Rice

RESEARCH

Open Access

A Novel Function of *GW5* on Controlling the Early Growth Vigor and its Haplotype Effect on Shoot Dry Weight and Grain Size in Rice (*Oryza sativa* L.)

Tifeng Yang[1†], Jingfang Dong[1†], Xijuan Xiong[1,2], Longting Zhang[1,2], Jian Wang[1], Haifei Hu[1], Lian Zhou[1], Wu Yang[1], Yamei Ma[1], Hua Fu[1], Jiansong Chen[1], Wenhui Li[1], Shuai Nie[1], Ziqiang Liu[2], Bin Liu[1], Feng Wang[1*], Junliang Zhao[1*] and Shaohong Zhang[1*]

Abstract

Strong early growth vigor is an essential target in both direct seeded rice breeding and high-yielding rice breeding for rice varieties with relatively short growth duration in the double-cropping region. Shoot dry weight (SDW) is one of the important traits associated with early growth vigor, and breeders have been working to improve this trait. Finding stable QTLs or functional genes for SDW is crucial for improving the early growth vigor by implementing molecular breeding in rice. Here, a genome-wide association analysis revealed that the QTL for SDW, *qSDW-5*, was stably detected in the three cultivation methods commonly used in production practice. Through gene-based haplotype analysis of the annotated genes within the putative region of *qSDW-5*, and validated by gene expression and knockout transgenic experiments, *LOC_Os05g09520*, which is identical to the reported *GW5/GSE5* controlling grain width (GW) and thousand grain weight (TGW) was identified as the causal gene for *qSDW-5*. Five main haplotypes of *LOC_Os05g09520* were identified in the diverse international rice collection used in this study and their effects on SDW, GW and TGW were analyzed. Phenotypic comparisons of the major haplotypes of *LOC_Os05g09520* in the three subpopulations (*indica*, *japonica* and *aus*) revealed the same patterns of wider GW and higher TGW along with higher SDW. Furtherly, the haplotype analysis of 138 rice varieties/lines widely used in southern China showed that 97.8% of the cultivars/lines carry $Hap2^{LOC\_Os05g09520}$. These results not only provide a promising gene source for the molecular breeding of rice varieties with strong early growth vigor, but also elucidate the effect of the *LOC_Os05g09520* haplotypes on SDW, GW, and TGW in rice. Importantly, this

†Tifeng Yang and Jingfang Dong contributed equally to this work.

*Correspondence:
Feng Wang
fwang1631@163.com
Junliang Zhao
zhao_junliang@gdaas.cn
Shaohong Zhang
szhanggz@tom.com

Full list of author information is available at the end of the article

study provides direct genetic evidence that these three traits are significantly correlated, and suggests a breeding strategy for developing high-yielding and slender grain-shaped *indica* cultivars with strong early growth vigor.

Keywords Genome-wide association analysis (GWAS), Quantitative trait loci (QTL), Shoot dry weight, Causal gene, Rice

Background

Rice is a staple food for more than half the world's population. Direct seeding of rice is a simple, labor-saving, low-cost and efficient rice cultivation technique, which is gradually replacing the traditional transplanting, but it still faces severe problems such as uneven seedling emergence and serious weed infestation (Li et al. 2023). To deal with these problems, it is necessary for the direct-seeded rice to have strong early growth vigor. In breeding systems where transplanting is the main cultivation mode, the evaluation and screening of early growth vigor traits are usually neglected due to the fact that there is a relatively suitable environment for seeds and seedlings growth during the process of seedling cultivation. However, early growth vigor is crucial in the direct seeding cultivation mode. Strong early growth vigor makes seedling emerge rapidly and uniformly, and quickly cover the surface of soil, which is conducive to competing with weeds for soil water and nutrients, thereby inhibiting the growth of weeds and laying a good foundation for the establishment of subsequent rice populations (Zhang et al. 2005; Rao et al. 2007; Diwan et al. 2013; Gulshan et al. 2013). Therefore, developing varieties with strong early growth vigor is an essential target in direct-seeded rice breeding. In addition, the early growth vigor of rice is also crucial to develop high-yielding varieties in South China, the double-cropping rice region due to the relatively short growth duration. Therefore, Huang (1990) proposed the breeding strategies of "short and early growth" and "cluster and early growth" for super high-yielding rice.

Early growth vigor of rice is a complex trait, and the shoot dry weight (SDW) is one of the important indexes of early growth vigor. SDW is affected not only by genotype, but also by environmental conditions and genotype-environment interactions. The SDW measurement in the field is time-consuming and disruptive to the plant, so it is difficult to effectively improve this trait through conventional breeding methods. Understanding its genetic basis, identifying the QTLs/functional genes and carrying out molecular breeding are the efficient and effective way to improve SDW in rice.

With the development of molecular marker technology, some progress has been made in dissecting QTLs associated with SDW in rice. Cui et al. (2002a) firstly used a RIL population to evaluate morphological and physiological indicators related to seedling growth vigor. Combined with the linkage map of RFLP marker, seven QTLs for SDW were mapped. Subsequently, SSR markers were used to dissect QTLs for SDW or shoot fresh weight in diverse populations such as RIL, BL, IL and DH, and more than 60 QTLs associated with SDW in rice were mapped by bi-parental QTL analysis (Zhang et al. 2005, 2017; Xu et al. 2004; Lu et al. 2007; Zhou et al. 2007; Cairns et al. 2009; Cordero-Lara et al. 2016; Singh et al. 2017; Yang et al. 2019, 2021; Dimaano et al. 2020). In recent years, genome-wide association analysis (GWAS) has been widely used to dissect the genetic basis of complex traits in rice. Several cases of QTL mapping of SDW in rice were also reported (Dang et al. 2014; Anandan et al. 2016; Chen et al. 2019; Guo et al. 2019; Zhao et al. 2019; Xu et al. 2021; Zeng et al. 2021). The early GWAS analysis mostly used SSR markers, and the number of markers was limited (commonly less than 1,000), so relatively few loci significantly associated with traits were identified (Dang et al. 2014; Anandan et al. 2016; Zhao et al. 2019); with the development of genome sequencing technology, the density of molecular markers gradually increased (commonly more than 20 K), and some closely adjacent significant loci appeared in QTL detection (Chen et al. 2019; Guo et al. 2019; Zeng et al. 2021). Although great progress has been made in the mapping QTLs for early growth vigor in previous studies, most of the studies were conducted in a single cultivation method or environment, so the reliability and stability of identified QTLs was unclear. So far, only *SBM1*controlling seedling biomass has been cloned in rice (Xu et al. 2021). Therefore, identifying additional stably expressed QTLs and cloning their causal genes will be crucial for rice breeding to improve early growth vigor.

In order to find stable QTLs and their causal genes for SDW, we measured the SDW of a subset of the Rice Diversity Panel 2 (RDP2) using three cultivation methods according to the production practice of direct seeding. A genome-wide association analysis revealed that only one QTL for SDW, *qSDW-5*, could be stably detected in the all three cultivation methods. Through gene-based haplotype analysis of the annotated genes within the *qSDW-5* region, and validated by gene expression and knockout transgenic experiments, *LOC_Os05g09520*, identical to the reported *GW5/GSE5* that controls grain width (GW) and thousand grain weight (TGW) (Duan et al. 2017; Liu et al. 2017), was identified as the causal gene underlying *qSDW-5*, and its pleiotropism in controlling SDW, GW and TGW was firstly confirmed using the diverse international rice collection in this study.

Moreover, phenotypic comparisons of the major haplotypes of *LOC_Os05g09520* in the three subpopulations (*indica*, *japonica* and *aus*), and the haplotype analysis of *LOC_Os05g09520* in 138 cultivars/lines widely used in South China were performed. These results not only provide a promising gene source for the molecular breeding of rice with strong seedling vigor but also elucidate the effect of the *LOC_Os05g09520* haplotypes on SDW, GW and TGW in rice. Based on the results of this study, we also propose a breeding strategy for developing high-yielding and slender grain-shaped *indica* cultivars with strong early growth vigor.

Results

Phenotypic Variations of Shoot Dry Weight under Three Cultivation Methods

According to the production practice, the SDW of the 391 rice accessions were measured using the three cultivation methods: seeds were pre-germinated and sown in plastic trays (GST), seeds were pre-germinated and sown in the paddy field (GSF), and seeds were directly sown in plastic trays without pre-germination (DST). Large variations in SDW were observed (Table S1), ranging from 22.9 to 71.2 mg, with an average of 45.0 mg and a variation coefficient of 19.8% under the GST; from 16.0 to 49.0 mg, with an average of 32.2 mg and a variation coefficient of 19.7% under the GSF; from 7.9 to 25.8 mg, with an average of 17.0 mg and a variation coefficient of 19.0% under the DST. The SDW in 391 rice accessions displayed a continuous and normal distribution (Fig. 1A-C), suggesting that SDW was a quantitative trait controlled by multiple genes.

The phenotype comparisons among different groups revealed that the SDW of *indica* and *aus* group were significantly higher than that of *japonica* group ($P<0.05$) under GST and GSF, but no significant difference among the three groups was detected ($P>0.05$) under DST (Fig. 1D-F).

Correlation analysis of SDW among different cultivation methods exhibited high correlation ($P<0.01$), with correlation coefficients of 0.79, 0.74, and 0.81 between GST and GSF, GST and DST, GSF and DST, respectively, suggesting there were stably QTLs associated with SDW.

Identification of QTLs for Shoot Dry Weight through GWAS under Three Cultivation Methods

Based on the criteria of having less than 30% missing data and minor allele frequency (MAF) more than 5% in the whole population, 446,536 SNPs were selected for GWAS from the 700 K SNPs dataset in the Open Rice GWAS Platform (McCouch et al. 2016; Yang et al. 2023a). Population structure estimated by admixture software, principal component analysis and kinship analysis suggested that there were three subpopulations in this panel, i.e. *indica*, *japonica* and *aus* subpopulation (Yang et al. 2020). Therefore, GWAS of SDW was conducted in the whole population, *indica*, *japonica* and *aus* subpopulation, respectively, using phenotypes generated under the three cultivation methods. Based on about 100 kb of the linkage disequilibrium (LD) decay in tested rice accessions (Yang et al. 2020), a QTL was delimited to a 200-kb region centered on the peak SNP, in which contained three or more than three significant SNPs ($P<0.0001$) (Yang et al. 2020, 2023a). Accordingly, a total of eighteen QTLs for SDW were identified in the present study (Fig. 2; Table 1).

