© The Author(s), 2020. Published by Cambridge University Press on behalf of NIAB
ISSN 1479-2621

Plant Genetic Resources: Characterization and Utilization (2020) **18**(5); 382–395
doi:10.1017/S1479262120000404

Bioinformatics studies on structures, functions and diversifications of rolling leaf related genes in rice (*Oryza sativa* L.)

Md. Jahangir Alam[1], Md. Alamin[1], Most. Humaira Sultana[1], Md. Asif Ahsan[1], Md. Ripter Hossain[1], S.M. Shahinul Islam[2] and Md. Nurul Haque Mollah[1]*

[1]*Bioinformatics Laboratory, Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh and* [2]*Plant Biotechnology and Genetic Engineering Laboratory, Institute of Biological Sciences, University of Rajshahi, Rajshahi 6205, Bangladesh*

Received 27 May 2020; Accepted 13 October 2020 – First published online 11 November 2020

Abstract

Leaf morphology of crop plants has significant value in agronomy. Leaf rolling in rice plays a vital role to increase grain yield. However, collective information on the rolling leaf (RL) genes reported to date and different comparative bioinformatics studies of their sequences are still incomplete. This bioinformatics study was designed to investigate structures, functions and diversifications of the RL related genes reported till now through several studies. We performed different comparative and functional analyses of the selected 42 RL genes among 103 RL genes using different bioinformatics techniques including gene structure, conserved domain, phylogenetic, gene ontology (GO), transcription factor (TF), Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein–protein network. Exon-intron organization and conserved domain analysis showed diversity in structures and conserved domains of RL genes. Phylogenetic analysis classified the proteins into five major groups. GO and TF analyses revealed that regulation-related genes were remarkably enriched in biological process and 10 different TF families were involved in rice leaf rolling. KEGG analysis demonstrated that 14 RL genes were involved in the KEGG pathways, among which 50% were involved in the metabolism pathways. Of the selected RL proteins, 55% proteins were non-interacting with other RL proteins and OsRL9 was the most interacting RL protein. These results provide important information regarding structures, conserved domains, phylogenetic revolution, protein–protein interactions and other genetic bases of RL genes which might be helpful to the researchers for functional analysis of new candidate RL genes to explore their characteristics and molecular mechanisms for high yield rice breeding.

Keywords: diversification, gene ontology and transcription factor analysis, gene structure and function, KEGG pathway analysis, phylogenetic analysis, protein–protein interaction network, rice, rolling leaf related genes

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple food crops all over the world, particularly in Asian countries. More than half of the world's population consume rice as a primary source of food and over 3.5 billion people are solely dependent upon rice for at least 20% of their daily required calories (Khush, 2013). Under the condition of limited farmland, the main challenge to meet the food demand is increasing rice yield per unit area which leads to the improvement in overall rice yields.

Leaf is the major photosynthetic organ in rice which is directly related to biomass accumulation and grain yield

*Corresponding author. E-mail: mollah.stat.bio@ru.ac.bd

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

production (Alamin *et al.*, 2017). It has been proved that appropriate leaf shape improves photosynthesis rate resulting in the higher grain yield in rice (Wu *et al.*, 2010). As a result, leaf morphology is becoming one of the key concerns in the study of high yield plant breeding, especially high yield rice breeding (Li *et al.*, 2014). The flag leaf or uppermost three leaves were proposed to be long, narrow, moderately rolled (V-shaped), erect and thick for the hybrid rice to be super high yielding (Zhou *et al.*, 1995; Yuan, 1997; Wu, 2009). Particularly, a moderately rolling leaf (RL) of rice can increase the grain yield by improving photosynthetic efficiency and increasing stress resistance through reduced leaf transpiration during drought resulting in the compact tailing and vertical position of the leaf (Lang *et al.*, 2004; Zhang *et al.*, 2009; Wu *et al.*, 2010; Zou *et al.*, 2011). Thus, identifying moderately RL rice mutants and isolating the responsible genes for controlling leaf rolling will be advantageous for rice breeding with the desired traits (e.g. high yield) and architecture (Zhang *et al.*, 2009; Xu *et al.*, 2014; Alamin *et al.*, 2017).

To date, at least 103 genes/quantitative trait loci (QTLs) for RL trait of rice have been identified by several studies. Most of the existing studies of RL trait of rice are related to the comparison of a RL mutant with a wild type, RL genes/QTLs isolation/identification and investigation of their molecular functions, and cloning/mapping of the identified RL genes/QTLs for developing a rice mutant with desired trait and architecture, which are very time consuming and laborious. So, collective information on RL genes and their comparative analysis are very essential and helpful before starting the experiment and functional studies of a desired RL rice mutant. To our knowledge, there is no collective information on the RL genes reported to date and different comparative bioinformatics analyses of their sequences are still incomplete. Therefore, this bioinformatics study was designed to compare the structures and functions of all RL genes/QTLs, characterized so far throughout several studies, using various bioinformatics analyses including gene structure (exons/introns), conserved domain, phylogenetic, gene ontology (GO), transcription factor (TF), Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein–protein interaction network. Altogether, this study will provide collective information about RL genes and their molecular characteristics till date, and might be helpful for geneticists and rice breeders to develop the rice mutants with desired RL trait and architecture.

Materials and methods

The whole process of this study, from data collection to data analysis, has been illustrated through a schematic diagram in Fig. 1. First, we have searched and collected the RL related genes/QTLs in rice from different sources (published papers, databases and search engines) and then finally selected those genes for which locus identifiers (IDs) was available. After this, we have downloaded the sequences (genomic, CDS and protein sequences) of the finally selected genes and performed various bioinformatics analyses with the sequence data to investigate the structures, functions and diversifications of the selected RL genes. The data collection procedures for RL related genes and the bioinformatics analysis techniques used to analyse the data have been discussed in the next subsections in detail.

Data collection procedures for RL related genes

Identification of RL related genes was done in several ways in this study (Fig. 1). First, we searched and collected all the published papers which were related to RL of different varieties (japonica, indica and hybrid) of rice (*Oryza sativa* L.). From these collected papers we extracted all the RL related genes. We also used some most popular and widely used publicly available databases such as Oryzabase (http://shigen.nig.ac.jp/rice/oryzabase), GRAMENE (http://archive.gramene.org/), Rice Genome Browser (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/), Rice Annotation Project (RAP) database (RAP-DB: https://rapdb.dna.affrc.go.jp/), PubMed (https://www.ncbi.nlm.nih.gov/pubmed) and some most popular search engines (e.g. Google Scholar, Google, Yahoo and Bing) to find out the RL genes of rice along with locus ID. In most of the papers that reported locus identifiers (IDs) of RL genes, the Michigan State University (MSU) locus IDs were reported, while some papers (Fujino *et al.*, 2008; Segami *et al.*, 2012; Yang *et al.*, 2016) mentioned the RAP locus IDs instead of MSU ID for RL related genes. Also, in some papers, only the names of the RL genes were reported instead of locus ID. MSU/RAP locus IDs of the RL genes, for which locus IDs were not reported, were found using various web databases of rice including Oryzabase (http://shigen.nig.ac.jp/rice/oryzabase), GRAMENE (http://archive.gramene.org/), Rice Genome Browser (http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/) and RAP database (RAP-DB: https://rapdb.dna.affrc.go.jp/). RAP IDs of RL genes were converted to MSU IDs using the RAP database (http://rapdb.dna.affrc.go.jp).

A total of 103 RL related genes/QTLs reported in different kinds of literature and databases were identified in this study. The detailed information of the identified 103 RL genes/QTLs are available in online Supplementary Table S1. Among the total 103 RL genes/QTLs, 9, 11, 16, 9, 7, 5, 14, 3, 12, 9, 2 and 8 genes/QTLs were identified on chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12, respectively. Note that the genes *RL1*, *RL3* and *RL11(t)* have been identified on two different chromosomes (1,

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

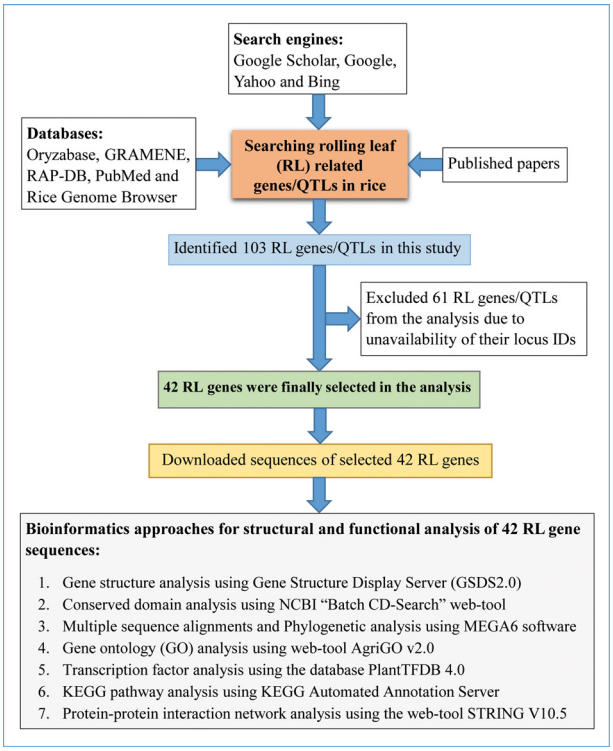


Fig. 1. Schematic diagram of the study.

