Sandhu, N., Ankush, A. P., Singh, J., Raigar, O. P., Bains, S., Jindal, T., Singh, M. P., Sethi, M., Pruthi, G., Augustine, G., Verma, V. K., Goyal, S., Kumar, A., Panwar, H., Sihag, M. K., Kaur, R., Kurup, S. and Kumar, A. 2023. Integrating Association Mapping, Linkage Mapping, Fine Mapping with RNA Seq Conferring Seedling Vigor Improvement for Successful Crop Establishment in Deep Sown Direct-Seeded Rice. *Rice.* 16, p. 46. https://doi.org/10.1186/s12284-023-00665-w

Rice

RESEARCH Open Access

Integrating Association Mapping, Linkage Mapping, Fine Mapping with RNA Seq Conferring Seedling Vigor Improvement for Successful Crop Establishment in Deep Sown Direct-Seeded Rice

Nitika Sandhu[1*†], Ade Pooja Ankush[1†], Jasneet Singh[1†], Om Prakash Raigar[1], Sutej Bains[1], Taveena Jindal[1], Mohini Prabha Singh[1], Mehak Sethi[1], Gomsie Pruthi[1], Gaurav Augustine[1], Vikas Kumar Verma[1], Shivani Goyal[1], Aman Kumar[1], Harsh Panwar[2], Manvesh Kumar Sihag[2], Rupinder Kaur[1], Smita Kurup[3] and Arvind Kumar[4,5]

Abstract

Background Ongoing large-scale shift towards direct seeded rice (DSR) necessitates a convergence of breeding and genetic approaches for its sustenance and harnessing natural resources and environmental benefits. Improving seedling vigour remains key objective for breeders working with DSR. The present study aims to understand the genetic control of seedling vigour in deep sown DSR. Combined genome-wide association mapping, linkage mapping, fine mapping, RNA-sequencing to identify candidate genes and validation of putative candidate genes were performed in the present study.

Results Significant phenotypic variations were observed among genotypes in both $F_{3:4:5}$ and $BC_2F_{2:3}$ populations. The mesocotyl length showed significant positive correlation with %germination, root and shoot length. The 881 kb region on chromosome 7 reported to be associated with mesocotyl elongation. RNA-seq data and RT-PCR results identified and validated seven potential candidate genes. The four promising introgression lines free from linkage drag and with longer mesocotyl length, longer root length, semi-dwarf plant height have been identified.

Conclusion The study will provide rice breeders (1) the pre breeding material in the form of anticipated DSR adapted introgression lines possessing useful traits and alleles improving germination under deep sown DSR field conditions (2) the base for the studies involving functional characterization of candidate genes. The development and utilization of improved introgression lines and molecular markers may play an important role in genomics-assisted breeding (GAB) during the pyramiding of valuable genes providing adaptation to rice under DSR. Our results offer a robust and reliable package that can contribute towards enhancing genetic gains in direct seeded rice breeding programs.

Keywords Direct seeded rice, Gene, Mesocotyl elongation, QTL mapping, Seedling vigor, Transcriptome

†Nitika Sandhu, Ade Pooja Ankush and Jasneet Singh are contributed equally.

*Correspondence:
Nitika Sandhu
nitikasandhu@pau.edu
Full list of author information is available at the end of the article

Background

Ongoing large-scale shift towards direct seeded rice (DSR) necessitates a convergence of breeding and genetic approaches for its sustenance and harnessing natural resources and environmental benefits. The mechanized direct-seeded rice cultivation is emerging as a natural resource-conserving, economically viable and climate smart strategy compared to the traditional puddled system of rice cultivation with huge potential to address the water-labour shortage and ensure sustainable rice production. The crop yield and resource-use efficiency mainly depend on the successful plant establishment in the field. The seed vigour defines the ability of the seeds to germinate and establish seedlings uniformly, rapidly, and robustly across variable environmental conditions. However, the poor emergence and non-uniform seedling establishment can lead to significant yield losses under DSR.

Improvement in the seedling vigour remains a breeding challenge in rice research (Finch-Savage and Bassel 2016) as it is not only important to enhance the crop yield but also can improve the crop resilience against the changing climatic conditions and the biotic impediments to rice yields. The sowing at recommended depth of 2–3 cm may lead to poor establishment and uneven crop stand due to heavy rain splashing, drought and heat stresses, more vapour pressure gradient, and predation (Kumar and Ladha 2011; Yamauchi and Winn 1996). In order to preserve the seeds and allow them to obtain moisture, deep sowing is an efficient strategy.

Deep seeding is generally not advised because it can be difficult for seedlings to penetrate the soil. Deep sowing may result in shallower crown placement, low dry matter build-up, delayed seedling emergence, and the poor seedling establishment. One limiting factor is the elongation of mesocotyl and coleoptile, during seedling germination, the mesocotyl, an embryonic structure located between coleoptile node and the basal portion of the seminal root, is crucial in pushing the shoot tip across the soil surface. The short mesocotyl of the modern, high-yielding and semidwarf rice varieties generally restricts their ability to emerge from the deeper sowing depths. Thus, varieties with longer mesocotyl are useful to overcome issues faced by direct seeding. A longer mesocotyl will minimize the sensitivity to deep sowing depth in the drill seeding and improve seedling establishment. The environmental and genetic factors both played major role in mesocotyl and coleoptile elongation. The genetic studies showed that the mesocotyl length is under the control of multiple genes and could be stably inherited (Dilday et al. 1990).

The mechanistic knowledge of the basis of regulation of seed germination, seedling establishment and vigour from deeper sowing depth is limited. No doubt, the past domestication has provided an incremental improvement in seed characteristics but there is an urgent need to understand the regulatory mechanism behind the seed characteristics that serve as adaptive responses to the seed environment if sown at deeper soil depth. Many genetic loci associated with it have been reported (Regan et al. 1992; Redona and Mackill 1996; Cui et al. 2002, Miura et al. 2001; Zhang et al. 2005; Fujino et al. 2008; Wang et al. 2010; Xie et al. 2014; Dang et al. 2014; Liu et al. 2014; Zhang et al. 2017; Zhao et al. 2019) but only a few have been validated and very few have been assessed for their impact mechanism on germination from depth. To breed rice varieties capable of germinating from depth under direct-seeding, there is need to explore the genomic regions or candidate genes for mesocotyl and coleoptile length that can be further used in molecular breeding by marker assisted selection.

The double digest restriction-site associated sequencing (ddRAD-seq) is a very flexible and cost-effective genotyping strategy providing in-depth insights into the genetic variations existing in the rice germplasm. A total of 39,137 polymorphic single nucleotide polymorphisms (SNPs) obtained through double digest restriction site associated DNA (dd-RAD) sequencing of 40 genotypes consisting of 20 rice landraces, and 20 released high yielding rice varieties used for association analyses revealed 188 stable SNPs associated with six traits across three environments. Mesocotyl and coleoptile length, along with seedling emergence and establishment are three important traits for determining high rice yields in DSR systems (Lee et al. 2017). It is critical to identify and exploit the genetic variations for seedling emergence from deep soil layer.

The next generation sequencing technologies such as dd-RAD provides SNP markers that can be utilized for the identification of alleles associated with the target traits (Poland and Rife 2012) utilizing different mapping methods such as genome-wide association study (GWAS) and the QTL mapping. Recently, a number of studies used the combined GWAS and QTL mapping approaches to identify and validate the genomic regions/marker-trait associations (MTAs) associated with complex traits (Sallam et al. 2022; Zhang et al. 2015; Gardiner et al. 2020). The GWAS is performed on a population constituting the unrelated individuals, while the QTL mapping used a biparental population (Alqudah et al. 2020; Zimmer et al. 2018).

Sufficient ground work in form of identification of donors, marker-trait association (MTAs)/QTL and segregating material as well as phenotyping platform tested on a wide scale is available. To understand the genetic control of rice seedling vigour under DSR, genome wide

association studies for multiple seedling traits in 684 accessions from the 3000 Rice Genomes (3K-RG) population in both the laboratory and in the field at three planting depths (4, 8 and 10 cm) was carried out (Menard et al. 2021). Significant MTAs/QTL for mesocotyl and coleoptile length, percentage seedling emergence and shoot biomass in this panel were identified. The potential donors to be used in genomics-assisted breeding programs have been identified. The bi-parental mapping population using the identified donors from 3K-RGP subset (3000 rice genome project) have been developed for the targeted traits. The present study designed to validate the identified genomic regions associated with seeding vigour from deep sowing depth under DSR. The study involves the genetic analysis of F_3 and F_4 mapping population derived from a cross between PR126 and IRGC 128442 for mesocotyl and coleoptile elongation/emergence from deeper soil depths and the development of near isogenic lines in background of PR126.

Results

Selection of Plant Material for the Development of Mapping Population

A sub-set of 684 re-sequenced accessions including the 266 tropical and temperate *japonica,* 266 *indica*, 131 *aus/boro,* 7 aromatic, and 14 admix subgroups from the 3K RGP representing the genetic and phenotypic diversity were selected. The phenotypic characterization of the accessions for seedling emergence traits at three planting depths (4, 8 and 10 cm) under field conditions and laboratory conditions was carried out. Genome wide association study (GWAS) showed a hotspot on the short arm chromosome 7 for mesocotyl and coleoptile length, percentage seedling emergence and shoot biomass in this panel (Menard et al. 2021). The seeds of the 6 selected accessions (Aus344, N22, Kula Karuppan, NCS237, Ashmber and IRGC 128442) were procured from IRRI and used as potential donors in genomics-assisted breeding program at PAU. The six accessions were evaluated first by drilling at 8 cm soil depth under dry direct seeded cultivation conditions under field conditions (Fig. 1) and the best donor IRGC 128442 was chosen for the development of mapping populations in background of PR126.

