PCP
PLANT & CELL PHYSIOLOGY

LARGE GRAIN Encodes a Putative RNA-Binding Protein that Regulates Spikelet Hull Length in Rice

Wan-Yi Chiou[1,2], Tadafumi Kawamoto[3], Eiko Himi[2], Kazuhide Rikiishi[2], Manabu Sugimoto[2], Mika Hayashi-Tsugane[4], Kazuo Tsugane[4,5,*] and Masahiko Maekawa[2,*]

[1]Graduate School of Environmental and Life Science, Okayama University, Okayama, 700-8530 Japan
[2]Institute of Plant Science and Resources, Okayama University, Kurashiki, 710-0046 Japan
[3]Radioisotope Research Institute, School of Dental Medicine, Tsurumi University, Yokohama, 230-8501 Japan
[4]Department of Evolutionary Biology and Biodiversity, National Institute for Basic Biology, Okazaki, 444-8585 Japan
[5]Department of Basic Biology in the School of Life Science, The Graduate University for Advanced Studies, Okazaki, 444-8585 Japan

*Corresponding authors: Tsugane Kazuo, E-mail, tsugane@nibb.ac.jp; Fax, +81564-55-7521; Masahiko Maekawa, E-mail, mmaekawa@rib.okayamau.ac.jp; Fax, +8186-434-1249.
(Received November 29, 2018; Accepted January 11, 2019)

Grain size is a key determiner of grain weight, one of the yield components in rice (*Oryza sativa*). Therefore, to increase grain yield, it is important to elucidate the detailed mechanisms regulating grain size. The *Large grain* (*Lgg*) mutant, found in the *nonautonomous DNA-based active rice transposon1* (*nDart1*)-tagged lines of Koshihikari, is caused by a truncated *nDart1–3* and 355 bp deletion in the 5′ untranslated region of *LGG*, which encodes a putative RNA-binding protein, through transposon display and cosegregation analysis between grain length and *LGG* genotype in F2 and F3. Clustered regularly interspaced short palindromic repeats/CRISPR-associated 9-mediated knockout and overexpression of *LGG* led to longer and shorter grains than wild type, respectively, showing that LGG regulates spikelet hull length. Expression of *LGG* was highest in the 0.6-mm-long young panicle and gradually decreased as the panicle elongated. *LGG* was also expressed in roots and leaves. These results show that LGG functions at the very early stage of panicle development. Longitudinal cell numbers of spikelet hulls of *Lgg*, knockout and overexpressed plants were significantly different from those of the wild type, suggesting that LGG might regulate longitudinal cell proliferation in the spikelet hull. RNA-Seq analysis of 1-mm-long young panicles from *LGG* knockout and overexpressing plants revealed that the expressions of many cell cycle-related genes were reduced in knockout plants relative to *LGG*-overexpressing plants and wild type, whereas some genes for cell proliferation were highly expressed in knockout plants. Taken together, these results suggest that LGG might be a regulator of cell cycle and cell division in the rice spikelet hull.

Keywords: DNA transposon • Grain size • Oryza sativa • RNA-binding protein • Tagged lines.

Abbreviations: AP2, APETALA2; *APO1*, *ABERRANT PANICLE ORGANIZATION1*; *Bdt1*, *Bushy dwarf tiller1*; CDK, cyclin-dependent kinase; CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated 9; CYC, cyclin; DAPI, 4′,6-diamidino-2-phenylindole; eCLIP, enhanced cross-linking immunoprecipitation; EL2, elliptocytosis type 2; ERF, ethylene responsive factor; GFP, green fluorescent protein; GO, gene ontology; *GSN1*, *GRAIN SIZE AND NUMBER 1*; hnRNP1, heterogeneous nuclear ribonucleoprotein 1; HoLAMap, observed heterozygosity per locus index unified linkage and association mapping; Msi1, Musashi1; *nDart1*, *nonautonomous DNA-based active rice transposon1*; NGS, next-generation sequencing; NIL, near-isogenic line; *pyl*, *pale-yellow-leaf*; qRT-PCR, quantitative real-time PCR; QTL, quantitative trait loci; RACE, rapid amplification of cDNA ends; RBP, RNA-binding protein; RNA-Seq, RNA-sequencing; RRM, RNA recognition motif; SMR, SIAMESE-RELATED; *SOMS1*, *SMALL ORGAN SIZE 1*; SPL, *SQUAMOSA* PROMOTER BINDING PROTEIN-LIKE; *SRS*, *SMALL AND ROUND SEED*; *swl1*, *snow-white leaf1*; *TAW1*, *TAWAWA1*; TD, transposon display; *thl*, *thumbelina*; TIR, terminal inverted repeat; TIS, transcription initiation site; TSD, target site duplication; UTR, untranslated region; WT, wild type.

Introduction

Rice (*Oryza sativa* L.) is a staple food for over half of the human population. To meet a growing population, it is estimated that rice yields need to increase by approximately 42% by 2050 (Ray et al. 2013). Rice production consists of four yield components: panicle number per unit area, spikelet number per panicle, percentage of filled grains per panicle and grain weight. Of these factors, grain weight is determined by the length, width and thickness of the grain and the grain's specific gravity (Takeda 1997). Standard partial regression coefficient analysis of grain weights in many varieties has shown that the length of the grain provides the greatest contribution to grain weight because most varieties have constant specific gravity values (Takeda 1997). Thus, it is important to elucidate the genetic mechanisms underlying grain size in order to increase yield.

Plant Cell Physiol. 60(3): 503–515 (2019) doi:10.1093/pcp/pcz014, Advance Access publication on 25 January 2019, available online at www.pcp.oxfordjournals.org

The grain size of rice is strictly regulated by the size of the spikelet hull (Takeda 1997). To date, more than 40 quantitative trait loci (QTL) and genes for grain size have been reported in rice, as reviewed by Li and Li (2016), McCouch et al. (2016) and Li et al. (2018). The final size of tissues like rice grains is determined by total cell number and cell elongation. Genes for cell number of the spikelet hull, *GW2* (Song et al. 2007), *qSW5*/*GW5* (Shomura et al. 2008, Weng et al. 2008), *GW8*/*OsSPL16* (Wang et al. 2012) and *GRAIN SIZE AND NUMBER 1* (*GSN1*) (Guo et al. 2018) have been identified. Meanwhile, *GLW7*/*OsSPL13* (Si et al. 2016), *SMALL AND ROUND SEED3* (*SRS3*) (Deng et al. 2015) and *SMALL AND ROUND SEED5* (*SRS5*) (Segami et al. 2012), have been reported to control hull cell elongation.

RNA-binding proteins (RBPs) play crucial roles in post-transcriptional gene regulation and RNA metabolism, such as mRNA processing, export, localization, translation and degradation. RBPs are also involved in development, stress responses and genome organization in plants (Fedoroff 2002, Lorković 2009, Ambrosone et al. 2012). Most RBPs contain one or more conserved domains, such as an RNA recognition motif (RRM), K homology (KH) domain, zinc finger domain, DEAD/DEAH (Asp-Glu-Ala-Asp/Asp-Glu-Ala-His) box, Pumilio/FBF domain, double-stranded RNA-binding domain (DS-RBD) or Piwi/Argonaute/Zwille (PAZ) domain (Fedoroff 2002, Ambrosone et al. 2012). Of these domains, the RRM has high RNA-binding affinity and specificity, and two or more RRMs significantly increase binding strength. The RRM also plays an essential role in protein–protein interaction in RNA regulation (Maris et al. 2005). Rice has six RBPs paralogous to the Musashi family that have two RRMs. Musashi was first discovered in the external sensory organ in the peripheral nervous system of *Drosophila* and is required for cell division of the sensory organ precursor cell (Nakamura et al. 1994). However, the functions of Musashi family members in plants have remained largely unknown because of a lack of mutants to analyze.

Mutants are useful for identifying the genes underlying a trait of interest (Wang et al. 2013, Howard et al. 2014). A transposon tagging system is a cost-effective and time-saving method for mutant creation and gene discovery. In rice, a tagging system using *Tos17*, an endogenous retrotransposon, has been developed, and a *Tos17*-tagged mutant pool has been established, as reviewed by Hirochika (2010). In addition, the maize transposons, *Ac*/*Ds* (Chin et al. 1999, Greco et al. 2003) and *En*/*Spm* (Greco et al. 2004) have been used to generate a large pool of insertional mutants in rice (Jiang and Ramachandran 2010, Wang et al. 2013). Tsugane et al. (2006) discovered an active DNA transposon, *nonautonomous DNA-based active rice transposon1* (*nDart1*) controlled by *aDart1-27*, an autonomous element. To date, several novel genes, including *pale-yellow-leaf* (*pyl*), *thumbelina* (*thl*) (Tsugane et al. 2006) and *snow-white leaf1* (*swl1*) (Hayashi-Tsugane et al. 2014), have been identified in *nDart1*-insertion mutant lines. In particular, some gain-of-function mutants, *aberrant panicle organization1-D* (*apo1-D*) (Ikeda-Kawakatsu et al. 2009), *tawawa1-D* (*taw1-D*) (Yoshida et al. 2013) and *Bushy dwarf tiller1* (*Bdt1*) (Hayashi-Tsugane et al. 2015) have been obtained.

In this study, we characterize the *Large grain* (*Lgg*) mutant. *Lgg* was found in *nDart1*-insertion lines of Koshihikari and showed incomplete dominance. Transposon display analysis and cosegregation between *nDart1* insertion and grain length demonstrated that *Lgg* was caused by an *nDart1* insertion and possibly subsequent 355-bp deletion at the 5′ untranslated region (UTR) of the putative RBP-encoding gene *LGG* (*Os11g0637700*), resulting in decreased expression of *LGG*. This is the report that a putative RBP belonging to the Musashi family controls spikelet hull length in rice. RNA-Seq analysis of overexpression and clustered regularly interspaced short palindromic repeats/CRISPR-associated 9 (CRISPR/Cas9)-mediated knockout plants suggested that *LGG* might control several genes for cell division in spikelet hull differentiation.