12), (3, 12) and (7, 4), respectively. Among the 103 RL genes, *ND1* (Li *et al.*, 2009), *DNL3* (Shi *et al.*, 2016) and *NRL1* (Hu *et al.*, 2010; Wu *et al.*, 2010) are identical; *RL9* (Yan *et al.*, 2008) and *SLL1* (Zhang *et al.*, 2009) are identical; and *COW1* (Woo *et al.*, 2007) and *NAL7* (Fujino *et al.*, 2008) are identical (online Supplementary Table S1). So, we got a total of 96 distinct RL genes/QTLs after considering one gene from each group of similar genes. Among 96 distinct genes/QTLs, all the QTLs (19 QTLs) were excluded from the analysis since locus ID was not available for any of the QTLs. Then among the remaining 77 genes, we excluded 35 genes for which locus IDs were not found. Finally, we included only 42 RL genes in our analysis for which locus IDs were available. As a result, we confined all of our analyses with only those 42 RL genes: *OsACL1* (Li *et al.*, 2010), *OsADL1* (Hibara *et al.*, 2009), *OsAGO1a* (Li *et al.*, 2013a), *OsAGO7* (Shi *et al.*, 2007), *OsARF18* (Huang *et al.*, 2016), *OsARVL4* (Wang *et al.*, 2016), *OsAS2* (Ma *et al.*, 2009), *OsCFL1* (Wu *et al.*, 2011), *OsDCL1* (Liu *et al.*, 2005), *OsHB4* (Zhang *et al.*, 2018), *OsI_14279* (Wang *et al.*, 2017), *OsLBD3-7* (Li *et al.*, 2016), *OsLC2* (Zhao *et al.*, 2010), *OsLRRK1* (Zhou *et al.*, 2018), *OsMYB103L* (Yang *et al.*, 2014), *OsNAL1* (Qi *et al.*, 2008), *OsNAL11* (Wu *et al.*, 2016; Zhao *et al.*, 2017), *OsNAL2* (Cho *et al.*, 2013), *OsNAL3* (Cho *et al.*, 2013), *OsNAL7* (Fujino *et al.*, 2008), *OsNAL9* (Li *et al.*, 2013b), *OsNRL1* (Hu *et al.*, 2010; Wu *et al.*, 2010), *OsNRL4* (Liang *et al.*, 2016), *OsREL1* (Chen *et al.*, 2015), *OsREL2* (Yang *et al.*, 2016), *OsRFS* (Cho *et al.*, 2018; Yoshimura *et al.*, 1997), *OsRL14* (Fang *et al.*, 2012), *OsRL15* (Lee *et al.*, 2016), *OsRL16* (Liu *et al.*, 2015), *OsRL9* (Yan *et al.*, 2008), *OsRRK1* (Ma *et al.*, 2017), *OsRoc5* (Zou *et al.*, 2011), *OsSCL1* (Shen, 2011), *OsSFL1* (Alamin *et al.*, 2017), *OsSLL2* (Zhang *et al.*, 2015), *OsSND2* (Ye *et al.*, 2018), *OsSRL1* (Xiang *et al.*, 2012), *OsSRL2* (Liu *et al.*, 2016), *OsSRS5* (Segami *et al.*, 2012), *OsYABBY1* (Dai *et al.*, 2007), *OsYABBY6* (Xia *et al.*, 2017) and *OsZHD1* (Xu *et al.*, 2014). Then we downloaded the genomic sequence, coding sequence (CDS) and protein sequence from the Rice Genome Annotation Project database (http://rice.plantbiology.msu.edu/index.shtml) for each of the 42 RL genes included in our study. For convenience, we defined these 42 finally selected RL genes as 'rolling leaf genes of interest' in our study. The collected RL dataset that is used in this research is deposited at http://www.bbcba.org/softwares/Rolling_Leaf_Dataset.zip.

Bioinformatics analysis of RL genes

We analysed the genomic sequences, CDS sequences and protein sequences of the RL genes using various software packages and web-based tools. The gene structure analysis was done based on genomic sequences and CDS using the web-based bioinformatics tool Gene Structure Display Server (GSDS2.0: http://gsds.cbi.pku.edu.cn) (Hu *et al.*, 2015). In this analysis, we investigated the exon-intron structure along with the number of exons and introns of 42 RL genes. The NCBI 'Batch CD-Search' web-based tool (Marchler-Bauer and Bryant, 2004; Marchler-Bauer *et al.*, 2011; Marchler-Bauer *et al.*, 2015) in Conserved Domain Database (CDD: http://www.ncbi.nlm.nih.gov/cdd) was used to perform conserved domain analysis based on the protein sequences of the identified RL genes. In our study, MEGA V6 (Tamura *et al.*, 2013) software was used for the multiple sequence alignment and phylogenetic analysis. Multiple alignments of protein sequences of RL genes was conducted using the 'ClustalW' method (Thompson *et al.*, 1994) and was viewed in a compatible printable format using GeneDoc Version 2.6.002 software. Similar to Rahman *et al.* (2016), the maximum likelihood method was used to construct the phylogenetic tree based on the multiple aligned protein sequences.

GO, TF and KEGG pathway analysis

The web-based tool AgriGO v2.0 (http://systemsbiology.cau.edu.cn/agriGOv2/) was used for GO analysis to investigate the functional enrichments of the RL genes of interest. *Oryza sativa* japonica was selected for species and Singular Enrichment Analysis was selected as an analysis tool with

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

different settings of reference backgrounds. A GO term was considered significantly enriched among a set of genes if *P*-value (*P*) is <0.05. We constructed the heatmap of GO term versus Gene along with the bar plot of GO term versus −log10(*P*) for all significant GO terms. TFs were identified by using the PlantTFDB 4.0 database (http://planttfdb.cbi.pku.edu.cn/) (Jin *et al.*, 2017). KEGG Automated Annotation Server (KAAS: http://www.genome.jp/tools/kaas/) (Yoshizawa *et al.*, 2007) was used for the KEGG pathway enrichment analysis to get the summary of the gene pathway network.

Protein network association construction

Functional protein association networks analysis was done using the web-based resource STRING V10.5 (http://string-db.org/) (Szklarczyk *et al.*, 2017) to examine the protein–protein interactions among the RL genes. The prediction of interactions was done with the default settings of the STRING database.

Results

Summary of genomic information of identified RL genes

The comprehensive information of 42 RL genes of rice including locus ID, gene location, genomic sequence length, CDS length, number of introns and exons, predicted protein length/size, molecular weight (Mol. Wt., kDa) and isoelectric point (pI), is shown in Table 1. Among the 42 RL genes, 3, 6, 9, 3, 1, 3, 5, 1, 3, 3, 2 and 3 genes were located on chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12, respectively. The length of RL genes varied from 315 to 16638 base pairs (or nucleotides). The RL gene *OsREL1* has the minimum genomic length (315 nucleotides) and *OsADL1* has the maximum genomic length (16638 nucleotides). The RL gene *OsSLL2* has the lowest CDS length (225 nucleotides) and protein length (74 amino acids) while the gene *OsRFS* has the highest CDS length (6579 nucleotides) and protein length (2192 amino acids). Gene *OsRFS* has the highest molecular weight (243660 kDa) and *OsSLL2* has the lowest molecular weight (8408.14 kDa). The gene *OsREL1* corresponds to the minimum value of pI (4.2637) whereas the gene *OsNRL4* exhibits the maximum value of pI (11.8634).

Gene structure analysis of RL genes

The gene structure and exon-intron pattern of 42 RL genes of interest were comparatively analysed and presented in online Supplementary Fig. S1. The number of introns in the RL genes of interest ranges from 0 to 29. The gene *OsADL1* has the maximum number of exons (30) whereas the genes *OsACL1*, *OsI_14279*, *OsNAL3*, *OsNRL4*, *OsREL1*, *OsSCL1*, *OsSLL2* and *OsZHD1* have the lowest number of exon(s) (only one exon). *OsAGO1a* and *OsSRL2* have the second and third highest number of exons (23 and 19, respectively). Each of the RL genes *OsDCL1*, *OsHB4* and *OsRL16* gene has 18 exons whereas *OsRFS* has 11 exons; and the *OsLBD3-7* gene has 10 exons. Furthermore, every gene in the group (*OsNAL9*, *OsRL15* and *OsRoc5)*; the gene *OsRL9*; and each gene in the gene set (*OsRRK1*, *OsYABBY1* and *OsYABBY6*) have 9, 7 and 6 exons, respectively, in our study. Moreover, each candidate of *OsNAL1* and *OsSRL1* genes; every single gene in the set (*OsARVL4*, *OsLC2*, *OsLRRK1*, *OsNAL7*, *OsREL2* and *OsSRS5*); each of the genes in the group (*OsAGO7*, *OsARF18*, *OsMYB103L* and *OsSND2*); and each candidate of the gene cluster (*OsAS2*, *OsCFL1*, *OsNAL11*, *OsNAL2*, *OsNRL1*, *OsRL14* and *OsSFL1*) have 5, 4, 3, 2 exons, respectively. Although each of the genes *OsAS2*, *OsCFL1*, *OsNAL11*, *OsNAL2*, *OsNRL1*, *OsRL14* and *OsSFL1* has two exons and one intron, the length of the exons and introns are different from each other indicating diversity in genes structure. The genes *OsNAL1*, *OsSFL1*, *OsRL9*, *OsSRL1*, *OsYABBY1* and *OsYABBY6* contain an intron with an unusually large size (online Supplementary Fig. S1). Genes *OsADL1*, *OsNAL1*, *OsNAL7*, *OsSLL2* and *OsSRL2* have long upstream; *OsAS2*, *OsI_14279*, *OsDCL1*, and *OsSRL1* longer downstream; *OsCFL1* and *OsDCL1* do not have upstream; and *OsNAL2*, *OsNAL3*, *OsNRL4*, *OsREL1*, *OsRFS* and *OsRL9* do not have upstream and downstream. The gene structure analysis shows that the genes *OsADL1*, *OsDCL1*, *OsHB4*, *OsRRK1* and *OsSRL2* are conserved in the sense that they have short length introns (online Supplementary Fig. S1).