Population Development and Phenotypic Characterization

The complete steps involved in development and evaluation of mapping populations is presented in Fig. 1. The four traits (%germination, mesocotyl and coleoptile length, root and shoot length) potentially linked to the improved germination and seedling vigor of rice from deep sowing depth and 7 traits {days to 50% flowering (DTF), plant height (PHT), panicle length (PL), number of panicles per plant (P/P), 1000-grain weight (g), filled grains/panicle (F/P), and grain yield (GY)} associated with grain yield were qualified in the present study. The significant phenotypic variations were observed among genotypes across different depths under screenhouse and field conditions in both $F_{3:4:5}$ and $BC_2F_{2:3}$ populations (Table 1). The frequency distribution of the traits measured in the present study were determined for each year at different sowing depths under both screenhouse and field conditions. Most of the traits segregated continuously and almost fitted a normal distribution under different depths, years, and conditions (Additional file 1: Figs. S1, S2). Compared to 10 cm sowing depth, better germination was observed under 4 cm sowing depth.

The % germination and mesocotyl and coleoptile length of IRGC 128442 at 4 cm and 10 cm sowing depth were significantly higher than the recipient parent (PR126). IRGC 128442 showed early flowering compared to PR126. The plant height of IRGC 128442 was significantly higher than the PR126 across years. The mean and range values of the traits measured in both $F_{3:4:5}$ and $BC_2F_{2:3}$ populations under screenhouse and DSR field conditions in 2021 and 2022 are presented in Table 1. The correlation between different traits at different depths considering the pooled mean data is presented in Fig. 2a. The mesocotyl length at 4 cm sowing depth found to be significantly and positively correlated with root length ($r=0.85$, $p<0.001$) and shoot length at 4 cm ($r=0.93$, $p<0.001$) (Fig. 2a). Similarly, mesocotyl length at 10 cm sowing depth showed significantly positive correlation with root length ($r=0.91$, $p<0.001$) and shoot length at 10 cm ($r=0.96$, $p<0.001$) (Fig. 2a). The phenotypic correlation coefficient analysis showed that the %germination at 4 and 10 cm of sowing depth is significantly and positively correlated with mesocotyl length at 4 ($r=0.80$, $p<0.001$) and 10 cm ($r=0.69$, $p<0.001$) of sowing depth, respectively (Fig. 2a).

SNP Filtering and Linkage Map Construction

A total of 5,256,378 variants [SNPs + Indels], 4,824,612 SNPs and 4,779,408 biallelic SNPs were identified in the PR126/IRGC 128442 mapping population and in the parental lines. The filtering was carried out considering parameters such as minor allele frequency (MAF), genotype depth, breeding lines and allele missingness and parental polymorphism. After filtering, the number of SNPs reduced to 11,196. Of the 420 breeding lines, 350 $F_{3:4:5}$ breeding lines were retained after filtering the breeding lines with high genotype missingness. Next, removing 898 SNPs that showed segregation distortion ($p<0.01$), finally generated a set of 10,298 SNPs. The SNPs were finally grouped on 12 chromosomes (Fig. 2b). The genetic linkage map spanned 1561.609 cM with an average SNP interval of 0.2075 cM (Additional file 1:

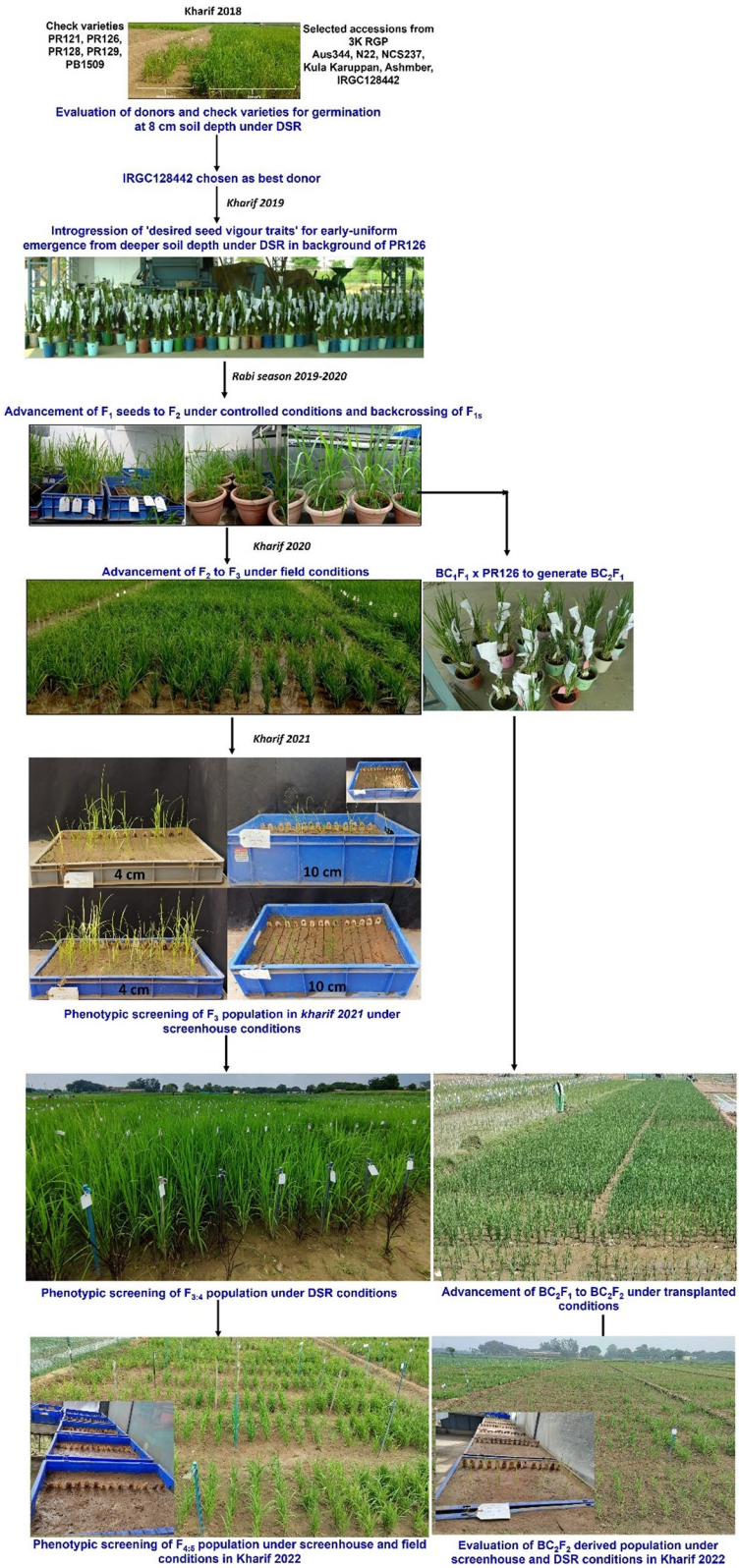

Fig. 1 The details of the steps involved in development and evaluation of mapping populations across seasons under field and screenhouse conditions

Table 1 Detailed information on analysis of variance (ANOVA), mean, range and heritability of the traits measured across years in different growing environments in both $F_{3:4:5}$ and $BC_2F_{2:3}$ mapping populations in background of PR126

