Genetic analysis of grain size using DNA transposon-tagged lines in rice

2019, March

Wan-Yi Chiou

Graduate School of Environmental and Life Science

(Doctor's Course)

OKAYAMA UNIVERSITY

Table of Contents

Abstract

Grain size is a key character for grain weight, one of the yield components in rice. To increase the grain yield of rice, it is important to elucidate the detailed regulatory mechanism for grain size. Although many genes for grain size have been reported, the precise regulation networks are still unrevealed. *Large grain* (*Lgg*) mutant found in the *non-autonomous DNA-based active rice transposon1* (*nDart1*)-tagged lines of Koshihikari, showed a long grain phenotype (109.3%) with long panicle (106.9%) and heavy 100-grain weight (109.2%) compared to WT. The heavier 100-grain weight was considered to be mainly due to the longer grains of *Lgg*. F2 segregation of *Lgg* and F3 test demonstrated that *Lgg* was inherited in an incomplete dominance manner. The causal gene for *Lgg* was revealed to encode a putative RNA-binding protein (RBP) with two RNA recognition motifs (RRMs) was located on chromosome 11 through transposon display analysis. Further, *Lgg* was suggested to be caused by truncated *nDart1-3* and 355 bp deletion at 5′ untranslated region (UTR) region of a putative RBP gene (*LGG*). Meanwhile, CRISPR/Cas9-mediated knockout and overexpressed plants for the RBP showed longer and shorter grains, respectively, than WT, proving that LGG regulated spikelet hull length. Expression of *LGG* was found to show the highest level at 0.6-mm-long young panicle and was gradually decreased depending on the length of young panicle, though *LGG* expressed at root and leaf. These results showed that LGG strongly functioned 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 WTs, suggesting that LGG might regulate cell division or cell proliferation of spikelet hull in the longitudinal direction. Through RNA-Seq analysis of 1-mm-long young panicles of knockout and overexpressed plants, expressions of many cell cycle-related genes were reduced in the knockout plants compared to those of overexpressed plants and WT, whereas some genes for cell proliferation expressed highly in knockout plants. Taken together, LGG was suggested to be a potential pivotal regulator for cell

cycle and cell division.

As a result, this study shows that *LGG* encodes a putative RBP for regulating cell cycle-related genes in spikelet hull primordium of rice and suggests that LGG might be useful for improving rice yield by combining with some gene for phytohormonal regulation for cell elongation in the spikelet hull of rice.

Abbreviations

ABA	abscisic acid
aDart	*autonomous DNA-based active rice transposon*
AFLP	amplified fragment-length polymorphism
AP2	APETALA2
APO1	*ABERRANT PANICLE ORGANIZATION 1*
BR	brassinosteroid
CDK	cyclin-dependent kinase
CYC	cyclin
DAPI	4′,6-Diamidino-2-phenylindole
DEP1	*DENSE AND ERECT PANICLE1*
DS-RBD	double-stranded RNA binding domain
EJC	exon junction complex
EMS	ethyl methanesulfonate
EREBP	ethylene-responsive element binding protein
ERF	ethylene response factor
GIF	GRF-interacting factor
GO	gene ontology
GRF	GROWTH -REGULATING FACTOR
GWAS	genome-wide association studies
KH	K homology
LGD1	*LAGGING GROWTH AND DEVELOPMENT 1*
Lgg	*Large grain* mutant
MAPK	mitogen-activated protein kinase
MEL	*MEIOSIS ARRESTED AT LEPTOTENE*
miRNA	microRNA
MNU	methyl nitrosourea
nDart1	*non-autonomous DNA-based active rice transposon1*
PAZ	Piwi/Argonaute/Zwille
PPKL	protein phosphatase kelch
pyl-v	*pale-yellow-leaf variegated*
QTL	quantitative trait locus
RACE	rapid amplification of cDNA ends

RBP	RNA-binding protein
RNP	ribonucleoprotein
RRM	RNA recognition motif
SBP	*SQUAMOSA* PROMOTER BINDING PROTEIN
SLG	*SLENDER GRAIN*
SMOS	*SMALL ORGAN SIZE*
SPL	*SQUAMOSA* PROMOTER BINDING PROTEIN-LIKE
SRS	*SMALL AND ROUND SEED*
SWL1	*SNOW-WHITE LEAF1*
TAW1	*TAWAWA1*
TD	transposon display
TE	transposable element
TF	transcription factor
TIR	terminal inverted repeat
TIS	transcript initiation site
TSD	target site duplication
UTR	untranslated region

Chapter 1 General Introduction

Rice (*Oryza sativa* L.) was domesticated around 10,000 years ago (Huang et al., 2012). It is a staple food for half of the world's population that supplies around one-fifth of dietary calories for the global population (Fitzgerald et al., 2009). Additionally, the rice whole genome sequence was released in 2005 and shared synteny with other cereal crops (IRGSP, 2005). Thus, rice becomes a model plant of monocotyledon for functional genetic analysis. Since the 'green revolution' in the 1960s that introduced semi-dwarf varieties and fertilizer application, rice production had been increased during the 30 years and had lower the production cost (Khush, 1999). However, it is still necessary to increase food production and quality to meet the rapid growth of the global population (Ray et al., 2013).

There are four yield components in rice: panicle number per unit area, spikelet number per panicle, percent of filled grains per panicle, and grain weight. Of these factors, grain weight is mainly determined by grain size, which has high heritability and is thus suitable for genetic analysis (Li et al., 2004). Over 50 quantitative trait loci (QTLs) and genes involved in the regulation of grain size in rice have been reported in the past two decades (Table1-1). Understanding the genetic mechanism responsible for grain size is very important in the attempt to modify the grain size and increase rice yield. Several rice insertional mutant libraries have been established, such as T-DNA, *Ds*/*dSpm*, and *Tos17* (Wang et al., 2013) and have been useful for identifying the genes underlying a trait of interest. Further, the active DNA transposon, *non-autonomous DNA-based active rice transposon1* (*nDart1*) is an active mutagen and *aDart/nDart* tagging system is useful for generating natural variations (Tsugane et al., 2006). Therefore, it is worth to use *nDart1*-tagged lines to screen for grain size related genes.

The main objective of this study is to identify the gene controlling grain size in rice using *Large*

grain (*Lgg*) mutant found in *nDart1*-tagged lines and to analyze the structure and function of the gene together with its possible regulation pathway.

1-1 Genes and QTLs related to grain size in rice

The grain size is determined by the coordination of caryopsis growth and the size of the spikelet hull or the husk, which consists of lemma and palea (Yoshida, 1981). Since the lemma and palea formed a closed space for caryopsis growth when the caryopsis was fully developed, the spikelet hull limits the grain size (Deng et al., 2015). Many QTLs and genes related to grain size have been fine mapped or cloned for functional analysis (Table1). Some genes, such as *GS3* (Takano-Kai et al., 2009) and *qSW5*/*GW5* (Shomura et al., 2008), are presumed to be selected during domestication. Here, I summarize three signaling pathways, transcription factors, microRNAs, phytohormones, and transporters for grain filling implicated in rice grain size regulation network.

Heterotrimeric G-protein signaling, mitogen-activated protein kinase signaling, and ubiquitous-proteasome pathway

The negative regulator, *GS3* (*Os03g0407400*), which is a major QTL on chromosome 3, is the first identified gene of grain size (Fan et al., 2006). *GS3* is a homolog of *AGG3* (a plant-specific G protein γ subunit) in *Arabidopsis* and contains four domains, the plant-specific organ size regulator (OSR) domain, the transmembrane domain, tumor necrosis factor receptor/nerve growth factor receptor (TNFR/NGFR) family cysteine-rich domain, and the von Willebrand factor type C (VWFC) domain (Mao et al., 2010; Chakravorty et al., 2011). A single nucleotide substitution on the second exon of *GS3*, cytosine (C) to adenine (A) causes premature termination and makes grain length increased (A allele). This allele is a target during domestication through haplotype analysis using 235 accessions of *O. sativa* and 401 accessions

of other *Oryza* species (Takano-Kai et al., 2009). *DEP1* (*DENSE AND ERECT PANICLE1*) regulated panicle and grain yield (Huang et al., 2009b). Although functions of both *GS3* and *DEP1* were not elucidated at that time, they were confirmed belonging to C-type γ subunit of G protein (Trusov et al., 2012).

Mitogen-activated protein kinase (MAPK) modules, which involve in environmental and developmental signal transduction, consist of several serine/threonine kinases, MAP kinase kinase kinase (MKKK), MAP kinase kinase (MKK;), and MAPK (Jagodzik et al., 2018). The OsMKKK10 (Guo et al., 2018), OsMKK4 (Duan et al. 2014), and OsMAPK6 (Liu et al., 2015b) cascade is a positive regulator for rice grain size. Additionally, MAPK phosphatase OsMPK1 is a negative regulator that directly inhibits the activity of OsMAPK6 (Guo et al., 2018; Xu et al., 2018).

GW2 is a major QTL for grain width and weight, encoding a RING-type E3 ubiquitin ligase, which takes part in the ubiquitous-proteasome pathway. Song et al. (2007) demonstrated that GW2 is a negative regulator, through analysis of loss of function near-isogenic line showing larger spikelet hull and a higher rate of grain milk filling.

Transcription factors and microRNAs

The transcription factors (TFs) are the proteins containing the DNA-binding domain(s), regulating the downstream gene expressions for developmental processes and stress responses. MicroRNAs (miRNAs) are 20 to 24 nucleotides small RNAs that control gene expression by silencing of target mRNAs (Rogers and Chen, 2013). *SQUAMOSA* PROMOTER BINDING PROTEIN-LIKE (SPL) and GROWTH -REGULATING FACTORs (GRFs) are plant-specific TFs that have been reported to regulate plant development and are controlled by miR156 and miR396, respectively. *SPL* genes contain a DNA-binding domain, which is the *SQUAMOSA* PROMOTER BINDING PROTEIN (SBP) domain (Xie et al., 2006). *GW8*/*OsSPL16* regulates grain shape, and down-regulation of *GW8*/*OsSPL16* results in slender grain (Wang et al., 2012).