Results

Characterization of the *Lgg* mutant

The *Lgg* mutant found in *nDart1*-tagged lines of cv. Koshihikari showed a long grain phenotype (**Fig. 1a, b**). Although *Lgg* plants did not show any obvious differences in culm length, panicle number, and days to heading, compared to wild type (WT), *Lgg* had significantly larger values for panicle length (106.9%), 100-grain weight (109.2%) and grain length (109.3%) and significantly lower values for grain width (95.1%), spikelet fertility (80%) and total grain weight per plant (84%) than those of the WT (**Fig. 1g–j**; Supplementary Table S1). Because the increased values for 100-grain weight were proportional to that of the grain length in the mutant, the heavier 100-grain weight was mainly due to the longer grains of *Lgg*. Since grain length is determined by spikelet hull length, the cell size and cell number of the spikelet hull were examined using longitudinal cryo-sections of the lemmas in *Lgg* and WT (**Fig. 1c, d**). The spikelet hull consists of five layers, including the innermost chlorophyll-contained parenchyma cell layer (**Fig. 1e, f**). The *Lgg* spikelet hull cell number was 125% that of the WT, whereas the spikelet hull cell length of *Lgg* was not significantly different from that of the WT (**Fig. 1k, l**). This result showed that increased spikelet hull length in *Lgg* was caused by increased cell number, suggesting that more cell divisions occurred in *Lgg* than in the WT, and that the cell growth in the lemma of *Lgg* might be comparable to that of WT.

Inheritance of the large grain phenotype in *Lgg*

To investigate the inheritance of the large grain phenotype in *Lgg*, F2 plants of the cross between *Lgg* and WT were examined for the segregation of grain length. In the F2, three distinctive peaks were observed: one peak for the WT, another for the *Lgg* phenotype and the third represented an intermediate phenotype, although clear boundaries between each phenotype were not observed (**Fig. 2a**). To confirm the F2 segregation pattern of grain length, we tested progeny for grain length using F3 lines randomly selected from the F2 population. The relationship

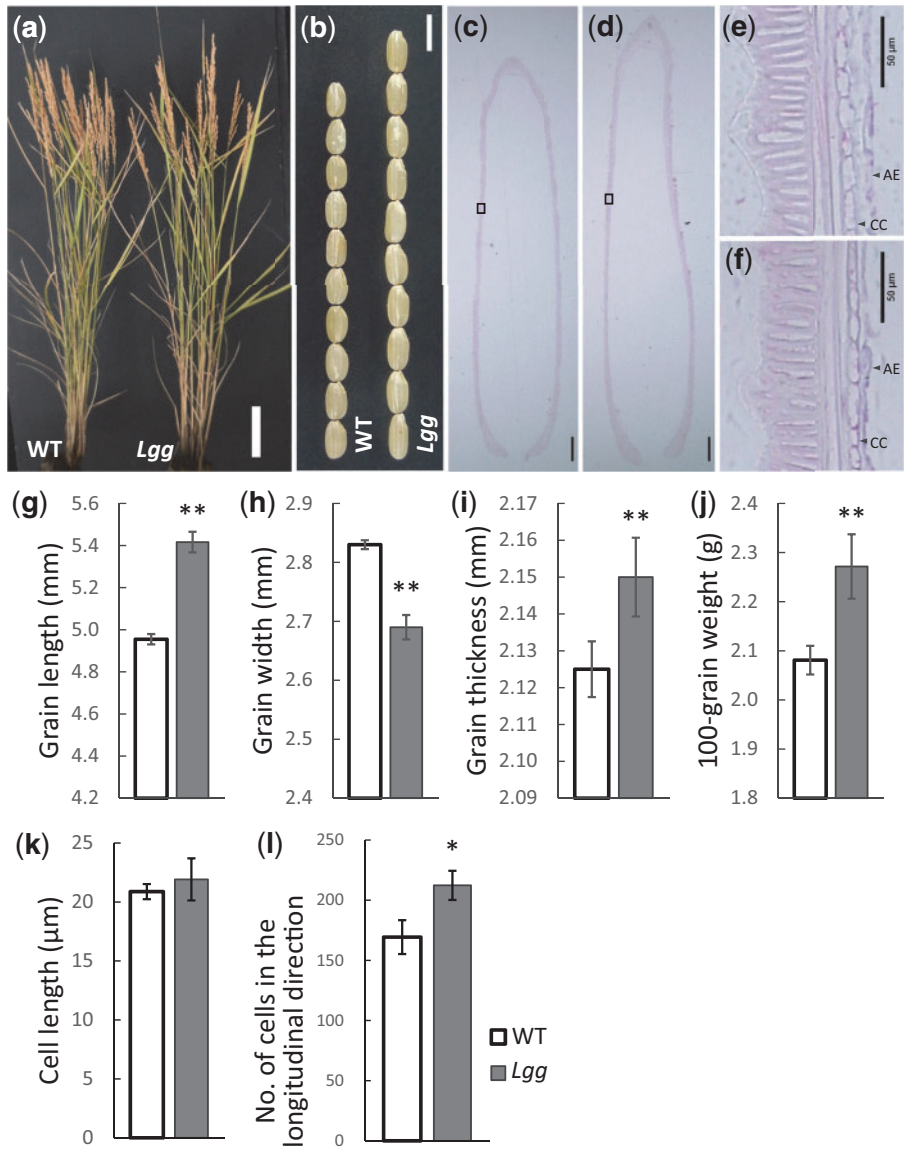

Fig. 1 Plant phenotype and grain morphology of *Lgg* and WT. (a) The entire plants. Bar = 10 cm. (b) The rice grains. Bar = 5 mm. (c) Longitudinal section of lemma of WT. Bar = 400 μm. (d) Longitudinal section of lemma of *Lgg*. Bar = 400 μm. (e) Magnified view of the section in (c) in WT. Bar = 50 μm. (f) Magnified view of the section in (d) in *Lgg*. Bar = 50 μm. (g) Grain length. $n = 8$. mean ± SD. (h) Grain width. $n = 8$. mean ± SD. (i) Grain thickness. $n = 8$. mean ± SD. (j) 100-grain weight. $n = 8$. mean ± SD. (k) The cell length of parenchyma cells of the longitudinal section of spikelet hull. (l) The cell number in longitudinal direction of the longitudinal section of spikelet hull. * and ** indicate significance at $P < 0.05$ and $P < 0.01$ by Student's *t*-test, respectively. AE, adaxial epidermis; CC, chlorophyll-contained parenchyma cell.

between grain length in F2 plants and mean grain length in the F3 line is shown in Supplementary Fig. S1. F3 lines derived from F2 plants with a WT- or *Lgg*-type grain length showed a narrow range of grain lengths reflective of their parents. On the other hand, F3 lines derived from F2 plants with an intermediate grain length phenotype showed wide variation in grain length (Supplementary Fig. S1). The segregation ratio of the WT, intermediate and the *Lgg* grain length phenotype in 40 F3 lines was 5:22:13 (**Fig. 2b**). These segregation numbers fitted well to a ratio of 1:2:1 ($\chi^2 = 0.178$, $P = 0.083$), suggesting that the long grain trait of *Lgg* was controlled by an incompletely dominant gene.

Identification of the *LGG* gene through transposon display

We performed transposon display (TD) using 10 *Lgg*-type plants from the F2 cross and *Lgg* to find the *nDart1* insertion sites in *Lgg* (Supplementary Fig. S2). Fourteen *nDart1* insertions were observed only in *Lgg*, and one of the insertions was detected by two restriction enzyme treatments on chromosome 2 (Supplementary Table S2). This identified 13 insertions distributed on chromosomes 1, 2, 3, 5, 8, 9 and 11. Then, we examined the relationship between these 13 insertions and grain length in 40 F2 plants. The genotype of one insertion detected at 25 Mb on chromosome 11 cosegregated with the grain length

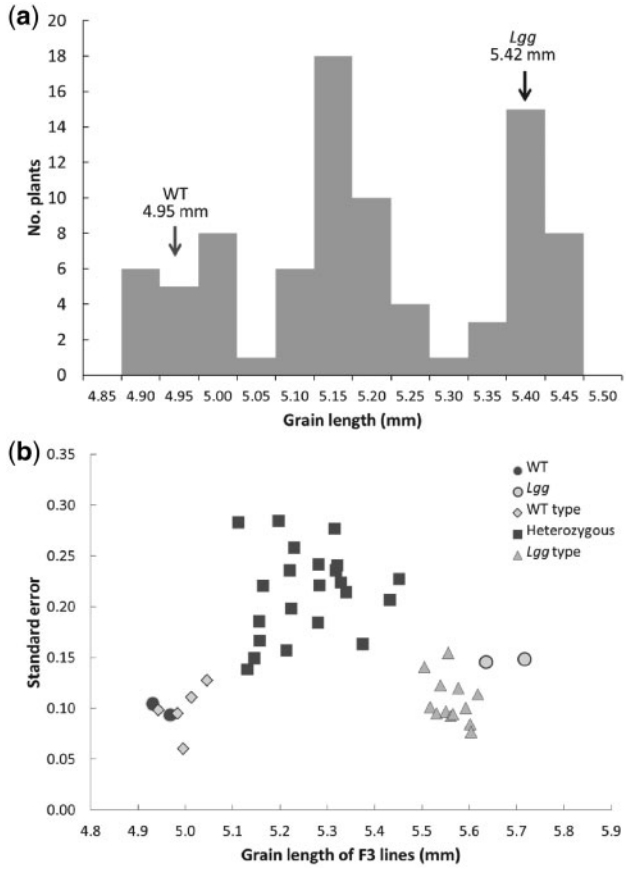

Fig. 2 The segregation pattern of grain length in F2 and F3 plants. (a) Frequency distribution of grain length in F2 plants of the cross between *Lgg* and WT. $n = 85$. (b) Relation between grain length and the standard errors of F3 lines. $n = 40$.

phenotype (**Fig. 2b**), suggesting that this insertion was the causal factor for increased grain length in *Lgg*. Through genomic sequence analysis of *nDart1* insertions in *Lgg*, we identified the causal *nDart1* element as *nDart1–3*, 1 of 13 *nDart1* homologs. We determined that only 500 bp of the 5′ region of the *nDart1–3* remained and that the remaining 107 bp of *nDart1–3* was deleted together with 355 bp flanking the insertion site at the 5′ UTR of *Os11g0637700* (**Fig. 3a;** Supplementary Fig. S3). These results suggested that *Os11g0637700* might be the candidate gene for *Lgg*. This gene was designated *LGG*.