Environment	Generation	Trait	Year	PR126	R146	PR126/R146		F value	*p* value	Heritability
				Mean + SE	Mean + SE	Mean + SE	Range			
Screenhouse	F_3:F_4	%germination_4 cm	2021	0	100 ± 4.2	10.11 ± 3.10	0–100	2.166	***	0.71
	F_4:F_5		2022	16.67 ± 3.31	90.30 ± 5.8	39.98 ± 8.21	0–100	2.9267	***	0.64
	F_3:F_4	%germination_10 cm	2021	0	96.97 ± 1.51	19.15 ± 4.45	0–100	35.7209	***	0.82
	F_4:F_5		2022	0	92.21 ± 2.19	17.64 ± 3.44	0–100	1.4026	*	0.77
	F_3:F_4	Mesocotyl length_4 cm	2021	0	5.64 ± 0.027	4.40 ± 0.12	0–6	3.342	**	0.62
	F_4:F_5		2022	0	4.21 ± 0.65	4.00 ± 1.52	0–6	1.6436	**	0.55
	F_3:F_4	Mesocotyl length_10 cm	2021	0	8.21 ± 0.23	2.75 ± 0.75	0–10	29.948	***	0.76
	F_4:F_5		2022	0	8.17 ± 0.94	3.52 ± 1.52	0–10	2.0515	**	0.69
	F_3:F_4	Root length_4 cm	2021	0.00	4.80 ± 0.1	4.43 ± 0.44	0–15.6	108.059	***	0.69
	F_4:F_5		2022	0.27 ± 0.08	5.27 ± 0.79	4.61 ± 1.06	0–16.2	2.1414	***	0.53
	F_3:F_4	Root length_10 cm	2021	0.00	8.60 ± 0.1	1.650 ± 0.23	0–18	76.9385	***	0.55
	F_4:F_5		2022	0.00	7.30 ± 0.75	2.28 ± 1.12	0–13.7	1.4029	*	0.49
	F_3:F_4	Shoot length_4 cm	2021	0.00	22.87 ± 0.35	10.23 ± 1.66	0–26.35	68.46	***	0.72
	F_4:F_5		2022	0.00	23.84 ± 2.36	10.69 ± 1.54	0–33.17	2.0776	**	0.62
	F_3:F_4	Shoot length_10 cm	2021	0.00	24.07 ± 1.15	15.28 ± 0.70	0–36.6	694.327	***	0.77
	F_4:F_5		2022	0.00	27.54 ± 2.80	10.42 ± 0.89	0–33.0	2.0179	**	0.77
Field	F_3:F_4	Days to 50% flowering	2021	80 ± 0.91	66.09 ± 0.91	69.36 ± 4.28	53–89	4.9315	***	0.76
	F_4:F_5		2022	77 ± 1.19	65.64 ± 1.19	67.93 ± 5.54	55–89	3.006	***	0.67
	F_3:F_4	Plant height	2021	82 ± 3.49	104.10 ± 4.13	113.09 ± 10.16	64–161	2.652	***	0.7
	F_4:F_5		2022	75 ± 3.88	91.45 ± 3.21	109.72 ± 10.51	60–155	2.351	***	0.66
	F_3:F_4	Panicles per plant	2021	8 ± 0.45	9.91 ± 0.62	9.45 ± 2.00	4–32	4.5217	***	0.61
	F_4:F_5		2022	6 ± 0.44	6.73 ± 0.48	5.92 ± 2.01	1–21	2.1306	***	0.67
	F_3:F_4	Panicle length	2021	24 ± 0.63	18.91 ± 0.54	21.86 ± 2.95	16–30	1.2973	*	0.72
	F_4:F_5		2022	19 ± 0.86	16.55 ± 0.89	20.99 ± 3.57	13–29	1.0498	*	0.65
	F_3:F_4	1000-grain weight	2021	16.23 ± 0.88	17.46 ± 0.76	13.78 ± 0.83	5.3–28.83	3.336	**	0.62
	F_4:F_5		2022	12.32 ± 0.69	13.22 ± 0.55	10.33 ± 0.49	2.6–23.58	3.016	***	0.61
	F_3:F_4	Filled grains/panicle	2021	52.0 ± 2.2	45.0 ± 1.6	38.0 ± 2.5	25.0–64.0	3.121	***	0.58
	F_4:F_5		2022	45.0 ± 2.5	40.0 ± 1.9	32.0 ± 2.8	22.0–58.0	3.332	**	0.55
	F_3:F_4	Grain yield	2021	1006 ± 87.88	1031.82 ± 92.34	1439 ± 268	217–3183	5.8333	***	0.82
	F_4:F_5		2022	509 ± 60.92	568.18 ± 54.46	717 ± 181	176–2812	4.989	***	0.70

Table 1 (continued)

Environment	Generation	Trait	Year	PR126	R146	PR126/R146		F value	*p* value	Heritability
				Mean + SE	Mean + SE	Mean + SE	Range			
Screenhouse	$BC_2F_{2:3}$	% of germination_4 cm	2022	16.67 ± 3.88	100 ± 3.6	47.97 ± 5.56	8.33–91.67	2.251	**	0.74
	$BC_2F_{2:3}$	% of germination_10 cm	2022	0	98.86 ± 1.8	37.03 ± 4.45	0.0–91.67	1.475	***	0.61
	$BC_2F_{2:3}$	Mesocotyl length_4 cm	2022	0.88 ± 0.06	4.56 ± 1.5	5.14 ± 1.15	0.0–6.0	3.344	**	0.66
	$BC_2F_{2:3}$	Mesocotyl length_10 cm	2022	0	8.8 ± 1.32	6.32 ± 0.99	0.0–9.43	2.853	***	0.69
	$BC_2F_{2:3}$	Root length_4 cm	2022	0.85 ± 0.9	5.5 ± 1.03	3.9 ± 0.92	0.0–11.36	2.14	***	0.62
	$BC_2F_{2:3}$	Root length_10 cm	2022	0	7.7 ± 1.1	2.6 ± 0.87	10.0–10.9	10.248	***	0.57
	$BC_2F_{2:3}$	Shoot length_4 cm	2022	0	20.02 ± 1.15	12.80 ± 1.51	0.0–22.27	2.575	***	0.79
	$BC_2F_{2:3}$	Shoot length_10 cm	2022	0	22.45 ± 1.03	11.96 ± 1.72	0.0–29.0	3.227	**	0.76

%germination: percentage germination, days to 50% flowering (days), plant height (cm), panicle length (cm), grain yield (kg/ha), root length (cm), shoot length (cm), mesocotyl length (cm), 1000-grain weight (gm), 4 cm: when the seed sown at 4 cm below soil surface (control), 10 cm: when the seed sown at 10 cm below soil surface (treatment)

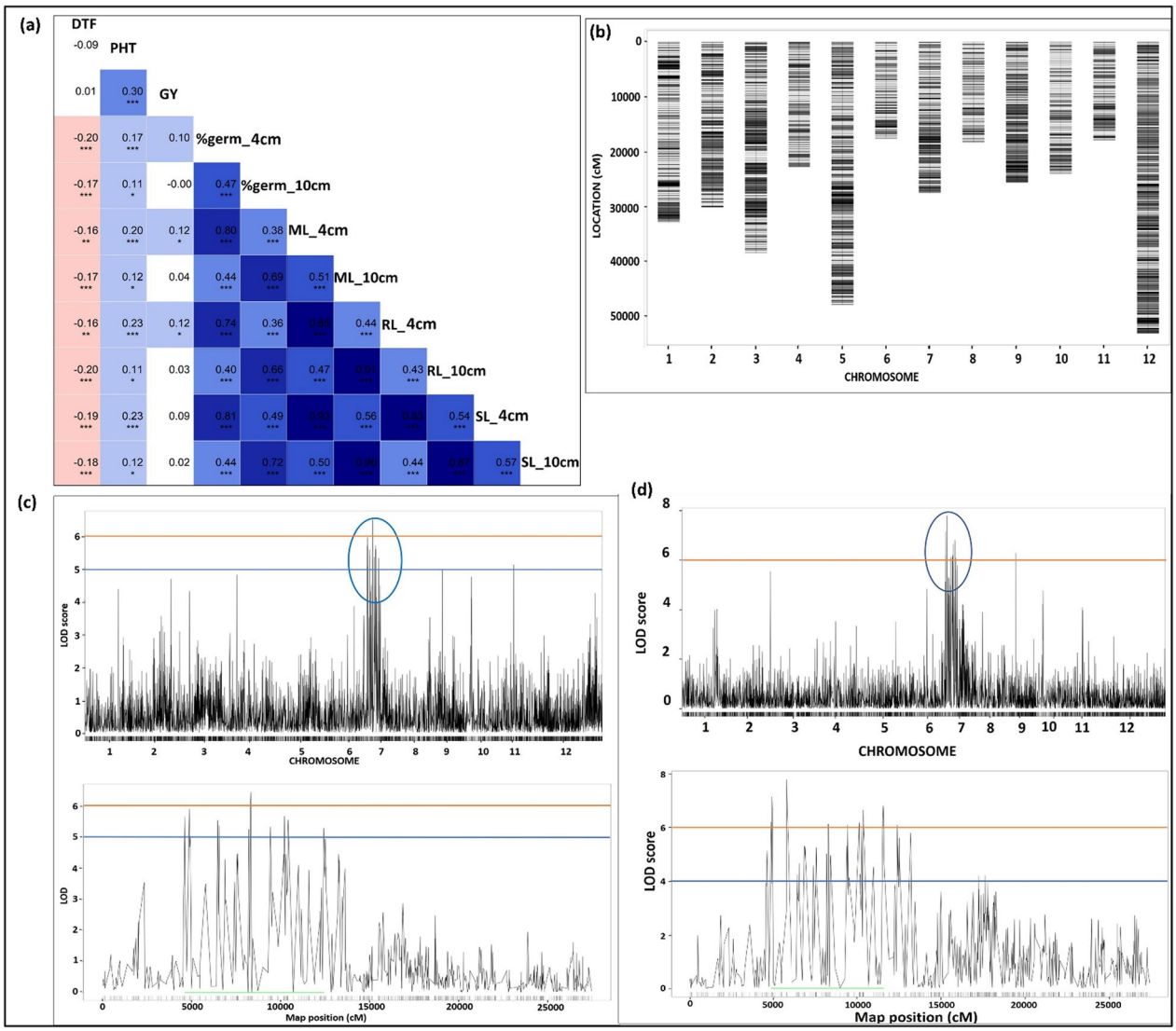

Fig. 2 (**a**) Plots of Pearson's *r*-values showing the correlation among traits measured at 4 and 10 cm of sowing depths. The blue colour indicates positive correlation and red colour indicated the negative correlation among different traits, the variation in colour intensity is representing the strength of the correlation among the traits. *Significance at < 5% level, **significance at < 1% level, ***significance at < 0.1% level (**b**) SNP density plot showing the distribution of SNPs along the chromosomes (**c**) QTL likelihood curves of LOD scores for %germination at 10 cm sowing depth considering pooled mean analysis across years on chr 7 in PR126 × IRGC 128442 mapping population (**d**) QTL likelihood curves of LOD scores for mesocotyl length at 10 cm sowing depth considering pooled mean analysis across years on chr 7 in PR126 x IRGC 128442 mapping population

Table S1). The genetic length of each linkage group (LG) ranged from 97.91 (LG11) to 172.2 cM (LG1), with an average SNP distance of 0.21–0.19 cM (Additional file 1: Table S1).