In the whole population, eleven QTLs located on chromosome 1, 2, 3, 4, 5, 6, 9 and 11 were identified under three cultivation methods (Fig. 2; Table 1), among which *qSDW-5* on chromosome 5 could be identified under all three cultivation methods of GST, GSF and DST, *qSDW-3b* on chromosome 3, *qSDW-4a* on chromosome 4 and *qSDW-9* on chromosome 9 could be identified under GSF and DST, while the other seven QTLs could only be identified under one cultivation method.

In *indica*, *aus* and *japonica* subpopulations, seven, three and one QTLs were identified on chromosomes 1, 2, 3, 7 and 9, chromosomes 3 and 8, and chromosomes 5, respectively. Among them, *qSDW_IND-3b* could be identified under GST and DST, *qSDW_AUS-3a* could be identified under GSF and DST, while the others could only be identified under one cultivation method (Fig. 2; Table 1).

Comparisons of the QTLs identified in different populations indicated that *qSDW_IND-9* from the *indica* subpopulation and *qSDW_JAP-5* from the *japonica* subpopulation overlapped with *qSDW-9* and *qSDW-5* identified in the whole population, respectively, while *qSDW_AUS-3a* and *qSDW_AUS-3b* from the *aus* subpopulation overlapped with *qSDW-3b* and *qSDW-3c* identified in the whole population, respectively.

Among the eighteen QTLs identified in the present study, the ten QTLs (*qSDW-1*, *q SDW-3c/qSDW_AUS-3b*, *qSDW-4a*, *qSDW-4b*, *qSDW-6*, *qSDW_IND-2*, *qSDW_IND-3b*, *qSDW_IND-7a*, *qSDW_IND-7b* and *qSDW_AUS-8*) are firstly reported in the present study, and the other eight co-localized with the previously identified QTLs for SDW (Table 1). It is noteworthy that the *qSDW-5* could be stably detected using the three cultivation methods, and had larger contribution to SDW variation in 391 rice accessions. More importantly, *qSDW-5* co-localized with the SDW QTLs previously identified using different rice germplasm under different cultivation environments (Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007; Zhao et al. 2019), but the functional genes underlying these QTLs remain unclear.

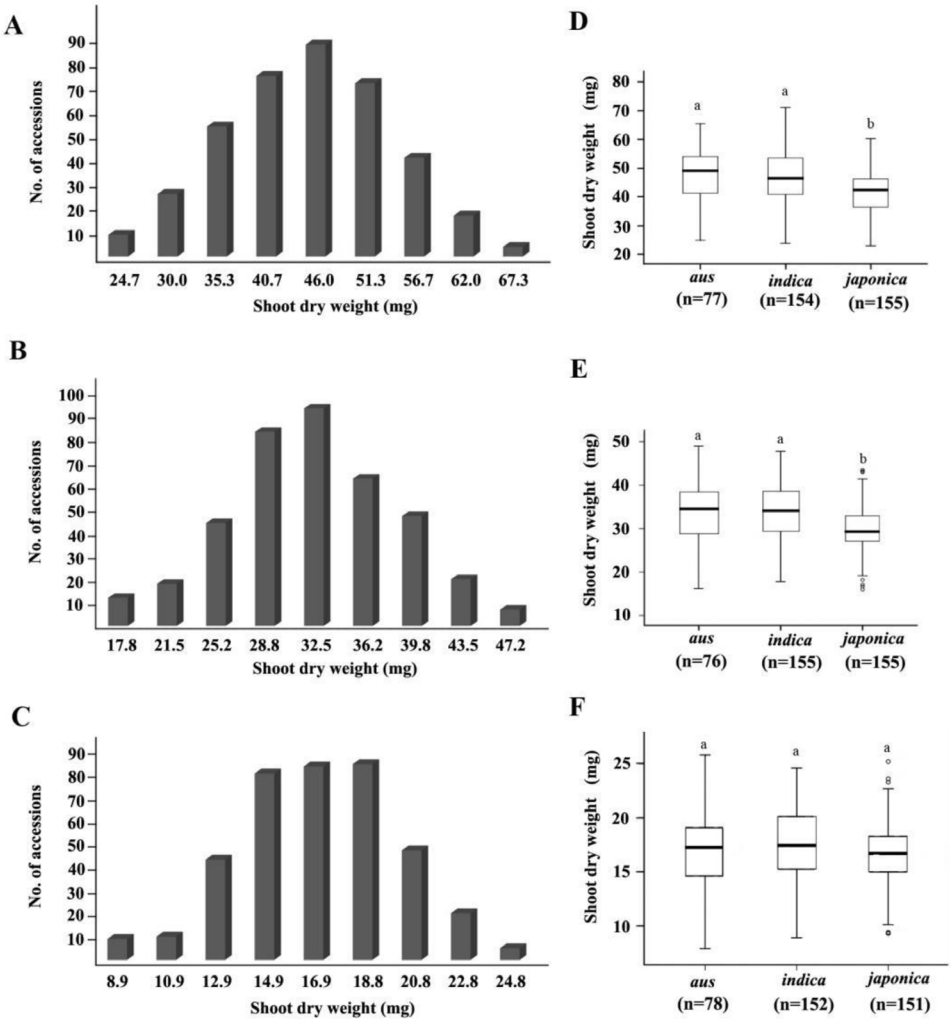

Fig. 1 Distribution and variations of shoot dry weight in the 391 rice accessions under three cultivation methods. **A**-**C**, Distribution of shoot dry weight in the 391 rice accessions under the cultivation method of GST (**A**), GSF (**B**) and DST (**C**). **D**-**F**, Boxplots of the shoot dry weight variation in the three subpopulations under the cultivation method of GST (**D**), GSF (**E**) and DST (**F**). The black horizontal lines represent the median value; the upper side and lower side of the box represent the upper quartile and lower quartile, respectively; the whiskers represent the range of data, and small circles represent outliers. The same letter upon the boxplot means no significant difference in shoot dry weight at $P=0.05$, based on Duncan's multiple range test

Haplotype Analysis of *qSDW-5*

Being detectable in different genetic background and cultivation environments, it is believed that *qSDW-5* is a stably expressed QTL for SDW and has great potential value in rice breeding for improving early growth vigor. To search for the favorable haplotype, haplotype analysis was performed based on the three significant SNPs within the *qSDW-5* interval, and three main haplotypes were identified (Fig. 3A).

Analysis of the SDW in the three main haplotypes showed significant differences between Hap1^{qSDW-5} and Hap2^{qSDW-5}, as well as Hap3^{qSDW-5} ($P<0.05$), but no significant difference in SDW between Hap2^{qSDW-5} and Hap3^{qSDW-5} ($P>0.05$) (Fig. 3B, C and D) under the all three cultivation methods. The average SDWs of the lines carrying Hap1^{qSDW-5} were 50.8, 36.8 and 18.8 mg, which were significantly higher than that of the lines carrying Hap2^{qSDW-5} and Hap3^{qSDW-5}, with the average SDWs of 43.6 and 43.4, 30.7 and 31.1, 16.8 and 16.3 mg under the cultivation method of GST, GSF and DST, respectively.

Candidate Genes Analysis of *qSDW-5*

The LD decay analysis in the QTL region indicated that an approximately 236.2 kb region at the associated locus was the putative region for *qSDW-5* (Fig. 4). Based on release seven of the MSU Rice Genome Annotation Project on rice IRGSP-1.0 genome (http://rice.plantbiology.msu.edu/) (Kawahara et al. 2013), there are 33 annotated genes within the putative region.

Using the genome re-sequencing information (50×) of 256 rice accessions used in this study (Wang et al. 2023; Yang et al. 2023a), the haplotypes of all annotated genes

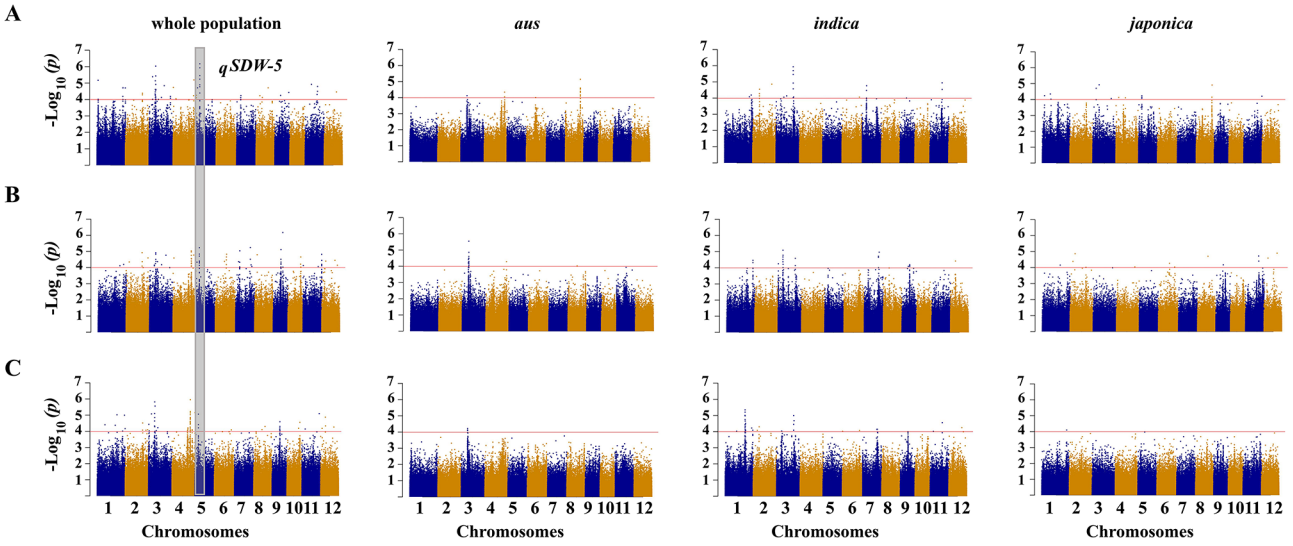

Fig. 2 Genome-wide association study of shoot dry weight in the whole population and three sub-populations using the three cultivation methods. **A**-**C**, Manhattan plots of GWAS for shoot dry weight in 12 chromosomes under GST (**A**), GSF (**B**) and DST (**C**)

within the putative region were analyzed according to the variation of each gene, and the significance of differences in SDW among the main haplotypes (containing more than ten lines) were tested.