Domain analysis of rice RL proteins

Online Supplementary Table S2 and Fig. S2(b) display the results of the domain analysis of the 42 RL proteins identified in this study using the CDD web-tool of NCBI. Conserved domain analysis shows that the proteins OsACL1, OsNRL4, OsREL1, OsSLL2, OsSRL1, OsSRL2 and OsNAL1 do not contain any database-recognized domain in spite of having an important role in leaf rolling of rice. The RL proteins OsARVL4, OsAS2, OsCFL1, OsLRRK1, OsNAL11, OsNAL2, OsNAL3/OsWOX3, OsNAL9, OsRL14, OsRL15, OsRL16, OsRL9/OsSLL1, OsRRK1, OsSCL1, OsSFL1, OsSND2, OsSRS5, OsYABBY1 and OsYABBY6 contain one conserved domain; proteins OsAGO1a, OsARF18, OsI_14279, OsLC2, OsREL2 and OsZHD1 contain two conserved domains; OsADL1, OsLBD3-7, OsNAL7/OsCOW1, OsNRL1 and OsRoc5 proteins contain three conserved domains; OsHB4 and OsMYB103L contain

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

Table 1. Gene name, Gene location, Gene length (number of nucleotides), CDS length (number of nucleotides), No. of introns and exons, protein length (number of amino acids), Molecular Weight (Mol.Wt.) in kDa, isoelectric point (pI) of rolling leaf genes

Gene name	Gene location (Chr: CDS coordinates 5′-3′)	Gene length	CDS length	Protein length	No. of Introns (Exons)	Mol. Wt. (kDa)	pI
OsACL1	Chr04:20508616-20509616	1001	351	116	0 (1)	12,705.1	8.5047
OsADL1	Chr02:29341254-29357891	16,638	6489	2162	29 (30)	239,859	6.1645
OsAGO1a	Chr02:27341836-27330144	11,693	3249	1082	22 (23)	120,449	9.7493
OsAGO7	Chr03:19243920-19248524	4605	3147	1048	2 (3)	117,498	9.7883
OsARF18	Chr06:28586445-28590358	3914	2103	700	2 (3)	75,882.2	7.574
OsARVL4	Chr04:20321514-20319029	2486	522	173	3 (4)	19,406.1	9.9632
OsAS2	Chr01:38676728-38679707	2980	810	269	1 (2)	27,621.9	8.0637
OsCFL1	Chr02:18641888-18639136	2753	825	274	1 (2)	27,695.7	6.9898
OsDCL1	Chr03:1204839-1195075	9765	5652	1883	17 (18)	210,203	6.6399
OsHB4	Chr03:24657650-24651800	5851	2589	862	17 (18)	93,747.6	6.6202
OsI_14279	Chr03:35443615-35440975	2641	951	316	0 (1)	34,798.5	4.6582
OsLBD3-7	Chr03:32859060-32866608	7549	2151	716	9 (10)	77,448.9	6.7414
OsLC2	Chr02:2882177-2876553	5625	2250	749	3 (4)	82,634.6	7.5743
OsLRRK1	Chr06:3360524-3363523	3000	1104	367	3 (4)	40,923.5	7.366
OsMYB103L	Chr08:2951372-2948522	2851	1080	359	2 (3)	39,954.7	6.6241
OsNAL1	Chr04:31203525-31214741	11,217	1749	582	4 (5)	63,252.2	4.8546
OsNAL11	Chr07:4981656-4977967	3690	339	112	1 (2)	12,030	10.737
OsNAL2	Chr11:68344-70989	2646	705	234	1 (2)	25,921.2	8.8891
OsNAL3	Chr12:64393-65004	612	612	203	0 (1)	22,352.1	8.7607
OsNAL7	Chr03:3355041-3360485	5445	1266	421	3 (4)	45,720.3	9.1475
OsNAL9	Chr03:16994268-16997742	3475	780	259	8 (9)	28,186.7	7.9423
OsNRL1	Chr12:22607315-22602880	4436	3648	1215	1 (2)	132,161	7.9039
OsNRL4	Chr03:11127352-11126732	621	621	206	0 (1)	20,837.2	11.8634
OsREL1	Chr01:37365967-37365653	315	315	104	0 (1)	10,872.1	4.2637
OsREL2	Chr10:22202335-22198114	4222	2304	767	3 (4)	85,721.4	7.1569
OsRFS	Chr07:18638679-18625785	12,895	6579	2192	10 (11)	243,660	6.8132
OsRL14	Chr10:21997584-21995563	2022	438	145	1 (2)	16,325.9	8.6486
OsRL15	Chr01:21175744-21179966	4223	1338	445	8 (9)	50,605.7	6.4233
OsRL16	Chr09:5006931-5020013	13,083	3288	1095	17 (18)	120,570	8.7902
OsRL9	Chr09:13758215-13765052	6838	1599	532	6 (7)	54,267.3	9.4739
OsRRK1	Chr06:28941271-28943704	2434	1179	392	5 (6)	43,716.1	5.6404
OsRoc5	Chr02:27494914-27487865	7050	2415	804	8 (9)	86,047.2	5.4935
OsSCL1	Chr02:26841585-26844331	2747	2130	709	0 (1)	74,248.2	5.9887
OsSFL1	Chr10:14565737-14560601	5137	1572	523	1 (2)	56,867.7	9.6336
OsSLL2	Chr07:23217095-23220951	3857	225	74	0 (1)	8408.14	10.378
OsSND2	Chr05:28003165-28005001	1837	945	314	2 (3)	34,100.4	9.4414
OsSRL1	Chr07:143578-134162	9417	1101	366	4 (5)	39,053	6.09
OsSRL2	Chr03:10979478-10970763	8716	2970	989	18 (19)	110,441	6.6818
OsSRS5	Chr11:7960506-7963390	2885	1356	451	3 (4)	49,737	4.5682
OsYABBY1	Chr07:3221592-3229297	7706	618	205	5 (6)	23,325.5	11.2851
OsYABBY6	Chr12:26477633-26487388	9756	624	207	5 (6)	22,777.8	9.042
OsZHD1	Chr09:17705611-17703754	1858	840	279	0 (1)	29,541.1	7.4822

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

four conserved domains; protein OsAGO7 contains five conserved domains; OsDCL1 contains six conserved domains; and OsRFS contains seven conserved domains. So, the RL protein OsRFS contains the maximum number of domains (seven domains), and the protein OsARVL4, OsAS2, OsCFL1, OsLRRK1, OsNAL11, OsNAL2, OsNAL3/OsWOX3, OsNAL9, OsRL14, OsRL15, OsRL16, OsRL9/OsSLL1, OsRRK1, OsSCL1, OsSFL1, OsSND2, OsSRS5, OsYABBY1 and OsYABBY6 contain the minimum number of domains (one domain). The OsAGO1a and OsAGO7 proteins encode a common conserved domain PIWI. The RL proteins OsRL9/OsSLL1 and OsMYB103L contain one similar domain 'Myb-like DNA-binding domain'. The OsAGO7 and OsDCL1 RL proteins contain a PAZ domain. A single similar Homeobox domain was encoded by each of the proteins OsHB4, OsNAL2, OsNAL3/OsCOW1, OsRoc5 and OsZHD1. The RL protein OsMYB103L contains two SANT domains and two Myb_DNA_binding domains. OsRFS contains two CHROMO domains. Proteins OsRoc5 and OsHB4 encodes one Basic leucine zipper (bZIP) domain. One common domain YABBY belongs to both the proteins OsYABBY1 and OsYABBY6. The RL proteins OsADL1, OsARF18, OsARVL4, OsAS2, OsCFL1, OsI_14279, OsLBD3-7, OsLC2, OsLRRK1, OsNAL11, OsNAL7, OsNAL9, OsNRL1, OsREL2, OsRL14, OsRL15, OsRL16, OsRRK1, OsSCL1, OsSFL1, OsSND2 and OsSRS5 do not have any common domain.