Another report further indicated that OsSPL16 targets the promoter of *GW7* (*LOC_Os07g41200*) and negatively regulates the expression of *GW7* for the grain size (Wang et al., 2015a). On the other hand, *GLW7*/*OsSPL13* (*LOC_Os07g32170*) was identified through genome-wide association study (GWAS) using 381 varieties, and small-grain haplotypes with two CACTTC tandem repeat in the 5′ untranslated region (UTR) led to a reduction of the expression of *OsSPL13* (Si et al., 2016). OsSPL13 targets *SMALL AND ROUND SEED 5* (*SRS5*) and *DEP1*, positively regulating their expression level to modulate the panicle architecture and grain size (Si et al., 2016).

GRFs contain two conserved regions. One is QLQ (Gln, Leu, Gln) motif, functions in protein-protein interaction, another is WRC (Trp, Arg, Cys) required for DNA binding. GRFs cooperate with the transcriptional coactivator, GRF-interacting factors (GIFs), and form a complex. Moreover, miR396 targets several GRFs and represses their expression levels (Omidbakhshfard et al., 2015). The OsGRF4 is a positive regulator of grain size and the miR396c-OsGRF4-OsGIF1 module was reported to control grain size and involved in brassinosteroid (BR) signaling (Che et al., 2015; Duan et al., 2015; Hu et al., 2015). Additionally, GRF6 as a target of miR396d positively affects panicle architecture, and the downstream pathway involved in auxin biosynthesis and signaling (Gao et al., 2015). Two other microRNAs modulate grain size, OsmiR397-*OsLAC* (Zhang et al., 2013b) and miR159 (Zhao et al., 2017) are positive regulators of grain size.

Phytohormones

Hormones are chemical compounds that interact with specific receptors for signal transduction to regulate growth and development. Currently, there are nine major hormones known in the plant: auxins, gibberellins, cytokinins, ethylene, abscisic acid (ABA), brassinosteroids (BRs), jasmonates, salicylic acid, and strigolactones. Many identified grain size genes related to phytohormones synthesis or signaling and mainly are BRs and auxins. For instance, *D2* (Hong

et al., 2003), *DWARF11* (Tanabe et al., 2005), *DWARF61*/*OsBRI1* (Morinaka et al., 2006; Zhao et al., 2013), and *PGL2*/*OsBUL1* (Heang and Sassa, 2012a; Jang et al 2017) are in BRs regulation pathway. *qSW5*/*GW5* encoding a calmodulin-binding protein with 1,212-bp deletion was also reported to regulate the grain size via BRs pathway (Shomura et al., 2008; Weng et al., 2008; Liu et al., 2017; Duan et al 2017). In auxins pathway, *tgw6,* encoding an indole-3-acetic acid (IAA)-glucose hydrolase (Ishimaru et al., 2013) and *BG1*, a novel membrane-localized protein, both are involved in auxin transport pathway (Liu et al., 2015a). However, *D2*, *DWARF11*, and *DWARF61* are directly involved in BR biosynthesis and signaling have pleiotropic effects on plant height and leaf angle (Hong et al., 2003; Tanabe et al., 2005; Morinaka et al., 2006)

Transporters for grain filling

The grain filling may determine the grain yield and grain quality. Over-filling will cause notched-bally grain (Deng et al., 2015; Lin et al., 2014); however, low-filling grain can cause chalky rice and reduce the yield (Wang et al., 2008). Transporters play crucial roles in the sink to source transport in grain filling. For example, *OsSWEET4* functioned in sugar allocation (Sosso et al., 2015).

Some identified genes are involved in several signaling pathways, heterotrimeric G-protein signaling, mitogen-activated protein kinase (MAPK) signaling, and ubiquitin-proteasome pathway. The other genes are transcriptional regulatory factors, microRNAs, phytohormones, and other grain size regulators (Zheng et al., 2015; Li and Li, 2015). Collectively, cell number and cell size of the spikelet hull are highly important for modification of the grain shape, and the identified genes are unevenly distributed in the genome, mainly located on chromosome 2, 3, 5 and 8, but none of them is on chromosome 12. Nevertheless, these pathways form a complicated network, and the regulatory network remains unclear.

Table 1-1 List of genes and QTLs related to grain size in rice

Locus / Gene / QTL Region	Chr	RAP-DB ID	MSUv7 ID	Reference
D2/CYP90D2/ebisu dwarf	1	Os01g0197100	LOC_Os01g10040	Hong et al 2003
BG3/OsPUP4	1	Os01g0680200	LOC_Os01g48800	Xiao et al., 2018
DWARF61/OsBRI1	1	Os01g0718300	LOC_Os01g52050	Morinaka et al 2006; Zhao et al 2013
qGRL1.1	1		LOC_Os01g66820 to LOC_Os01g66950*	Singh et al 2012
OsSGL	2	Os02g0134200	LOC_Os02g04130	Wang et al 2016
FUWA	2	Os02g0234200	LOC_Os02g13950	Chen et al 2015
GW2	2	Os02g0244100	LOC_Os02g14720	Song et al 2007
qLWR2	2		LOC_Os02g34530 to LOC_Os02g34680	Bai et al 2010
OsSWEET4	2	Os02g0301100	LOC_Os02g19820	Sosso et al 2015
GS2/OsGRF4/GL2	2	Os02g0701300	LOC_Os02g47280	Zhang et al 2013a; Hu et al 2015; Duan et al 2015; Che et al 2015
PGL2/OsBUL1	2	Os02g0747900	LOC_Os02g51320	Heang and Sassa 2012a; Jang et al 2017
SMG1/OsMKK4	2	Os02g0787300	LOC_Os02g54600	Duan et al 2013
TH1/BSG1/BLS1/AFD1	2	Os02g0811000	LOC_Os02g56610	Li et al 2012c; Peng et al., 2017
PGL1	3	Os03g0171300	LOC_Os03g07510	Heang and Sassa 2012 b
OsFBK12	3	Os03g0171600	LOC_Os03g07530	Chen et al 2013
BG1	3	Os03g0175800	LOC_Os03g07920	Liu et al 2015a
OsMADS87	3	Os03g0582400	LOC_Os03g38610	Chen et al 2016
qGL3/GL3.1/OsPPKL1	3	Os03g0646900	LOC_Os03g44500	Zhang et al 2012; Qi et al 2012.; Hu et al 2012
RGB1	3	Os03g0669100	LOC_Os03g46640	Utsunomiya et al 2011
OsGIF1	3	Os03g0733600	LOC_Os03g52320	Duan et al 2015
GS3	3	Os03g0407400	N / A	Fan et al 2006; Mao et al 2010
An-1	4	Os04g0350700	LOC_Os04g28280	Luo et al 2013
qGL4b	4		LOC_Os04g32740 to LOC_Os04g38150*	Kato et al 2011
GIF1/OsCIN2	4	Os04g0413500	LOC_Os04g33740	Wang et al 2008
XIAO	4	Os04g0576900	LOC_Os04g48760	Jiang et al 2012
DWARF11/CYP724B1	4	Os04g0469800	LOC_Os04g39430	Tanabe et al 2005

Table 1-1 Continued

Locus / Gene / QTL Region	Chr	RAP-DB ID	MSUv7 ID	Reference
OsMKKK10/YDA1/OsYODA-2	4	Os04g0559800	LOC_Os04g47240	Guo et al., 2018
CYP704A3	4	Os04g0573900	LOC_Os04g48460	Tang et al 2016
GSN1/OsMKP1	5	Os05g0115800	LOC_Os05g02500	Guo et al., 2018
APG	5	Os05g0139100	LOC_Os05g04740	Heang and Sassa 2012b; Heang and Sassa 2012c
SRS3/SGL/OsKinesin-13A	5	Os05g0154700	LOC_Os05g06280	Kitagawa et al 2010; Wu et al 2014; Deng et al 2015
GS5	5	Os05g0158500	LOC_Os05g06660	Li et al 2011; Xu et al 2015
qSW5/GW5	5	Os05g0187500	LOC_Os05g09520	Shomura et al 2008; Weng et al 2008; Liu et al 2017; Duan et al 2017
GSK2, *GSK3/SHAGGY-like*	5	Os05g0207500	LOC_Os05g11730	Tong et al 2012; Che et al 2015
OsAGSW1	5	Os05g0323800	LOC_Os05g25840	Li et al 2015
D1/RGA1	5	Os05g0333200	LOC_Os05g26890	Ashikari et al 1999; Fijisawa et al 1999
SMOS1	5	Os05g0389000	LOC_Os05g32270	Aya et al 2014; Hirano et al 2017
OsLAC	5	Os05g0458600	LOC_Os05g38420	Zhang et al 2013b
DLT/OsGRAS32/SMOS2	6	Os06g0127800	LOC_Os06g03710	Tong et al 2009; Hirano et al 2017
OsMAPK6	6	Os06g0154500	LOC_Os06g06090	Liu et al 2015b
HGW/UBA domain protein	6	Os06g0160400	LOC_Os06g06530	Li et al 2012a
BU1	6	Os06g0226500	LOC_Os06g12210	Tanaka et al 2009
GW6	6		LOC_Os06g33930 to LOC_Os06g35860*	Guo et al 2009
TGW6, *IAA-glucose hydrolase*	6	Os06g0623700	LOC_Os06g41850	Ishimaru et al 2013
GW6a, *H4 acetyltransferase*	6	Os06g0650300	LOC_Os06g44100	Song et al 2015
GL7/GW7/qSS7/GS7/SLG7	7	Os07g0603300	LOC_ Os07g41200	Shao et al 2012; Qiu et al 2012; Wang et al 2015b; Wang et al 2015a; Zhou et al 2015
GLW7/OsSPL13	7	Os07g0505200	LOC_Os07g32170	Si et al 2016
*OsBZR*1	7	Os07g0580500	LOC_Os07g39220	Zhu et al 2015
GE/CYP78A13	7	Os07g0603700	LOC_Os07g41240	Nagasawa et al 2013; Yang et al 2013
SRS1/DEP2	7	Os07g0616000	LOC_Os07g42410	Abe et al 2010
FZP/qGL7	7	Os07g0669500	LOC_Os07g47330	Bai et al 2017; Bai et al 2010
OsFIE1	8	Os08g0137250	LOC_Os08g04290	Folsom et al 2014
defective glume 1 (*dg1*)	8	Os08g0162800	LOC_Os08g06550	Yu et al 2017a