Throughout the putative region, haplotype analysis revealed that four genes (*LOC_Os05g09510*, *LOC_Os05g09520*, *LOC_Os05g09660*, *LOC_Os05g09700*) showed significant differences in SDW corresponding to their haplotypes under the three cultivation methods. Among the various haplotypes of the four genes, however, the non-Nipponbare haplotype ($Hap5^{LOC\_Os05g09520}$) of *LOC_Os05g09520* had the highest SDW, while the Nipponbare haplotype of the other three genes had the highest SDW (Fig. 5, Table S3, S4). For *LOC_Os05g09520*, there were three major PAVs variations in the promoter region, and four Indel or SNP variations in the CDS region that cause amino acid variation. Five main haplotypes were identified based on these variations, in which $Hap3^{LOC\_Os05g09520}$ is corresponding to Nipponbare (Fig. 5A). SDW analysis of the five haplotypes showed consistently and highly significant differences between $Hap5^{LOC\_Os05g09520}$ and $Hap1^{LOC\_Os05g09520}$ ($P<0.01$), but there was no significant difference ($P>0.05$) or inconsistently significant difference ($P<0.05$) in SDW among $Hap2^{LOC\_Os05g09520}$, $Hap3^{LOC\_Os05g09520}$ and $Hap4^{LOC\_Os05g09520}$ under the three cultivation methods. The lines carrying $Hap5^{LOC\_Os05g09520}$ exhibited the consistently highest SDW, while the lines carrying $Hap1^{LOC\_Os05g09520}$ exhibited the consistently lowest SDW under the three cultivation methods (Fig. 5B).

To find additional evidence to support the possible function of the candidates, gene differential expression analysis was conducted using two sets of contrasting lines. Firstly, according to the haplotype analysis of *qSDW-5* (Fig. 3), three lines (accession 463, 521 and 1008) carrying $Hap3^{qSDW-5}$ with low SDW and three lines (accession 684, 941 and 1245) carrying $Hap1^{qSDW-5}$ with high SDW were selected from the low and high SDW haplotype panel for RNA-seq. The results indicated that *LOC_Os05g09510* and *LOC_Os05g09700* exhibited no expression (Data not shown), while *LOC_Os05g09660* was rarely expressed in shoots (Fig. 6B). These finding imply that these genes may not play a role in the regulation of seedling growth. In contrast, *LOC_Os05g09520* exhibited higher expression levels in shoots (Fig. 6A). Furtherly, we verified the expression levels of *LOC_Os05g09520* and *LOC_Os05g09660* by qRT-PCR assays using additional nine lines (accession 1089,1166,1216,1235,1302,1321,1329,1377 and 1420) carrying $Hap3^{qSDW-5}$ with low SDW and nine lines (accession 885, 893, 897, 905, 953, 989, 1006, 1281 and 1320) carrying $Hap1^{qSDW-5}$ with high SDW (Fig. 6C, D). The results showed that the expression of *LOC_Os05g09660* in the two groups was hard to be detected at the three time points (Fig. 6D); while the expression level of *LOC_Os05g09520* in the high SDW group was significantly lower than that in the low SDW group at the 3rd d and 9th d (Fig. 6C), implying that *LOC_Os05g09520* may negatively regulate SDW.

Combining the results of haplotype analysis and expression levels of genes, *LOC_Os05g09520* was considered as the most possible candidate gene underlying *qSDW-5*.

Functional Confirmation of *LOC_Os05g09520*

To validate the effect of the candidate gene on SDW, CRISPR/Cas9 was applied to knock out *LOC_Os05g09520* in Nipponbare. Two homozygous lines of

Table 1 QTLs for shoot dry weight identified using the three cultivation methods and their co-location QTLs identified in the previous studies

QTL	Chromosome	GST			GSF			DST			Co-location QTL[#]	Reference
		Position (bp)†	*P*-value	PVE (%)*	Position (bp)†	*P*-value	PVE (%)*	Position (bp)†	*P*-value	PVE (%)*		
whole												
qSDW-1	1	39,523,080	1.93E-05	3.32								
qSDW-2	2				24,656,727	1.19E-05	3.37				Loci48	Zeng et al. 2021
qSDW-3a	3	9,594,696	7.27E-06	3.67							*qSV3b*	Chen et al. 2019
qSDW-3b	3				9,798,479	1.21E-05	3.36	9,786,867	1.52E-06	4.65	*qSV3b*	Chen et al. 2019
qSDW-3c	3	10,112,776	9.33E-07	4.41								
qSDW-4a	4				28,241,595	9.21E-06	3.45	28,274,458	1.47E-05	3.75		
qSDW-4b	4							28,425,835	1.11E-06	4.78		
qSDW-5	5	5,372,955	6.72E-07	4.53	5,372,955	5.85E-06	3.61	5,359,520	8.72E-06	3.96	*qSV-5*, *qSEV-5-1*, *qFV-5-1*, *qSW5-1*	Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007; Zhao et al. 2019
qSDW-6	6				17,247,415	2.20E-05	3.15					
qSDW-9	9				12,151,617	7.66E-06	3.52	12,091,270	2.57E-05	3.54	*qFW9*	Cordero-Lara et al. 2016
qSDW-11	11				28,863,460	1.50E-05	3.29				*qRV-11*, *Loci202*	Yang et al. 2021; Zeng et al. 2021
indica												
qSDW_IND-1	1							30,734,959	4.52E-06	8.92	*qSEV-1-2*, *qSW1-4*	Lu et al. 2007; Zhao et al. 2019
qSDW_IND-2	2	10,505,720	2.84E-05	8.1								
qSDW_IND-3a	3				7,591,381	8.13E-06	8.19				*qSV3b*	Chen et al. 2019
qSDW_IND-3b	3	26,545,143	1.16E-06	11.15				26,580,401	1.02E-05	8.22		
qSDW_IND-7a	7	6,623,534	1.73E-05	8.56								
qSDW_IND-7b	7				21,992,988	1.91E-05	7.47					
qSDW_IND-9	9				12,169,164	6.41E-05	6.49				*qFW9*	Cordero-Lara et al. 2016
aus												
qSDW_AUS-3a	3				9,739,198	3.82E-05	16.39	9,739,198	6.10E-05	14.18	*qSV3b*	Chen et al. 2019
qSDW_AUS-3b	3				10,285,003	2.85E-06	21.94					
qSDW_AUS-8	8	23,436,199	2.54E-05	16.00								
japonica												
qSDW_JAP-5	5	5,372,955	5.81E-05	8.97							*qSV-5*, *qSEV-5-1*, *qFV-5-1*, *qSW5-1*	Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007; Zhao et al. 2019

GST: Pre-germinated seeds were sown in plastic trays; GSF: Pre-germinated seeds were sown in the paddy field; DST: Seeds were directly sown in plastic trays without pre-germination;†Position of the peak SNP at the QTL region; *Percentage of variance explained; #QTLs for shoot dry weight identified by the previous studies

A

Haplotype	5359520 #	5362484	5371772
Hap1$^{qSDW\text{-}5}$	G	A	G
Hap2$^{qSDW\text{-}5}$	A	G	G
Hap3$^{qSDW\text{-}5}$	A	G	A

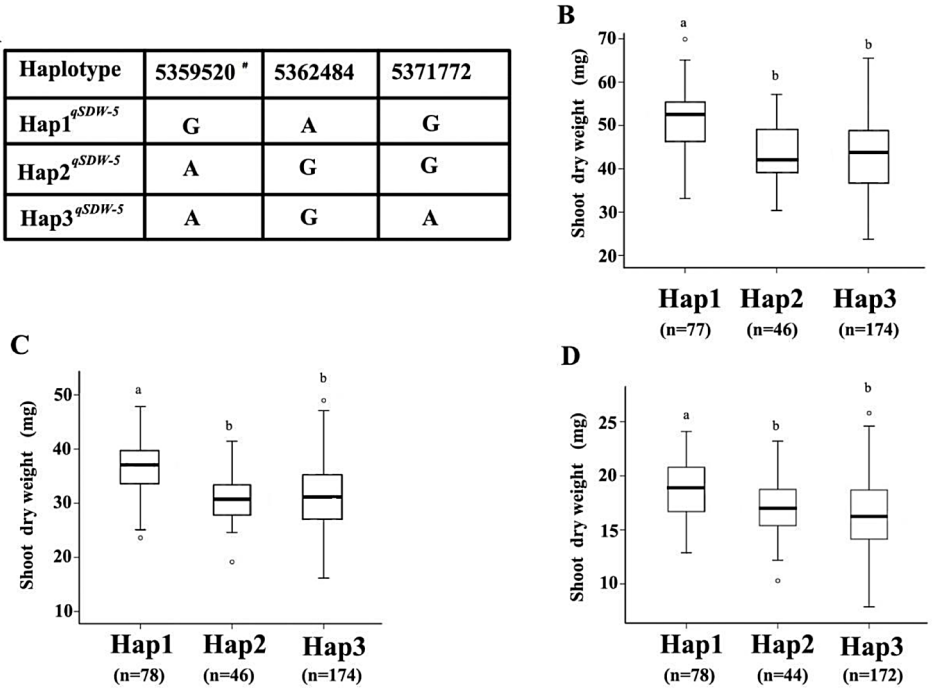

Fig. 3 Differences in shoot dry weight among the main haplotypes of *qSDW-5*. **A**, The main haplotypes of *qSDW-5*; # The SNP position (bp); **B**-**D**, Boxplots for shoot dry weight based on the haplotypes of *qSDW-5* under GST (**B**), GSF (**C**) and DST (**D**); Numbers in parenthesis indicate the number of rice accessions with the haplotype. The black horizontal lines represent the median value; the upper side and lower side of the box represent the upper quartile and lower quartile, respectively; the whiskers represent the range of data, and small circles represent outliers. The values with the same lower letter indicate no significant difference at $P = 0.05$ based on Duncan's multiple range test