To verify that *LGG* was responsible for the long grain phenotype, knockout (GE) and overexpression (OE) lines were produced by transforming a CRISPR/Cas9 construct targeting the RRM region of *LGG* (Supplementary Fig. S4a, b) and an *LGG* construct with its WT promoter (pLGG:LGG^{Koshi}; OE) into cv. Nipponbare (NP) (**Fig. 4a–c**), respectively, because we could not obtain any transformants in the Koshihikari genetic background. Two of three independent GE lines showed significantly increased grain length, and three independent OE lines had shorter grains than NP (**Fig. 4c, d**; Supplementary Table S3). Any significant differences of panicle length, number of panicles per plant and number of spikelets per panicle were not observed among NP, GE and OE except for culm length (Supplementary Table S3). Observations of longitudinal sections from lemmas of GE and OE lines revealed that the cell sizes of GE and OE lines were comparable to NP (**Fig. 4e–h**), suggesting that *LGG* might regulate the longitudinal cell number of spikelet hull and thus grain length.

Expression pattern of *LGG*

To elucidate the effect of *nDart1–3* insertion and deletion at the 5′ UTR on transcription of *LGG* in the *Lgg* mutant, 5′- and 3′-RACE assays were carried out. As shown in **Fig. 3b**, the transcription initiation site (TIS) in *Lgg* mutant was shifted to the first intron of -53 bp upstream from the translation start site with adenine (+1) of first methionine codon of WT, resulting in another start codon created 29 bp upstream of the start codon of WT in the *Lgg* allele. On the other hand, exons 3 and 4 of the *Lgg* allele were identical to the WT, suggesting that the transcript of the *Lgg* allele contained the full coding region and that the splicing sites between introns 2 and 3 were not affected.

The expression levels of *LGG* were examined by quantitative real-time PCR (qRT-PCR) using the third leaf and the root in 2-week-old seedlings and four different lengths of young panicles (0.6, 1, 2 and 3 mm). It has reported that the beginning of spikelet hull primordia occurs when the young panicle reaching to 1.5 mm (Hoshikawa 1989). *LGG* transcripts were detected at all tested tissues in both WT and *Lgg*. The expression level of *LGG* in the leaf was about 2-fold higher than in the root (**Fig. 3c**). Further, it was found that *LGG* expression level decreased with increasing panicle length. The 0.6-mm-long young panicle showed the highest expression level among different lengths of young panicles in the WT (**Fig. 3d**). The *Lgg* mutant showed significantly lower *LGG* expression compared to the WT. *LGG* expression in the *Lgg* mutant background was 36.2% and 15.5% of WT expression levels in the leaf and the root, respectively. The 0.6-, 1-, 2- and 3-mm-long young panicles showed *LGG* expression levels 24.5, 4.5, 22.6 and 19.9% that of the WT, respectively. The reduced expression level of *LGG* in 1-mm-long young panicle was particularly remarkable (**Fig. 3d**). These results showed that *LGG* was highly expressed in young panicles and the *Lgg* allele reduced *LGG* expression.

To investigate the subcellular localization of LGG, green fluorescent protein (GFP)-fused constructs driven by 35S promoter, 35S:LGG^{NP}-GFP was transformed into rice calli. GFP fluorescence spots were observed in the nuclei in calli (**Fig. 5**). These results suggest that LGG is localized to the nucleus.

Phylogenetic analysis of LGG homologs

LGG consists of four exons and three introns with a 1,404-bp coding sequence (**Fig. 3a**; Supplementary Fig. S4a). *LGG* encodes a heterogeneous nuclear ribonucleoprotein 1 (hnRNP1) 467 amino acids in length with two RRMs at the N-terminus (Supplementary Fig. S4a; Supplementary Table S4). Database searches revealed that LGG has five homologs in rice, and homologs of LGG are present across plants, fungi and animals (**Fig. 6a, b**). LGG was grouped into a grass clade through phylogenetic analysis of amino acid sequences of LGG orthologs in

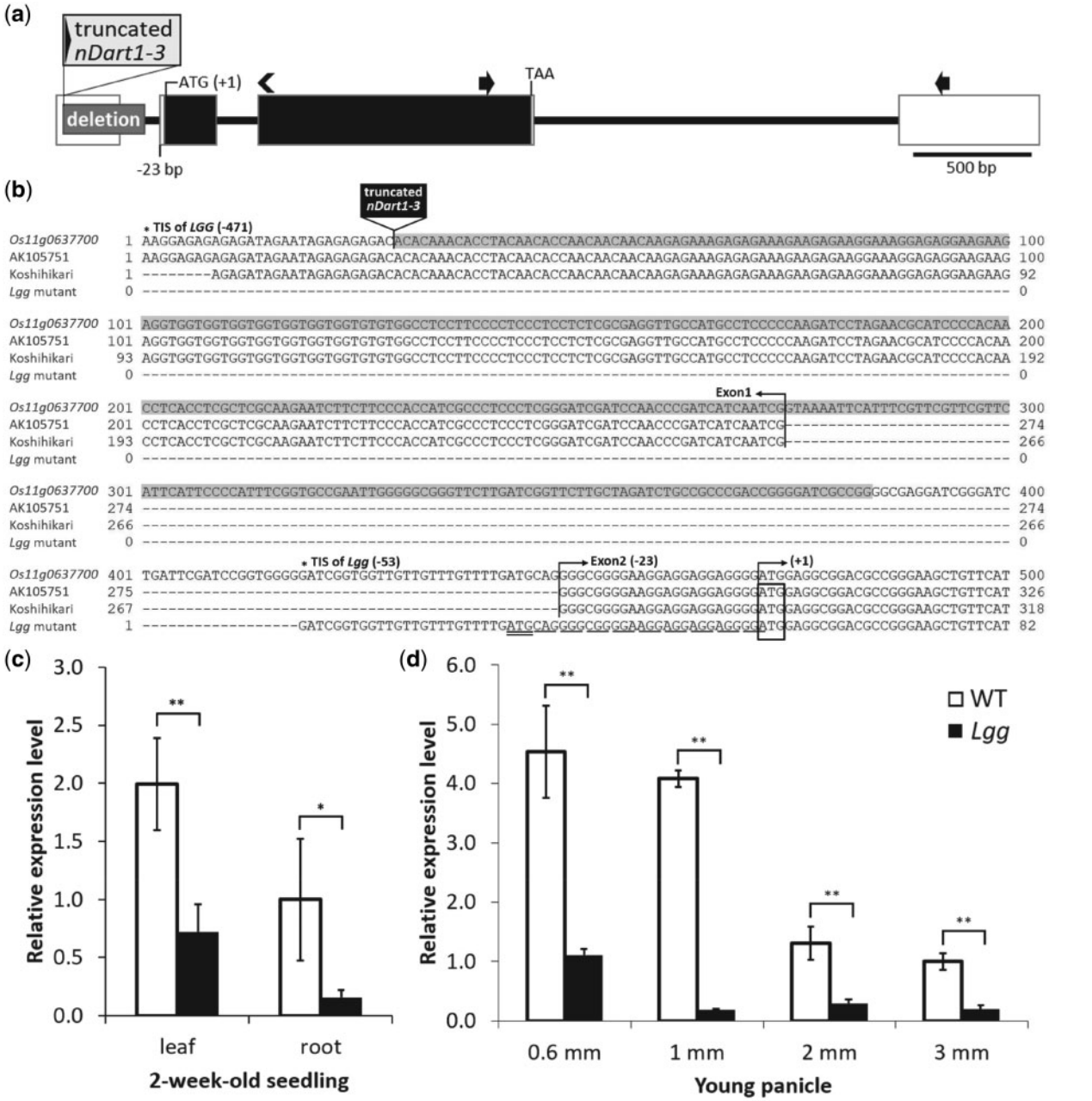

Fig. 3 The structure, transcript sequence and expression levels of *LGG*. (a) *LGG* structure with truncated *nDart1–3* insertion and deletion in *Lgg*. Arrow head and arrows indicate the primer positions for 5′-RACE assay and qRT-PCR, respectively. (b) Transcript sequences of 5′-region of *LGG* in Koshihikari and *Lgg* together with genomic sequence (*Os11g0637700*) of NP. Gray box indicates the genomic deletion in *Lgg*. Black lined box and double underline show the original translation start codon in Koshihikari and predicted translation start codon in *Lgg*, respectively. Underline represents triplet codon. * shows TIS. (c) The expression level of *LGG* in leaf and root of 2-week-old seedling of WT and *Lgg*. (d) The expression level of *LGG* in 0.6, 1, 2 and 3 mm young panicles of WT and *Lgg*. * and ** indicate significance at $P < 0.05$ and $P < 0.01$ by Student's *t*-test, respectively. $n = 3$. Mean ± SD. *Ubiquitin* was used as internal control.