Phylogenetic analysis of RL related genes

To investigate the evolution and diversification of RL genes, we constructed a maximum likelihood based phylogenetic tree on the alignment of full length RL proteins (Fig. 2). The phylogenetic tree showed that the 42 RL proteins clustered into five major groups (I–V) with the groups III and IV further divided into two subgroups while group V separated into four subgroups with robust bootstrap support. The RL proteins OsAS2, OsNAL1, OsSND2, OsRL9, OsCFL1 and OsNRL4 gathered into the group I, while the RL proteins OsRRK1, OsLRRK1, OsSRS5 and OsACL1 formed the group II. RL proteins OsNRL1, OsLBD3-7, OsRL14, OsNAL11 and OsRL16 constructed the subgroup IIIa while the proteins OsRoc5 and OsHB4 congregate in subgroup IIIb. Proteins OsRFS, OsRL15, OsNAL7, OsZHD1, OsNAL9, OsADL1 and OsSRL2 clustered into subgroup IVa whereas OsSFL1 and OsSRL1 grouped into subgroup IVb. The RL proteins OsREL2, OsARVL4, OsI_14279, OsDCL1 and OsSLL2; OsMYB103L and OsSCL1; OsAGO1a, OsAGO7, OsYABBY1 and OsYABBY6; and OsNAL2, OsNAL3, OsREL1, OsLC2 and OsARF18 appeared in subgroup Va, Vb, Vc and Vd, respectively. These results can be helpful for further evolutionary and functional studies of RL related genes.

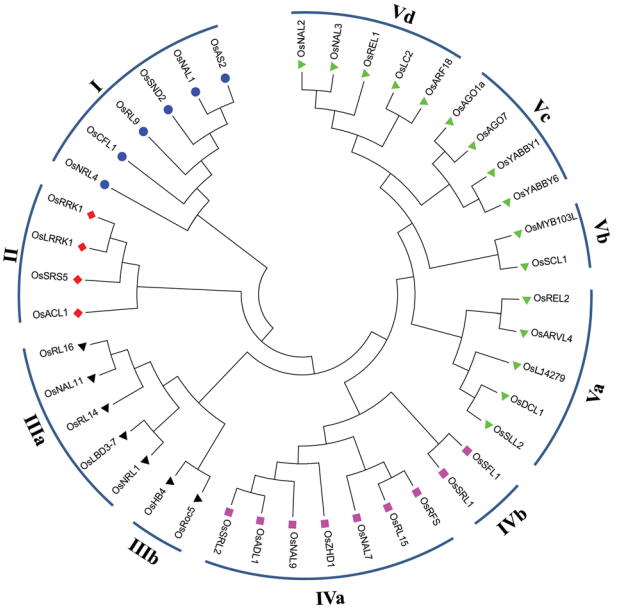


Fig. 2. Phylogenetic tree of 42 rolling leaf genes of interest identified in this study. The tree was constructed based on multiple aligned sequences by maximum likelihood method with bootstrap of 1000 in MEGA6. Multiple sequence alignment was performed using ClustalW program in MEGA6. The coloured shapes indicate different clusters of RL proteins. The roman numerals I-Vd indicate groups and subgroups of RL proteins.

GO, TF and KEGG analysis of RL genes

In order to better understand the functional involvement and to characterize the selected RL genes, different types of gene enrichment analyses were conducted. The GO analysis was performed to explain the biological importance of RL genes. The 'GO analysis toolkit' of the agriGO v2.0 database was used to perform the GO analysis. Fig. 3(a) represents the heatmap of 'GO term versus gene' along with the bar plot of –log10(P) against all significant GO terms ($P < 0.05$). All of the significant GO terms enriched among the RL genes and their description, the number in the input list and P-value are shown in online Supplementary Table S3. From the analysis result, it was observed that 38 GO terms were significantly enriched for the biological process whereas very few significant overrepresentations were found for cellular component and molecular function (Fig. 3(a) and online Supplementary Table S3). We found that the 42 RL genes clustered into three groups (Group I, Group II and Group III) based on the enriched GO terms. The genes in three groups were enriched in various biological processes. The genes in group I were mostly involved in different types of regulations including regulation of transcription (GO:0045449), regulation of nitrogen compound metabolic process (GO:0051171), regulation of cellular biosynthetic

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

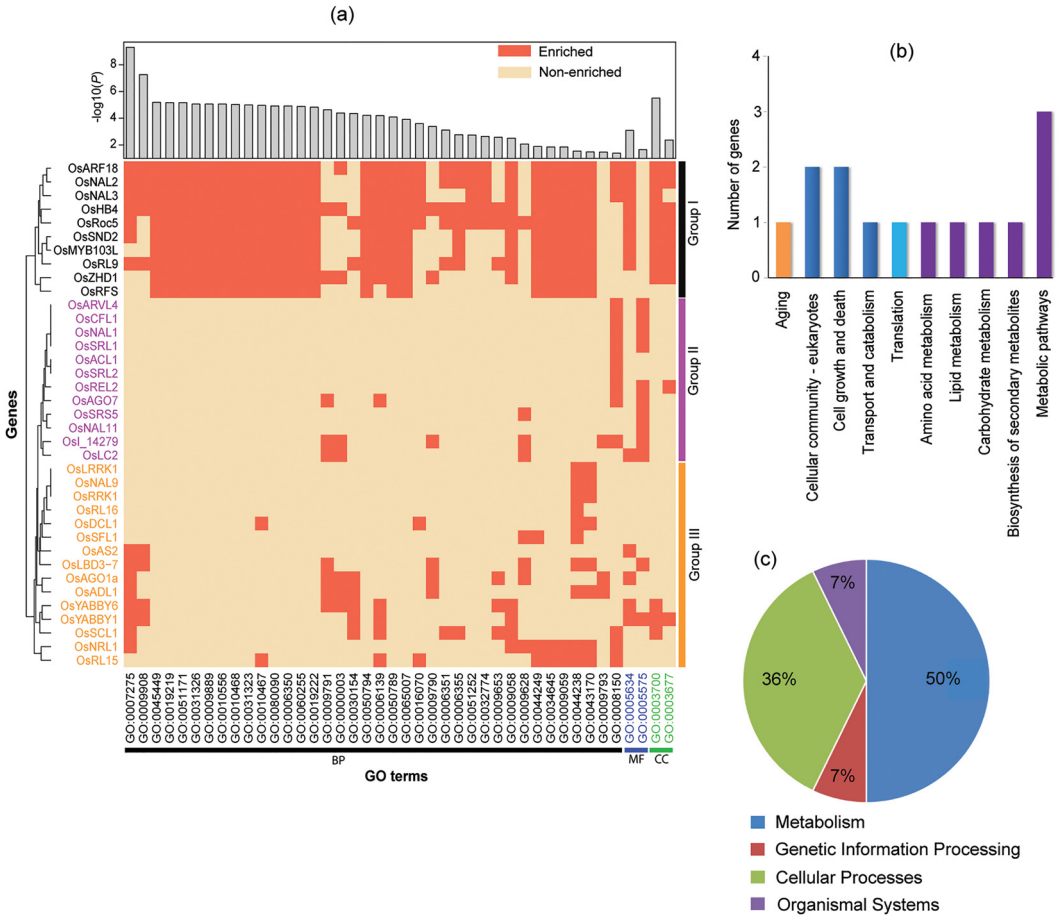


Fig. 3. The enriched GO terms for all RL genes and KEGG analysis results. (a) The GO terms indicated by BP are involved in the biological process, the GO terms indicated by MF are involved in molecular function and the GO terms indicated by CC are involved in the cellular component. (b) Bar chart for all significant KEGG pathways of all RL genes where *X*-axis represents significant KEGG pathway terms and *Y*-axis represents the number of genes. (c) Pie chart for all allocated KEGG pathways of all RL genes where the significant pathway terms are grouped and indicated by different colour.

process (GO:0031326), regulation of biosynthetic process (GO:0009889), regulation of primary metabolic process (GO: 0080090) and so on. The genes of this group were also involved in response to the abiotic stimulus (GO:0009628) and various biosynthetic processes (GO:0044249 – cellular biosynthetic process, GO:0034645 – cellular macromolecule biosynthetic process, GO:0009059 – macromolecule biosynthetic process). This observation indicated that the genes in group I are more effective and biologically more functional. The genes in group II were mainly involved in the biological process (GO:0008150) whereas genes in group III were mostly involved in multicellular organismal development (GO:0007275), cell differentiation (GO:0030154), primary metabolic process (GO:0044238) and macromolecule metabolic process (GO:0043170). The biological process 'multicellular organismal development' (GO:0007275; $P < 4.92 \times 10^{-10}$) was the most representative process that enriched for a large number of genes belonging to group I and group III.