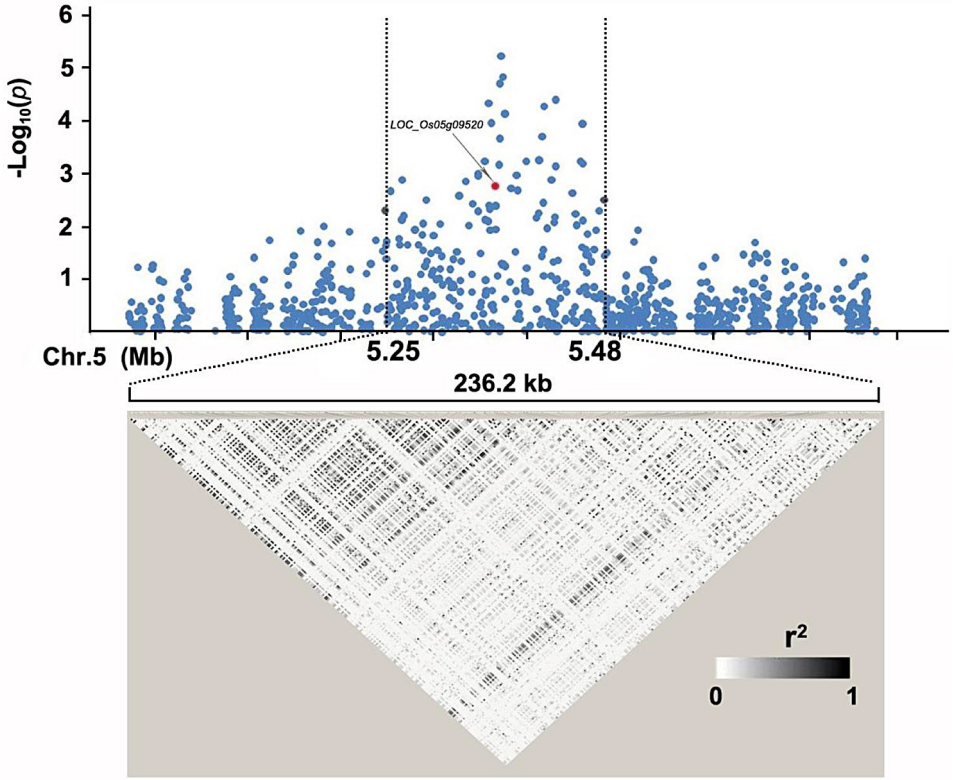

Fig. 4 Candidate region of *qSDW-5* on chromosome 5. Local Manhattan plot (top) and LD heat map (bottom) of *qSDW-5*, indicating the candidate region between 5.25 and 5.48 Mb

A

Haplotype	-6627	-6127	-5747	-4927	-700	+ 771	+ 1239	+ 1362	+ 1378
1	Ref *	Ref	Ref	Ref	~370 deletion	A	T	ACCGGCGGCGCCGGCG	G
2	Ref	Ref	Ref	Ref	~370 deletion	G	TGCG	ACCGGCGGCGCCGGCG	C
3	Ref	~1212 bp deletion			~370 deletion	G	TGCG	ACCGGCGGCGCCGGCG	C
4	Ref	Ref	Ref	Ref	~370 deletion	A	T	ACCGGCGGCGCCGGCG	C
5	~ 950 bp deletion			Ref	Ref	G	TGCG	A	/

B

Haplotype	Shoot dry weight (mg)			Grain width (mm)	Thousand grain weight (g)	Subpopulation		
	GST	GSF	DST			*indica*	*aus*	*japonica*
Hap1 $^{LOC\_Os05g09520}$	37.3±6.5 a (58)	26.5±4.8 a (58)	15.1±2.8 a (58)	2.8±0.4 b (58)	26±4.1 b (56)	3/3/3/3/3$^{\#}$	0/0/0/0/0	55/55/55/55/53
Hap2 $^{LOC\_Os05g09520}$	43.7±7.6 b (52)	32.3±5.2 b (52)	16.6±3.1 b (51)	2.6±0.2 a (52)	22.8±3.3 a (49)	49/49/48/49/46	0/0/0/0/0	3/3/3/3/3
Hap3 $^{LOC\_Os05g09520}$	45.5±6.3 bc (41)	32.2±4.6 b (41)	18.1±2.8 c (40)	3.5±0.2 d (40)	28.8±4.8 c (38)	0/0/0/0/0	0/0/0/0/0	41/41/40/40/38
Hap4 $^{LOC\_Os05g09520}$	47.2±9.6 c (40)	32.8±7.1 b (40)	16.7±3.8 b (40)	2.7±0.3 a (40)	21.5±3.4 a (38)	2/2/2/2/2	38/38/38/38/36	0/0/0/0/0
Hap5 $^{LOC\_Os05g09520}$	50.9±8.1 d (64)	37.3±4.8 c (65)	19.1±2.8 c (64)	3.1±0.2 c (65)	25.2±3.7 b (63)	51/52/51/52/50	10/10/10/10/10	3/3/3/3/3

Fig. 5 Gene structure and haplotype analysis of *LOC_Os05g09520*. **A**, The haplotypes of *LOC_Os05g09520*; * A pangenome was used as the reference genome; / represents no base information. **B**, The shoot dry weight, grain width and thousand grain weight of various haplotypes for *LOC_Os05g09520*. The values were presented in mean ± SD. The values with the same lower letter indicate no significant difference at $P = 0.05$ based on Duncan's multiple range test; Numbers in parenthesis indicate the number of rice accessions; # indicates the number of rice accessions used to measure shoot dry weight (GST), shoot dry weight (GSF), shoot dry weight (DST), grain width and thousand grain weight, respectively

transgenic plants were selected for SDW measurement using two cultivation methods (Fig. 7A-F).

Since differences in shoot length (SL) may be an essential factor contributing to differences in SDW, the SL were also measured along with SDW. After 14 days of growth in the nutrient solution, there was no significant difference in SL between the KO lines and Nipponbare, with 22.84 cm, 22.51 cm and 22.94 cm for Nipponbare and the two KO lines, respectively (Fig. 7E); while the SDW of the KO lines were significantly higher than that of Nipponbare, with SDW of 15.24 mg for Nipponbare, and 19.57 mg and 16.67 mg for the two KO lines, respectively (Fig. 7C); and also, the growth rate of SDW of the KO lines were significantly higher than that of Nipponbare, with growth rate of SDW of 1.09 mg/day for Nipponbare, and 1.40 mg/day and 1.19 mg/day for the two KO lines, respectively (Fig. 7D).

To investigate the phenotypic stability of the gene, the SL and SDW were also measured after 14 days of growth in soil. The results showed that the SL of KO lines were significantly shorter than that of Nipponbare, with 37.15 cm for Nipponbare, 34.72 cm and 34.57 cm for the two KO lines, respectively (Fig. 7I), but the SDW of the KO lines were still significantly higher than that of Nipponbare, with Nipponbare having a SDW of 33.69 mg and the KO lines having a SDW of 37.53 mg and 38.37 mg, respectively (Fig. 7G); and also, the growth rate of SDW of the KO lines were significantly higher than that of Nipponbare, with growth rate of SDW of 2.41 mg/day for Nipponbare, and 2.68 mg/day and 2.74 mg/day for the two KO lines, respectively (Fig. 7H).

Conclusively, although there was no significant difference in SL between the KO lines and Nipponbare in nutrient solution (Fig. 7E), or the SL of the KO lines were significantly shorter than that of Nipponbare in soil culture (Fig. 7I), the SDW and the growth rate of SDW of the KO lines were significantly increased under both cultivation methods, compared with that of their wild-type Nipponbare (Fig. 7C, D, G, H), indicating that *LOC_Os05g09520* is the causal gene for SDW in rice.

LOC_Os05g09520 is the reported *GW5/GSE5* controlling GW and TGW (Duan et al. 2017; Liu et al. 2017), so GW and TGW of the KO lines were measured, and it was found that the KO lines had wider grains and higher TGW compared with that of their wild-type Nipponbare (Fig. S1).

Various Haplotypes of *LOC_Os05g09520* Cause Significant Differences in the Shoot Dry Weight, Grain Width and Thousand Grain Weight

Since the upstream sequence of about 5 kb has a regulatory effect on the expression of *LOC_Os05g09520* (Duan et al. 2017; Liu et al. 2017), and the CDS region of *LOC_Os05g09520* also has sequence variation that leads to amino acid variation, we analyzed the haplotype of *LOC_Os05g09520* based on these variations (Fig. 5A),

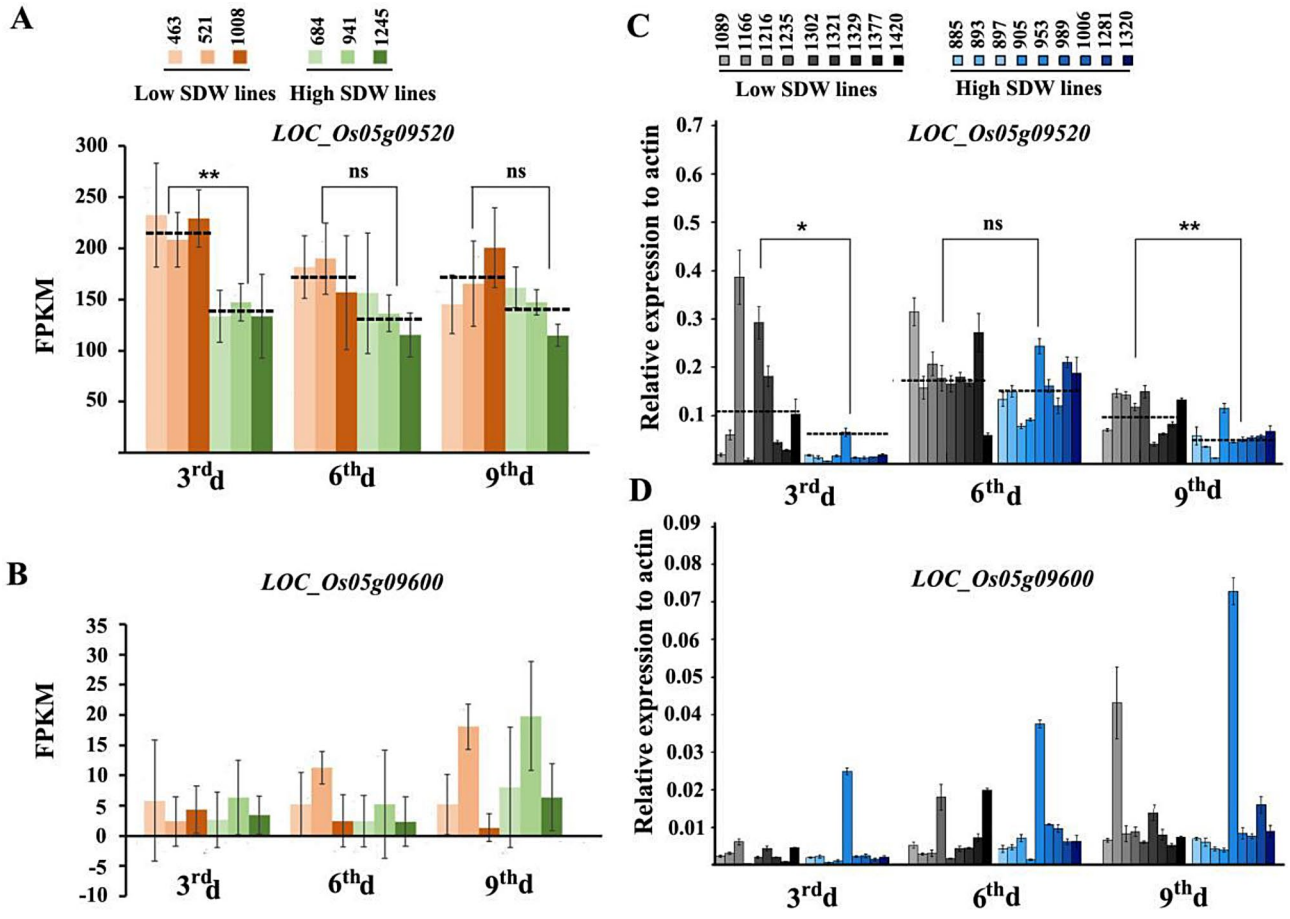

Fig. 6 Temporal expression patterns of candidate genes measured by RNA-seq and qRT-PCR. Temporal expression patterns of *LOC_Os05g09520* (A) and *LOC_Os05g09660* (B) measured by RNA-seq using three lines (accessions 463, 521 and 1008) carrying Hap3^{qSDW-5} with low SDW and three lines (accessions 684, 941 and 1245) carrying Hap1^{qSDW-5} with high SDW; temporal expression patterns of *LOC_Os05g09520* (C) and *LOC_Os05g09660* (D) measured by qRT-PCR using nine lines with low SDW (accessions 1089,1166,1216,1235,1302,1321,1329, 1377 and 1420) and nine lines with high SDW (accessions 885,893,897,905,953,989,1006,1281 and 1320). ns means no significant difference at $P=0.05$, based on *t*-test. * and ** means significant difference at $P=0.05$ and $P=0.01$