17 species (**Fig. 6b**). Plants hosted a greater number of LGG homologs than did animals (Supplementary Table S5).

LGG affects expression of cell cycle-related genes

To investigate the effects of LGG on the expression of other genes, RNA-Seq analysis was performed on 1-mm-long young panicles of GE, OE and NP plants. We selected 1-mm-long panicles for analysis because *LGG* expression was the highest in 0.6-mm-long young panicles in the WT, and we expected LGG to regulate affected genes at later development stages. As shown in Supplementary Figs. S5 and S6, the expression of 477 genes in GE were reduced compared to NP, whereas 351 genes of GE were expressed at higher levels than those of NP.

According to defined gene ontology (GO) terms, DNA binding, protein binding, cytosol and chromosomal organization genes as well as cell cycle-related genes such as DNA replication

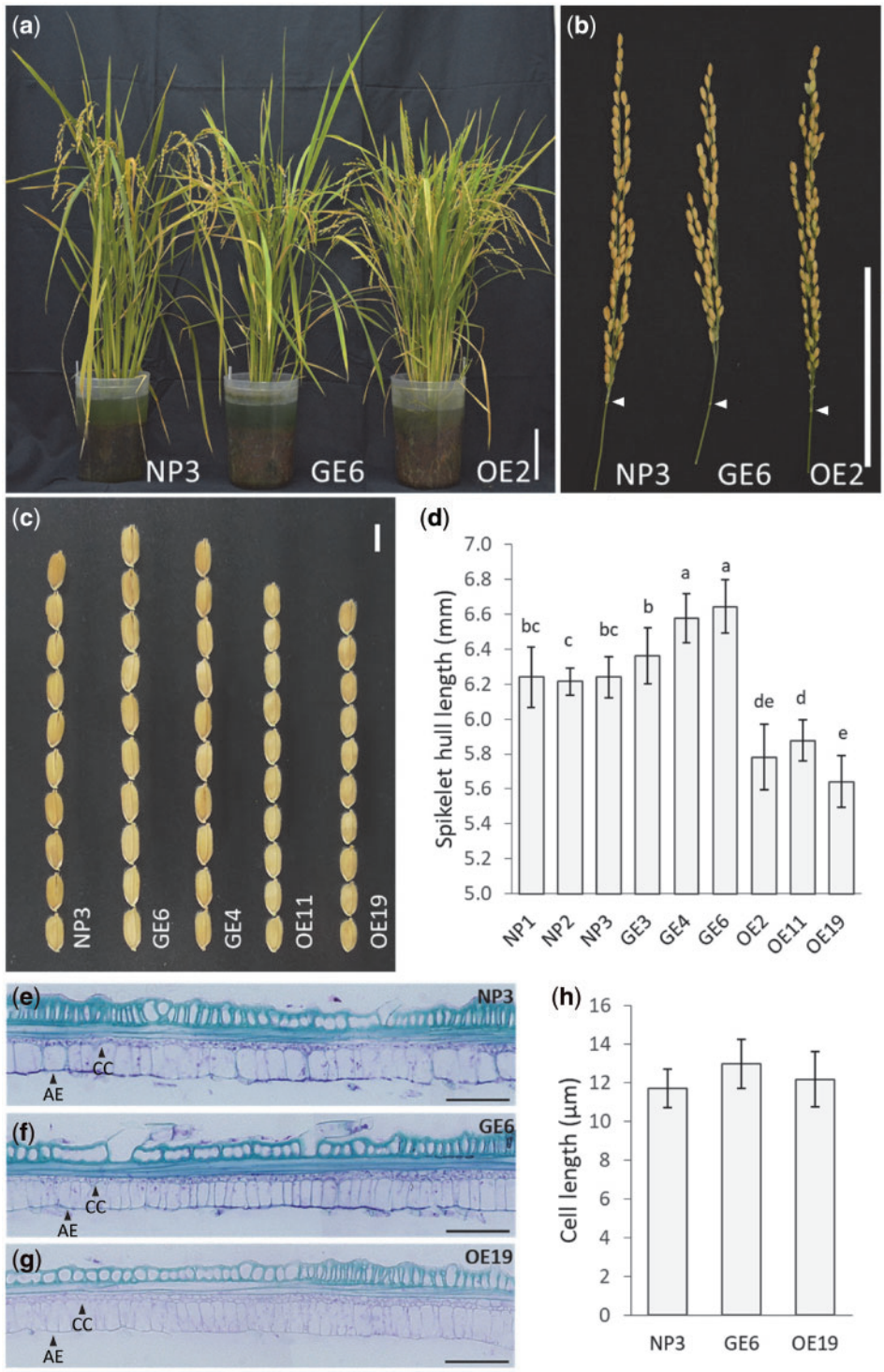

Fig. 4 Phenotypes of NP, the genome-editing line (GE) and the pLGG:LGG^{Koshi} overexpression line (OE). (a) The entire plants. Bar = 10 cm. (b) The panicles. Arrow head depicts panicle base as the base for panicle length. Bar = 10 cm. (c) The spikelets. Bar = 5 mm. (d) Spikelet hull length of NP, GE and OE plants. Different letters indicate $P < 0.05$ by Fisher's Least Significant Difference test. $n = 10$. Mean ± SD. (e, f, g) Longitudinal sections of lemma of NP, GE and OE, respectively. Bar = 100 μm. (h) Cell length of lemma in NP, GE and OE. $n = 3$. Mean ± SD. AE, adaxial epidermis; CC, chlorophyll-contained parenchyma cell.

and cell proliferation genes were found to be reduced in GE (Supplementary Fig. S5). On the other hand, GE showed increased expressions of genes related to transcription, transcription factor activity and response to abiotic stimulus (Supplementary Fig. S6). Further, we examined the expression levels of cell cycle, cell division and cell proliferation genes in GE, OE and NP through RNA-Seq analysis, because LGG appeared to regulate longitudinal cell number in the spikelet hull, and cell

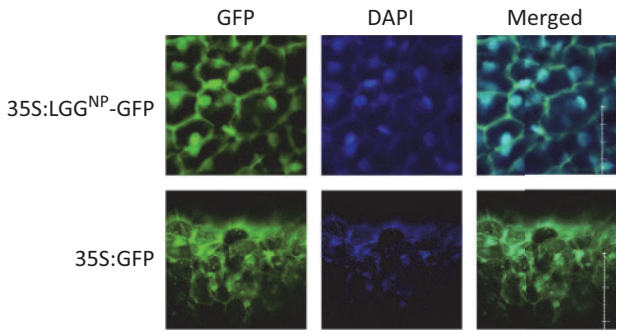

Fig. 5 Subcellular localization of LGG. 35S:LGG^{NP}-GFP is expressed in rice calli. 4′,6-Diamidino-2-phenylindole is used for nuclear staining. Bar = 20 μm.

number is controlled this class of genes. We found that 38 of 46 cell cycle-related genes showed lower expression and that 8 genes showed higher expression levels in GE than those of OE and NP (**Fig. 7a**). In particular, the order of expression levels of *CYCD3-2* and *Os04g0488100* (*OsRad21-2*) were GE < NP < OE and GE > NP > OE, respectively. Expression levels of half the cell division and cell proliferation-related genes were reduced in GE compared to OE and NP. *Os05g0389000* (*OsAP2/ERF142/SMOS1*) showed the highest expression among GE, OE and NP with the order of GE > NP > OE (**Fig. 7b**). Further, *EL2* and *SMR*, cyclin-dependent kinase (CDK) inhibitors, showed increased expression in GE compared to OE and NP (**Fig. 7a**). These results suggested that LGG might regulate cell cycle-related genes in the spikelet hull. On the other hand, the causal genes behind the higher longitudinal cell number in the spikelet hull of *Lgg* remain unknown.

Discussion

This study showed that a putative RBP containing RRMs regulates grain length in rice. Grain length is the most important determinant of grain weight, which is a crucial component of yield in rice (Takeda 1997). Recently, 99 QTL for grain length were reported using the observed heterozygosity per locus index unified linkage and association mapping method for 504 varieties of rice (Yu et al. 2017). The length dimension is basically controlled by two factors: cell number and cell length. Thus, it is reasonable that many QTL for grain length were detected, because cell number and cell length are controlled by different genetic systems. Segami et al. (2017) reported that elucidated grain length genes were categorized into two types based on gene functions: genes for cell length (*GL7/GW7*, *D2*, *D11*, *D61*, *BRD1*, *SRS1*, *SRS3*, *SRS5* and *GLW7*) and gene for cell division (*GS3*, *GW2*, *qSW5/GW5*, *GS5*, *GW8*, *GL3.1*, *TGW6*, *D1*, *SG1* and *TUD1*) in addition to *LG3* (Yu et al. 2017) and *GSN1* (Guo et al. 2018). While *GS5* (Li et al. 2011) and *GW8* (Wang et al. 2012) were positive regulators of the cell cycle, *GW2* was reported to be a negative regulator of cell division (Song et al. 2007). In the case of *LG3*, 19 of 27 cell cycle-related genes analyzed were found to be upregulated in the near-isogenic line carrying *LG3*. Generally, gene function relies on several transcript-processing steps mediated by gene-specific RBPs. Although many causal genes responsible for grain length have been reported, so far no RBPs controlling grain length have been identified. In this study, the *Lgg* mutant found in *nDart1*-tagged lines of Koshihikari was revealed to be caused by deletion and *nDart1–3* insertion in the 5′ UTR of *Os11G0637700*, which encodes a putative RBP.