QTL Mapping and Fine Mapping

To identify the genomic regions associated with traits improving germination and seedling vigor and grain yield/yield associated traits under DSR, QTL mapping study was performed on a set of 350 $F_{3:4:5}$ breeding lines using 10,298 SNPs. In our previous genome wide association mapping, we found that the main QTL for mesocotyl and coleoptile length, percentage seedling emergence and shoot biomass in the 684 accessions selected from 3K-RGP (rice genome project) are co-located on 4.84 Mb region on the short arm of chromosome 7 (Menard et al. 2021). The earlier identified genomic regions in the 3K-RG subset using genome wide association mapping approach were validated in the present study on $F_{3:4:5}$ mapping population developed using the donor (IRGC 128442) selected from 3K-RG subset in background of popular *parmal* rice variety (PR126)

grown in Punjab. The genomic region extended from 8.83 to 15.44 Mb (6.61 Mb) on chromosome 7 showed association with % germination, mesocotyl length (Table 2, Fig. 2c, d), root and shoot length at different sowing depths across seasons (Table 2, Additional file 1: Fig. S3). The striking finding of the present study is that there is a convergence of genomic region associated with %germination and mesocotyl and coleoptile length on chromosome 7 using QTL mapping approach with the genomic region identified earlier on chromosome 7 using genome wide association mapping approach (Fig. 3). Concerning, the grain yield and yield associated traits, a genomic region spanning 8.99 Mb region on chromosome 1 reported to be associated with DTF, PHT and GY traits (Table 2, Additional file 1: Fig. S4).

The backcross generation was finally applied to fine map the genomic region associated with traits improving germination of rice under DSR. The 45 KASP markers which were designed in the 6.6 Mb genomic region on chromosome 7, were first checked for parental polymorphism. A total of 31 KASP assays showed satisfactory results on parents used to genotype the $BC_2F_{2:3}$ plants (Fig. 4a). Based on the genotypic and phenotypic analysis of 339 $BC_2F_{2:3}$ plants, the genomic region from 10131516

Table 2 Detailed information on the genomic regions associated with different traits of interest identified in the different mapping populations

Mapping population	Trait	Chr	QTL interval (bp)	QTL span (Mb)	Left flanking marker	Peak marker (bp)	Right flanking marker	LOD	R^2
$F_{3:4:5}$	%germination_4 cm	7	8906211–14769869	5.86	7:8906211	7:10322112	7:14769869	10.10	12.43
	%germination_10 cm	7	8831295–14769869	5.94	7:8831295	7:12711103	7:14769869	5.60	7.10
	Mesocotyl length_4 cm	7	8906211–15425220	6.52	7:8906211	7:11758772	7:15425220	9.90	12.19
	Mesocotyl length_10cm	7	8831295–15438381	6.61	7:8831295	7:9530368	7:15438381	7.70	9.58
	%germination_4cm	2	25640406–30681013	5.04	2:25640406	2:26193032	2:30681013	7.80	9.77
	%germination_4cm	9	8088106–8164448	0.08	9:8088106		9:8164448	6.30	7.93
	Mesocotyl length_4 cm	10	4770677–4770755	–	10:4770677	10:4770686	10:4770755	6.10	7.66
	Mesocotyl length_4 cm	11	15550458–15550470	–	11:15550458	11:15550468	11:15550470	4.70	5.95
	Root length_4cmShoot length_4cm	7	8906211–10369837	1.46	7:8906211	7:10369837	7:10369837	5.20	11.93
	Root length_10cm	3	547685–1730359	1.18	3:547685	3:777138	3:1730359	8.80	10.88
	Days to 50% floweringPlant heightGrain yield	1	25344969–34329979	8.99	1:25344969	1:28134927	1:34329979	11.80	14.41
	Days to 50% flowering	3	858902–1730359	0.87	3:858902	3:894084	3:1730359	9.30	11.51
	Days to 50% flowering	8	6986267–7377701	0.39	8:6986267	8:7232778	8:7377701	7.20	9.03
	Days to 50% flowering	9	8088106–8088106	–	–	9:8088106	–	6.20	7.82
$BC_2F_{2:3}$	%germination_4 cm	7	10923664	–	–	K_10923664	–	3.30	9.6
	%germination_10 cm	7	10131516–11112584	0.98	K_10131516	K_11112584	K_11112584	3.34	19.74
	Mesocotyl length_4 cm	7	10923664–14713452	3.79	K_10923664	K_14713452	K_14713452	8.68	11.28
	Mesocotyl length_10cm	7	10923664–14713452	0.88	K_13832487	K_14713452	K_14713452	9.00	12.80
	Shoot length_10 cm	7	10923664–14713452	3.79	K_10923664	K_14713452	K_14713452	3.30	11.20

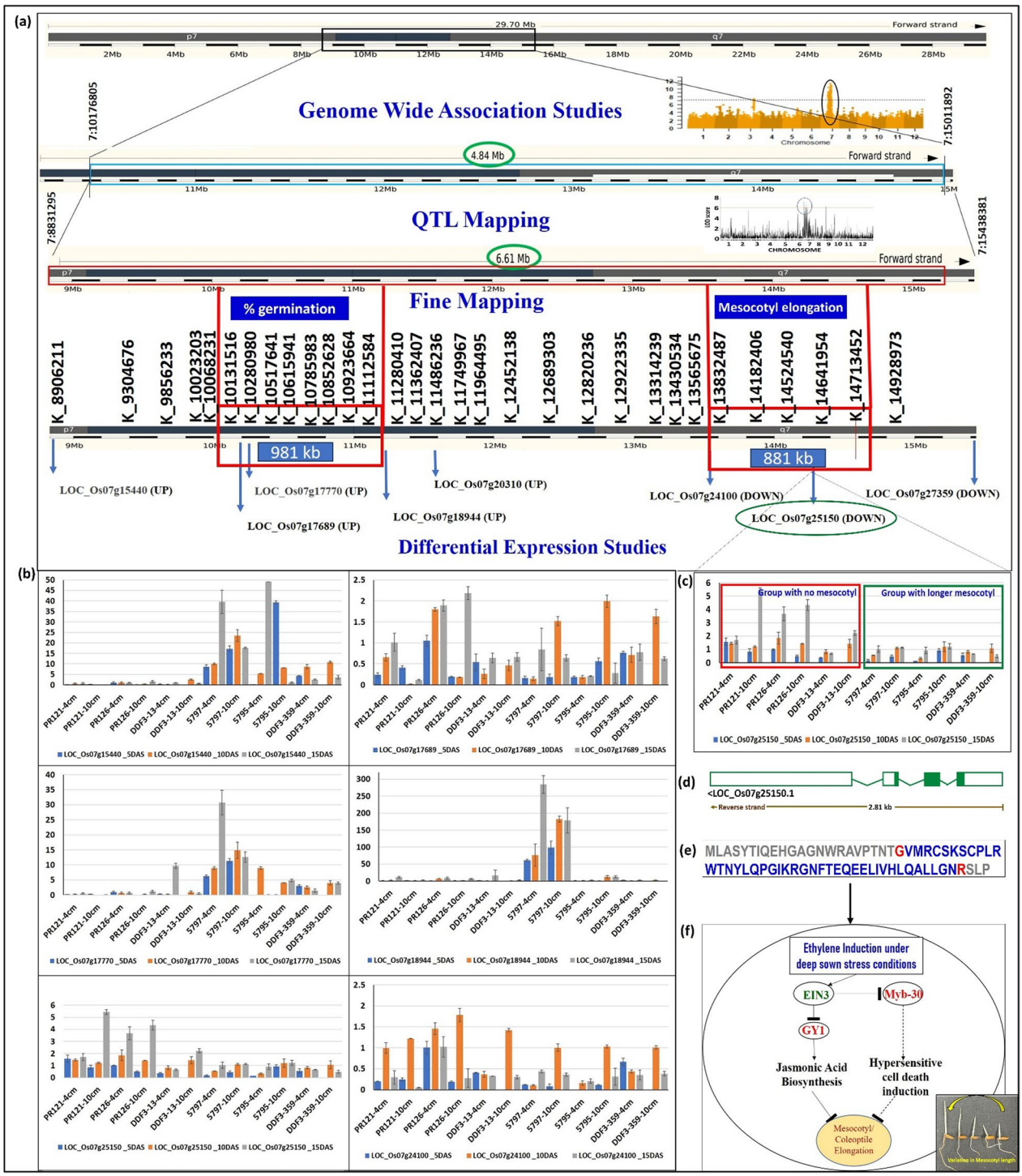

Fig. 3 (**a**) The combined approaches involving genome-wide association mapping, linkage mapping and fine mapping followed by transcriptome based RNA-seq analysis identifying genomic region and putative candidate genes associated with mesocotyl elongation on chromosome 7 under deep sown direct-seeded rice cultivation conditions (**b**) validation of 6 putative candidate genes by RT-qPCR studies (**c**) validation of *LOC_Os07g25150* on a panel of 6 genotypes constituting 3 with longer and 3 with short/no mesocotyl length by RT-qPCR studies (**d**) gene structure of the validated candidate gene *LOC_Os07g25150* (codes for Myb-30 transcription factor) (**e**) protein sequence of the validated candidate gene *LOC_Os07g25150* (codes for *Myb-30* transcription factor) (**f**) Interplay between Ethylene induced *EIN-3* and *Myb30* regulates mesocotyl elongation