and compared the SDW, GW and TGW in the five main haplotypes (Fig. 5B).

For SDW, the Hap5^{LOC_Os05g09520} exhibited the consistently highest SDW under the three cultivation methods, among which *indica*, *aus* and *japonica* accounted for 79.7%, 15.6% and 4.7%, respectively, while the Hap1^{LOC_Os05g09520} exhibited the consistently lowest SDW under the three cultivation methods, among which *indica* and *japonica* accounted for 5.2% and 94.8%, respectively. For GW and TGW, the Hap3^{LOC_Os05g09520} containing only *japonica* exhibited the widest GW and the highest TGW, while both Hap2^{LOC_Os05g09520} and Hap4^{LOC_Os05g09520} exhibited the narrowest GW and the lowest TGW, among which the Hap2^{LOC_Os05g09520} contained 94.2% *indica* and 5.8% *japonica*, Hap4^{LOC_Os05g09520} contained 95.0% *aus* and 5.0% *indica*. It is worth noting that in the subpopulations, two major haplotypes were identified in *aus* (Hap4^{LOC_Os05g09520} and Hap5^{LOC_Os05g09520}), *indica* (Hap2^{LOC_Os05g09520} and Hap5^{LOC_Os05g09520}) and *japonica* (Hap1^{LOC_Os05g09520} and Hap3^{LOC_Os05g09520}), and the SDW, GW and TGW of the one haplotype were consistently higher than those of the other haplotype in all three subpopulations, which showed the same patterns of wider GW and higher TGW along with higher SDW in all three subpopulations. Further correlation analysis among the three traits in the GWAS population showed that SDW was significantly positive correlated with GW and TGW ($P<0.01$), with correlation coefficients of 0.42 and 0.34, 0.4 and 0.37, 0.49 and 0.53 under cultivation method of GST, GSF and DST, respectively.

Haplotype Analysis of *LOC_Os05g09520* in Cultivars of South China Rice Region

Numerous *indica* cultivars bred in the South China rice region are characterized by early and fast growth, especially the Super Rice Wuyou 308 and Tianyou 998. In order to clarify how *LOC_Os05g09520* is selectively utilized by breeders in the South China rice region, we

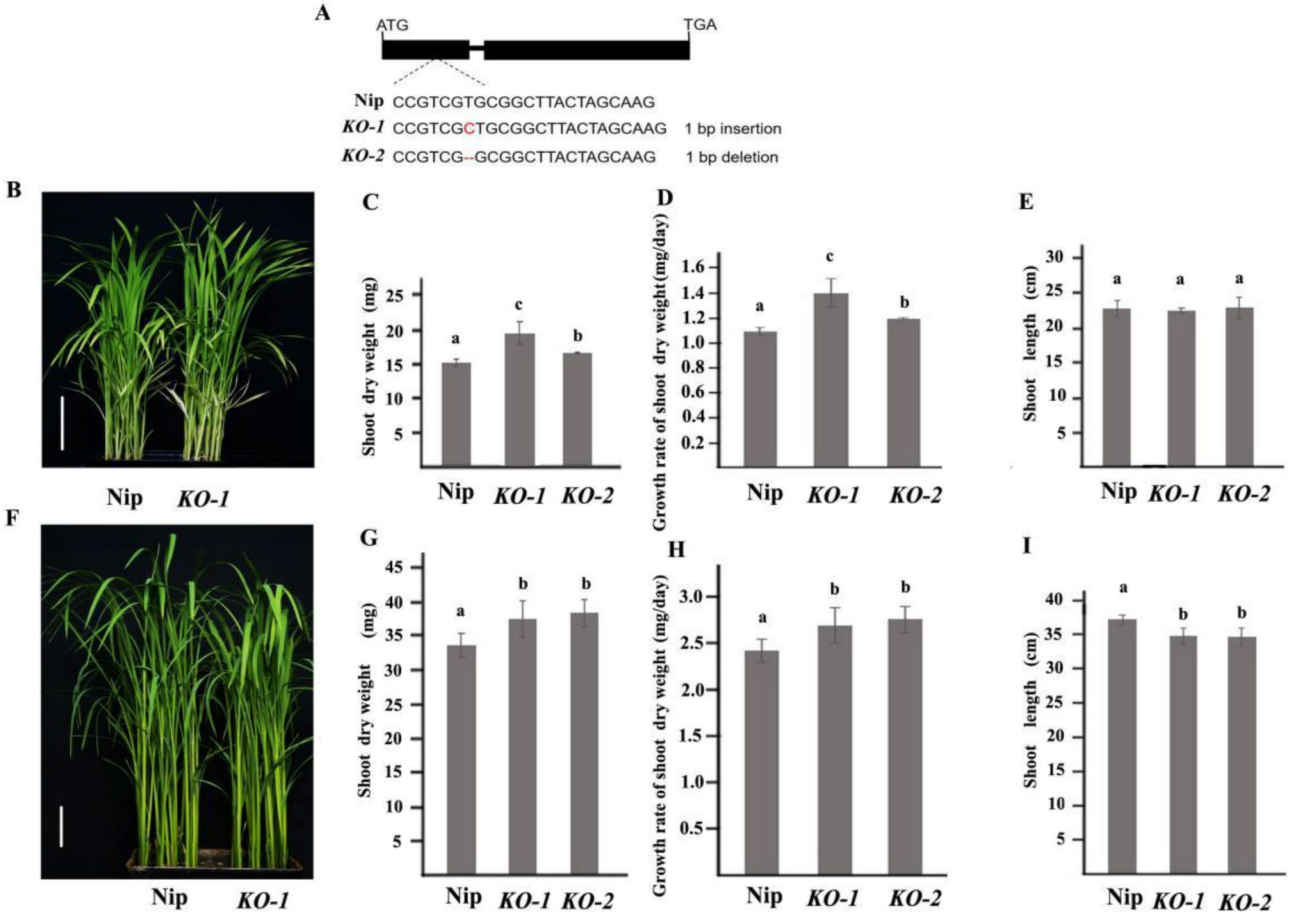

Fig. 7 Mutation types, shoot dry weights and shoot lengths of the knockout transgenic lines of *LOC_Os05g09520*. **A**, The mutation types of *LOC_Os05g09520*. **B** and **F**, The phenotypes of the knockout transgenic (KO) lines and their wild-type Nipponbare (Nip) in nutrition solution (**B**) and soil culture (**F**) after 14 days of growth. **C** and **G**, The shoot dry weight of the knockout transgenic (KO) lines were significantly higher than that of their wild-type Nipponbare (Nip) in nutrition solution (**C**) and soil culture (**G**). **D** and **H**, The growth rate of shoot dry weight of the knockout transgenic (KO) lines were significantly higher than that of their wild-type Nipponbare (Nip) in nutrition solution (**D**) and soil culture (**H**). **E** and **I**, The shoot length of the knockout transgenic (KO) lines and their wild-type Nipponbare (Nip) in nutrition solution (**E**) and soil culture (**I**). The different letter upon the histogram indicates the significant difference at $P = 0.05$ based on Duncan's multiple range test. Scale bar, 5 cm

analyzed the haplotypes of *LOC_Os05g09520* in 138 cultivars/lines widely used in South China. The results revealed that 97.8% of the cultivars/lines carried Hap-$2^{LOC\_Os05g09520}$, including the maintainer lines(Wufeng B and Tianfeng B) and restorer lines (Guanghui 308 and Guanghui 998) of Wuyou 308 and Tianyou 998 (Table S2). The Hap$2^{LOC_Os05g09520}$ was characterized by narrower GW, lower TGW and SDW between the two major haplotypes of *indica* (Fig. 5B).

Discussion

The *qSDW-5* has a Great Potential Value in Rice Breeding

The trait with sufficiently large phenotypic variation is an ideal breeding target in rice breeding. In this study, large variations in SDW were found in 391 diverse germplasms (Table S1), with the lowest SDW of 22.9 mg (GST), 16.0 mg (GSF), and 7.9 mg (DST) and the highest SDW of 71.2 mg (GST), 49.0 mg (GSF), and 25.8 mg (DST), which makes it possible to select different SDW according to different breeding goals.

In the practice of rice production, the cultivation methods and environment of direct seeding are changeable, searching the early growth vigor QTL that can be stably expressed under different cultivation methods may be an economical and effective way to address this issue. In this study, a total of eighteen QTLs for SDW were identified using the three cultivation methods (Fig. 2; Table 1), but thirteen of them could only be identified under one cultivation method, indicating that QTLs for SDW were significantly affected by environments. However, it was found the QTL, *qSDW-5* could be detected in all three cultivation methods and four QTLs, *qSDW-3b*, *qSDW-4a*, *qSDW-9* and *qSDW_IND-3b* could be detected in two cultivation methods, which could be the important QTLs for SDW in rice. Notably, *qSDW-5* identified in this study overlaps with the SDW QTLs identified in the

previous studies (Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007; Zhao et al. 2019). Using the RIL population derived from Lemont/Teqing, three overlapping QTLs associated with SDW at the different seedling ages were identified under different cultivation methods and environments (Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007), among which *qSV-5* controlling dry matter weight of 15-day-old seedlings was identified by the paper-roll tests (Zhang et al. 2005), *qSEV-5-1* controlling dry weight of 34-day-old seedlings was identified using pot culture with soil (Lu et al. 2007), and *qFV-5-1b* controlling seedling dry weight of 7-day-old seedlings was identified under field conditions (Zhou et al. 2007). Through GWAS of a mini core collection consisting of 273 cultivated rice accessions in hydroponic culture, *qSW5-1* associated with shoot weight of 21-day-old seedlings was also identified (Zhao et al. 2019). Although these studies were conducted using different rice germplasm at different seedling growth periods, and under different cultivation methods or environments, the QTL intervals on chromosome 5 that overlapped with *qSDW-5* could be detected. It can be seen that the expression of *qSDW-5* is stable across different genetic backgrounds and environments. Larger effect on SDW and stable expression make *qSDW-5* have a great potential value in molecular breeding for early growth vigor in rice.