RBPs are key regulators of gene expression. Recently, new methods for identifying RBPs genome-wide have been developed: system-wide identification of RBPs in vitro, identification of RBP repertoires in vivo by RNA interactome capture and enhanced crosslinking immunoprecipitation (eCLIP) (Hentze et al. 2018). In *Mus musculus*, *Homo sapiens* and *Arabidopsis thaliana*, there are 1,914, 1,393 and 719 RBPs, respectively. In rice, the RBP database (RiceRBP) containing 257 experimentally identified RBPs was established (Morris et al. 2011). The precise function of a given RBP was directly confirmed by phenotyping the RBP mutant. However, to date, RBP mutants have rarely been reported, possibly because of the essential roles played by many RBPs. Transposon insertion mutagenesis is one method for generating mutants (Jiang and Ramachandran 2010, Wang et al. 2013). Further, transposon insertion mutagenesis often generates gain-of-function mutants, depending on the insertion site of the transposon. Some gain-of-function mutants such as *Apo1* (Ikeda-Kawakatsu et al. 2009), *TAW1* (Yoshida et al. 2013) and *Btd1* (Hayashi-Tsugane et al. 2015) have already been identified in *nDart1*-tagged lines in rice. In this study, the new incomplete dominant *Lgg* was discovered in *nDart1*-tagged lines of Koshihikari. A combination of transposon display analysis of *Lgg* and *Lgg*-type F2 plants and the analysis of CRISPR/Cas9-mediated knockout and overexpressed lines established that the causal gene for the large grain phenotype was a putative RBP with two RRMs. The causal gene, of *Os11G0637700*, was altered in *Lgg* by *nDart1–3* insertion following sequence deletion in the 5′ UTR. This deletion event was possibly induced by microhomology (CGG)-mediated homologous recombination between *nDart1–3* and the first intron of *Os11G0637700*, as shown in Supplementary Fig. S7, resulting in a shift of the TIS in the mutant. This deletion event was not assumed to be specific to the transposon insertion, although transposon insertion was reported to induce chromosomal rearrangements, including deletion (Zhang et al. 2009, Xuan et al. 2012).

We found that the long grain phenotype of the mutant was due to an increased cell number in the spikelet hull in the mutant compared to WT, suggesting that *LGG* might regulate cell division in the spikelet hull primordium. The expression level of cell division genes in the spikelet hull should reach a maximum in the primordium before tissue differentiation begins. The spikelet hull differentiation stage is strictly defined by young panicle length. The 1.5-mm-long young panicle is the beginning stage of spikelet hull primordium differentiation and the 2-mm-long young panicle is the beginning stage of stamen and pistil primordium differentiation (Hoshikawa 1989). Intriguingly, the expression of *LGG* was the highest in the 0.6-mm-long young panicle and gradually reduced with the increase in young panicle length (**Fig. 3d**). This result suggested

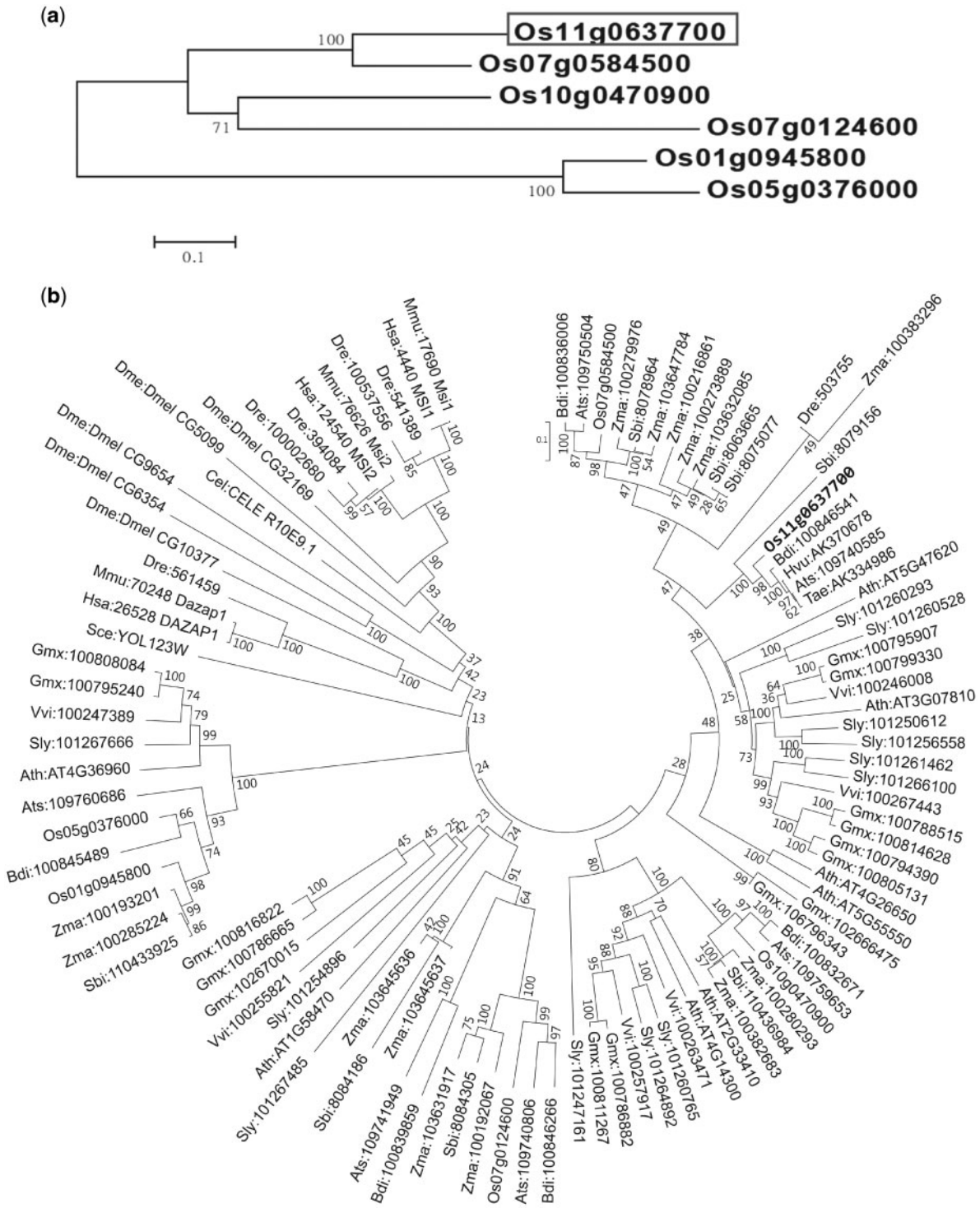

Fig. 6 The phylogenetic analyses of LGG (Os11g0637700) homologs. (a) Homologs of LGG in rice. (b) Phylogenetic tree of 102 homologs of LGG in 17 species. Amino acid sequences were aligned, and phylogenetic tress was constructed by MEGA 6.0. Ath, *Arabidopsis thaliana*; Ats, *Aegilops tauschii*; Bdi, *Brachypodium distachyon*; Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Dre, *Danio rerio*; Gmx, *Glycine max*; Hsa, *Homo sapiens*; Hvu, *Hordeum vulgare*; Mmu, *Mus musculus*; Os, *Oryza sativa japonica*; Sbi, *Sorghum bicolor*; Sce, *Saccharomyces cerevisiae*; Sly, *Solanum lycopersicum*; Tae, *Triticum aestivum*; Vvi, *Vitis vinifera*; Zma, *Zea mays*.

that *LGG* might control primordium differentiation in the spikelet hull. However, RNA-Seq analysis of 1-mm-long young panicles in CRISPR/Cas9-mediated knockout (GE) and overexpressed (OE) plants did not lead to distinctive results for increased transcripts for cell cycle-related genes as shown in **Fig. 7a**. Through RNA-Seq analysis of 1-mm-long young panicles of GE, OE and NP plants, the expression levels of many cell cycle-related genes were found to be significantly reduced in GE