Table 1-1 Continued

Locus / Gene / QTL Region	Chr	RAP-DB ID	MSUv7 ID	Reference
OsBAK1/OsSERK1	8	Os08g0174700	LOC_Os08g07760	Li et al 2009; Yuan et al 2017
GAD1	8	Os08g0485500	LOC_Os08g37890	Jin et al 2016
GW8/OsSPL16	8	Os08g0531600	LOC_Os08g41940	Wang et al 2012
WTG1/OsOTUB1	8	Os08g0537800	LOC_Os08g42540	Huang et al 2017; Wang et al 2017
SLENDER GRAIN (*SLG*)	8	Os08g0562500	LOC_Os08g44840	Feng et al 2016
DEP1	9	Os09g0441900	LOC_Os09g26999	Huang et al 2009b; Zhou et al 2009
SG1	9	Os09g0459200	LOC_Os09g28520	Nakagawa 2012
OsFH15	9	Os09g0517600	LOC_Os09g34180	Sun et al 2017
gw9.1	9		LOC_Os09g36735 to LOC_Os09g36790*	Xie et al 2008
BRD2	10	Os10g0397400	LOC_Os10g25780	Hong et al 2005
DEL1	10	Os10g0457200	LOC_Os10g31910	Leng et al 2017
Cyclin-T1;3	11	Os11g0157100	LOC_Os11g05850	Qi et al 2012
SP1/OsNPF4.1	11	Os11g0235200	LOC_Os11g12740	Li et al., 2009b
SRS5	11	Os11g0247300	LOC_Os11g14220	Segami et al 2012
tgw11	11		LOC_Os11g45740 to LOC_Os11g47070*	Oh et al 2011
OsmiR156				Wang et al 2012
miR159				Zhao et al 2017
OsmiR396				Zhang et al 2013a; Hu et al 2015; Duan et al 2015; Che et al 2015
OsmiR397				Zhang et al 2013b
osa-miRf10422-akr				Tang et al 2016

*; Physical positions of QTL regions were estimated based on marker locations using BLAST.

The table includes the lists described by McCouch et al. (2016) and Li et al. (2018).

1-2 Active DNA transposon and *nDart1*-tagged lines

For gene functional analysis, one of the approaches is to use mutants or activate the genes. To disrupt genes in a random manner, physical (ion beam, γ-ray, fast neutron), chemical (ethyl methanesulfonate (EMS), sodium azide, methyl nitrosourea (MNU)), and insertional mutagenesis (T-DNA, transposable elements (TEs)) (Wei et al., 2013) have been used. Using endogenous TEs to generate insertional mutants have several merits, such as high efficiency for identifying the insertion sites though transposon display (TD) (Van den Broeck et al., 1998) or inverse PCR, without concerning about environmental issues, e.g. genetically modified pollen spread, and direct application of the mutants as breeding resource.

TEs are classified into two classes; class I or retrotransposon, that transposition mechanism is RNA-mediated, called 'copy-and-paste' and class II or DNA transposon that is able to transpose DNA in 'cut-and-paste' manner (Wicker et al., 2007). TEs occupy many portions in the genomes, and it was estimated to account for 13% of DNA transposons and 22% of retrotransposons in the rice genome (Rice Annotation Project et al., 2007). Although most of the TEs are silenced, some endogenous TEs were reported to be transposable in rice, such as *Tos17* (Hirochika et al., 1996) and *Karma* (Komatsu et al., 2003) of the retrotransposons, *mPing* (Kikuchi et al., 2003; Nakazaki et al., 2003), *nDart1* (*nonautonomous DNA-based active rice transposon1*) (Tsugane et al., 2006), *dTok* (Moon et al., 2006), and *nDaiZ* (Huang et al., 2009a) of the DNA transposons. Of them, *nDart1* is well characterized and has many advantages for functional analysis.

The *nDart1* elements with the 19-bp terminal inverted repeat (TIR) are about 0.6-kb long and belong to *hAT* superfamily, and 8-bp target site duplication (TSD) is formed after insertion. *nDart1* elements are divided into *nDart1-3*, *nDart1-101* and *nDart1-201* subgroups based on the subterminal sequences and their size (Fig. 1-1). *nDart1-0*, which belongs to the *nDart1-3* subgroup, was found in *OsClp5* on chromosome 3 in the *pale-yellow-leaf variegated* (*pyl-v*)

plant (Tsugane et al., 2006). The active autonomous *Dart* element, *aDart1-27* was also identified in the *pyl-v* plants. *aDart1-27* is 3.6-kb long and located on chromosome 6. Homologs of *aDart1-27*, which contain potential transposase, are highly methylated in Nipponbare (Eun et al., 2012). The elements of the *nDart1-3* subgroup, especially *nDart1-0*, are transposed frequently across the genome and are preferred to integrate into intragenic regions under the presence of *aDart1-27* (Takagi et al., 2010). One feature of the *nDart1* inserted mutants is showing the variegated phenotype. The mutant phenotype could be fixed when active *aDart1-27* is removed by recombination after crossing. Another one is that the later developed tillers that accumulate more *de novo* insertions of *nDart1* than the main culm in a rice plant (Hayashi-Tsugane et al., 2011).

The endogenous DNA transposon *aDart1*/*nDart1* tagging system is cost-effective (Maekawa et al., 2011). The *nDart1* can transpose under the presence of the active autonomous element *aDart1* without any other treatment, and the revertants of a *nDart1* insertion mutant often occur and leave transposon footprints after *nDart1* transposes. Thus, it provides additional evidence to prove the causal gene (Tsugane et al., 2006; Nishimura et al., 2008; Shimatani et al., 2009). Furthermore, the *nDart1*-tagged lines are useful to capture dominant phenotype for gene functional analysis, such as reported for *ABERRANT PANICLE ORGANIZATION 1* (*APO1*) (Ikeda-Kawakatsu et al., 2009), *TAWAWA1* (*TAW1*) (Yoshida et al., 2013), *SNOW-WHITE LEAF1* (*SWL1*) (Hayashi-Tsugane et al., 2013). Therefore, *aDart*/*nDart* tagging system is a powerful tool for functional analysis and possible improvement for variety in rice.

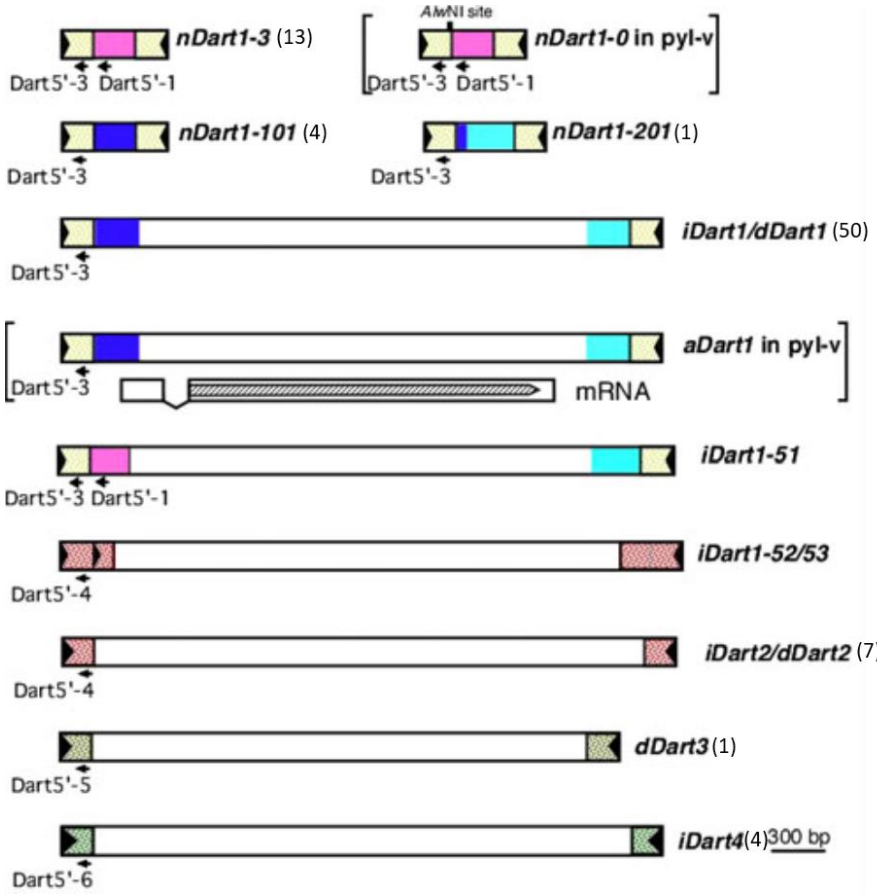

Fig. 1-1. *nDart1*-related elements in the Nipponbare genome. Numbers in parentheses indicate the homologs number (modified from Takagi et al., 2010).