TF analysis showed that a total of 13 genes were related to 10 different TF families (Table 2). Gene sets (*OsRoc5*, *OsHB4*), (*OsNAL3*, *OsNAL2*) and (*OsYABBY1*, *OsYABBY6*) belonged to the TF families HD-ZIP, WOX and YABBY, respectively. Genes *OsARF18*, *OsRL9*, *OsSCL1*, *OsAS2*, *OsMYB103L*, *OsSND2* and *OsZHD1* were the members of the TF families ARF, G2-like, GRAS, LBD, MYB, NAC and ZF-HD, respectively. The KEGG pathway analysis was carried out using the KEGG database to know the biological process in the RL genes of interest (Fig. 3(b) and (c)). Results showed that only 14 out of 42 RL genes were involved in the KEGG pathways (Fig. 3(b)). Of those 14 genes, 50% genes were involved in the metabolism pathway, followed by cellular processes (36%) and 7% genes were involved in both genetic information processing and organismal system in this study (Fig. 3(c)). KEGG pathway

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

Table 2. Summary of transcription factors (TFs) identified in this study

Genes name	Transcription factors	Description	Features
OsARF18	ARF	Auxin response factors	ARF are TFs that regulate the expression of auxin response genes
OsRL9	G2-like	Golden2-like	GLK proteins are members of the newly classified GARP superfamily of TFs
OsSCL1	GRAS	GRAS	The GRAS family of putative transcriptional regulators is found throughout the plant kingdom, and these proteins have diverse roles in plant development, including root development, axillary shoot development and maintenance of the shoot apical meristem
OsRoc5, *OsHB4*	HD-ZIP	Homeodomain leucine Zipper	HD-ZIP gene family are vital regulators of plant development
OsAS2	LBD	LOB Domain	The LBD genes encode a novel class of DNA-binding TFs
OsMYB103L	MYB	MYB	The encoded proteins are crucial to the control of proliferation and differentiation in a number of cell types and share the conserved MYB DNA-binding domain
OsSND2	NAC	NAM, ATAF, and CUC	NAC TFs are involved in various aspects of plant development
OsNAL3, *OsNAL2*	WOX	WUS homeobox-containing	WOX family members fulfil specialized functions in key developmental processes in plants, such as embryonic patterning, stem-cell maintenance and organ formation
OsYABBY1, *OsYABBY6*	YABBY	YABBY	The YABBY gene are expressed in a polar manner in all lateral organs produced by apical and flower meristems
OsZHD1	ZF-HD	Zinc Finger Homeodomain protein	ZF-HD proteins involved in the mesophyll-specific expression of the C4 and C3 plants

enrichment analysis (Fig. 3(b)) also showed that three RL genes functioned in the metabolic pathways; two genes functioned in the cell growth and death, cellular community-eukaryotes pathways; and one gene functioned in the pathways: biosynthesis of secondary metabolites, carbohydrate metabolism, lipid metabolism, amino acid metabolism, translation, transport and catabolism, and ageing. Online Supplementary Table S4 represented the KEGG orthologous of the RL genes of interest along with their description.

Network analysis of RL related proteins

Network analysis (Fig. 4) showed that there is no protein–protein interaction for the 23 candidate proteins OsREL2, OsNAL2, OsYABBY1, OsLRRK1, OsACL1, OsARVL4, OsLC2, OsREL1, OsSLL2, OsARF18, OsRL16, OsNRL4, OsNAL3, OsRL15, OsZHD1, OsSRL1, OsNAL1, OsI_14279, OsRFS, OsRRK1, OsRL14, OsNAL11 and OsSFL1. The RL protein OsRL9 is associated with the proteins OsRoc5, OsADL1, OsNAL7, OsAS2, OsNRL1, OsYABBY6, OsHB4 and OsAGO7. The RL proteins within each of the pairs (OsMYB103L and OsSND2), (OsSRS5 and OsNAL9), (OsRoc5 and OsCFL1), (OsRoc5 and OsADL1), (OsADL1 and OsAGO7), (OsLBD3-7 and OsAGO1a), (OsSCL1 and OsDCL1) and (OsSRL2 and OsYABBY6) are associated with each other. Proteins OsAGO1a, OsAGO7 and OsDCL1 are associated with each other. Network analysis predicted the association of OsRL9 with OsRoc5, OsADL1, OsNAL7, OsAS2, OsNRL1, OsYABBY6, OsHB4 and OsAGO7; and association of OsRoc5 with OsADL1 by text mining. The interaction between OsRoc5 and OsCFL1 was experimentally determined as well as predicted by text mining. The interaction between the RL proteins within each of the pairs (OsMYB103L and OsSND2) and (OsSRS5 and OsNAL9) was predicted by text mining and co-expression analysis. Also, the association between OsLBD3-7 and OsAGO1a was experimentally determined as well as predicted by co-expression analysis. The RL protein OsAGO1a and OsAGO7 exhibited association which was known from curated databases as well as predicted by gene text mining and protein homology. The association of OsAGO1a with OsDCL1 and OsAGO7 with OsDCL1 was experimentally determined and known from curated databases, and also predicted by text mining and co-expression analysis.

Discussion

The ultimate goal of rice breeding is the development of rice mutants with super-high yield and stress tolerance

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

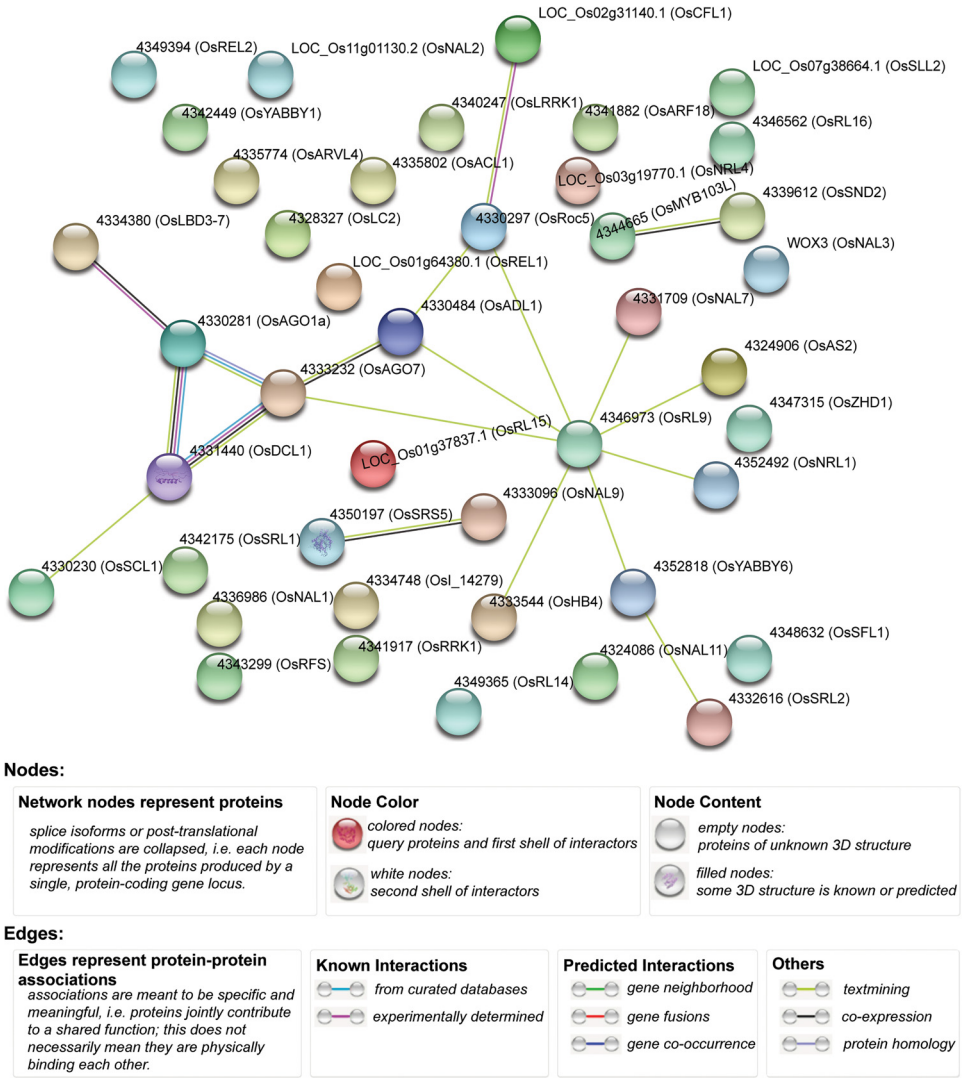


Fig. 4. Protein–protein interaction network of 42 rolling leaf genes of interest. The different methods of prediction of interactions have been indicated with the different coloured connective lines. Protein–protein interaction network was built using the web-based tool STRING V10.5 (http://string-db.org/).

traits. In rice, appropriate leaf rolling is considered to be an important agronomic element, particularly, a moderate leaf rolling is regarded as a crucial phenotypic trait of the ideal rice plant as it is important for increasing grain yield (Yuan, 1997; Lang *et al.*, 2004; Wu, 2009; Zhang *et al.*, 2009). However, to the best of our knowledge, there is no study where all the RL genes/QTLs and their genomic information have been put together along with their genome-wide comparative analysis from the different bioinformatics point of view. In this study, we listed up 103 RL genes/QTLs along with their genomic information reported in various studies till date and performed various comparative analyses from different bioinformatics standpoints with the RL genes of interest.