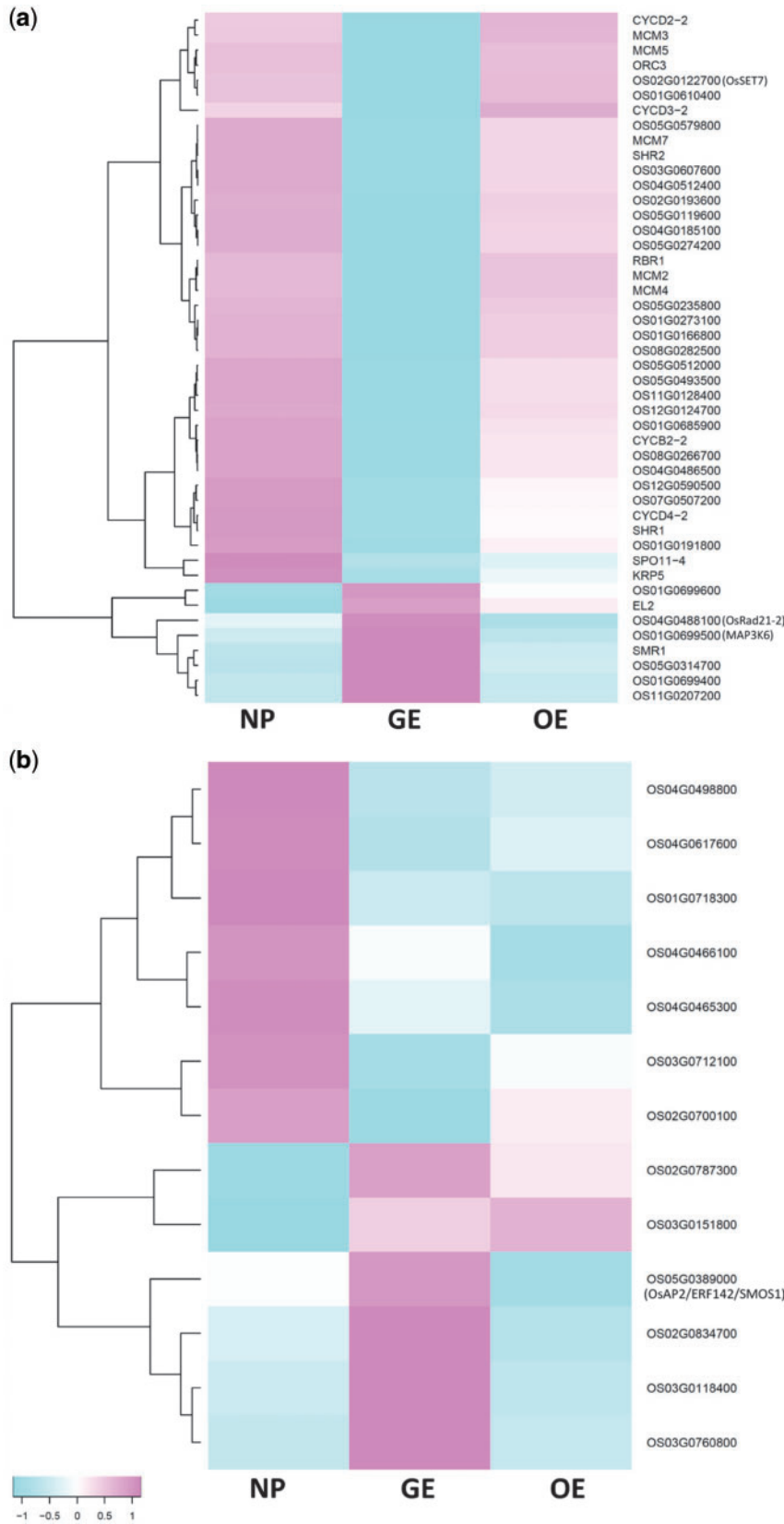

Fig. 7 Differential expression of genes related to cell cycle, cell division and cell proliferation in GE, OE and NP using RNA-Seq. (a) Heatmap of cell cycle-related gene expression. (b) Heatmap of cell division and cell proliferation-related gene expression.

lines, whereas CKD inhibitor genes, *EL2* and *SMR5*, were highly expressed in GE compared to OE and NP. These genes are responsible for the cell cycle controlled by the antagonistic interaction of the CYC–CDK complex and the inhibitors, EL2 (Peres et al. 2007) and SMR (Komaki and Sugimoto 2012). On the other hand, the expression levels of some cell cycle-related genes, e.g. three serine/threonine protein, kinase domain genes, *Rad21*-like gene and *MAP3Ka*, were significantly increased in GE compared to OE and NP. Of the six genes showing increased expression, four genes, including *MAP3Ka*, were protein kinase-related. *Rad21*/*Rec8*-like is also connected to cell division (da Costa-Nunes et al. 2006, Gong et al. 2011).

The reason for reduced expression levels of many cell cycle-related genes and some cell division/cell proliferation genes in GE remains unknown. It is likely that *LGG* controls a key player(s) in a cell cycle-related gene network. Then GE plants might reduce the expression of many cell cycle-related and some cell division/cell proliferation-related genes. In particular, *CYCD3-2* was differentially expressed between GE, OE and NP in the order GE $<$ NP $<$ OE. Since CYCD3 is the target of SMR (Hamdoun et al. 2016), it is plausible that opposite trends in expression of *SMR1* and *CYCD3-2* were observed in GE and OE. Recently, CYCD3 was reported to regulate cell proliferation and secondary growth in the cambium of Arabidopsis (Collins et al. 2015). However, *Os04g0488100* (*Rad21-2*/*Rec8*-like) and *Os05g0389000* (*AP2*/*ERF*/*SOMS1*) showed differential expression in the order of GE $>$ NP $>$ OE (**Fig. 7a, b**). The expression of these genes in GE and OE were consistent with the spikelet hull length in these lines (**Fig. 4c, d**). OsRAD21-2 controls cell division and growth in actively dividing tissues, including premeiotic flowers, stem intercalary meristems and leaf meristems (Gong et al. 2011), whereas Aya et al. (2014) reported that AP2/ERF/SMOS1 functioned as an auxin-dependent regulator of cell expansion. On the other hand, *OsLG3*, encoding AP2/ERF125, functions as a positive regulator of grain length, increasing the expression of cell cycle-related genes, such as *CDKA1*, *CDC3*, *MCM3* and *CYCA2.1* (Yu et al. 2017). Recently, Guo et al. (2018) demonstrated that *GSN1* which encodes mitogen-activated protein kinase phosphatase is a negative regulator of OsMKK10-OsMKK4-OsMPK6 cascade for cell differentiation and proliferation in spikelet hull and panicle. Meanwhile, loss of function of *TGW6* gene for endosperm length and grain filling in rice was demonstrated to lead to long grain through reduced expressions of core cell cycle genes including *CycB2;2* (Ishimaru et al. 2013). In this study, expressions of several cell cycle genes including *CYCB2-2* was also found to be reduced through RNA-Seq analysis of CRISPR/Cas9-mediated *LGG* knockout lines (**Fig. 7a**). Since *CycB2;2* works with the highest level at the G2-M transition and during M phase (de Veylder et al. 2007), it was likely that cell division in the spikelet hull primordium might delay with lower expressions of crucial cell cycle genes including *CYCB2-2* in CRSPR/Cas9-mediated knockout lines, resulting in increase of cell number of the spikelet hull.

Phylogenetic analysis indicated that Os11g0637700 (LGG) and AtRBP1, which was first identified in Arabidopsis (Suzuki et al. 2000), belong to the Musashi RBP family (Okabe et al. 2001) (**Fig. 6b**), although Os11g0637700 was in a grass clade and distantly related to Musashi-1 (Msi1). *LGG* was highly expressed in the spikelet hull primordium. The spikelet hull is a striking characteristic in the Gramineae (Kellogg 2001), and the originating organ for bract or sepal has been a controversial topic. Recently, Lombardo and Yoshida (2015) reported that the spikelet hull is a modified sepal based on expressions of key genes in the sepals of dicots. Thus, RBPs in the grass clade, including LGG, are likely expressed in the grass-specific spikelet hull. In this study, *LGG* transcripts were detected at root and third leaf of 2-week-old seedling together with young spikelet. RiceXPro (http://ricexpro.dna.affrc.go.jp/field-development.php?featurenum=26064) demonstrated that *LGG* (*Os11g0637700*) is expressed at vegetative and reproductive tissues, suggesting that *LGG* expressed globally. Okabe et al. (2001) reported that Msi1 controlled asymmetric cell division in the nervous system of *Drosophila*. AtRBP1 was proposed to regulate cell proliferation progression in meristematic tissues, such as the shoot apex and root tip (Suzuki et al. 2000), and it was reported that underexpression and overexpression plants showed similar stunted root growth, although the reason remained unknown (Shida et al. 2015). In this study, RNA-Seq analysis of GE and OE lines suggested that LGG regulates the gene(s) for cell division specifically during spikelet hull differentiation through. Further, small interfering RNA knockdown of Msi1 downregulated cell-cycle genes and upregulated cell-cycle inhibitor genes in Ishikawa cells (Götte et al. 2011). This antagonistic relationship was also observed in the relation between CYCD3-2 and SMR1 in GE, OE and NP in this study. These results suggest that the Musashi RBP family might regulate genes controlling the progression of cell cycle.

Taken together, it is suggested that *LGG* might regulate target genes for cell cycle and cell division in spikelet hull primordium. The mutant gene, *Lgg* induced by *nDart1*-insertion and deletion event is presumed to alter the target genes expressions quantitatively. Cosegregation between grain length and *Lgg* in the F2 suggested that *Lgg* might be a hypomorphic allele in an incomplete dominance manner. Of 42 genes for rice grain size (Li et al. 2018), *GW8* (Wang et al. 2012), *SRS5* (Segami et al. 2012), *GL7* (Wang et al. 2015) and *SLG* (Feng et al. 2016) were reported to be inherited in an incomplete dominance manner. Springer et al. (2010) demonstrated that although at least 80% of 730 genes tested in yeast showed dosage effect for protein production, most of the genes produced sufficient protein for normal phenotype in heterozygote. Meanwhile, it is well known that Waxy protein for amylose content of rice endosperm is regulated by dosage effect of *Wx* gene (Sano 1984). Johzuka-Hisatomi et al. (2011) quantitatively elucidated incomplete dominance nature of *CHS-D* dosage effect for anthocyanin pigmentation in flower of the morning glory through levels of transcript and CHS protein together with anthocyanin content of flower. Bollmann et al. (1991) reported that *niv-525* and *niv-572* alleles of *Niv* locus were semidominant with different nature from dosage effect in *Antirrhinum*. Since hypomorphic allele of RBP have not been so far reported in agronomically important trait, the incomplete dominance nature of *Lgg* could be proved through segregation analysis of grain size in F2 of the cross between GE and WT.

