular mapping of indica high-tillering dwarf mutant htd4, a mild phenotype allelic mutant of D14 in rice (*Oryza sativa* L.). Plant Biol. 2017;19(6):851–8.

107. Liang R, Qin R, Yang C, Zeng D, Jin X, Shi C. Identification and characterization of a novel strigolactone-insensitive mutant, Dwarfism with high tillering ability 34 (dhta-34) in rice (*Oryza sativa* L.). Biochem Genet. 2019;57:403–20.

108. Hoang TV, Vo KTX, Rahman MM, Zhong R, Lee C, Ketudat Cairns JR, Ye ZH, Jeon JS. SPOTTED-LEAF7 targets the gene encoding β-galactosidase9, which functions in rice growth and stress responses. Plant Physiol. 2023;193(2):1109–25.

109. Zhang Q, Peng Y, Li X, Chen B, Liu J. β-galactosidase is involved in rice seed germination. Seed Sci Technol. 2021;49(3):261–74.

110. Sohail A. Genetic and signaling pathways of flowering regu lation in rice (Oryza sativa L). Braz J Bot. 2023;46(3):599–608.

111. Shah SS, Kumar S. Adaptors as the regulators of HECT ubiquitin ligases. Cell Death Differ. 2021;28(2):455–72.

112. Wang Z, Spoel SH. HECT ubiquitin ligases as accessory proteins of the plant proteasome. Essays Biochem. 2022;66(2):135–45.

113. Yang Y, Fu D, Zhu C, He Y, Zhang H, Liu T, Li X, Wu C. The RING-finger ubiquitin ligase HAF1 mediates heading date 1 degradation during photoperiodic flowering in rice. Plant Cell. 2015;27(9):2455–68.

114. Du Y, He W, Deng C, Chen X, Gou L, Zhu F, Guo W, Zhang J, Wang T. Flowering-related RING protein 1 (FRRP1) regulates flowering time and yield potential by affecting histone H2B monoubiquitination in rice (*Oryza sativa*). PLoS One. 2016;11(3):e0150458.

115. Zhou H, Yang M, Zhao L, Zhu Z, Liu F, Sun H, Sun C, Tan L. HIGH-TILLERING AND DWARF 12 modulates photosynthesis and plant architecture by affecting carotenoid biosynthesis in rice. J Exp Bot. 2021;72(4):1212–24.

116. Liu X, Hu Q, Yan J, Sun K, Liang Y, Jia M, Meng X, Fang S, Wang Y, Jing Y, Liu G. ζ-Carotene isomerase suppresses tillering in rice through the coordinated biosynthesis of strigolactone and abscisic acid. Mol Plant. 2020;13(12):1784–801.

117. Lu Y. Identification and roles of photosystem II assembly, stability, and repair factors in *Arabidopsis*. Front Plant Sci. 2016;7:168.

118. Mabbitt PD, Wilbanks SM, Eaton-Rye JJ. Structure and function of the hydrophilic Photosystem II assembly proteins: Psb27, Psb28 and Ycf48. Plant Physiol Biochem. 2014;81:96–107.

119. Gemperline DC, Marshall RS, Lee KH, Zhao Q, Hu W, McLoughlin F, Scalf M, Smith LM, Vierstra RD. Proteomic analysis of affinity-purified 26S proteasomes identifies a suite of assembly chaperones in *Arabidopsis*. Journal of Biological Chemistry. 2019;294(46):17570–92.

120. Kang J, Yim S, Choi H, Kim A, Lee KP, Lopez-Molina L, et al. Abscisic acid transporters cooperate to control seed germination. Nat Commun. 2015;6:8113.

121. Zhao G, Shi J, Liang W, Xue F, Luo Q, Zhu L, Qu G, Chen M, Schreiber L, Zhang D. Two ATP binding cassette G transporters, rice ATP binding cassette G26 and ATP binding cassette G15, collaboratively regulate rice male reproduction. Plant Physiol. 2015;169(3):2064–79.

122. Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, Griffin JL, Bessant C, Brandizzi F, Hawes C, Watson RB. Mapping the *Arabidopsis* organelle proteome. Proc Natl Acad Sci. 2006;103(17):6518–23.

123. Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiol. 2008;148(3):1201–11.

124. Xu M, Hu T, Zhao J, Park MY, Earley KW, Wu G, Yang L, Poethig RS. Developmental functions of miR156-regulated SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes in *Arabidopsis thaliana*. PLoS Genet. 2016;12(8):e1006263.

125. Pan R, Hu J. The Arabidopsis E3 ubiquitin ligase SP1 targets to chloroplasts, peroxisomes, and mitochondria. Plant Physiol. 2018;176(1):480–2.

126. Rajjou L, Belghazi M, Huguet R, Robin C, Moreau A, Job C, Job D. Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. Plant Physiol. 2006;141(3):910–23.

127. Fichman Y, Koncz Z, Reznik N, Miller G, Szabados L, Kramer K, Nakagami H, Fromm H, Koncz C, Zilberstein A. SELENOPROTEIN O is a chloroplast protein involved in ROS scavenging and its absence increases dehydration tolerance in *Arabidopsis thaliana*. Plant Sci. 2018;270:278–91.

Sachdeva *et al. BMC Plant Biology*   (2024) 24:124

Page 18 of 18

128. Tanaka M, Takahashi R, Hamada A, Terai Y, Ogawa T, Sawa Y, et al. Distribution and functions of monodehydroascorbate reductases in plants: comprehensive reverse genetic analysis of *Arabidopsis thaliana* enzymes. Antioxidants. 2021;10(11):1726.

129. Stanley L, Yuan YW. Transcriptional regulation of carotenoid biosynthesis in plants: so many regulators, so little consensus. Front Plant Sci. 2019;9(10):1017.

130. Sathasivam R, Radhakrishnan R, Kim JK, Park SU. An update on biosynthesis and regulation of carotenoids in plants. South Afr J Bot. 2021;140:290–302.

131. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet. 2011;43(12):1266–9.

132. Xu C, Liu Y, Li Y, Xu X, Xu C, Li X, Xiao J, Zhang Q. Differential expression of GS5 regulates grain size in rice. J Exp Bot. 2015;66(9):2611–23.

133. Hruba P, Honys D, Twell D, Čapková V, Tupý J. Expression of β-galactosidase and β-xylosidase genes during microspore and pollen development. Planta. 2005;220:931–40.

134. Ahn YO, Zheng M, Bevan DR, Esen A, Shiu SH, Benson J, Peng HP, Miller JT, Cheng CL, Poulton JE, Shih MC. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. Phytochemistry. 2007;68(11):1510–20.

135. Nibbering P, Petersen BL, Motawia MS, Jørgensen B, Ulvskov P, Niittylä T. Golgi-localized exo-β1, 3-galactosidases involved in cell expansion and root growth in *Arabidopsis*. J Biol Chem. 2020;295(31):10581–92.

Publisher's Note