Chapter 2 Characterization of *Large grain* mutant

2-1 Introduction

Owing to the increase in world population and climate change, we have more concern about food security than before. Rice (*Oryza sativa* L.) is a staple food of the world, especially in Asian countries. Rice has the smallest genome among monocotyledonous crops, and its complete genomic sequence has been available since 2005 (IRGSP, 2005). Rice grain yield consists of four main components, namely panicle number per plant, spikelet number per panicle, spikelet fertility and grain weight (Chandraratna, 1964; Yoshida, 1981). Grain weight is mainly determined by grain size, which is also an important trait for grain quality.

Wide variations in the length (3-11mm) and width (1.2-3.8mm) of grain size are observed in the rice germplasms (Fitzgerald et al., 2009). In general, the grain size is a quantitative trait and controlled by multiple genes, but grain length is a high heritable trait, suggesting that it is possibly determined by major genes and the phenotype is less affected by environmental influences (McCouch et al., 2016). There are more than 50 QTLs and genes related to grain size published and gave some insights to the complex network of rice grain development (Table 1-1; Zuo and Li, 2014; Li et al., 2018). The grain size is limited by the size of the spikelet hull, which composed of lemma and palea. The cell number and cell size are the factors determining the final organ size. It is necessary to dissect the spikelet hull.

Some QTLs or genes could not always be identified by map-based cloning method and genome-wide association studies (GWAS) because of their minor effects or rare alleles (Barabaschi et al., 2016). Unlike the map-based cloning method, *nDart1* transposon tagging system can provide another effective way to identify the causal gene of a mutant phenotype and analyze the function of uncharacterized genes (Tsugane et al., 2006).

In this chapter, to characterize *Large grain* (*Lgg*) mutant, which was found in endogenous DNA transposon *nDart1*-tagged lines of Koshihikari, the agronomic performance and the lemma morphology of *Lgg* and Koshihikari were examined.

2-2 Materials and Methods

2-2-1 Plant materials and traits

Koshihikari and *Lgg* were grown in the paddy field at IPSR, Kurashiki, Okayama, Japan, in 2014 and 2015. Culm length, panicle length, number of panicles per plant, culm-base diameter, and panicle-base diameter of the tallest tiller were measured at the mature stage (about 45 days after heading). Number of 1st branches per panicle, number of 2nd branches per panicle, number of spikelets per panicle, spikelet fertility (number of fertile grains/total grains/panicle), 100-grain weight, total grain weight per plant, and total panicle weight per plant were examined after drying. Days to heading was calculated from sowing date to the emergence of the first panicle. Grain length, grain width, and grain thickness were measured using a rice grain quality checker (RGQI 20A, Satake). Especially, part of the phenotypic data in 2014 was measured by Prof. M. Maekawa.

2-2-2 Cryosectioning and analysis

The spikelets were sampled in a day after heading, then the lemmas were observed by scanning electron microscope (TM3030Plus, Hitachi) or subjected to cryomicrotomic sectioning using Kawamoto method (Kawamoto and Kawamoto, 2014). The mature and dried hulled-grains of Koshihikari and *Lgg* were imbibed overnight at room temperature in dark before sectioning. Kawamoto method procedures were conducted as follows;

1. The samples were dipped into dry-ice hexane to freeze quickly.

2. SCEM medium (SECTION-LAB Co. Ltd., Japan) for embedding was poured into a stainless steel container.
3. The frozen samples were buried in the medium.
4. The container was quickly subjected to freezing in dry-ice hexane.
5. The frozen specimen block was taken out of the container.
6. The block was fixed on the cryomicrotome sample holder.
7. The sample holder was set to sample holder chuck of cryomicrotome.
8. The specimen block was trimmed to the required surface with a disposable blade, 30UF at -25°C.
9. The adhesive film (Cryofilm type 3C (16UF) (SECTION-LAB Co. Ltd., Japan)) was applied to the surface using a fitting tool and then the specimen was cut into 4 μm thickness.
10. The Cryofilm with the section after cutting was picked up carefully.
11. The section was left for approx. 20 seconds to dry slightly at room temperature.
12. The section was stained with hematoxylin for 60 seconds.
13. After being rinsed with running water for 4 minutes, the section was stained with 0.2 % eosin for 10 seconds.
14. After being rinsed with water for a few seconds and with 100% ethanol for 10 seconds, the section was mounted between Cryofilm and glass slide with SCMM-R2 (SECTION-LAB Co. Ltd., Japan).
15. The excessive mounting medium was removed with a filter paper.
16. The mounting medium was polymerized with UV light.

The images of the sections were taken by light microscope (BZ-800, Keyence), then the photos were stocked using Photoshop (Adobe) and analyzed using Image J.

2-3 Results

2-3-1 Comparison of agronomic traits between Koshihikari and *Lgg*

Lgg was found in over 3,000 *nDart1*-tagged lines in the genetic background of cv. Koshihikari. *Lgg* shows long spikelet and hulled grains, but it has no obvious phenotypic difference in aerial part from Koshihikari, the wild type (Fig. 2-1). To characterize *Lgg*, the agronomic traits and hulled grain size were examined in the year 2014 and 2015 (Table 2-1 and 2-2). The mutant had significantly longer grain length and panicle length in both 2014 and 2015, larger 100-grain weight in 2014, but significantly lower in grain width, spikelet fertility, total grain weight per plant and total panicle weight per plant than those of Koshihikari. Although the mutant did not show any significant differences in the number of spikelets per panicle from Koshihikari in 2014, a fewer number of spikelets were observed together with a smaller number of first branches and second branches in *Lgg* mutant in 2015. There were no significant differences in culm length, culm base diameter, panicle base diameter and days to heading between the mutant and Koshihikari (Table 2-1 and 2-2). The longer grain size mainly contributed to the large grain trait in the mutant.

Fig. 2-1. Plant phenotype and grain morphology of Koshihikari and *Lgg*.

Table 2-1 Agronomic traits of Koshihikari and *Lgg* in 2014

Trait	Koshihikari	*Lgg*
Grain length (mm)	4.955 ± 0.024	5.416 ± 0.049 * (109.3%)
Grain width (mm)	2.830 ± 0.008	2.690 ± 0.021 * (95.1%)
Grain thickness (mm)	2.125 ± 0.008	2.150 ± 0.011 * (101.2%)
100-grain weight (g)	2.081 ± 0.029	2.272 ± 0.066 *(109.2%)
Culm length (cm)	75.5 ± 3.3	74.3 ± 3.1
Panicle length (cm)	20.0 ± 1.1	21.3 ± 0.9 * (106.9%)
No. panicles per plant	10.6 ± 0.9	11.3 ± 2.4
Culm base diameter (mm)	4.729 ± 0.241	4.421 ± 0.374
Panicle base diameter (mm)	1.554 ± 0.164	1.608 ± 0.053
No. spikelets per panicle	125.9 ± 20.6	111.1 ± 16.0
Spikelet fertility (%)	94.58 ± 2.18	75.62 ± 5.63 * (80%)
Total grain weight per plant (g)	18.247 ± 2.465	15.332 ± 1.600 * (84%)
Days to heading	101.9 ± 1.6	102.8 ± 2.3

The percentage in parentheses is the rate of trait value of *Lgg* to Koshihikari.
Values are mean ± SD, n = 8. *; significant at 5% level by Student's *t*-test.

Table 2-2 Agronomic traits of Koshihikari and *Lgg* in 2015

Trait	WT	*Lgg*
Grain length (mm)	4.963 ± 0.093	5.694 ± 0.150 * (114.7%)
Grain width (mm)	2.720 ± 0.109	2.602 ± 0.093 * (95.7%)
Grain thickness (mm)	2.042 ± 0.042	2.052 ± 0.041
Culm length (cm)	72.34 ± 2.05	70.99 ± 3.41
Panicle length (cm)	21.04 ± 1.65	22.46 ± 1.16 * (106.7%)
No. Panicles	12.1 ± 1.6	11.8 ± 1.1
No. 1st branches	12.3 ± 2.6	9.5 ± 0.7 * (77.6%)
No. 2nd branches	31.2 ± 5.4	26.8 ± 4.5 * (85.8%)
No. Spikelets per panicle	148.4 ± 25.6	129.5 ± 13.9 * (87.2)
Spikelet fertility (%)	95.92 ± 1.87	86.35 ± 5.11 * (90%)
Total panicle weight (g) per plant	32.576 ± 4.210	26.492 ± 2.550 * (81.3%)
Days to heading	97.3 ± 2.8	99.0 ± 2.1

The percentage in parentheses is the rate of trait value of *Lgg* to Koshihikari.
Values are mean ± SD, n=16. *; significant at 5% level by Student's *t*-test.

2-3-2 The morphology of spikelet hull and grain in Koshihikari and *Lgg*

The spikelet hull size is the key trait affecting grain size. To investigate the cell size and the cell number of the spikelet hull, the inner surface image of lemma from Koshihikari and *Lgg* were taken by the scanning electron microscope. However, it was difficult to analyze the cell size through these images (Fig. 2-2). Subsequently, the lemmas of the mutant and Koshihikari were subjected to cryomicrotomic sectioning for longitudinal sections (Fig. 2-3). As shown in Fig. 2-3 and 2-4, at the cell layer under the adaxial epidermis, cell length and cell number were measured. The cell length was similar in *Lgg* and Koshihikari, but the average number of cells in the longitudinal direction was 212 in the mutant, which was 125% more than 169 in Koshihikari (Fig. 2-5 and Table 2-3). This result showed that longer spikelet hull in *Lgg* was caused by the increase of cell number, suggesting that the cell proliferation is elevated during spikelet hull development in the mutant.

Furthermore, to examine the cell morphology in embryo and endosperm of Koshihikari and *Lgg*, the imbibed mature grains were subjected to cryomicrotomic sectioning. As shown in Fig. 2-6, the longitudinal sections of the mutant showed longer endosperm and bigger embryo than in Koshihikari. In the transversion sections, however, *Lgg* showed smaller section area than Koshihikari (Fig. 2-6c and d). Overall, it was observed that the cell size, the cell pattern, and the ratio between embryo and endosperm were similar between Koshihikari and the mutant.