LOC_Os05g09520 is a Promising Target for Improvement of Shoot Dry Weight in Rice

SDW is one of the important traits associated with early growth vigor in rice, but its measurement in the field is destructive and time-consuming and labor-intensive. It is difficult to improve this trait efficiently by traditional breeding methods. Identifying stable QTL for SDW and conducting molecular breeding is the best way to break through early growth vigor breeding. Although *qSDW-5* overlaps with the SDW QTLs identified in the previous studies using different rice germplasm under different cultivation methods and environments (Zhang et al. 2005; Lu et al. 2007; Zhou et al. 2007; Zhao et al. 2019), none of functional gene is obtained. In this study, *LOC_Os05g09520* was identified as the functional gene for SDW and its favorable haplotype was also obtained, which is beneficial to haplotype-based breeding for early growth vigor.

LOC_Os05g09520 is identical to *GW5/GSE5*, a known gene controlling GW and TGW, and encodes a putative protein containing IQ domains (IQD) (Duan et al. 2017; Liu et al. 2017). IQD proteins are an ancient calmodulin-binding proteins family, which can regulate plant stress response and plant development (Abel et al. 2005; Xiao et al. 2008). Previous study demonstrated that *LOC_Os05g09520* is a positive regulator of brassinosteroid (BR) signaling. LOC_Os05g09520 protein is localized to the plasma membrane, and can physically interact with GSK2 (glycogen synthase kinase 2) and repress its kinase activity in rice, resulting in accumulation of unphosphorylated OsBZR1 (*Oryza sativa* BRASSINAZOLE RESISTANT1) and DLT (DWARF AND LOW-TILLERING) proteins in the nucleus to mediate brassinosteroid (BR)-responsive gene expression and growth responses (including grain width and weight) (Liu et al. 2017). Since BR mainly regulates the normal growth, development and morphogenesis of plants by promoting processes such as cell extension, division and differentiation, including seed germination, root extension, stem elongation, leaf extension, and xylem differentiation, etc. (Steven et al. 1998), it could be inferred that the protein encoded by *LOC_Os05g09520* ultimately affects the BR response, resulting in differences in SDW. It is the first time to report the biological function of *LOC_Os05g09520/GW5* on SDW in the present study.

In order to obtain some information on the molecular mechanism of *GW5* in regulating SDW, a bioinformatic analysis based on the RNA-seq data were performed (Fig.S2). The results suggest that the gene *LOC_Os05g09520* may enhance SDW and resilience through hormone regulation and environmental stress adaptation, and high SDW accessions carrying the relevant haplotype of *qSDW-5* demonstrate higher expression levels in genes that orchestrate a fine-tuned balance of hormone-mediated growth processes and adaptive stress responses across different developmental stages. Initially, hormone regulation is prioritized to kickstart growth (Fig.S2C). As seedlings develop, the focus shifts towards enhancing resilience against environmental stresses while sustaining growth-promoting pathways (Fig.S2E and G). Conversely, low SDW accessions downregulate crucial growth and stress response mechanisms, potentially leading to impaired growth and reduced seedling vigor (Fig.S2D, F and H). The dynamic regulation observed here underscores the multifaceted role of genes in controlling growth and adaptation processes, providing valuable insights for crop improvement strategies.

The utilization of natural variation greatly contributes to improvement of important agronomic traits in crops. Using the re-sequencing information of 256 accession in the GWAS population, five main haplotypes in *LOC_Os05g09520* were identified based on the variations. Among them, $Hap5^{LOC\_Os05g09520}$ exhibited the highest SDW, and $Hap1^{LOC\_Os05g09520}$ exhibited the lowest SDW under the three cultivation methods (Fig. 5B). There were three major PAV variations in the promoter region, and four Indel or SNP variations in the CDS region that cause amino acid variations between $Hap1^{LOC\_Os05g09520}$ and $Hap5^{LOC\_Os05g09520}$ (Fig. 5B). The expression level of *LOC_Os05g09520* in the high SDW group exhibited significantly lower than that in the low SDW group

(Fig. 6C), which is similar to the effects of *GW5* on regulation of GW and TGW (Duan et al. 2017), the relatively lower expression levels of *GW5* may be one of the reasons for the relatively higher SDW (Fig. 6A, C).Whether the amino acid variations in CDS of *LOC_Os05g09520* also contribute to SDW diversity remains to be further studied.

The Suitable Haplotypes of *LOC_Os05g09520* is Beneficial to Achieve the Desired Shoot Dry Weight and/or Grain Size in Rice Breeding

Phenotypic correlation analysis showed that seed size/weight was positively correlated with traits related to early growth vigor in rice (Roy et al. 1996; Fauzi et al. 2021), and the genetic basis analysis of these traits also indicated that seed size/weight was closely associated with seedling vigor (Lu et al. 2007; Cui et al. 2002b). However, there is few direct genetic evidence to elucidate their correlation. In this study, *LOC_Os05g09520*, a known gene controlling GW and TGW (Duan et al. 2017; Liu et al. 2017), was firstly identified as a causal gene for SDW in rice. By investigating the phenotypes of *LOC_Os05g09520* knockout transgenic plants, it was also confirmed that *LOC_Os05g09520* controls SDW, GW and TGW in rice (Fig. 7, Fig. S1), which provides a direct genetic evidence for the correlation between these traits.

By analyzing the effect of *LOC_Os05g09520* haplotype variations on SDW, GW and TGW in the GWAS population, the same pattern of wider GW and higher TGW along with higher SDW was found between the two or two major haplotypes in *aus*, *indica* or *japonica* subpopulation (Fig. 5B). Given this pattern, $Hap3^{LOC\_Os05g09520}$ and $Hap5^{LOC\_Os05g09520}$ are the favorable haplotypes for *japonica* and *indica*/*aus*, respectively, if we consider the optimal haplotype for breeding in terms of yield (GW and TGW) and seedling vigor (SDW). However, through the haplotype analysis of *LOC_Os05g09520* in 138 cultivars/lines (*indica*) from the South China rice region, it was found that 97.8% of the cultivars/lines carried $Hap2^{LOC\_Os05g09520}$ (Table S2). Among the two major haplotypes of *indica*, $Hap2^{LOC\_Os05g09520}$ and $Hap5^{LOC\_Os05g09520}$, $Hap2^{LOC\_Os05g09520}$ had narrower GW, lower TGWs and SDW (Fig. 5B). This may be mainly due to the fact that cultivars are the product of a comprehensive trait selection, achieving a high balance of yield, quality and resilience, taking into account the satisfaction of market consumption needs. In the *indica* rice growing region, the rice consumption market needs cultivars with slender grains, and the trait of grain slenderness is easy to be recognized and selected by breeders in breeding, so $Hap2^{LOC\_Os05g09520}$ is always selected based on the breeding orientation of slender grain, because *GW5/GSE5* is a core gene controlling grain slenderness (Yang et al. 2023b). In the previous study, we found that *GS3*, *GW5*, and *GW7/GL7* were the key genes regulating grain size in Guangdong simiao rice (slender grains), and these three genes significantly contributed to grain length, grain width, and length/width ratio. Based on the criteria of length/width ratio more than 3.5, 75% of Guangdong simiao rice carried the gene combination of $Hap3^{GS3}$+$Hap2^{LOC\_Os05g09520/GW5}$+$Hap2^{GL7}$, which has the distinctive characteristics of slender and long grains (Yang et al. 2023b). However, $Hap5^{LOC\_Os05g09520/GW5}$ with wider GW, higher TGW and SDW in *indica* should be emphasized and applied in the current demand for yield improvement and the increasing emphasis on early growth and fast development traits. Fortunately, some of the early high-yielding varieties/lines in Guangdong Province (including Guangluai4hao, Teqing and Guanghui3550, etc.) carry $Hap5^{LOC\_Os05g09520}$, which can used as excellent donors for developing high-yielding varieties with strong early growth vigor. Therefore, using *GS3* and *GL7* to satisfy the consumption habit of long-grain size, and combining with *GW5* to develop varieties with slender and larger grain and improved early growth and fast development characteristics, the three genes with the haplotype combination of $Hap3^{GS3}$+$Hap2^{GL7}$+$Hap5^{GW5/LOC\_Os05g09520}$ would be beneficial to achieve the desired traits for *indica* rice production in South China, as well as and the popularization of rice direct seeding.

Although 97.8% of these 138 cultivars/lines carried the $Hap2^{LOC\_Os05g09520}$, but Wuyou 308 and Tianyou 998, the Super Hybrid Rice derived from parents carried the $Hap2^{LOC\_Os05g09520}$ (Table S2) had better early growth and fast development characteristics. Therefore, it can be hypothesized that in addition to *LOC_Os05g09520*, there are other QTLs for SDW that have a greater effect on early growth and fast growth in *indica*, such as *qSDW-3b/qSDW_AUS-3a*, *qSDW_IND-3b*, *qSDW-4a* and *qSDW-9* identified in this study. Further cloning of the causal genes underlying these QTLs and analysis of their haplotypes would help to comprehensively elucidate the genetic basis of the early growth and fast development trait in rice. In addition, heterosis may also be responsible for their early growth and fast development characteristics, which is worthy to further study.

Conclusions

In this study, *qSDW-5* was identified as a stable QTL for SDW in rice. *LOC_Os05g09520*, identical to *GW5*, was the causal gene underlying *qSDW-5* and its main haplotypes were also identified. The effects of *LOC_Os05g09520* on SDW, GW and TGW are validated through gene-based haplotype analysis and knockout transgenic experiment. Analyzing the effect of *LOC_Os05g09520* haplotype variations on SDW, GW and TGW in the GWAS population revealed the same

patterns of wider GW and higher TGW along with higher SDW in the subpopulations of *aus*, *indica* or *japonica*.