We selected 42 genes as the genes of interest in our analysis among the total 103 identified RL genes/QTLs due to the availability of their locus IDs in this study (Table 1). The maximum number of RL genes was found on chromosome 3 and minimum genes were found on chromosomes 5 and 8 among the 42 RL genes. The RL gene *OsADL1* has the longest genomic length and *OsREL1* has the shortest genomic length. *OsSLL2* has the lowest value of CDS and protein length while the gene *OsRFS* has the highest value of CDS and protein length. *OsRFS* has the highest molecular weight and *OsSLL2* has the lowest molecular weight. The gene *OsREL1* corresponds to the minimum value of pI whereas the gene *OsNRL4* exhibits the maximum value of pI. The comparative gene structure analysis revealed that some RL genes contained only one exon, some genes did not have upstream and some genes did not contain any upstream and downstream (online Supplementary Fig. S1). Most of the genes (60%) contained 0 to 3 introns.

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

Although some genes contained an equal number of exons and introns the length of exons and introns were different from each other indicating structural diversity. From the gene structure analysis (online Supplementary Fig. S1) it can be inferred that RL genes are diverse in their structures from each other in terms of conserveness and structure of exons and introns.

Domain analysis revealed that very few of the RL proteins contain the domains of the same family. Both the RL genes proteins OsNAL2 and OsNAL3 contained only one and identical domain named 'Homeobox domain' (online Supplementary Table S2 and Fig. S2(b)). According to Cho *et al.* (2013), the proteins NAL2 and NAL3 were paralogs and encoded an identical homeobox 3A (*OsWOX3A*) protein which supported our findings. Both the proteins OsYABBY1 and OsYABBY6 contained only one domain (YABBY domain) which is identical in both proteins (online Supplementary Table S2 and Fig. S2(b)). As reported by Toriba *et al.* (2007), both the proteins OsYABBY1 and OsYABBY6 belonged to the YABBY family which is in favour of our findings. Most of the RL proteins contained the domains of various families indicating the diversity of RL proteins. This study showed that RL proteins were diverse in terms of domains they contained and their domain structure. Based on the phylogenetic analysis of 42 RL protein sequences, we grouped the RL proteins into five major groups. Again groups III, IV and V were divided into two, two and four subgroups, respectively. We found that both the RL proteins OsNAL2 and OsNAL3 belonged to same subgroup Vd, and the proteins OsYABBY1 and OsYABBY6 belonged to the same subgroup Vc. These findings are supported by Cho *et al.* (2013) and Toriba *et al.* (2007).

To explore functional similarity and characterize a set of genes various enrichment analyses, such as GO and KEGG analyses, were performed. The biological process 'multicellular organismal development' (GO:0007275; $P < 4.92E\text{-}10$) was found as the most representative process that enriched for a large number of genes that belonged to groups I and III. The GO term GO:0007275 (multicellular organismal development) was found as the most significant. From the heatmap of the GO terms against gene, three distinct clusters of genes were observed and the genes that belonged to different clusters were involved in different biological processes. We found that *OsNAL2* and *OsNAL3* are involved in the regulation of the biosynthetic process, which is similar to the findings by Cho *et al.* (2016). TFs have significant functions for controlling plant propagation, maturation and react to unfavourable situations including drought, chill, salinity and high temperature (Zhang *et al.*, 2013). For example, myb-protein encoding genes have been reported to function as regulators of cell differentiation (Kirik and Bäumlein, 1996) and NAC TFs involved in various mechanisms include developmental process, flower creation, seed growth, hormone signalling, responses to abiotic stress and ageing in plants (Nuruzzaman *et al.*, 2010). TF analysis showed that 10 different TFs families (such as HD-ZIP, WOX, ARF, G2-like, GRAS, LBD, MYB, NAC and ZF-HD) involved in leaf rolling, where all of these TFs were responsible for the different aspects of plant developments (Table 2). Metabolomics plays a significant role in fundamental plant biology and applied biotechnology. Several current studies have found that the analysis of the metabolite composition of mutants can assist the assignment of functions to genes (Schauer and Fernie, 2006). Results of KEGG pathway analysis showed that half of the RL genes were involved in the metabolism pathway including biosynthesis of secondary metabolites, carbohydrate metabolism, lipid metabolism, amino acid metabolism and metabolic pathways (Fig. 3(b) and (c)). According to Li *et al.* (2017), the proteins responsible for leaf rolling in rice are involved in different pathways including biosynthesis of phenylpropanoid, phenylalanine metabolism, and sucrose and starch metabolism. As reported by Chen *et al.* (2018), RL proteins in *Brassica napus* are involved in different pathways including 'metabolic pathways', 'biosynthesis of secondary metabolites' and 'starch and sucrose metabolism'. So, the results of previous studies of RL support our finding.

The network analysis was used for investigating the association among the 42 RL proteins. Results showed that the RL protein OsRL9 was associated with a maximum number of proteins (9 proteins). We found that the proteins OsAGO1a, OsAGO7 and OsDCL1 were associated with each other. Protein homology and text mining predicted the association between the proteins OsAGO1a and OsAGO7 which was also known from the curated databases. All the four methods (experimentally determination, determination from curated databases, prediction using text mining and prediction by co-expression analysis) identified the association between the genes of each of the pairs of RL proteins (OsAGO1a and OsDCL1) and (OsAGO7 and OsDCL1). The protein OsRoc5 was associated with OsCFL1 and OsADL1; OsMYB103L was associated with OsSND2; and OsLBD3-7 was associated with OsAGO1a. Chen *et al.* (2019) reported direct interaction between the proteins OsNAL9 and OsNRL3 whereas due to the unavailability locus ID for the gene *OsNRL3* investigation of the interaction between the proteins OsNAL9 and OsNRL3 was not possible. However, in this study, we found the interaction of OsNAL9 with OsSRS5. We found 23 RL proteins (OsREL2, OsNAL2, OsYABBY1, OsLRRK1, OsACL1, OsARVL4, OsLC2, OsREL1, OsSLL2, OsARF18, OsRL16, OsNRL4, OsNAL3, OsRL15, OsZHD1, OsSRL1, OsNAL1, OsI_14279, OsRFS, OsRRK1, OsRL14, OsNAL11 and OsSFL1) which were not associated with any other RL proteins. This indicates that more than 50% of the RL proteins of interest are not associated with each other.

Altogether, this is the first study where we have studied and put the genetic information of all identified RL genes/

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

QTLs till now and their genomic information along with their genome-wide comparative analysis from the different bioinformatics point of view. This study will enable the rice breeders and researchers to get collective information on RL genes/QTLs along with their comparative analysis results. As a result, the rice breeders will be able to develop a super-high-yield rice mutant with moderate leaf rolling and desired architecture.

Conclusion

Our present study identified 103 RL genes/QTLs in rice genome characterized by several studies to date. Among 103 RL genes/QTLs, 42 genes, for which locus IDs were available, were finally selected in our analysis. This study provides a comparative genomic analysis of the selected 42 RL genes to explore their structures, functions and diversifications using different bioinformatics techniques. Gene structure analysis shows that the selected RL genes are diverse in structures in terms of intron-exon. Domain analysis reveals that the RL proteins contain different types of domains except for some proteins indicating their functional diversification. Phylogenetic analysis has clustered the RL proteins into five major groups with groups III, IV and V divided into some subgroups. GO analysis indicates that a total of 42 significant GO terms enriched, TF analysis shows 10 different TFs families and KEGG analysis shows that 14 genes are involved in the KEGG pathways from all RL genes. Protein–protein interaction network shows that 23 RL proteins are not associated with any other RL proteins and the protein OsRL9 is associated with maximum proteins. Therefore, we may conclude that the RL genes have different types of functions for controlling leaf rolling in rice. These results provide insight into the structures, predicted functions, evolutionary information and diversifications of RL related genes in rice which will be a basis for further functional analysis of selected candidate genes. The application of machine learning methods has been booming in the past several years (Khatun *et al.*, 2019a, b; Hasan *et al.*, 2020a, b, c, d, e; Mosharaf *et al.*, 2020), prompting researchers to design machine learning architectures to predict diverse plant species. As for future studies on structures, functions and diversifications of plant genes, machine learning methods will be employed to propel the RL related genes prediction to the next level.

Supplementary material
To view supplementary material for this article, please visit https://doi.org/10.1017/S1479262120000404.

Acknowledgements

We would like to thank the learned reviewers whose valuable comments and suggestions helped to strengthen this paper. Also we would like to acknowledge HEQEP subproject (CP-3603,W2, R3), and Bioinformatics Laboratory, Department of Statistics, University of Rajshahi, for providing access to use all the required resources during this research work.

References

Alamin M, Zeng D-D, Qin R, Sultana MH, Jin X-L and Shi C-H (2017) Characterization and fine mapping of *SFL*1, a gene controlling *Screw Flag Leaf* in rice. *Plant Molecular Biology Reporter* 35: 491–503.

Chen Q, Xie Q, Gao J, Wang W, Sun B, Liu B, Zhu H, Peng H, Zhao H and Liu C (2015) Characterization of *Rolled and Erect Leaf* 1 in regulating leave morphology in rice. *Journal of Experimental Botany* 66: 6047–6058.

Chen W, Wan S, Shen L, Zhou Y, Huang C, Chu P and Guan R (2018) Histological, physiological, and comparative proteomic analyses provide insights into leaf rolling in *Brassica napus*. *Journal of Proteome Research* 17: 1761–1772.