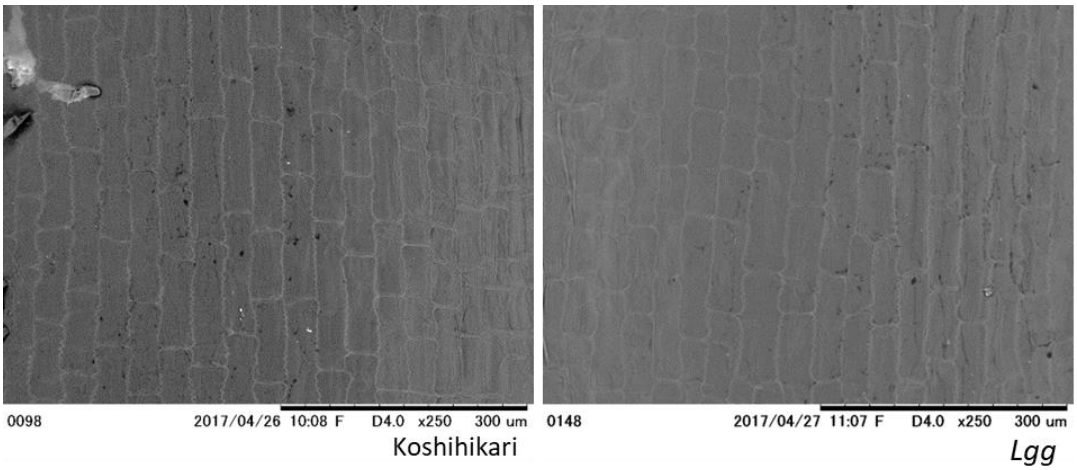

Fig. 2-2. Scanning electron microscope image of the inner epidermis of the lemma of Koshihikari and *Lgg*. Bar = 300 μm.

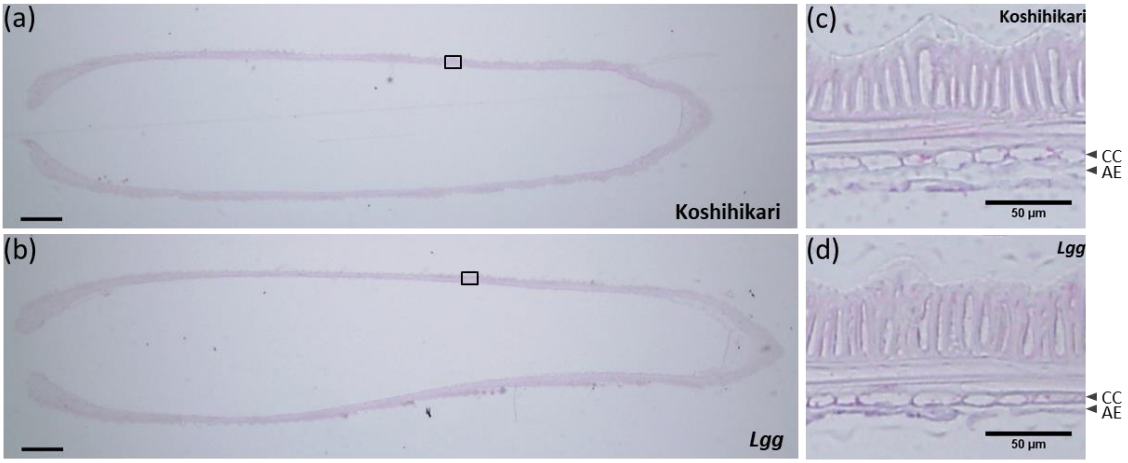

Fig. 2-3. The longitudinal section of the lemma of Koshihikari (a) and *Lgg* (b), and the magnified view of Koshihikari (c) and *Lgg* (d). CC; Chlorophyll contain parenchyma cell layer. AE; Adaxial epidermis. (a), (b); Bar = 400 μm. (c), (d); Bar = 50 μm.

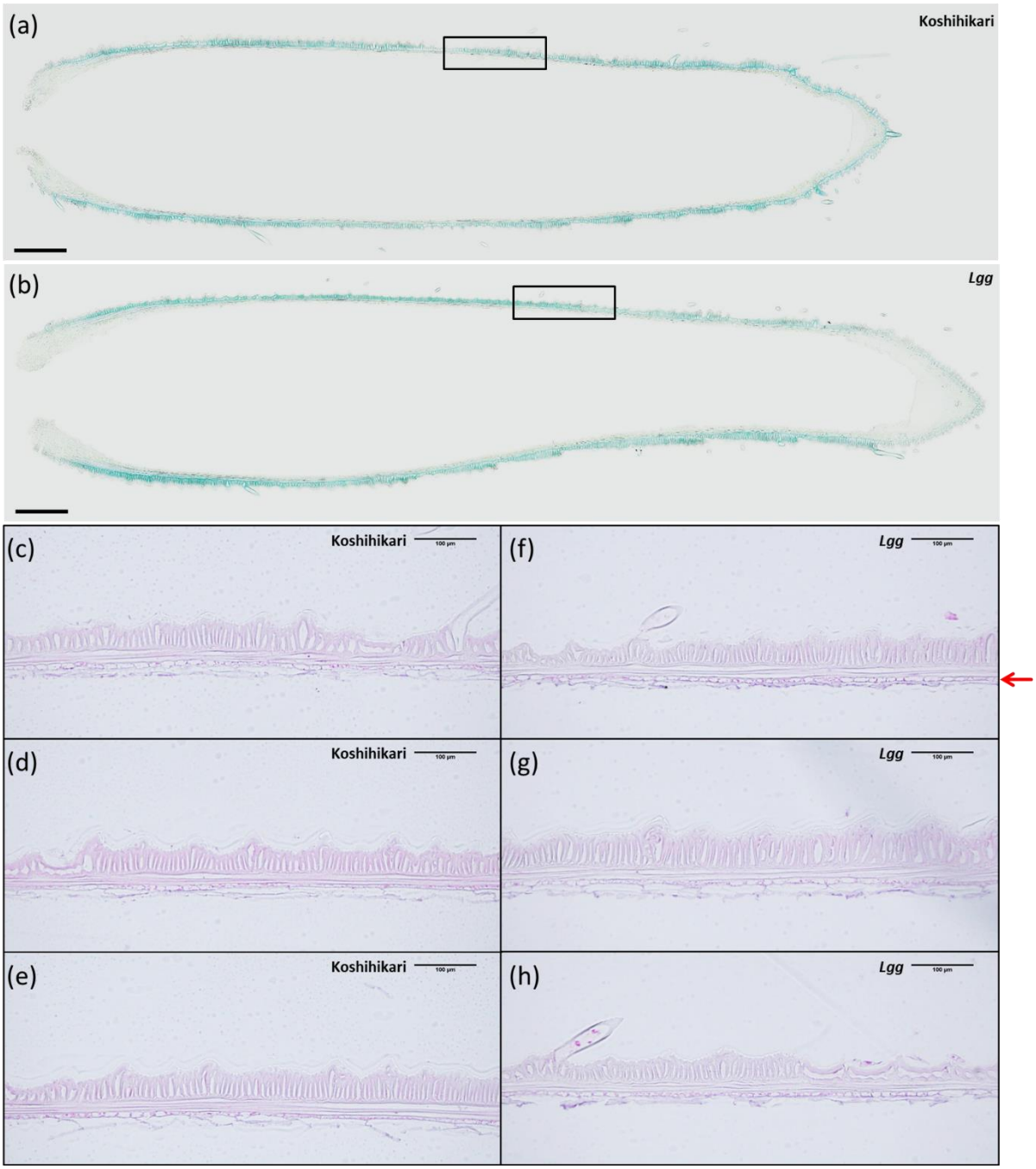

Fig. 2-4. The longitudinal section of the lemma of Koshihikari (a) and *Lgg* (b). Black lined box indicates the magnified view in (c) and (f). Bar = 400 μm. (c)-(f); Magnified view of longitudinal lemma section of Koshihikari and *Lgg*. Red arrow indicates the chlorophyll contain parenchyma cell layer for measuring cell length. Bar = 100 μm

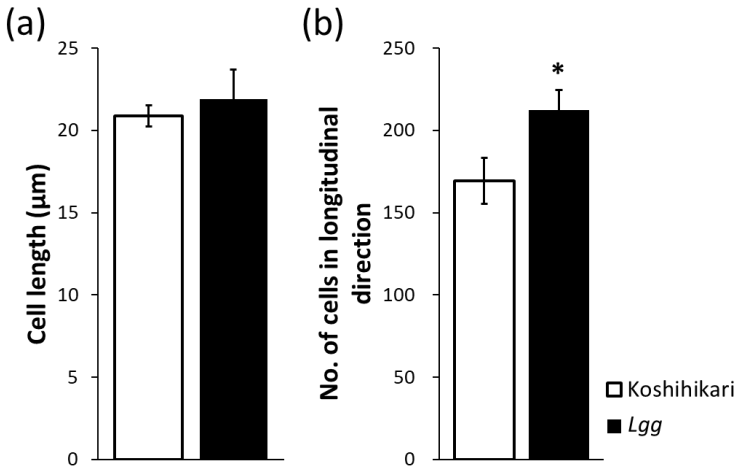

Fig. 2-5. Cell length (a) and cell number (b) in the longitudinal section of the lemma of Koshihikari and *Lgg*. *; significant at 5% level by Student's *t*-test.

Table 2-3 Cell length and cell number in the longitudinal section of the lemma of Koshihikari and *Lgg*

Line	Cell length (μm)	No. cells in the longitudinal direction
Koshihikari	20.88 ± 0.64	169.3 ± 14.0
Lgg	21.92 ± 1.78	212.3 ± 12.1 *

Values are mean ± SD, n=3. *; significant at 5% level by Student's *t*-test.