Experimental Procedures

Plant materials and trait measurement

Lgg was found among over 3,000 *nDart1*-tagged lines in the genetic background of cv. Koshihikari. Eighty-five F2 plants of the cross between *Lgg* and Koshihikari together with the parents were grown in a paddy field at Kurashiki, Okayama, Japan, in 2014, and 40 F3 lines with 10 plants in each line selected randomly from the F2 population were grown for progeny testing in 2015. Culm length, panicle length, number of panicles per plant, culm-base diameter and panicle-base diameter of the longest tiller were measured at the mature stage. Number of spikelets per panicle, spikelet fertility, 100-grain weight and total grain weight per plant were examined after harvesting and drying. Days to heading were counted from sowing date to the emergence date of the first panicle. The length, width and thickness of hulled grains were measured by rice grain quality checker (RGQI 20A, Satake). Spikelet hull length was measured by a vernier caliper.

Histological analysis

To measure cell size and number in the spikelet hull, lemmas of the spikelet hull sampled at the heading day were subject to longitudinal cryotomic sectioning by the Kawamoto method (Kawamoto and Kawamoto 2014), as described by Chiou et al. (2018). Photos of the sections were taken on a light microscope (BZ-800, Keyence), and Image J was used for analyzing the cell size and cell number.

DNA extraction and transposon display

Genomic DNA was extracted from leaf tissues, as described by Gichuhi et al. (2016). Transposon display was performed as described by Takagi et al. (2007).

RNA extraction, rapid amplification of cDNA ends (RACE) and qRT-PCR

RNA was extracted from the third leaves and roots of 2-week-old seedlings and 0.6, 1, 2 and 3 mm young panicles using RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. DNase (TURBO DNA-free Kit, Ambion) was used for removing contaminated DNA. 5′ RACE was performed by SMARTer RACE 5′/3′ Kit (Clontech) using the gene-specific primers, 15De02_R and 15Nv01_R, according to the manufacturer's instructions. 3′ RACE was performed using the reverse transcription kit (PrimeScript RT Reagent Kit, Takara) using an adapter primer and gene-specific primers, 15De03_F and 15Nv05_F. The RACE products were amplified through TA cloning (pGEM-T Easy vector, Promega) for sequencing. qRT-PCR was conducted using PrimerScript RT reagent Kit (Perfect Real Time) (Takara) and SYBR Premix Ex Taq (Perfect Real Time) (Takara) by LightCycler 2.0 (Roche) with three biological replicates and three technical replicates. *Ubiquitin* was used as an internal control, and the gene-specific primers 15Nv05_F and 15Nv06_R were used.

Vector construction and plant transformation

The overexpression vector pLGG:LGG^{Koshi} carried a 7.3 kb fragment including the 2.5 kb promoter and 4.8 kb LGG coding region in pCAMBIA1305.1. The 7.3 kb LGG fragment was assembled by In-fusion cloning kit (Takara) from four fragments (Supplementary Table S6) amplified by high fidelity Taq, PrimeSTAR MAX (Takara), and was sequenced to confirm its identity to the genomic region. The subcellular localization vector was derived from pCAMBIA1305.1. p35S:LGG-GFP was assembled from LGG cDNA (AK105751) and EGFP fragments (Supplementary Table S6) with the In-fusion cloning kit. The LGG native promoter vector, pLGG:LGG^{NP}-GFP was based on p35S:LGG-GFP, and the assembled 2.5 kb promoter region of LGG. The genome-editing vector, pZH_gYSA_MMC as9_LGG600 used the guide sequence **CCG**CGC CCGGGGATT C-GGCTTCA (protospacer-adjacent motif in bold) according to the construction protocol described in Mikami et al. (2015). The vectors were transformed into *Agrobacterium tumefaciens* strain EHA105, and *Agrobacterium*-mediated transformations were carried out as described by Shimatani et al. (2009).

Amino acid sequence and phylogenetic analysis

Amino acid sequences of 102 LGG orthologs from 17 species using the full length of amino acid of LGG were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/) and the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/). Sequence alignment and phylogenetic analysis were performed with MEGA 6.0.

cDNA library preparation and RNA sequencing

RNA samples were isolated from 1-mm-long young panicles of NP, GE and OE with three biological replicates. RNA integrity and quality were checked by 2100 Bioanalyzer (Agilent Technologies). Isolated total RNAs were treated with DNase I and enriched mRNAs using magnetic beads with Oligo(dT) (BGI). Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments. The first strand-cDNA was synthesized using random hexamer-primers. For size selection, short fragments, around 200 bp were purified and resolved for end reparation and adenine addition, then connected with adapters (Zhang et al. 2010). The cDNA libraries were prepared according to Illumina's protocol. Sequencing was performed on Illumina HiSeq 4000 platform generating 100-bp paired-end reads (BGI).

Mapping of NGS result to the rice genome and annotation of expressed genes

The rice genome sequence, Os-Nipponbare-Reference-IRGSP-1.0 (https://rapdb.dna.affrc.go.jp) was used for data analysis in this study. Data sets of next-generation sequencing (NGS) RNA sequences were analyzed using following protocol (Trapnell et al. 2012). After sequencing, the raw reads were filtered by removing adaptor sequences, contamination and low-quality reads. TopHat version 2.1.1 was used for aligning the reads to the genome and determine the splicing sites of transcript.

For assembling the reads to the genome, Cufflinks version 2.2.1(http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/) was used as reference genome annotation, the Ensemble Plants *O. sativa japonica* (https://plants.ensembl.org/Oryza_sativa/Info/Index). Then, Cuffmerge was used for assembling the transcripts from individual reads. Finally, Cuffdiff calculated differentially expressed reads. Results of Cuffdiff were analyzed by R statistical analysis and plotting package version 3.4.3 with CummeRbund 2.22.0 (https://bioconductor.org/packages/release/bioc/html/cummeRbund.html). GO analysis were used g:profiler (https://biit.cs.ut.ee/gprofiler/).

Supplementary Data

Supplementary data are available at PCP online.

Acknowledgments

We thank Ms. Midori Kawamoto (Radioisotope Research Institute, School of Dental Medicine, Tsurumi University, Japan), and Ms. Hiroyo Nishide (NIBB) for her technical support. Some experiments were performed with equipment maintained by the Functional Genomics and Model Plant Facility, and computational resources were provided by the Data Integration and Analysis Facility, in NIBB.

Funding

The Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan [No. 16K07561]; the Joint Research Program implemented at the Institute of Plant Science and Resources, Okayama University in Japan [No. 2942 to K.T.] and the NIBB Cooperative Research Program [17-331 to M.M.].

Disclosures

The authors have no conflicts of interest to declare.

References

Ambrosone, A., Costa, A., Leone, A. and Grillo, S. (2012) Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. *Plant Sci.* 182: 12–18.

Aya, K., Hobo, T., Sato-Izawa, K., Ueguchi-Tanaka, M., Kitano, H. and Matsuoka, M. (2014) A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol.* 55: 897–912.

Bollmann, J., Carpenter, R. and Coen, E.S. (1991) Allelic interactions at the *nivea* locus of *Antirrhinum*. *Plant Cell.* 3: 1327–1336.

Chin, H.G., Choe, M.S., Lee, S.-H., Park, S.H., Park, S.H., Koo, J.C., et al. (1999) Molecular analysis of rice plants harboring an Ac/Ds transposable element-mediated gene trapping system. *Plant J.* 19: 615–623.

Chiou, W.-Y., Tsugane, K., Kawamoto, T. and Maekawa, M. (2018) Easy sectioning of whole grain of rice using cryomicrotome. *Breed. Sci.* 68: 381–384.

Collins, C., Maruthi, N.M. and Jahn, C.E. (2015) CYCD3 D-type cyclins regulate cambial cell proliferation and secondary growth in *Arabidopsis*. *J. Exp. Bot.* 66: 4595–4606.

da Costa-Nunes, J.A., Bhatt, A.M., O'Shea, S., West, C.E., Bray, C.M., Grossniklaus, U., et al. (2006) Characterization of the three *Arabidopsis thaliana RAD21* cohesins reveals differential responses to ionizing radiation. *J. Exp. Bot.* 57: 971–983.

Deng, Z.Y., Liu, L.T., Li, T., Yan, S., Kuang, B.J., Huang, S.J., et al. (2015) OsKinesin-13A is an active microtubule depolymerase involved in glume length regulation via affecting cell elongation. *Sci. Rep.* 5: 9457.

de Veylder, L., Beeckman, T. and Inzé, D. (2007) The ins and outs of the plant cell cycle. *Nat. Rev. Mol. Cell Biol.* 8: 655–665.

Fedoroff, N. (2002) RNA-binding proteins in plants: the tip of an iceberg? *Curr. Opin. Plant Biol.* 5: 452–459.

Feng, Z., Wu, C., Wang, C., Roh, J., Zhang, L., Chen, J., et al. (2016) *SLG* controls grain size and leaf angle by modulating brassinosteroid homeostasis in rice. *J. Exp. Bot.* 67: 4241–4253.

Gichuhi, E., Himi, E., Ahmed, N., Takahashi, H. and Maekawa, M. (2016) Preliminary QTL detection for improving basmati rice in a F_2 population derived from the cross between kernel basmati and pLIA-1 carrying *Oryza longistaminata* chromosome. *SABRAO J. Breed. Genet.* 48: 402–415.

Gong, C., Li, T., Li, Q., Yan, L. and Wang, T. (2011) Rice *OsRAD21-2* is expressed in actively dividing tissues and its ectopic expression in yeast results in aberrant cell division and growth. *J. Integr. Plant Biol.* 53: 14–24.

Götte, M., Greve, B., Kelsch, R., Müller-Uthoff, H., Weiss, K., Kharabi Masouleh, B., et al. (2011) The adult stem cell marker Musashi-1 modulates endometrial carcinoma cell cycle progression and apoptosis *via* Notch-1 and $p21^{WAF1/CIP1}$. *Int. J. Cancer* 129: 2042–2049.