Chen W, Sheng Z, Cai Y, Li Q, Wei X, Xie L, Jiao G, Shao G, Tang S, Wang J and Hu P (2019) Rice Morphogenesis and Chlorophyll Accumulation Is Regulated by the Protein Encoded by *NRL*3 and Its Interaction With *NAL*9. *Frontiers in Plant Science* 10:175.

Cho SH, Yoo SC, Zhang H, Pandeya D, Koh HJ, Hwang JY, Kim GT and Paek NC (2013) The rice *narrow Leaf*2 and *narrow Leaf*3 loci encode WUSCHEL-related homeobox 3A (OsWOX3A) and function in leaf, spikelet, tiller and lateral root development. *New Phytologist* 198: 1071–1084.

Cho S-H, Kang K, Lee S-H, Lee I-J and Paek N-C (2016) OsWOX3A is involved in negative feedback regulation of the gibberellic acid biosynthetic pathway in rice (*Oryza sativa*). *Journal of Experimental Botany* 67: 1677–1687.

Cho SH, Lee CH, Gi E, Yim Y, Koh HJ, Kang K and Paek NC (2018) The rice rolled fine striped (RFS) CHD3/Mi-2 chromatin remodeling factor epigenetically regulates genes involved in oxidative stress responses during leaf development. *Frontiers in Plant Science* 9: 364.

Dai M, Zhao Y, Ma Q, Hu Y, Hedden P, Zhang Q and Zhou DX (2007) The rice *YABBY*1 gene is involved in the feedback regulation of gibberellin metabolism. *Plant Physiology* 144: 121–133.

Fang L, Zhao F, Cong Y, Sang X, Du Q, Wang D, Li Y, Ling Y, Yang Z and He G (2012) Rolling-leaf14 is a 2OG-Fe (II) oxygenase family protein that modulates rice leaf rolling by affecting secondary cell wall formation in leaves. *Plant Biotechnology Journal* 10: 524–532.

Fujino K, Matsuda Y, Ozawa K, Nishimura T, Koshiba T, Fraaije MW and Sekiguchi H (2008) *NARROW LEAF 7* controls leaf shape mediated by auxin in rice. *Molecular Genetics and Genomics* 279: 499–507.

Hasan MM, Basith S, Khatun MS, Lee G, Manavalan B and Kurata H (2020a) Meta-i6mA: an interspecies predictor for identifying DNA N6-methyladenine sites of plant genomes by exploiting informative features in an integrative machine-learning framework. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbaa202

Hasan MM, Manavalan B, Khatun MS and Kurata H (2020b) i4mC-ROSE, a bioinformatics tool for the identification of DNA N4-methylcytosine sites in the Rosaceae genome. *International Journal of Biological Macromolecules* 157: 752–758.

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

Hasan MM, Manavalan B, Shoombuatong W, Khatun MS and Kurata H (2020c) i4mC-Mouse: improved identification of DNA N4-methylcytosine sites in the mouse genome using multiple encoding schemes. *Computational and Structural Biotechnology Journal* 18: 906–912.

Hasan MM, Manavalan B, Shoombuatong W, Khatun MS and Kurata H (2020d) i6mA-Fuse: improved and robust prediction of DNA 6mA sites in the Rosaceae genome by fusing multiple feature representation. *Plant Molecular Biology* 103: 225–234.

Hasan MM, Schaduangrat N, Basith S, Lee G, Shoombuatong W and Manavalan B (2020e) HLPpred-Fuse: improved and robust prediction of hemolytic peptide and its activity by fusing multiple feature representation. *Bioinformatics (Oxford, England)* 36: 3350–3356.

Hibara KI, Obara M, Hayashida E, Abe M, Ishimaru T, Satoh H, Itoh JI and Nagato Y (2009) The *ADAXIALIZED LEAF*1 Gene functions in leaf and embryonic pattern formation in rice. *Developmental Biology* 334: 345–354.

Hu J, Zhu L, Zeng D, Gao Z, Guo L, Fang Y, Zhang G, Dong G, Yan M and Liu J (2010) Identification and characterization of *NARROW AND ROLLED LEAF* 1, a novel gene regulating leaf morphology and plant architecture in rice. *Plant Molecular Biology* 73: 283–292.

Hu B, Jin J, Guo A-Y, Zhang H, Luo J and Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics (Oxford, England)* 31: 1296–1297.

Huang J, Li Z and Zhao D (2016) Deregulation of the OsmiR160 target gene *OsARF*18 causes growth and developmental defects with an alteration of auxin signaling in rice. *Scientific Reports* 6: 1–14.

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J and Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* 45: D1040–D1045.

Khatun MS, Hasan MM and Kurata H (2019a) PreAIP: computational prediction of anti-inflammatory peptides by integrating multiple complementary features. *Frontiers in Genetics* 10: 129.

Khatun S, Hasan M and Kurata H (2019b) Efficient computational model for identification of antitubercular peptides by integrating amino acid patterns and properties. *FEBS Letters* 593: 3029–3039.

Khush GS (2013) Strategies for increasing the yield potential of cereals: case of rice as an example. *Plant Breeding* 132: 433–436.

Kirik V and Bäumlein H (1996) A novel leaf-specific myb-related protein with a single binding repeat. *Gene* 183: 109–113.

Lang Y, Zhang Z, Gu X, Yang J and Zhu Q (2004) Physiological and ecological effects of crimpy leaf character in rice (*Oryza sativa* L.) I. Leaf orientation, canopy structure and light distribution. *Zuo Wu Xue Bao* 30: 806–810.

Lee YK, Woo MO, Lee D, Lee G, Kim B and Koh HJ (2016) Identification of a novel candidate gene for rolled leaf in rice. *Genes & Genomics* 38: 1077–1084.

Li M, Xiong G, Li R, Cui J, Tang D, Zhang B, Pauly M, Cheng Z and Zhou Y (2009) Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *The Plant Journal* 60: 1055–1069.

Li L, Shi ZY, Li L, Shen GZ, Wang XQ, An LS and Zhang JL (2010) Overexpression of *ACL*1 (*abaxially Curled Leaf* 1) Increased bulliform cells and induced abaxial curling of leaf blades in rice. *Molecular Plant* 3: 807–817.

Li L, Xue X, Zuo SM, Chen ZX, Zhang YF, Li QQ, Zhu JK, Ma YY, Pan XB and Pan CH (2013a) Suppressed expression of *OsAGO*1*a* leads to adaxial leaf rolling in rice. *Chinese Journal of Rice Science* 27: 223–230.

Li W, Wu C, Hu G, Xing L, Qian W, Si H, Sun Z, Wang X, Fu Y and Liu W (2013b) Characterization and fine mapping of a novel rice narrow leaf mutant *nal*9. *Journal of Integrative Plant Biology* 55: 1016–1025.

Li L, Xue X, Chen Z, Zhang Y, Ma Y, Pan C, Zhu J, Pan X and Zuo S (2014) Isolation and characterization of rl (t), a gene that controls leaf rolling in rice. *Chinese Science Bulletin* 59: 3142–3152.

Li C, Zou X, Zhang C, Shao Q, Liu J, Liu B, Li H and Zhao T (2016) *OsLBD*3-7 Overexpression induced adaxially rolled leaves in rice. *PLoS ONE* 11: e0156413.

Li W-Q, Zhang M-J, Gan P-F, Qiao L, Yang S-Q, Miao H, Wang G-F, Zhang M-M, Liu W-T, Li H-F, Shi C-H and Chen K-M (2017) *CLD*1/*SRL*1 modulates leaf rolling by affecting cell wall formation, epidermis integrity and water homeostasis in rice. The *Plant Journal* 92: 904–923.

Liang R, Qin R, Zeng DD, Zheng X, Jin XL and Shi CH (2016) Phenotype analysis and gene mapping of narrow and rolling leaf mutant *nrl*4 in rice (*Oryza sativa* L.). *Scientia Agricultura Sinica* 49: 3863–3873.

Liu B, Li P, Li X, Liu C, Cao S, Chu C and Cao X (2005) Loss of function of *OsDCL*1 affects microRNA accumulation and causes developmental defects in rice. *Plant Physiology* 139: 296–305.

Liu C, Kong WY, You SM, Zhong XJ, Jiang L, Zhao ZG and Wan JM (2015) Genetic analysis and fine mapping of a novel rolled leaf gene in rice. *Scientia Agricultura Sinica* 48: 2487–2496.

Liu X, Li M, Liu K, Tang D, Sun M, Li Y, Shen Y, Du G and Cheng Z (2016) *Semi-Rolled Leaf*2 modulates rice leaf rolling by regulating abaxial side cell differentiation. *Journal of Experimental Botany* 67: 2139–2150.

Ma Y, Wang F, Guo J and Zhang XS (2009) Rice *OsAS*2 gene, a member of LOB domain family, functions in the regulation of shoot differentiation and leaf development. *Journal of Plant Biology* 52: 374–381.

Ma Y, Zhao Y, Shangguan X, Shi S, Zeng Y, Wu Y, Chen R, You A, Zhu L, Du B and He G (2017) Overexpression of *OsRRK*1 changes leaf morphology and defense to insect in rice. *Frontiers in Plant Science* 8: 1–14.

Marchler-Bauer A and Bryant SH (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Research* 32: W327–W331.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C and Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39: D225–D229.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C and Bryant SH (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research* 43: D222–D226.