Materials and methods

Plant Materials

In this study, 391 diverse rice accessions from 56 countries were used for GWAS (designated as GWAS population, Table S1). These rice accessions were selected from the Rice Diversity Panel 2 consisting of 1568 accessions based on their diversity and origins (McCouch et al. 2016). These accessions were obtained from the International Rice Research Institute genebank, and all seeds used in this study were newly increased in the experimental year. The harvested seeds were stored at room temperature for three months before phenotypic evaluation. In addition, 138 cultivars/lines from the South China rice region were used for *LOC_Os05g09520* haplotype analysis (Table S2).

Evaluation of Shoot Dry Weight

For the GWAS population, the healthy and filled seeds were incubated at 49℃ for 96 h to break dormancy. After sterilization in 3% sodium hypochlorite solution, the seeds were soaked in distilled water for 24 h and sown using three cultivation methods according to the production practice, as described in our previous study (Yang et al. 2023a). Briefly, Method 1 (GST): seeds were pre-**g**erminated and **s**own in plastic **t**rays (35.0 cm×23.0 cm×6.0 cm); method 2 (GSF): seeds were pre-**g**erminated and **s**own in the paddy **f**ield; method 3 (DST): seeds were **d**irectly **s**own in plastic **t**rays without pre-germination. GST studies were conducted in May 2018, while GSF and DST studies were conducted between late April and early May 2019. The average air temperature was 28.7℃, 23.6℃ and 24.1℃under GST, GSF and DST, respectively. After 14 days of growth in natural environment, the seedlings were pulled out and cut at the junction between shoot and root. The shoots were dried to constant weight in an oven at 80 ℃, and then the SDWs were weighed using a 0.001-g electronic balance. In this study, three replicates with ten seedlings per accession were used for SDW evaluation, and the measurements were transformed into shoot dry weight per plant (mg) for further analysis.

For knockout transgenic (KO) lines, the healthy and filled seeds of the KO lines and their wild-type line were incubated at 49℃ for 96 h to break dormancy. After sterilization in 3% sodium hypochlorite solution, the seeds were soaked in distilled water for 24 h. The pre-germinated seeds were sown in black plastic culture boxes (12 cm×8.6 cm×11 cm) filled with 0.1% Yoshida nutrient solution or fine soil, then put into a growth chamber set at 30℃, 70% relative humidity and a 12 h light/12 h dark cycle. After 14 days, the SDW and shoot length (SL) were measured. Three replicates with 30 seedlings per line were used for the evaluation of SDW and SL. The growth rate of SDW is calculated as an average of SDW at the14th day, i.e., SDW per day.

Evaluation of Grain Width and Thousand Grain Weight

For haplotype analysis, the re-sequenced accessions from the GWAS population (Table S1) were planted in paddy fields at the Guangzhou Experimental Station in Guangdong Province, China. The experiments were conducted in the second cropping season (July to November) in 2016 and arranged in a completely randomized block design with two replications. The germinated seeds were sown in a seedling bed and sixteen 15-day-old seedlings were transplanted into two rows in the field with an individual plant space of 20 cm×20 cm. The seeds were harvested on the 35th day after heading and dried naturally. GWs were measured using the Rice Appearance Quality Determination Instrument (SC-E, Hangzhou, Chian), and TGWs were weighed using a 0.001-g electronic balance.

For KO lines, the *LOC_Os05g09520* KO lines and their wild-type line were planted in the transgenic experimental field in the second cropping season (July to November) in Guangzhou (2023) of Guangdong Province, China, and arranged in a completely randomized block design with three replications. Twenty-four 15-day-old seedlings were transplanted into three rows in the field with an individual plant space of 20 cm×20 cm. The seeds were harvested on the 35th day after heading, dried naturally, then GWs and TGWs were measured.

The field management, including irrigation, fertilization, and disease and pest control, followed the conventional practice for rice production.

GWAS Analysis and QTL Delimitation

GWAS analysis was performed as described in our previous study by using software GAPIT version two and HDRA dataset consisting of 700 K single nucleotide polymorphisms (SNPs) (Zhao et al. 2018; Yang et al. 2020, 2023a). SNPs were filtered by the criteria of having less than 30% missing data and minor allele frequency (MAF) > 0.05. In order to reduce the effect of population structure on GWAS, the mixed linear model (MLM) was selected in which the kinship matrix was used jointly with PC in GAPIT, and three PCs were included when analyzing the whole population consisting of *indica*, *japonica* and *aus* (Yang et al. 2020). The Manhattan plot was produced using the R package qq-man. A region having three or more than three significant SNPs ($P < 0.0001$) within 200 kb is considered as one QTL, which was used in previous studies (Zhao et al. 2018; Yang et al. 2020, 2023a).

DNA Sequence Analysis

Two hundred and fifty-six rice accessions in the GWAS population and 138 cultivars/lines from the South China rice region were re-sequenced using the Illumina NovaSeq6000 platform and the details of sequencing data analysis were described in our previous study (Zhao et al. 2018; Yang et al. 2023a, b). All raw sequence data have been deposited in the NCBI sequence read archive (BioProject accession PRJNA820969).

Haplotype Analysis and Candidate Gene Identification

For the QTL haplotype analysis, three significant SNPs including the peak SNP in the QTL interval were used for analysis. For gene-based haplotype analysis, the indel (≤50 bp), SNP (with Nipponbare as the reference genome) and PAV (the presence/absence variation > 50 bp, with Pan-genome as the reference genome) within the QTL interval were firstly analyzed using the re-sequencing information (50×) for the 256 rice accessions (Wang et al. 2023; Yang et al. 2023a). Next, all annotated genes within the QTL interval were examined to identify their haplotypes based on their sequence variations, respectively. Then the accessions were grouped based on the haplotypes of each gene and the post hoc multiple comparison with Duncan function was performed to identify the significant differences in SDW between the major haplotypes (containing more than ten accessions). A gene was considered a candidate gene if the significant differences in SDWs were observed among haplotypes of a gene under all cultivation environments.

The haplotype analysis of *LOC_Os05g09520* in 138 cultivars/lines from the South China rice region was also performed based on the sequence variations of the gene.

RNA-Sequencing and Data Analysis

Three rice accessions with low SDW and three rice accessions with high SDW were selected for differential expression analysis of candidate genes based on the haplotype analysis. The germinating seeds were sown into trays filled with fine soil. Sampling was conducted on the 3rd, 6th and 9th d after sowing, respectively, with three biological replications. Total RNA was extracted from shoots using Trizol reagent (Invitrogen, Carlsbad, CA, USA) and purified using RNeasy Plant Mini Kit (Qiagen, Valencia, CA). RNA-Seq was performed at GENEWIZ Technology (Suzhou, China), and data analysis was conducted using HISAT2-Stringtie-Deseq2 pipeline. Raw counts of each sample exported from Stringtie were imported and normalized by DEseq2. Genes with read counts less than ten in all samples were filtered out for further analysis. Gene expression was quantified as mean FPKM (fragments per kilo-base of exon per million fragments mapped) of three biological replicates. Differential expression analyses were performed between the three stages with low and high SDW groups using DEseq2, with 0.05 as the FDR cut-off and a Log2 fold change (FC) cut-off of 1. Hypergeometric tests were performed to determine whether specific functional categories from Gene Ontology (GO) were significantly over-represented in differential expression gene sets using the R package cluster Profiler.

qRT-PCR Analysis

Nine rice accessions with low SDW and nine rice accessions with high SDW were selected for differential expression analysis of candidate genes based on the haplotype analysis. The germinating seeds were sown into trays filled with fine soil. Sampling was conducted on the 3rd, 6th and 9th d after sowing, respectively, with three biological replications.The methods for total RNA extraction, RNA reverse transcription reaction, and quantitative real-time PCR refer to Yang et al. (2023a). The primers for qRT-PCR were designed by Primer Blast (https://www.ncbi.nlm.nih.gov/). The *actin* was used as endogenous normalized genes for mRNA. All reactions were run in triplicate. Primers used to amplify the selected genes are listed in Table S5.

Validation of Candidate Genes for Shoot Dry Weight

In order to validate the candidate gene for SDW, we conducted the knockout transgenic experiments. To generate the CRISPR/Cas9 vectors, *LOC_Os05g09520* single guide RNA (sgRNA) sequences were cloned using pYLgRNA-OsU3, respectively, as described previously (Ma et al. 2015). The target site sequence of *LOC_Os05g09520* was 5'-CTTGCTAGTAAGCCGCACGA-3', which contained a protospacer adjacent motif (PAM) CGG at the 3' end. The positive plasmids were electroporated into *Agrobacterium tumefaciens* EHA105, then introduced into calli of the cultivar Nipponbare via Agrobacterium-mediated genetic transformation.

At the T_2 generation, the homozygous positive transgenic plants of the candidate gene were selected by gene cloning and sequencing. The primers used for screening of knockout lines were *KO*-F (5'-3') GGAGGGAGGAAG GAGCAGAA and *KO*-R (5'-3') AGAGCAAGAAGACG AGCACC. The seeds of the homozygous positive plants were used to evaluate SDW as described above, and their wild-type Nipponbare was used as control.

Data Analysis

A *t*-test or multiple comparisons was conducted using *SPSS*10.0 to detect the differences in SDW, GW, TGW and expression levels of the candidate genes between or among the tested rice accessions.

Abbreviations

BL	Backcrossing introgression line
CDS	Coding DNA Sequence

DH	Doubled haploid line
DST	Direct seeding into the trays
GST	Pre-germinated seeds were sown in trays
GSF	Pre-germinated seeds were sown in field
GW	Grain width
GWAS	Genome-wide association study
IL	Introgression line
KO line	Knockout transgenic line
LD	Linkage disequilibrium
PAV	The presence/absence variation
QTLs	Quantitative trait loci
RDP2	Rice diversity panel 2
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
SDW	Shoot dry weight
SNP	Single nucleotide polymorphism
SL	Shoot length
TGW	Thousand grain weight

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12284-024-00728-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Author Contributions

Tifeng Yang: Conceptualization, Data curation, Methodology, Investigation, Validation, Writing - original draft. Jingfang Dong: Investigation, Validation, Writing - original draft. Xijuan Xiong: Investigation, Validation. Longting Zhang: Investigation, Validation. Jian Wang: Sequence analysis. Haifei Hu: Sequence analysis. Lian Zhou: Investigation. Wu Yang: Investigation. Yamei Ma: Investigation. Hua Fu: Investigation. Jiansong Chen: Investigation. Wenhui Li: Investigation. Shuai Nie: Investigation. Ziqiang Liu: Investigation. Bin Liu: Resources, Supervision. Feng Wang: Resources, Supervision. Junliang Zhao: Data analysis, Supervision. Shaohong Zhang: Project administration, Supervision, Writing - review & editing.