Greco, R., Ouwerkerk, P.B.F., de Kam, R.J., Sallaud, C., Favalli, C., Colombo, L., et al. (2003) Transpositional behaviour of an *Ac*/*Ds* system for reverse genetics in rice. *Theor. Appl. Genet.* 108: 10–24.

Greco, R., Ouwerkerk, P.B.F., Taal, A.J.C., Sallaud, C., Guiderdoni, E., Meijer, A.H., et al. (2004) Transcription and somatic transposition of the maize *En*/*Spm* transposon system in rice. *Mol. Genet. Genomics* 270: 514–523.

Guo, T., Chen, K., Dong, N.-Q., Shi, C.-L., Ye, W.-W., Gao, J.-P., et al. (2018) *GRAIN SIZE AND NUMBER1* negatively regulates the OsMKKK10-OsMKK4-OsMPK6 cascade to coordinate the trade-off between grain number per panicle and grain size in rice. *Plant Cell.* 30: 871–888.

Hamdoun, S., Zhang, C., Gill, M., Kumar, N., Churchman, M., Larkin, J.C., et al. (2016) Differential roles of two homologous cyclin-dependent kinase inhibitor genes in regulating cell cycle and innate immunity in Arabidopsis. *Plant Physiol.* 170: 515–527.

Hayashi-Tsugane, M., Maekawa, M. and Tsugane, K. (2015) A gain-of-function *Bushy dwarf tiller 1* mutation in rice microRNA gene *miR156d* caused by insertion of the DNA transposon *nDart1*. *Sci. Rep.* 5: 14357.

Hayashi-Tsugane, M., Takahara, H., Ahmed, N., Himi, E., Takagi, K., Iida, S., et al. (2014) A mutable albino allele in rice reveals that formation of thylakoid membranes requires the *SNOW-WHITE LEAF1* gene. *Plant Cell Physiol.* 55: 3–15.

Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19: 327–341.

Hirochika, H. (2010) Insertional mutagenesis with *Tos17* for functional analysis of rice genes. *Breed. Sci.* 60: 486–492.

Hoshikawa, K. (1989) The Growing Rice Plant: An Anatomical Monograph. p. 301. Nobunkyo, Tokyo.

Howard, T.P., Hayward, A.P., Tordillos, A., Fragoso, C., Moreno, M.A., Tohme, J., et al. (2014) Identification of the maize gravitropism gene *lazy plant1* by a transposon-tagging genome resequencing strategy. *PLoS One* 9: e87053.

Ikeda-Kawakatsu, K., Yasuno, N., Oikawa, T., Iida, S., Nagato, Y., Maekawa, M., et al. (2009) Expression level of *ABERRANT PANICLE ORGANIZATION1* determines rice inflorescence form through control of cell proliferation in the meristem. *Plant Physiol.* 150: 736–747.

Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013) Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* 45: 707–711.

Jiang, S.-Y. and Ramachandran, S. (2010) Natural and artificial mutants as valuable resources for functional genomics and molecular breeding. *Int. J. Biol. Sci.* 6: 228–251.

Johzuka-Hisatomi, Y., Noguchi, H. and Iida, S. (2011) The molecular basis of incomplete dominance at the *A* locus of *CHS-D* in the common morning glory, *Ipomoea purpurea*. *J. Plant Res.* 124: 299–304.

Kawamoto, T. and Kawamoto, K. (2014) Preparation of thin frozen sections from nonfixed and undecalcified hard tissues using Kawamot's film method (2012). *Methods Mol. Biol.* 1130: 149–164.

Kellogg, E.A. (2001) Evolutionary history of the grasses. *Plant Physiol.* 125: 1198–1205.

Komaki, S. and Sugimoto, K. (2012) Control of the plant cell cycle by developmental and environmental cues. *Plant Cell Physiol.* 53: 953–964.

Li, N. and Li, Y. (2016) Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* 33: 23–32.

Li, N., Xu, R., Duan, P. and Li, Y. (2018) Control of grain size in rice. *Plant Reprod.* 31: 237–251.

Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43: 1266–1269.

Lombardo, F. and Yoshida, H. (2015) Interpreting lemma and palea homologies: a point of view from rice floral mutants. *Front. Plant Sci.* 6: 61.

Lorković, Z.J. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 14: 229–236.

Maris, C., Dominguez, C. and Allain, F.H.-T. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* 272: 2118–2131.

McCouch, S.R., Wright, M.H., Tung, C.-W., Maron, L.G., McNally, K.L., Fitzgerald, M., et al. (2016) Open access resources for genome-wide association mapping in rice. *Nat. Commun.* 7: 10532.

Mikami, M., Toki, S. and Endo, M. (2015) Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. *Plant Mol. Biol.* 88: 561–572.

Morris, R.T., Doroshenk, K.A., Crofts, A.J., Lewis, N., Okita, T.W. and Wyrick, J.J. (2011) RiceRBP: a database of experimentally identified RNA-binding proteins in *Oryza sativa* L. *Plant Sci.* 180: 204–211.

Nakamura, M., Okano, H., Blendy, J.A. and Montell, C. (1994) Musashi, a neural RNA-binding protein required for *Drosophila* adult external sensory organ development. *Neuron* 13: 67–81.

Okabe, M., Imai, T., Kurusu, M., Hiromi, Y. and Okano, H. (2001) Translational repression determines a neuronal potential in *Drosophila* asymmetric cell division. *Nature* 411: 94–98.

Peres, A., Churchman, M.L., Hariharan, S., Himanen, K., Verkest, A., Vandepoele, K., et al. (2007) Novel plant-specific cyclin-dependent kinase inhibitors induced by biotic and abiotic stresses. *J. Biol. Chem.* 282: 25588–25596.

Ray, D.K., Mueller, N.D., West, P.C. and Foley, J.A. (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8: e66428.

Sano, Y. (1984) Differential regulation of *waxy* gene expression in rice endosperm. *Theor. Appl. Genet.* 68: 467–473.

Segami, S., Kono, I., Ando, T., Yano, M., Kitano, H., Miura, K., et al. (2012) *Small and round seed 5* gene encodes alpha-tubulin regulating seed cell elongation in rice. *Rice (N Y)* 5: 4.

Segami, S., Takehara, K., Yamamoto, T., Kido, S., Kondo, S., Iwasaki, Y., et al. (2017) Overexpression of *SRS5* improves grain size of brassinosteroid-related dwarf mutants in rice (*Oryza sativa* L.). *Breed. Sci.* 67: 393–397.

Shida, T., Fukuda, A., Saito, T., Ito, H. and Kato, A. (2015) *AtRBP1*, which encodes an RNA-binding protein containing RNA-recognition motifs, regulates root growth in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 92: 62–70.

Shimatani, Z., Takagi, K., Eun, C.-H., Maekawa, M., Takahara, H., Hoshino, A., et al. (2009) Characterization of autonomous *Dart1* transposons belonging to the *hAT* superfamily in rice. *Mol. Genet. Genomics* 281: 329–344.

Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* 40: 1023–1028.

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016) *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48: 447–456.

Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z. and Lin, H.-X. (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39: 623–630.

Springer, M., Weissman, J.S. and Kirschner, M.W. (2010) A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* 6: 368.

Suzuki, M., Kato, A. and Komeda, Y. (2000) An RNA-binding protein, AtRBP1, is expressed in actively proliferative regions in *Arabidopsis thaliana*. *Plant Cell Physiol.* 41: 282–288.

Takagi, K., Ishikawa, N., Maekawa, M., Tsugane, K. and Iida, S. (2007) Transposon display for active DNA transposons in rice. *Genes Genet. Syst.* 82: 109–122.

Takeda, K. (1997). Inheritance of morphological characters: spikelets. *In* Science of the Rice Plant: Genetics. Edited by Matsuo, T., Futsuhara, Y., Kikuchi, F. and Yamaguchi, H. pp. 294–299. FAPRC, Tokyo.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7: 562–578.

Tsugane, K., Maekawa, M., Takagi, K., Takahara, H., Qian, Q., Eun, C.-H., et al. (2006) An active DNA transposon *nDart* causing leaf variegation and mutable dwarfism and its related elements in rice. *Plant J.* 45: 46–57.

Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012) Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* 44: 950–954.

Wang, N., Long, T., Yao, W., Xiong, L., Zhang, Q. and Wu, C. (2013) Mutant resources for the functional analysis of the rice genome. *Mol. Plant.* 6: 596–604.

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., et al. (2015) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat. Genet.* 47: 944–948.

Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008) Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res.* 18: 1199–1209.

Xuan, Y.H., Zhang, J., Peterson, T. and Han, C. (2012) *Ac*/*Ds*-induced chromosomal rearrangements in rice genomes. *Mob. Genet. Elements* 2: 67–71.

Yoshida, A., Sasao, M., Yasuno, N., Takagi, K., Daimon, Y., Chen, R., et al. (2013) *TAWAWA1*, a regulator of rice inflorescence architecture, functions through the suppression of meristem phase transition. *Proc. Natl. Acad. Sci. USA* 110: 767–772.

Yu, J., Xiong, H., Zhu, X., Zhang, H., Li, H., Miao, J., et al. (2017) *OsLG3* contributing to rice grain length and yield was mined by Ho-LAMap. *BMC Biol.* 15: 28.

Zhang, J., Yu, C., Pulletikurti, V., Lamb, J., Danilova, T., Weber, D.F., et al. (2009) Alternative *Ac*/*Ds* transposition induces major chromosomal rearrangements in maize. *Genes Dev.* 23: 755–765.

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., et al. (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20: 646–654.