Mosharaf MP, Hassan MM, Ahmed FF, Khatun MS, Moni MA and Mollah MNH (2020) Computational prediction of protein

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

ubiquitination sites mapping on *Arabidopsis thaliana*. *Computational Biology and Chemistry* 85: 107238.

Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H and Kikuchi S (2010) Genome-wide analysis of NAC transcription factor family in rice. *Gene* 465: 30–44.

Qi J, Qian Q, Bu Q, Li S, Chen Q, Sun J, Liang W, Zhou Y, Chu C and Li X (2008) Mutation of the rice Narrow leaf1 gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiology* 147: 1947–1959.

Rahman H, Yang J, Xu Y-P, Munyampundu J-P and Cai X-Z (2016) Phylogeny of plant CAMTAs and role of AtCAMTAs in non-host resistance to *Xanthomonas oryzae* pv. *oryzae*. *Frontiers in Plant Science* 7: 177.

Schauer N and Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science* 11: 508–516.

Segami S, Kono I, Ando T, Yano M, Kitano H, Miura K and Iwasaki Y (2012) *Small and Round Seed* 5 Gene encodes alpha-tubulin regulating seed cell elongation in rice. *Rice* 5: 4.

Shen NW (2011) Genetic analysis and gene mapping for the semi-curled leaf mutant in rice (*Oryza sativa* L.). *Beijing: Graduate School of Chinese Academy of Agricultural Sciences* 9: 1002–1013.

Shi Z, Wang J, Wan X, Shen G, Wang X and Zhang J (2007) Over-expression of rice *OsAGO7* gene induces upward curling of the leaf blade that enhanced erect-leaf habit. *Planta* 226: 99–108.

Shi L, Wei X, Adedze Y, Sheng Z, Tang S, Hu P and Wang J (2016) Characterization and gene cloning of the rice (*Oryza sativa* L.) dwarf and narrow-leaf mutant *dnl3*. *Genetics and Molecular Research* 15: gmr8731.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ and Mering CV (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* 45: D362–D368.

Tamura K, Stecher G, Peterson D, Filipski A and Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

Toriba T, Harada K, Takamura A, Nakamura H, Ichikawa H, Suzaki T and Hirano H-Y (2007) Molecular characterization the *YABBY* Gene family in *Oryza sativa* and expression analysis of *OsYABBY1*. *Molecular Genetics and Genomics* 277: 457–468.

Wang L, Xu J, Nian J, Shen N, Lai K, Hu J, Zeng D, Ge C, Fang Y, Zhu L, Qian Q and Zhang G (2016) Characterization and fine mapping of the rice gene *OsARVL4* regulating leaf morphology and leaf vein development. *Plant Growth Regulation* 78: 345–356.

Wang X, Wang F, Chen H, Liang X, Huang Y and Yi J (2017) Comparative genomic hybridization and transcriptome sequencing reveal that two genes, *OsI_*14279 (*LOC_Os*03*g*62620) and *OsI_*10794 (*LOC_Os*03*g*14950) regulate the mutation in the *γ-rl* rice mutant. *Physiology and Molecular Biology of Plants* 23: 745–754.

Woo Y-M, Park H-J, Su'udi M, Yang J-I, Park J-J, Back K, Park Y-M and An G (2007) Constitutively wilted 1, a member of the rice YUCCA gene family, is required for maintaining water homeostasis and an appropriate root to shoot ratio. *Plant Molecular Biology* 65: 125–136.

Wu X (2009) Prospects of developing hybrid rice with super high yield. *Agronomy Journal* 101: 688–695.

Wu C, Fu Y, Hu G, Si H, Cheng S and Liu W (2010) Isolation and characterization of a rice mutant with narrow and rolled leaves. *Planta* 232: 313–324.

Wu R, Li S, He S, Waßmann F, Yu C, Qin G, Schreiber L, Qu L-J and Gu H (2011) CFL1, A WW domain protein, regulates cuticle development by modulating the function of HDG1, a class IV homeodomain transcription factor, in rice and Arabidopsis. *The Plant Cell* 23: 3392–3411.

Wu Y, Luo L, Chen L, Tao X, Huang M, Wang H, Chen Z and Xiao W (2016) Chromosome mapping, molecular cloning and expression analysis of a novel gene response for leaf width in rice. *Biochemical and Biophysical Research Communications* 480: 394–401.

Xia ML, Tang DY, Yang YZ, Li YX, Wang WW, Lu H, Liu XM and Lin JZ (2017) Preliminary study on the rice *OsYABBY6* gene involving in the regulation of leaf development. *Life Science Research* 21: 23–30.

Xiang JJ, Zhang GH, Qian Q and Xue HW (2012) *SEMI-ROLLED LEAF1* Encodes a putative glycosylphosphatidylinositol-anchored protein and modulates rice leaf rolling by regulating the formation of bulliform cells. *Plant Physiology* 159: 1488–1500.

Xu Y, Wang Y, Long Q, Huang J, Wang Y, Zhou K, Zheng M, Sun J, Chen H and Chen S (2014) Overexpression of *OsZHD1*, a zinc finger homeodomain class homeobox transcription factor, induces abaxially curled and drooping leaf in rice. *Planta* 239: 803–816.

Yan S, Yan C-J, Zeng X-H, Yang Y-C, Fang Y-W, Tian C-Y, Sun Y-W, Cheng Z-K and Gu M-H (2008) *ROLLED LEAF* 9, Encoding a GARP protein, regulates the leaf abaxial cell fate in rice. *Plant Molecular Biology* 68: 239–250.

Yang C, Li D, Liu X, Ji C, Hao L, Zhao X, Li X, Chen C, Cheng Z and Zhu L (2014) OsMYB103L, an R2R3-MYB transcription factor, influences leaf rolling and mechanical strength in rice (*Oryza sativa* L.). *BMC Plant Biology* 14: 158.

Yang SQ, Li WQ, Miao H, Gan PF, Qiao L, Chang YL, Shi CH and Chen KM (2016) *REL2*, A gene encoding an unknown function protein which contains DUF630 and DUF632 domains controls leaf rolling in rice. *Rice* 9: 37.

Ye Y, Wu K, Chen J, Liu Q, Wu Y, Liu B and Fu X (2018) OsSND2, a NAC family transcription factor, is involved in secondary cell wall biosynthesis through regulating MYBs expression in rice. *Rice* 11: 36.

Yoshimura A, Ideta O and Iwata N (1997) Linkage map of phenotype and RFLP markers in rice. *Plant Molecular Biology* 35: 49–60.

Yoshizawa AC, Itoh M, Okuda S, Moriya Y and Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35: W182–W185.

Yuan L (1997) Hybrid rice breeding for super high yield. *Hybrid Rice* 12: 1–6.

Zhang G-H, Xu Q, Zhu X-D, Qian Q and Xue H-W (2009) SHALLOT-LIKE1 is a KANADI transcription factor that modulates rice leaf rolling by regulating leaf abaxial cell development. *The Plant Cell* 21: 719–735.

Zhang X, Rerksiri W, Liu A, Zhou X, Xiong H, Xiang J, Chen X and Xiong X (2013) Transcriptome profile reveals heat response mechanism at molecular and metabolic levels in rice flag leaf. *Gene* 530: 185–192.

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press

Zhang JJ, Wu SY, Jiang L, Wang JL, Zhang X, Guo XP, Wu CY and Wan JM (2015) A detailed analysis of the leaf rolling mutant *sll*2 reveals complex nature in regulation of bulliform cell development in rice (*Oryza sativa* L.). *Plant Biology* 17: 437–448.

Zhang J, Zhang H, Srivastava AK, Pan Y, Bai J, Fang J, Shi H and Zhu JK (2018) Knock-down of rice microRNA166 confers drought resistance by causing leaf rolling and altering stem xylem development. *Plant Physiology* 176: 2082–2094.

Zhou K, Ma Y, Liu T, Shen M and Pan S (1995) The breeding of subspecific heavy ear hybrid rice-exploration about super-high yield breeding of hybrid rice. *Journal of Sichuan Agricultural University* 13: 403–407.

Zhao SQ, Hu J, Guo LB, Qian Q and Xue HW (2010) Rice leaf inclination2, a VIN3-like protein, regulates leaf angle through modulating cell division of the collar. *Cell Research* 20: 935–947.

Zhao J, Luo H, Jiang Y, Yang X and Zha R (2017) Gene mapping for rice narrow leaf mutant *Narrow Leaf* 11 (*nal*11). *Journal of Southern Agriculture* 48: 1133–1138.

Zhou Y, Wang D, Wu T, Yang Y, Liu C, Yan L, Tang D, Zhao X, Zhu Y and Lin J (2018) LRRK1, A receptor-like cytoplasmic kinase, regulates leaf rolling through modulating bulliform cell development in rice. *Molecular Breeding* 38: 48.

Zou LP, Sun XH, Zhang ZG, Liu P, Wu JX, Tian CJ, Qiu JL and Lu TG (2011) Leaf rolling controlled by the homeodomain leucine zipper class IV gene *Roc*5 in rice. *Plant Physiology* 156: 1589–1602.

https://doi.org/10.1017/S1479262120000404 Published online by Cambridge University Press