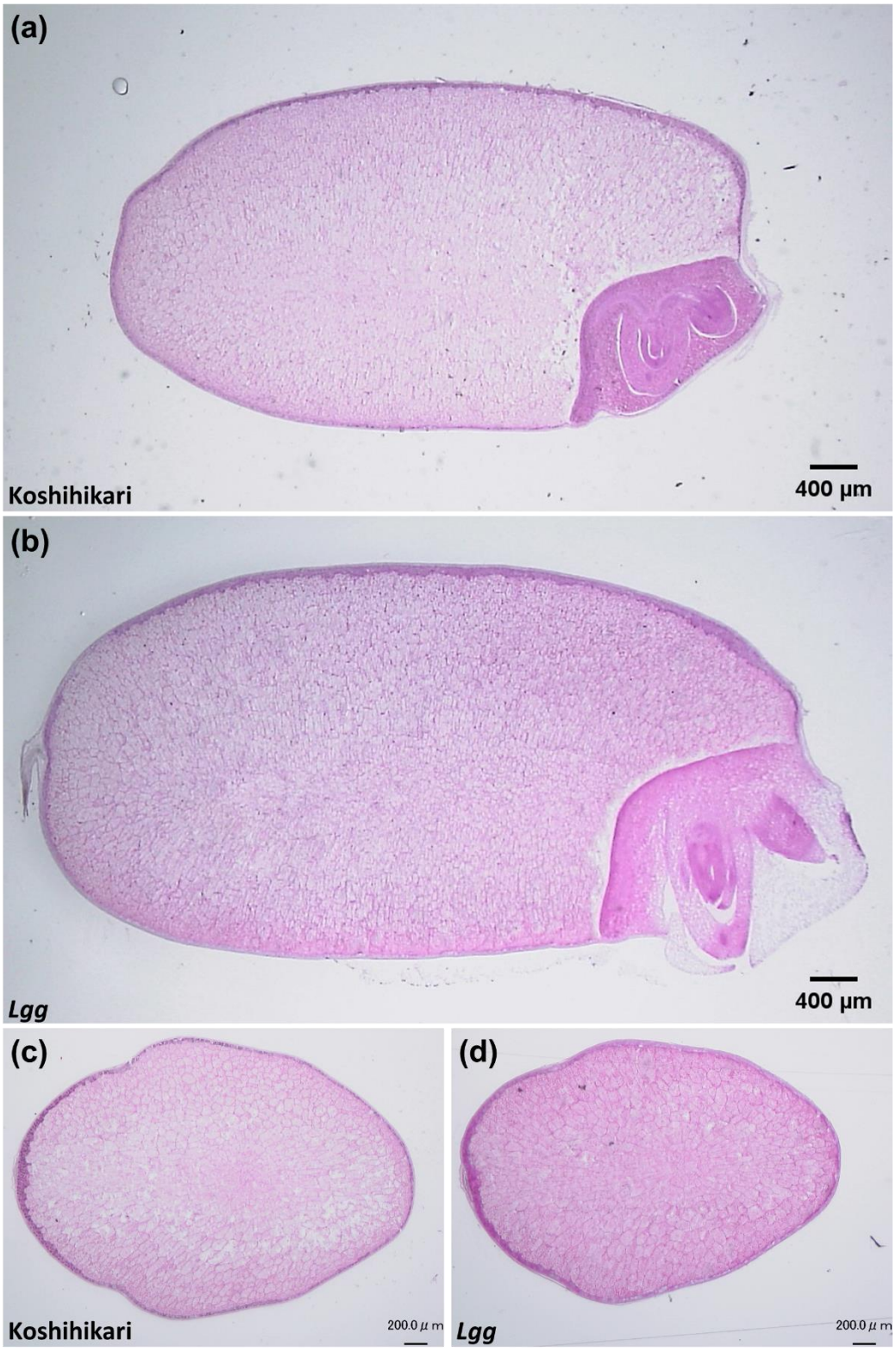

Fig. 2-6. The section of endosperm. (a) The longitudinal section of endosperm of Koshihikari. (b) The longitudinal of endosperm of *Lgg*. (c) The transverse section of endosperm of Koshihikari. (d) The transverse section of endosperm of *Lgg*. Bar = 400 μm in (a) and (b), 200 μm in (c) and (d).

2-4 Discussion

Grain yield in cereal crop is the most important trait for food and other material production. To increase grain yield, grain size is one of the targets for rice improvement. There are many genes related to grain size in rice (Table 1-1), however, the regulatory mechanisms have not been well understood. In this chapter, *Lgg* found in *nDart1*-tagged lines of Koshihikari was characterized having longer grain length, heavier 100-grain weight, and longer panicle length than those of Koshihikari (Table 2-1 and 2-2). The grain size is influenced by spikelet hull size and grain filling rate (Yoshida, 1981). Some genes related to grain size have been reported to be involved in cell cycle regulation that governs the grain size. *GL3.1* encodes a protein phosphatase kelch (PPKL) Ser/Thr phosphatase that interacts with Cyclin-T1;3 to accelerate cell division in the spikelet hull (Qi et al., 2012). A putative serine carboxypeptidase, *GS5* functions positively to regulate grain size as a modulator upstream of cell cycle genes (Li et al., 2011). OsMKKK10-OsMKK4-OsMAPK6 cascade plays a role in cell proliferation (Duan et al., 2014; Guo et al., 2018). The lemma section of Koshihikari and *Lgg* (Fig. 2-5; Table 2-3) suggests that the long grain phenotype of *Lgg* is due to an increase in the cell number in the spikelet hull. Furthermore, it also suggests that the causal gene for elongated grain length in the mutant may influence the cell division and proliferation.

Beside the lemma sections, the brown rice sections (Fig. 2-3, 2-4 and 2-6) using Kawamoto method (Kawamoto and Kawamoto, 2014; Chiou et al., 2018) also showed a very clear shape and cell arrangement. The pattern of cell arrangement in the central point of endosperm was similar and the obvious alteration of cell size was not observed between both lines (Fig. 2-6), suggesting that the increase of endosperm size resulted from cell proliferation.

In addition to increased grain length, *Lgg* exhibited lower spikelet fertility, total grain weight

per plant and total panicle weight per plant than those of Koshihikari (Table 2-1 and 2-2). Spikelet fertility is another component of rice yield and low fertility reduces the yield. It will be a problem when introducing a trait with side effects, such as low fertility. The high correlation between two traits can be due to many factors, one is genetic pleiotropy (Chen and Lübberstedt, 2010). Nevertheless, it is unclear whether the long grain phenotype is correlated with low fertility in *Lgg*. Therefore, further genetic analysis using segregation populations is required to evaluate the relation between grain length and fertility in the mutant.

Chapter 3 Inheritance of large grain phenotype in *Lgg*

3-1 Introduction

A quantitative trait, such as weight, height, and yield, is influenced by both genetic and environmental factors, and there could be multiple genes responsible for the phenotypic variation. Therefore, to determine how many genes control the trait between two parent lines, phenotyping of the quantitative trait and assessing the frequency distribution in the segregating progenies derived from the cross between parental lines, e.g. the primary segregating generations F2 and F3 (Tanksley, 1993), are needed.

In the breeding processes, especially during backcross breeding, it is necessary to consider the correlations between traits. When the target trait is highly correlated with another trait, it may result in improving the correlated trait together, or co-introduction of an undesirable trait. The high correlation between traits may be regulated by a single gene that involved in multiple metabolic pathways and/or can be due to the linkage drag, as reviewed by Chen and Lübberstedt (2010). Thus, the estimation of the level of the correlations between traits and the genetic effect is important for cross breeding.

In the previous chapter, *Lgg* was characterized with a longer grain length. To elucidate genetic basis on grain length and examine the correlation between grain length and other traits, this chapter will focus on genetic analysis using F2 and F3 lines from the cross between Koshihikari and the mutant.

3-2 Materials and Methods

3-2-1 Plant materials and trait measurement

Eighty-five F2 plants of the cross between *Lgg* and Koshihikari were grown together with the parental lines in the paddy field at IPSR, Kurashiki, Okayama, Japan, in 2014, and 40 selected F3 lines were grown for the progeny test in 2015. Several important agronomic traits were investigated, as described in Chapter 2. Part of the phenotypic data in 2014 was measured by Prof. M. Maekawa. Pearson's correlation coefficient was calculated, and the significance test was done by ANOVA for the collected data.

3-3 Results

3-3-1 The frequency distribution of the grain length in F2 and progeny test in F3 lines

In the F2 population of the cross between *Lgg* and Koshihikari, three peaks for grain length were observed; one likely corresponded to Koshihikari type grain length, the second to an intermediate between Koshihikari and *Lgg*, and the third to *Lgg* type (Fig. 3-1). Although the obvious segregation boundaries for grain length were not observed, the grain length of the mutant was suggested to be governed by a major gene.

Subsequently, the F3 progeny test was conducted to confirm the inheritance of grain length of the mutant. Five lines derived from Koshihikari type and 13 lines from the mutant type in F2 plants were fixed to the Koshihikari type and the mutant type, respectively (Fig. 3-2). On the other hand, 22 F3 lines from the intermediate type F2 plants showed intermediate grain length between Koshihikari and *Lgg*, and these lines showed wide variegations for length (Fig. 3-2), suggesting that all intermediate length type plants in F2 were heterozygous for the gene of interest. This result suggested that the long grain trait of *Lgg* is governed by an incomplete dominant gene.