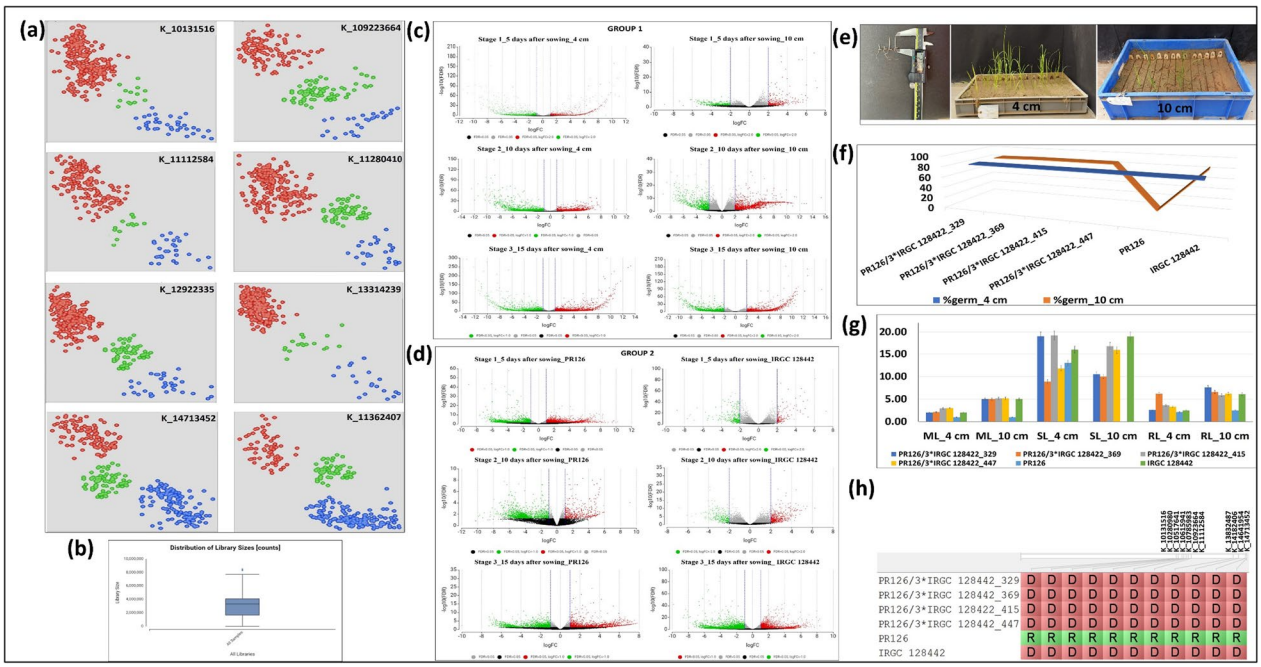

Fig. 4 (**a**) The pictorial representation of the 8 KASP assays conducted on the 339 $BC_2F_{2:3}$ plants including parents. Blue color indicates the donor allele, red color indicates the recipient allele and green color indicates the heterozygotes. (**b**) The schematic representation of the library size for 106 samples used to generate RNA-seq data (**c**) The volcano plot showing differentially expressed genes in group I (**d**) The volcano plot showing differentially expressed genes in group II (**e**) The pictorial representation of the measurement of mesocotyl length using vernier calliper, phenotypic screening of the backcross population under control (4 cm) and deep sown (10 cm) conditions (**f**) The performance of selected promising breeding in comparison to recipient parent (PR126) and donor parent (IRGC 128442) lines in terms of mesocotyl elongation (ML), root length (RL) and shoot length (SL) under control (4 cm) and deep sown (10 cm) conditions (**g**) The performance of selected promising breeding in comparison to recipient parent (PR126) and donor parent (IRGC 128442) lines in terms of %germination under control (4 cm) and deep sown (10 cm) conditions (**h**) The allelic constitution of the selected breeding lines, recipient parent (PR126) and donor parent (IRGC 128442) in the fine mapped region on chromosome 7. D: donor parent allele, R: recipient parent allele

to 11112584 Mb (981 kb) showed association with %germination and 13832487 to 14713452 Mb (880 kb) with mesocotyl length (Fig. 3) on chromosome 7.

Application of RNA-Sequencing to Filter DEGs

To identify the differentially expressed genes between IRGC 128442 and the control variety PR126, a total of 108 RNA sequencing libraries were developed for both the PR126 and IRGC 128442 sown at 2 different depths (4 cm and 10 cm), and 3-time points (5 DAS, 10 DAS, and 15 DAS) with each having 3 biological and 3 technical replicates. The library size for 106 samples was distributed from 1665409 to 7727934 with a median value of 4090419 (Fig. 4b). Two samples with a library size of 0 were discarded and were not further considered for the downstream analysis. Gene-level quantification analysis based on alignment with the reference genome annotation file was done for all the samples to generate a count table for differential expression analysis. The alignment results showed that an average of 93.45% reads for all 106 samples uniquely aligned to the *Oryza sativa* reference genome (Additional file 1: Table S2).

To compare the samples of the same rice cultivar under different treatment conditions [control (4 cm) and stress (10 cm)] and different rice cultivars (PR126 and IRGC 128442) under the same treatment at 3 different time points; a total of 2 groups were created with each having 6 subgroups. In group I, differential expression analysis between (IRGC 128442 vs PR126) was done at (Stage 1_5DAS—10 cm, Stage 2_10DAS–10 cm, Stage 3_15DAS—10 cm, Stage 1_5DAS–4 cm, Stage 2_10DAS—4 cm, Stage 3_15DAS—4 cm) subgroups, and in group II, differential expression among 2 treatments [control (4 cm) and stress (10 cm)] for the same rice cultivar was observed at (Stage 1_5DAS—IRGC 128442, Stage 2_10DAS—IRGC 128442, Stage 3_15DAS—IRGC 128442, Stage 1_5DAS—PR126, Stage 2_10DAS—PR126, Stage 3_15DAS—PR126) subgroups. The differentially expressed genes (DEGs) identified by RNA sequencing in both groups are summarized as a volcano diagram by estimating the expression of genes (Fig. 4c, d). Eight differentially expressed genes; *LOC_Os07g15440, LOC_Os07g17689, LOC_Os07g17770, LOC_Os07g18944, LOC_Os07g20310, LOC_Os07g24100, LOC_Os07g25150*

and *LOC_Os07g27359* were identified in the QTL regions reported in association and linkage mapping analysis (Fig. 3a). The expression (FPKM value) of 8 genes in 106 samples was analyzed to check whether the gene expression is consistent in three biological replicates and their corresponding 3 technical replicates. Out of the eight genes, 5 genes were upregulated at both Stage 2 and Stage 3 (*LOC_Os07g15440, LOC_Os07g17689, LOC_Os07g17770, LOC_Os07g18944* and *LOC_Os07g20310*), 1 gene was downregulated at both stage 1 and stage 2 (*LOC_Os07g27359*), 2 genes were downregulated at stage 3 (*LOC_Os07g24100*, and *LOC_Os07g25150*). The *LOC_Os07g17689, LOC_Os07g17770,* and *LOC_Os07g18944* were present in the 981 kb region extended from 10.1 to 11.1 Mb region associated with %germination on chr 7. The *LOC_Os07g24100,* and *LOC_Os07g25150* were present in the 880 kb region extended from 13.8 to 14.7 Mb region associated with mesocotyl length on chr 7.

Validation of the Selected Differentially Expressed Genes by RT-qPCR

To confirm the accuracy and reproducibility of transcriptome results of the above five DEGs (*LOC_Os07g17689, LOC_Os07g17770, LOC_Os07g18944*, *LOC_Os07g24100,* and *LOC_Os07g25150*) and 1 DEGs *LOC_Os07g15440* from the upstream region; qRT-PCR was performed using a panel of 6 samples (PR126, PR121, a line with very short mesocotyl from the PR126/IRGC 128442 mapping population, IRGC 128442, NCS237, a line with long short mesocotyl from the PR126/IRGC 128442 mapping population) (Fig. 3b, c). The results showed that the expression of six genes was significantly different in the contrasting genotypes. The qRT-PCR analysis results revealed the relative trends in the expression patterns of the seven genes that were consistent with the RNA-sequencing data.

Functional Annotation of the Validated Genes

Out of the six validated candidate genes, the potential candidate gene, the potential candidate *LOC_Os07g25150* gene transcript is 2329 bp in size, constitute four exons (Fig. 3d) and the protein sequence of 68 amino acids (Fig. 3e). The *LOC_Os07g17689* gene transcript is 2920 bp in size, constitute two exons, and the protein sequence of 60 amino acids. The transcript length of another putative candidate gene *LOC_Os07g17770* is 975 bp in length, constituting nine exons, and protein sequence of 324 amino acids. The transcript length, number of exons and protein sequence of *LOC_Os07g15440* and *LOC_Os07g18944* are 4194 and 2462 bp, 25 and 2 exons, and 1314 and 80 amino acids, respectively. The transcript length of *LOC_Os07g24100* is 5727 bp, constitutes 14 exons, and protein sequence of 1908 amino acids. The *LOC_Os07g25150* codes for *Myb 30*-related transcription factor (Fig. 3f), *LOC_Os07g17689* annotated as protein translation factor, *LOC_Os07g17770* as tryptophanyl-tRNA synthetase, *LOC_Os07g15440* as alanyl-tRNA synthetase family protein, *LOC_Os07g24100* as retrotransposon protein and the remaining others as expressed proteins.