Funding

This study was partially supported by the Scientific and Technological Plan of Guangzhou (201804020078), Guangdong Basic and Applied Basic Research Foundation (2022A1515012135), Guangdong Science and Technology Specialist Program (KTP20210269), Seed industry revitalization project of special fund for rural revitalization strategy in Guangdong Province (2022NPY00005), Guangdong Key Laboratory of New Technology in Rice Breeding(2023B1212060042), Project of agricultural science and technology development and resource and environmental protection management from the Special Rural Revitalization Funds of Guangdong Province (2022KJ153-01).

Data Availability

No datasets were generated or analysed during the current study.

Declarations

Ethics Approval and Consent to Participate

No applicable.

Consent for Publication

No applicable.

Competing Interests

The authors declare no competing interests.

Author details

[1]Rice Research Institute, Guangdong Key Laboratory of New Technology in Rice Breeding, Rice Engineering Laboratory, Key Laboratory of Genetics and Breeding of High Quality Rice in Southern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, Guangdong, China
[2]College of Agriculture, South China Agricultural University, Guangzhou 510642, China

Received: 12 April 2024 / Accepted: 1 August 2024
Published online: 10 August 2024

References

Abel S, Savchenko T, Levy M (2005) Genome-wide comparative analysis of the IQD gene families in Arabidopsis thaliana and *Oryza sativa*. BMC Evol Biol 5(1):1–25

Anandan A, Anumalla M, Pradhan SK, Ali J (2016) Population structure, diversity and trait association analysis in rice (*Oryza sativa* L.) germplasm for early seedling vigor (ESV) using trait linked SSR markers. PLoS ONE 11(3):0152406

Cairns JE, Namuco OS, Torres R, Simborio FA, Courtois B, Aquino GA, Johnso DE (2009) Investigating early vigour in upland rice (*Oryza sativa* L.): part II. Identification of QTLs controlling early vigour under greenhouse and field conditions. Field Crop Res 113(3):207–217

Chen K, Zhang Q, Wang C, Liu Z, Jiang Y, Zhai L, Zheng T, Xu J, Li Z (2019) Genetic dissection of seedling vigour in a diverse panel from the 3,000 rice (*Oryza sativa* L.) genome project. Sci Rep 9(1):4804

Cordero-Lara K, Kim H, Tai T (2016) Identification of seedling vigor-associated quantitative trait loci in temperate japonica rice. Plant Breed Biotech 4(4):426–440

Cui K, Peng S, Xing Y, Xu C, Yu S, Zhang Q (2002a) Molecular dissection of seedling-vigor and associated physiological traits in rice. Theor Appl Genet 105(5):745–753

Cui K, Peng S, Xing Y, Yu S, Xu C (2002b) Molecular dissection of relationship between seedling characteristics and seed size in rice. Acta Bot Sinica 44(6):702–707

Dang X, Thi T, Dong G, Wang H, Edzesi W, Hong D (2014) Genetic diversity and association mapping of seed vigor in rice (*Oryza sativa* L). Planta 239:1309–1319

Dimaano N, Ali J, Mahender A, Cruz P, Baltazar A, Diaz M, Pang Y Jr, Li BZ (2020) Identification of quantitative trait loci governing early germination and seedling vigor traits related to weed competitive ability in rice. Euphytica 216(10):159

Diwan J, Channbyregowda M, Shenoy V, Salimath P, Bhat R (2013) Molecular mapping of early vigour related QTLs in rice. Res J Biology 1:24–30

Duan P, Xu J, Zeng D, Zhang B, Geng M, Zhang G, Huang G, Huang K, Huang L, Xu, Ge S, Qian Q, Li Y (2017) Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. Mol Plant 10(5):685–694

Fauzi AR, Junaedi A, Lubis I, Ghulamahdi M, Aswidinnoor H (2021) Relationship of size and shape rice seed to early seedling vigor traits. IOP Conference Series:Earth and Environmental Science 694(1):12039

Gulshan M, Bhagirath SC (2013) The role of cultivars in managing weeds in dry-seeded rice production systems. Crop Prot 49:52–57

Guo T, Yang J, Li D, Sun K, Luo L, Xiao W, Wang J, Liu Y, Wang S, Wang H, Chen Z (2019) Integrating GWAS, QTL, mapping and RNA-seq to identify candidate genes for seed vigor in rice (*Oryza sativa* L). Mol Breed 39(6):87

Huang Y (1990) Reserach on super-high yields breeding of rice. Crops 04:1–2

Kawahara Y, Bastide M, Hamilton J, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim JS, Numa H, Itoh T, Buell C, Matsumoto T (2013) Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. Rice 6(1):4

Li X, Dong J, Zhu W, Zhao J, Zhou L (2023) Progress in the study of functional genes related to direct seeding of rice. Mol Breed 43(6):46

Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, Tian P, Cheng Z, Yu X, Zhou K, Zhang X, Guo X, Wang J, Wang H, Wan J (2017) GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat Plants 3:17403

Lu X, Niu A, Cai H, Yong Z, Liu J, Zhu Y, Zhang Z (2007) Genetic dissection of seedling and early vigor in a recombinant inbred line population of rice. Plant Sci 172:212–220

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, Xie Y, Shen R, Chen S, Wang Z, Chen Y, Guo J, Chen L, Zhao X, Dong Z, Liu Y (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant 8(8):1274–1284

McCouch S, Wright M, Tung C, Maron L, McNally K, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg A, Naredo M, Mercado S, Harrington S, Shi Y, Branchini D, Kuser-Falcão P, Leung H, Ebana K, Yano M, Eizenga G, McClung A, Mezey J (2016) Open access resources for genome-wide association mapping in rice. Nat Commun 7:10532

Rao AN, Johnson DE, Sivaprasad B, Ladha JK, Mortimer AM (2007) Weed management in direct-seeded rice. Adv Agron 93:153–255

Roy SKS, Hamid A, Giashuddin Miah M, Hashem A (1996) Seed size variation and its effects on germination and seedling vigour in rice. J Agron Crop Sci 176(2):29–82

Singh UM, Yadav S, Dixit S, Ramayya PJ, Devi MN, Raman KA, Kumar A (2017) QTL hotspots for early vigor and related traits under dry direct-seeded system in rice (*Oryza sativa* L). Front Plant sci 8:286

Steven DC, Jenneth MS (1998) BRASSINOSTEROIDS: essential regulators of plant growth and development. Annu Rev Plant Physiol Plant Mol Biol 49(1):427–451

Wang J, Yang W, Zhang S, Hu H, Yuan Y, Dong J, Chen L, Ma Y, Yang T, Zhou L, Chen J, Liu B, Li C, Edwards D, Zhao J (2023) A pangenome analysis pipeline provides insights into functional gene identification in rice. Genome Biol 24(1):19

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A Retrotransposon-m ediated gene duplication underlies morphological variation of tomato fruit. Science 319(5869):1527–1530

Xu C, Li X, Xue Y, Huang Y, Gao J, Xing Y (2004) Comparison of quantitative trait loci controlling seedling characteristics at two seedling stages using rice recombinant inbred lines. Theor Appl Genet 109(3):640–647

Xu J, Shang L, Wang J, Chen M, Fu X, He H, Wang Z, Zeng D, Zhu L, Hu J, Zhang C, Chen G, Gao Z, Zou W, Ren D, Dong G, Shen L, Zhang Q, Li Q, Guo L, Qian Q, Zhang G (2021) The Seedling Biomass 1 allele from indica rice enhances yield performance under low-nitrogen environments. Plant Biotechnol J 19(9):1681–1683

Yang J, Yang G, Yang M, Su L, Xia A, Li D, Huang C, Zhou D, Liu Y, Wang H, Chen Z, Guo T (2019) Quantitative trait Locus Analysis of seed germination and early seedling growth in rice. Front Plant sci 10:1582

Yang T, Zhou L, Zhao J, Dong J, Liu Q, Fu H, Mao X, Yang W, Ma Y, Chen L (2020) The candidate genes underlying a stably expressed QTL for low temperature germinability in rice (*Oryza sativa* L). Rice 13(1):1–15

Yang J, Guo Z, Luo L, Gao Q, Xiao W, Wang J, Wang H, Chen Z, Guo T (2021) Identification of QTL and candidate genes involved in early seedling growth in rice via high-density genetic mapping and RNA-seq. Crop J 9:360–371

Yang T, Dong J, Zhao J, Zhang L, Zhou L, Yang W, Ma Y, Wang J, Fu H, Chen J, Li W, Hu H, Jiang X, Liu Z, Liu B, Zhang S (2023a) Genome-wide association mapping combined with gene-based haplotype analysis identify a novel gene for shoot length in rice (*Oryza sativa* L). Theor Appl Genet 136(12):251

Yang T, Gu H, Yang W, Liu B, Liang S, Zhao J (2023b) Artificially selected grain shape gene combinations in Guangdong Simiao varieties of rice (*Oryza sativa* L). Rice 16(1):3

Zeng M, Yang J, Wu K, Wang H, Sun K, Chen Z, Guo T, Chen C (2021) Genome-wide association study reveals early seedling vigour-associated quantitative trait loci in indica rice. Euphytica 217:141

Zhang Z, Yu S, Yu T, Huang Z, Zhu Y (2005) Mapping quantitative trait loci (QTLs) for seedling-vigor using recombinant inbred lines of rice (*Oryza sativa* L). Field Crop Res 91(2):161–170

Zhang A, Liu C, Chen G, Hong K, Gao Y, Tian P, Peng Y, Zhang B, Ruan B, Jiang H, Guo L, Qian Q, Gao Z (2017) Genetic analysis for rice seedling vigor and fine mapping of a major QTL *qSSL1b* for seedling shoot length. Breed sci 67(3):307–315

Zhao J, Yang W, Zhang S, Yang T, Liu Q, Dong J, Fu H, Mao X, Liu B (2018) Genome-wide association study and candidate gene analysis of rice cadmium accumulation in grain in a diverse rice collection. Rice 11(1):15–61

Zhao Y, Jiang C, Rehman R, Zhang H, Li J, Li Z (2019) Genetic analysis of roots and shoots in rice seedling by association mapping. Genes Genom 41(1):95–105

Zhou L, Wang J, Yi Q, Wang Y, Zhu Y, Zhang R (2007) Quantitative trait loci for seedling vigor in rice under field conditions. Field Crop res 100(2–3):294–301

Publisher's Note