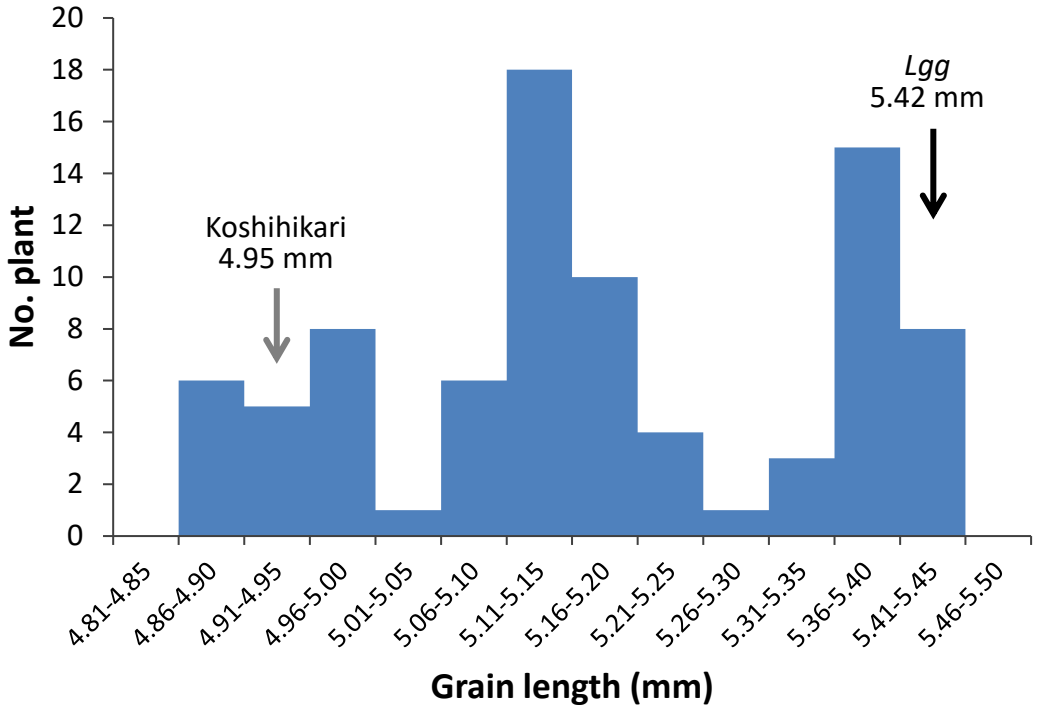

Fig. 3-1. Frequency distribution of grain length in the F2 of the cross between *Lgg* and Koshihikari. Arrows indicate means of grain length of the parents. n=85.

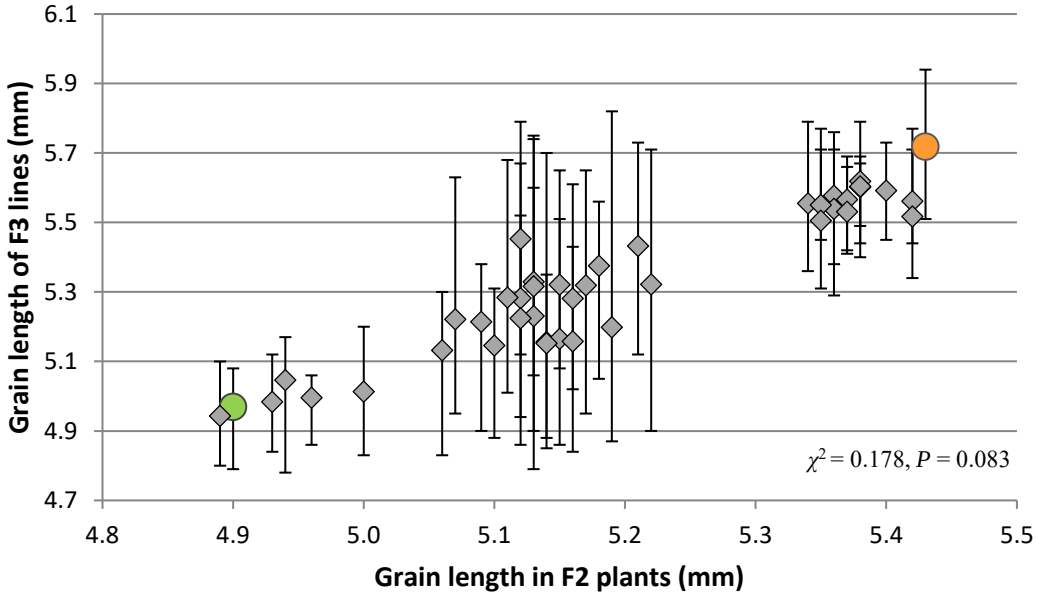

Fig. 3-2. Relation of grain length between 40 F2 plants and F3 lines (●; Koshihikari. ●; *Lgg*. ◆; F2-3 lines). Bar shows the range of the grain length.

3-3-2 Correlation between grain length and other traits in the F2 population

To check the pleiotropy effect of the long grain trait, the correlations between grain length and other traits are examined in F2, and the scatter plot graphs are shown with the correlation coefficients in Fig. 3-3. The grain length had significantly positive correlation with 100-grain weight (r = 0.816), and no significant correlations with grain thickness (r = 0.013), panicle length (r = 0.005), number of panicles per plant (r = 0.068), and days to heading (r = 0.035). However, grain length had significantly negative correlations with other seven traits, grain width (r = −0.702), culm length (r = −0.458), culm base diameter (r = −0.418), panicle base diameter (r = −0.446), number of spikelets per panicle (r = −0.544), fertility (r = −0.391), and total grain weight per plant (r = −0.328).

3-4 Discussion

The grain length in *Lgg* was 9.3% and 14.7% higher than that of Koshihikari in 2014 and 2015, respectively (Table 2-1 and 2-2). The frequency distribution of grain length in the F2 population and the progeny test using F3 line showed three types of grain length, Koshihikari type, intermediate type, and *Lgg* type, and fitted with 1:2:1 ratio (5:22:13, $\chi^2 = 0.178$, $P = 0.083$). These results suggest that the long grain phenotype is governed by an incomplete dominant gene (Fig. 3-1 and 3-2). Several genes showed semi-dominance have been reported, for example, *GW8* (Wang et al., 2012), *SRS5* (Segami et al., 2012), *GL7* (Wang et al., 2015b), and *SLENDER GRAIN* (*SLG*) (Feng et al., 2016). The incomplete dominance of grain length makes the long grain plants selection easier in the breeding process. However, further investigation of the regulatory mechanism of the causal gene in *Lgg* is needed.

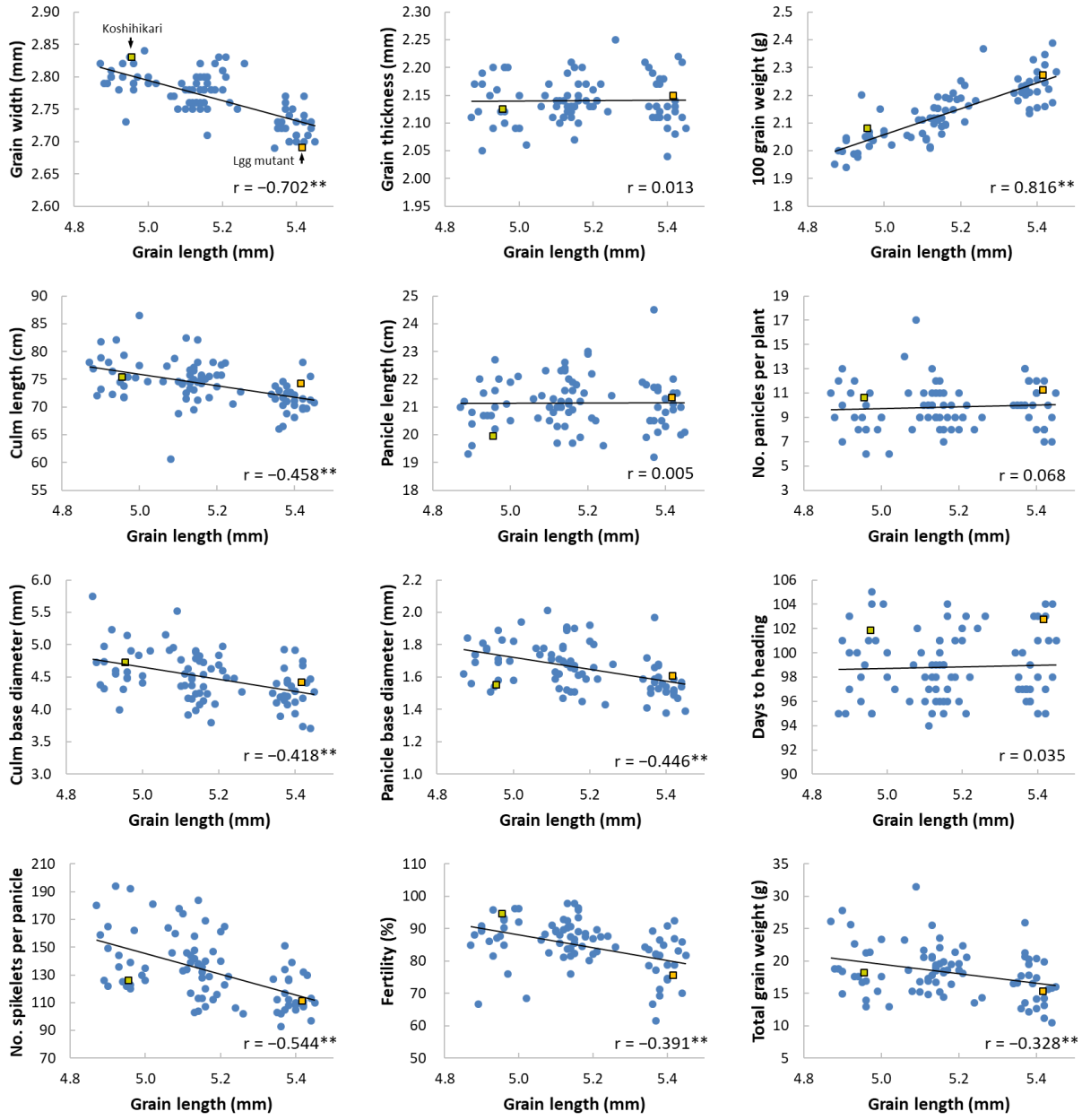

Fig. 3-3. The relations between grain length and other traits in the F2 population.