Selection of Promising Lines

The four promising introgression lines from $BC_2F_{2:3}$ populations with improved %germination (Fig. 4e, f), longer mesocotyl length, long root length, semi-dwarf seedling height (Fig. 4g) and possessing the favourable allele combinations (Fig. 4h) have been selected for further evaluation.

Discussion

Keeping water-food security as well as farmer's livelihood in mind, technologies aimed at saving water-labor-energy such as DSR system hold significance. However, poor emergence and inadequate seedling establishment are one among the major constrains led to substantial yield loss under DSR. The rice breeding programme at PAU has taken initiative to identify and characterize potential donors for improved seedling vigor and initiated backcross breeding to transfer and consolidate these traits with the help of precise phenotypic screening as well as use of molecular markers linked to these traits.

Due to the high resolution and dense genome coverage, many genomic regions associated with complex polygenic traits can be detected in GWAS compared to the biparental QTL mapping (Sallam et al. 2022). However, the core weakness of the GWAS mapping is the high false-positive rates resulted from the population structure (Larsson et al. 2013). The consideration of the principal component analysis (PCA), the structure (Q) as a fixed effects and the kinship matrix (K) (Mahuku et al. 2016) may solve the problem resulted in the rare to few false positives in the GWAS studies. In contrast, the QTL mapping has not the problem of the presence of the population structure. To understand the genetic control of rice seedling vigour under DSR; combined GWAS, QTL mapping and RNA-sequencing to identify candidate genes and KASP markers and validation of putative candidate genes have been performed in the present study. The integration of different genomic technologies is a perfect way to dissect the genetic basis of complex traits.

The GWAS for multiple seedling traits in 684 accessions from the 3000 Rice Genomes (3K-RG) was carried out by Menard et al. (2021). The significant MTAs/QTL for mesocotyl length, coleoptile length, seedling emergence and shoot biomass were identified in 4.84 Mb on chr 7 and validated on $F_{3:4:5}$ mapping population

developed using donor (IRGC 128442, selected from 3K-RG) in background of popular *parmal* rice variety (PR126) using linkage mapping 8.83–15.44 Mb (6.61 Mb)}. The fine mapping identified that the genomic region from 10131516 to 11112584 Mb (981 kb) showed association with %germination on chromosome 7 and 13832487 to 14713452 Mb (880 kb) with mesocotyl length. The candidate genes selected from the RNA-seq data were validated using RT-PCR expression studies.

According to the results of the GWAS, linkage mapping, RNA-seq and qRT_PCR studies, six genes involved in the protein synthesis, cellular response to DNA damage, cell division, elongation, hypersensitive cell death response, ubiquitination/deubiuitination, aminoacylation, metal-ion binding, membrane, nucleus, cytoplasm, mitochondria and biological metabolism of the plant hormones were selected as the high-confidence candidate genes (Additional file 1: Fig. S5). Out of the 6 validated potential candidate genes in response to deep sowing for mesocotyl elongation, the *LOC_Os07g17689, LOC_Os07g17770,* and *LOC_Os07g18944* were present in the 981 kb region associated with %germination; the *LOC_Os07g24100,* and *LOC_Os07g25150* were present in the 880 kb region associated with mesocotyl length on chr 7. The *LOC_Os07g25150* codes for a *Myb 30*-related transcription factor protein which binds specifically to 5'-AACAAAC-3' DNA sequence (Li et al. 2009a, b) and has been validated for initiating the hypersensitive cell death response in Arabidopsis (Daniel et al. 1999). Ethylene insensitive-3 (EIN-3) is a major factor regulating mesocotyl elongation by downregulating the *GY1* gene and inhibiting the jasmonic acid pathway (Xiong et al. 2017) (Fig. 3f). EIN3 has been found to directly interact with *Myb 30* transcription factor and hider its downstream cascade (Xiao et al. 2021) (Fig. 3f). Thus, the upregulation of *EIN3* and downregulation of *Myb-30* gene regulates the activation of hypersensitive cell death response leading to decrease in the mesocotyl elongation. This antagonistic relationship between *Myb-30* and *EIN-3* in regulating mesocotyl elongation remains unexplored. Our RNA-seq data and RT-PCR results suggest similar trends between *EIN-3* and *Myb-30* in the genotype possessing longer mesocotyl. This implies a strategic role and cross-talk between ethylene-mediated signalling pathway along with *Myb-30* in regulating mesocotyl elongation. The three genes reported in the genomic region associated with mesocotyl elongation, *LOC_Os07g17689* annotated as protein translation factor, *LOC_Os07g17770* as tryptophanyl-tRNA synthetase and *LOC_Os07g15440* as alanyl-tRNA synthetase family protein and thus involved in the protein synthesis. The protein synthesis and cell cycling both are the prime physiological tasks in cell multiplication. Cell division and proliferation exerts high demand on protein synthesis. The mesocotyl continues to grow utilizing a combination of cell division in the meristems and the cell elongation in subapical regions. The *LOC_Os07g24100* annotated as retrotransposon protein. The retrotransposons play key roles in plant genetic variation, gene regulation and usually activated by stress and environmental factors.

The knockout/overexpression of identified putative candidate genes may help to better understand the mechanism behind mesocotyl elongation. The knocking of candidate gene may provide a powerful way to mesocotyl elongation improving seedling emergence, seedling vigor and providing adaptation to rice under DSR. The knockout mutation in the negative regulator of gene may lead to the development of DSR adapted rice varieties with longer mesocotyl length as well as possessing the capability to germinate from deeper soil depth. As the best of our knowledge, four genes for mesocotyl elongation, *OsGY1* (Xiong et al. 2017), *OsGSK2* (Sun et al. 2018), *OsSMAX1* (Zheng et al. 2020), and *OsPAO5* (Lv et al. 2021) have been cloned. None of the genes located on same position as the genomic region we have detected on chr 7. The seedling vigour trait reported to be associated with plant height, increase in plant height may lead to lodging under DSR. The introgression lines possessing alleles associated with longer mesocotyl length and semi-dwarf plant type have been identified. The identified promising lines may provide improvements in seedling establishment over currently available un-adapted rice varieties with water-labour saving. Farmers does not need to increase seed rate or spend more money to compensate poor seedling establishment under DSR.

Conclusion and Future Perspectives

The integrated approach involving GWAS, QTL mapping and fine mapping led to the identification of 3.79 Mb region on chromosome 7 to be associated with mesocotyl elongation. A total of 7 potential genes including 4 genes in the fine mapped region and 3 genes in the upstream have been identified. The potential candidate gene, *LOC_Os07g25150* codes for *Myb 30*-related transcription factor. The present research work will provide rice breeders (1) the pre breeding material in the form of anticipated DSR adapted introgression lines possessing useful traits and alleles improving germination under deep sown DSR field conditions (3) the base for the studies involving functional characterization and knocking of *Myb-30* gene. The next step is to understand the genetics and validate the molecular mechanism behind the cross-talk between ethylene-mediated signalling (*EIN-3*) and *Myb-30* in regulating mesocotyl elongation in rice using genome editing. We are at the stage of testing the

cross-talk hypothesis through active research and development with analytical and laboratory studies.

Materials and Methods

The plant material of this study consisted of biparental mapping population constituting 420 recombinant inbred lines derived from crossing between PR126 (an early elite *indica* rice variety) and IRGC 128442 (an accession from 3K-RGP panel with better germination ability from deep sowing depth). The phenotypic characterization was carried out in the field and screenhouse areas of School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India. The six accessions selected from 3K-RGP including Aus344, N22, Kula Karuppan, NCS237, Ashmber and IRGC 128442 along with checks (PR121, PR126, PR128, PR129, PB1509) were evaluated first by drilling at 8 cm soil depth under direct seeded cultivation conditions under field conditions in 6 m plot in 3 replications with 0.2 m row to row and 0.2 m plant to plant spacing. The best donor was selected and used to develop the bi-parental mapping population in the background of PR126. The introgression of 'desired seed vigour trait' for early-uniform emergence from deeper soil depth under dry direct seeded cultivation conditions in background of PR126 was initiated in kharif 2019. A total of 200 F_1s seeds were produced. The F_1 seeds were advanced to F_2 and F_1 seeds were also backcrossed to PR126 in 2019–2020 rabi season under controlled conditions in screenhouse. The F_2s were advanced to F_3 under field conditions in 2020 Kharif season and the BC_1F_1s were backcrossed to PR126 to generate BC_2F_1s. The $F_{3:4}$ and $F_{4:5}$ populations were evaluated under screenhouse and field conditions in 2021 and 2022 *kharif* seasons, respectively. The BC_2F_1s were advanced to BC_2F_2s and then to BC_2F_3s in 2021 and 2022 *kharif* seasons, respectively.

Phenotypic Evaluation of Mapping Populations in Screenhouse

A set of 750 breeding lines derived from the $F_{3:4}$ population PR126/IRGC128442 was screened for the emergence from deeper soil depth under controlled screenhouse conditions. Based on 0–50 percent germination (212 lines), 50–80 percent germination (68 lines), and 80–100 percent germination (96 lines), a set of 420 breeding lines were selected. The $F_{3:4,}$ and $F_{4:5}$ and control checks including PR126, PR121, PR128, PR129 and PB1509 (low/no germination from deep sowing depth) Aus344, N22, Kula Karuppan, NCS237, Ashmber and IRGC 128442 (better germination from deep sowing depth) were screened for the emergence from deeper soil depth under controlled screenhouse conditions in 2021 *kharif* season and 2022 *kharif* season, respectively. The $BC_2F_{2:3}$ population constituting 339 plants along with above mentioned checks were screened for the emergence from deeper soil depth under controlled screenhouse conditions in 2022 *kharif* season.

Screening for Emergence from Deeper Soil Depth in Screenhouse

Six seeds of all the breeding lines was sown in the plastic trays filled with soil at 4 cm, and 10 cm depth (Fig. 1). In each tray, the donor possessing the trait of interest (IRGC 128442) was kept as a positive check and PR126 was be kept as a negative check. The seed sown at 4 cm depth was act as a control for soil depth at 10 cm. The phenotypic data on percent germination at a different soil depth, uniformity of germination, days to germination, mesocotyl and coleoptile length, root and shoot length was recorded. The total number of emerged seedlings (percent germination) per breeding line was recorded daily starting from the 4 DAS (days after sowing) until the 15 DAS. The destructive sampling of all the six plants per breeding line was performed at 16 DAS to evaluate primary root and shoot traits. The total root length (RL) and shoot length (SL) was measured with a centimetre scale (cm) and mesocotyl/coleoptile length (ML) with vernier calliper for the six plants sampled per breeding line.

Phenotypic Evaluation of Mapping Populations Under Field Conditions

The $F_{3:4}$, $F_{4:5}$ and $BC_2F_{2:3}$ populations were evaluated in the field under direct seeded cultivation conditions in an augmented design in 1.5 m paired row plot maintaining row to row spacing of 20 cm and plant to plant spacing of 20 cm in 2021 Kharif season and 2022 Kharif season, respectively. The observations on early vegetative vigour, days to 50% flowering, plant height, grain yield and related traits were recorded in both the years in $F_{3:4}$, $F_{4:5}$. The data on days to 50% flowering (DTF) was recorded in the field when 50% of the plants in the paired row plot exerted their panicles. The plant height (PHT) was measured as the mean height of three random plants, from the plant base to the tip of the highest panicle during the maturity stage. The panicle length (PL) was measured on centimeter (cm) scale and the number of panicles/plant (P/P) and filled grains/panicle were counted manually. At physiological maturity, the harvested grains were threshed, dried and 1000-grain weight, grain weight per breeding line was recorded in g (gram).

Statistical Analysis

The analysis of variance (ANOVA) and the year wise mean were calculated using mixed model analysis in PBTools V 1.4.0. The Fisher's t-test was used to determine the significant difference among the breeding lines.

The correlation analysis among different traits was performed in R. v.1.1.423.

Genotyping and Variant Calling

Genomic DNA of the $F_{3:4}$ mapping population including PR126 and IRGC 128442 was prepared from 21 days old seedling. The integrity was analyzed on gel electrophoresis and then subjected to high throughput ddRAD sequencing using Illumina HiSEQ 4000. The dual enzyme (ddRAD) digested libraries were prepared and run on bioanalyzer for the library profiling. The raw sequences were generated. The paired-end sequencing and the read processing were carried out at NGB diagnostics Private Limited, New Delhi (India). The Illumina adaptor sequences were removed first and quality trimming of the adaptor-clipped reads was performed. The reads with final length of < 20 bases were rejected from further analysis. The sequencing reads were mapped against the reference genome sequence of *O. sativa v7.0* (http://rice.plantbiology.msu.edu/pub/data/Eukaryotic Projects/osativa/annotationdbs/pseudomolecules/version_7.0/all.dir/) using bwa (version 0.7.17-r1188). The read pairs with both read aligning in the expected orientation were used for further analyses. The SAMtools *version 0.1.1* (Li et al. 2009a, b) was used for the conversion of mapping files from the sam alignment format to the bam binary format. Picard software (*version 1.48*) was used to mark the duplicates marked in the sorted bam files. The generated bam files were used for the variant calling using unified Genotyper of Genome Analysis Toolkit, version 3.6 (GATK pipeline). The Vcftools (version 0.1.17) was used for the filtering of variant calls. Finally, the samples with MAF 5% and missing 10% were kept.

Linkage Map Construction, QTL, and Fine Mapping

The genotypic data of the mapping population with filtered SNP markers was used for the linkage map construction using JoinMap 4.1 software (Van Ooijen 2006) and R/qtl package (Broman and Sen 2009). The underlying algorithm used for calculating the QTL genotype probabilities and to deal with missing or partially missing genotypic data was hidden Markov model (HMM) (Baum et al. 1970) by EM algorithm (Dempster et al. 1977; Lander and Botstein 1989). A permutation test using 1000 permutation (Churchill and Doerge 1994) was used for determining the LOD threshold significance level. The QTL confidence interval (flanking markers) was determined based on 95% Bayesian Credible Interval method with an interval expanded to the nearest markers (Sen and Churchill 2001; Manichaikul et al. 2006). Based on the highest LOD score in the LG, the peak marker for each QTL was identified using R/QTL's 'find.marker' method. In addition, the mapping was also performed using QTL IciMapping software v4.1 (Meng et al. 2015). For the fine mapping, the SNPs between parents in the mapped genomic regions were identified from the already available whole genome resequencing data of the parents.

Designing of KASP Markers and KASP Assay

A total 15 KASP (Kompetitive Allele-Specific PCR) markers in the identified genomic region on chromosome 7 associated with the traits improving germination of rice under deep sowing depth were designed following protocol as mentioned in Sandhu et al. [44]. The designed 15 KASP markers were used for the polymorphic survey of the donor (IRGC 128442) and the recipient backgrounds (PR126) used in the present study. The polymorphic markers were used to fine map the identified genomic region using $BC_2F_{2:3}$ population. The KASP genotyping assays were carried out following protocol as mentioned in Sandhu et al. (2022).

Identification and Validation of Candidate Genes

RNA isolation and sequencing: Mesocotyl and coleoptile sampling for both the varieties (PR126 and IRGC 128442) at three different sowing depths (4, 8 and 10 cm) was done at every alternate day, starting from 3 days after sowing (DAS) to 15 DAS. Mesocotyl and coleoptile length was measured using a vernier calliper and the data was recorded from 10 seedlings per treatment. The appropriate stage/timepoint when there were significant variations in mesocotyl and coleoptile elongation was observed, tissue samples for transcriptomics study were collected. The mesocotyl and coleoptile samples were collected with three biological replicates and immediately frozen in liquid nitrogen. Total RNA was isolated using TRIzol reagent (Chomczynski and Sacchi 1987). Three technical replicates were prepared.

The 108 RNA samples {2 genotypes (PR126, IRGC 128442), 2 treatments (4, 10 cm) and 3 stages (5DAS, 10DAS and 15DAS), 3 biological replicates and 3 technical replicates) were used for RNA sequencing using the Illumina HiSeq 4000 platform at NGB Diagnostics Private Limited, New Delhi (India). The raw reads obtained were pre-processed by removing adaptor sequences and discarding empty reads and low-quality sequences. All clean reads obtained were assembled and aligned to the reference genome (MUS Rice Genome Annotation Project Release 7, http://rice.plantbiology.msu.edu) using BWA alignment tool followed by sorting.

RNA-Seq Based Differential Expression anaLysis

The transcriptome based differential expression analysis was done using Transcriptomics Module of Omics box (version 2.0). A count table was generated using aligned

sorted bam files. Transcripts was normalized using TMM (Trimmed mean of M values) and filtered based on CPM values less than 0.5. Pairwise Differential expression analysis was done for all the possible combinations using GLM (Likelihood ratio test) using edgeR software in omics box. Time course expression analysis was done to understand the behaviour of selected candidate genes using maSigPro software in omics box itself.

Candidate gene filtration and analysis: Transcripts having a false discovery rate (FDR) > 0.05 and log2 (fold change) < 2 was filtered and not considered for the further analysis. Functional analysis was performed on the selected candidate genes. DEGs (Differentially expressing genes) extracted from pairwise Differential expression analysis and time course expression analysis were blasted against the non-redundant protein sequence (nr v5) database using the cloud blast tool and used for GO (gene ontology) mapping and functional annotation against the Gene Ontology annotated proteins database, to obtain their functional labels through BLAST2GO software.

Validation of the Candidate Genes

The shortlisted candidate genes were validated based on their expression profiles using a panel of 6 samples (PR126, PR121, a line with very short mesocotyl from the PR126/IRGC 128442 mapping population, IRGC 128442, NCS237, a line with long short mesocotyl from the PR126/IRGC 128442 mapping population) by performing RT-qPCR measurements.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12284-023-00665-w.

Additional file 1. **Fig. S1.** The QQ plots for different seedling vigor, grain yield and yield associated traits across years. **Fig. S2.** The frequency distribution curve for different seedling vigor, grain yield and yield associated traits across years. **Fig. S3.** QTL likelihood curves of LOD scores for (A) root length (B) shoot length at 6 cm sowing depth considering pooled mean analysis across years on chr 7 in PR126 x IRGC 128442 mapping population. **Fig. S4.** QTL likelihood curves of LOD scores for (A) days to 50% flowering (DTF) (B) Plant height (PHT) and (C) grain yield (GY) under direct-seeded rice field conditions considering pooled mean analysis across years on chr 7 in PR126 x IRGC 128442 mapping population. **Fig. S5.** The role of validated candidate genes associated with mesocotyl elongation in different (A) biological processes, (B) molecular functions and (C) cellular components at three different time points (5, 10 and 15 days after sowing) and two different sowing depths (4 cm as control and 10 cm as treatment). **Table S1.** Details on the chromosome wise number of SNPs, map distance and average marker distance. **Table S2.** Summary of the alignment of 106 samples aligned to the Oryza sativa reference genome.

AUTHOR CONTRIBUTIONS

NS conceptualized this study, provided resources, compiled the results, and drafted the manuscript; APA, OPR, MP, MS, GP, and SG. conducted field and screenhouse experiments, APA and JS performed mapping studies, JS, OPR, SB, GA, and VKV conducted the RNA seq experiments, J.S. conducted the transcriptome analysis; TJ, GA, and AK conducted the RT-PCR studies; HP, and MKS helped with RT-PCR studies; NS, RK, SK, and AK contributed to the critical revision of the manuscript.

Funding

The work was compiled under projects funded by the Department of Biotechnology, Govt. of India (Grant No. BT/PR31462/ATGC/127/6/2019), BBSRC GCRF (BB/P023428/1) and SERB, DST (Grant No. WEA/2021/000003).

Availability of Data and Materials

The required data has been included in the supplementary information of the manuscript.

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

The manuscript has been approved by all authors.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author details

[1]Punjab Agricultural University, Ludhiana, Punjab 141004, India. [2]Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, Punjab 141004, India. [3]Department of Plant Science, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK. [4]International Rice Research Institute (IRRI) South Asia Regional Centre (ISARC), Varanasi, Uttar Pradesh 221106, India. [5]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana 502324, India.

Received: 30 August 2023 Accepted: 12 October 2023
Published online: 17 October 2023

References

Alqudah AM, Haile JK, Alomari DZ, Pozniak CJ, Kobiljski B, Börner A (2020) Genome-wide and SNP network analyses reveal genetic control of spikelet sterility and yield-related traits in wheat. Sci Rep 10:1–2

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41(1):164–171

Broman KW, Sen S (2009) A guide to QTL MAPPING with R/qtl, vol 46. Springer, New York

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162(1):156–159

Churchill GA, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138(3):963–971

Cui K, Peng S, Xing Y et al (2002) Molecular dissection of seedling-vigor and associated physiological traits in rice. Theor Appl Genet 105:745–753

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

Daniel X, Lacomme C, Morel JB, Roby D (1999) A novel myb oncogene homologue in Arabidopsis thaliana related to hypersensitive cell death. Plant J 20:57–66

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). J R. Stat. Soc. B 39(1):1–38

Dilday RH, Mgonja MA, Amonsilpa SA, Collins FC, Wells BR (1990) Plant height versus mesocotyl and celeoptile elongation in rice: linkage or pleitropism? Crop Sci 30(4):815–818

Finch-Savage WE, Bassel GW (2016) Seed vigour and crop establishment: extending performance beyond adaptation. J Exp Bot 67(3):567–591
Fujino K, Sekiguchi H, Matsuda Y, Sugimoto K, Ono K, Yano M (2008) Molecular identification of a major quantitative trait locus, qLTG3-1, controlling low temperature germinability in rice. Proceed Nat Acad Sci USA 105:12623
Gardiner LJ, Bansept-Basler P, El-Soda M, Hall A, O'Sullivan DM (2020) A framework for gene mapping in wheat demonstrated using the Yr7 yellow rust resistance gene. PLoS ONE 15(4):e0231157
Kumar V, Ladha JK (2011) Direct seeding of rice: recent developments and future research needs. Adv Agron 111:297–413
Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121(1):185–199
Larsson SJ, Lipka AE, Buckler ES (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. PLoS Genet 9(2):pe1003246
Lee HS, Sasaki K, Kang JW, Sato T, Song WY, Ahn SN (2017) Mesocotyl elongation is essential for seedling emergence under deep-seeding condition in rice. Rice 10(1):1–11
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009a) The sequence alignment/mapformat and SAMtools. Bioinformatics 25:2078–2079
Li L, Yu X, Thompson A, Guo M, Yoshida S, Asami T, Chory J, Yin Y (2009b) Arabidopsis MYB30 is a direct target of BES1 and cooperates with BES1 to regulate brassinosteroid-induced gene expression. Plant J 58(2):275–286
Liu LF, Lai YY, Cheng JP, Wang L, Du WL, Wang ZF, Zhang HS (2014) Dynamic quantitative trait locus analysis of seed vigor at three maturity stages in rice. PLoS ONE 9:e115732
Lv Y, Shao G, Jiao G, Sheng Z, Xie L, Hu S et al (2021) Targeted mutagenesis of POLYAMINE OXIDASE 5 that negatively regulates mesocotyl elongation enables the generation of direct-seeding rice with improved grain yield. Mol Plant 14(2):344–351
Mahuku G, Chen J, Shrestha R, Narro LA, Guerrero KVO, Arcos AL, Xu Y (2016) Combined linkage and association mapping identifies a major QTL (qRtsc8-1) conferring tar spot complex resistance in maize. Theor Appl Genet 129:1217–1229
Manichaikul A, Dupuis J, Sen S, Broman KW (2006) Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics 174(1):481–489
Menard G, Sandhu N, Anderson D, Catolos M, Hassall KL, Eastmond PJ, Kumar A, Kurup S (2021) Laboratory phenomics predicts field performance and identifies superior indica haplotypes for early seedling vigour in dry direct-seeded rice. Genomics 113(6):4227–4236
Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J 3(3):269–283
Miura K, Lin SY, Yano M, Nagamine T (2001) Mapping quantitative trait loci controlling low temperature germinability in rice (*Oryza sativa* L.). Breed Sci 51:293–299
Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. Plant Genome 5:92–102
Redona ED, Mackill DJ (1996) Mapping quantitative trait loci for seedling vigor in rice using RFLPs. Theor Appl Genet 92:395–402
Regan KL, Siddique KHM, Turner NC, Whan BR (1992) Potential for increasing early vigor and total biomass in spring wheat II Characteristics associated with early vigor. Aust J Agric Res 43:541–553
Sallam A, Eltaher S, Alqudah AM, Belamkar V, Baenziger PS (2022) Combined GWAS and QTL mapping revealed candidate genes and SNP network controlling recovery and tolerance traits associated with drought tolerance in seedling winter wheat. Genomics 114:110358
Sandhu N, Singh J, Singh G, Sethi M, Singh MP, Pruthi G, Raigar OP, Kaur R, Sarao PS, Lore JS, Singh UM (2022) Development and validation of a novel core set of KASP markers for the traits improving grain yield and adaptability of rice under direct-seeded cultivation conditions. Genomics 114(2):110269
Sen Ś, Churchill GA (2001) A statistical framework for quantitative trait mapping. Genetics 159(1):371–387
Sun S, Wang T, Wang L, Li X, Jia Y, Liu C, Huang X, Xie W, Wang X (2018) Natural selection of a GSK3 determines rice mesocotyl domestication by coordinating strigolactone and brassinosteroid signaling. Nature Commun 9(1):p2523
Van Ooijen JW (2006) JoinMap® 4 Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV Wageningen 33(10):1371
Wang Z, Wang J, Bao Y, Wang F, Zhang H (2010) Quantitative trait loci analysis for rice seed vigour during the germination stage. J Zhejiang UnivSci B (Biomed and Biotechnol):11–58
Xiao F, Zhang Y, Zhao S, Zhou H (2021) *MYB30* and *ETHYLENE INSENEITIVE3* antagonistically regulate root hair growth and phosphorus uptake under phosphate deficiency in Arabidopsis. Plant Sig & Behavior 16(7):1913310
Xie L, Tan Z, Zhou Y, Xu R, Feng L, Xing Y, Qi X (2014) Identification and fine mapping of quantitative trait loci for seed vigor in germination and seedling establishment in rice. J Integr Plant Biol 56:749
Xiong Q, Ma B, Lu X, Huang YH, He S, Yang C et al (2017) Ethylene-inhibited jasmonic acid biosynthesis promotes mesocotyl/coleoptile elongation of etiolated rice seedlings. Plant Cell 29(5):1053–1072
Yamauchi M, Winn T (1996) Rice seed vigour and seedling establishment in anaerobic soil. Crop Sci 36(3):680–686
Zhang ZH, Yu SB, Yu T, Huang Z, Zhu YG (2005) Mapping quantitative trait loci (QTLs) for seedling-vigor using recombinant inbred lines of rice (*Oryza sativa* L.). Field Crop Res 91:161–170
Zhang X, Pérez-Rodríguez P, Semagn K, Beyene Y, Babu R, López-Cruz MA, San Vicente F, Olsen M, Buckler E, Jannink JL et al (2015) Genomic prediction in biparental tropical maize populations in water-stressed and well-watered environments using low-density and GBS SNPs. Heredity 114:291–299
Zhang A, Liu C, Chen G, Hong K, Gao Y, Tian P, Peng Y et al (2017) Genetic analysis for rice seedling vigor and fine mapping of a major QTL qSSL1b for seedling shoot length. Breed Sci 67:307
Zhao Y, Jiang CH, Rehman RMA, Zhang HL, Li J, Li ZC (2019) Genetic analysis of roots and shoots in rice seedling by association mapping. Genes Genom 41(1):95
Zheng JS, Hong K, Zeng LJ, Wang L, Kang SJ, Qu MH et al (2020) Karrikin signaling acts parallel to and additively with strigolactone signaling to regulate rice mesocotyl elongation in darkness. Plant Cell 32(9):2780–2805
Zimmer CM, Ubert IP, Pacheco MT, Federizzi LC (2018) Molecular and comparative mapping for heading date and plant height in oat. Euphytica 214:101

Publisher's Note