Based on the correlations between the grain length and the agronomic traits in the F2 population, there are some features associated with the long grain phenotype. The grain length is negatively correlated with grain width, culm length, culm base diameter, number of spikelets per panicle, fertility, and total grain weight. In particular, the negative correlation between grain length and total grain weight per plant, which is a very important trait of yield, is presumed to be caused by the negative correlation between grain length and both number of spikelets per panicle and fertility. In fact, the trade-off between large grain and number of spikelets per panicle has been reported (Guo et al., 2018). This phenomenon might be related to the timing of the expression of the responsible gene at the stage of spikelet hull differentiation. Meanwhile, fertility is also very important for the yield. The negative correlation between grain length and fertility has not been reported, so far. Since some plants showing high fertility with long grain phenotype were observed in the F2 (Fig. 3-3), it is likely that low fertility in the F2 was caused by a different genetic factor from the large grain gene. It requires further investigation of the negative correlations between grain length and other agronomic traits in another population.

Chapter 4 Identification of *LGG*

4-1 Introduction

To identify the causal gene in transposon-tagged mutants, transposon display (TD) is a method with high screening efficiency and commonly used (Takagi et al., 2007). TD is a modified method for identifying high copy number transposons based on amplified fragment-length polymorphism (AFLP) (Van de Broeck et al., 1998). In the *aDart*/*nDart* tagging system, *nDart1-3* subgroup contains 13 homologous elements in the reference genomic sequence of cv. Nipponbare and *nDart1-3* homologs were transposed most frequently and tend to insert into regions corresponding to the promoter or 5′ UTR (Takagi et al., 2010). Additionally, the *nDart*-specific TD was developed to visualize approximately 87% of *nDart* homologous elements and all *nDart1-3* subgroup elements by using polymorphisms at the 5′ subterminal region (Takagi et al., 2007). This protocol is essential to establish the *nDart1*-tagging system for rice functional genomics and analyze the *nDart1*-tagged mutants more efficiently.

The transposon-tagging system is a powerful tool to identify and analyze the function of a given gene because transposon insertion is not only to change genomic sequence but also to affect the gene expression around insertion site (Hsing *et al*., 2007). For example, the expression levels of *TAW1* (Yoshida et al., 2013) and *APO1* (Ikeda-Kawakatsu et al., 2009) were increased in the *nDart1*-inserted mutants. Similarly, the transcription initiation sites of *OsClp5* (Tsugane et al., 2006) and *SWL1* (Hayashi-Tsugane et al., 2014) were altered due to *nDart1* insertion.

To utilize the long grain trait as breeding material, it is important to elucidate the inheritance and identify the causal gene of *Lgg*, which was found in *nDart1*-tagged lines of Koshihikari. Thus, the gene of *Lgg* is identified through TD, and the validation of candidate gene is by

cosegregation between the identified marker and grain length and the transformation assay.

4-2 Materials and Methods

4-2-1 Plant materials and phenotyping

A total of 40 F3 lines were randomly selected from the 85 F2 plants and grown together with the parental lines, Koshihikari and *Lgg*, in the paddy field at IPSR, Kurashiki, Okayama, Japan. The transgenic plants, CRISPR/Cas9-mediated knockout lines (GE) and overexpression line (OE) were grown together with Nipponbare in the transgenic greenhouse in NIBB, Okazaki, Aichi, Japan. Grain length, grain width, and grain thickness were measured by a grain quality inspector (RGQI 20A, SATAKE). The spikelet hull length was measured by a vernier caliper. The ploidy data was measured using 1 mm-long young panicles of GE, OE, and Nipponbare with a flow cytometer (Cyflow SL, Sysmex Partec), and conducted at NIBB.

4-2-2 DNA extraction and PCR reaction

Leaf samples collected in 2 mL tubes were crashed with bead into powder (Multiple-beadshacker, Yasui Kikai) and 400 μL extraction buffer (200 mM Tris-HCl (pH7.5), 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was applied. After centrifuging at 150,000 rpm at 4°C for 10 minutes, the supernatant was transferred into a new 1.5 mL tube and 300 μL of cold isopropanol was added. Then, the supernatant was gently inverted and mixed. After centrifuging at 150,000 rpm at 4°C for 10 minutes to remove the supernatant, the pellet was rinsed with 500 μL of 70% ethanol and centrifuged at 150,000 rpm at 4°C for 10 minutes. Then the supernatant was discarded, and the DNA was dried in the tube for 50 minutes at room temperature. Finally, 100 mL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was added, and the DNA sample was stored at 4°C. PCR reactions were prepared in total 15 μL according to the manual instructions of the polymerase (LA Taq, TaKaRa). Primers are listed as Table 4-1, and PCR conditions are

shown as follows;

1 cycle	35 cycles	1 cycle	1 cycle
94°C 1 min.	94°C 1 min. 60°C 30 sec. 72°C 1 min.	72°C 7 min.	12°C ∞

The products were checked in 2% agarose gel (Agarose-LE, Classic Type, NACALAI TESQUE) at 100 V for 30 minutes.

4-2-3 Transposon display

Transposon display was conducted using four restriction enzymes with 4-bp recognition sequences, *Taq*I, *Msp*I, *Mse*I and *Cvi*AII, adapters and transposon primers as described by Takagi *et al*. (2007). The procedure is shown in Fig. 4-1. PCR reactions were performed using a polymerase that is able to amplify GC-rich regions (Advantage GC 2 Polymerase, TaKaRa), and the products were checked by 6% polyacrylamide gel electrophoresis at 2000 V, 50 mA for 2 hours.

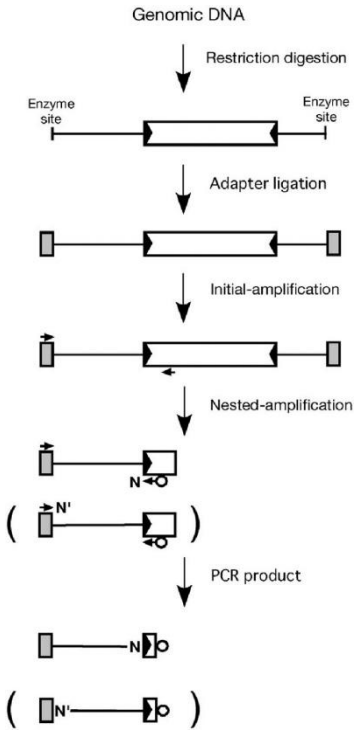

Fig. 4-1 Schematic representation of transposon display (TD) procedure by Takagi et al., 2007.

4-2-4 RNA extraction

The RNAs were extracted from leaf and root of 2-week-old seedlings and 0.6 mm-, 1 mm-, 2 mm- and 3 mm-long young panicles of the mutant and Koshihikari using an RNA extraction kit (RNeasy Plant Mini Kit, Qiagen) according to the manual instructions. DNase I (TURBO DNA-free™ Kit, Ambion) was used for removing contaminated DNA. The RNA quality was checked by spectrophotometer (NanoDrop™ 1000, Thermo Fisher Scientific).

4-2-5 Rapid amplification of cDNA ends (RACE)

Samples for rapid amplification of cDNA ends (RACE) were collected from the leaves of the mutant and Koshihikari. 5′ RACE was performed by SMARTer RACE 5′/3′ Kit (Clontech) using the gene-specific primers, 15De02_R and 15Nv01_R, followed by the manual instructions. 3′ RACE was done by reverse transcription kit (PrimeScript™ RT Reagent Kit, TaKaRa) using adapter primer and gene-specific primers, 15De03_F and 15Nv05_F. The RACEs products were amplified through TA cloning (pGEM-T Easy vector, Promega) for sequencing.

4-2-6 Vector construction and transformation

The genome editing vector, pZH_gYSA_MMCas9_LGG600, and the guide RNA sequence **CCG**CGC CCGGGGATTC-GGCTTCA (the protospacer-adjacent motif in bold) were used according to the construction protocol described in Mikami et al. (2015). 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 4 fragments (Table 4-1) amplified by high fidelity Taq, PrimeSTAR MAX (Takara), and was sequenced to confirm its identity to the genomic region. The vectors were transformed into *Agrobacterium tumefaciens* strain EHA105, and *Agrobacterium*-mediated transformations were carried out as described by Shimatani et al. (2009). The vector

constructions and *Agrobacterium*-mediated transformation experiments were conducted at NIBB with the collaboration of Dr. K. Tsugane.

Table 4-1 The list of the primers used in this study

Primer	Forward sequence	Reverse sequence
15Ju05	GGGAACGTGTCTAGCAGGAG	GAAGCTGCTCAACGACCTGT
15Ju02	CTGCTGATGATGCTGAATGC	GATGATTAGGAGCCAGTAGTTGTC
15Ju06	GGTTGGATTTGGAATGGATG	AAGATATCCCGTTCCCGTTC
15Ju03	CGAACGTCCGGTAGTAGTGC	GTGAAAGAGACCGGACAAGC
15Ap05	AATTTTCAGCCCATCAAACG	GGGAATGAGGGTTTTGGATT
15Ju15	GATCGGTATGGGAGTGGAGA	TGGGAAGCGTTGGATTTTTA
15Ju16	AGGCAGGGGTGGAGAAAG	TTGGGGTTTAAATCATTAGGG
15Ju18	CATCTTCCTCTGTCCCCAAA	CTGGTTATCCTTGCCGTGAT
15Ju12	CCTGTCAGCCACATTTAACG	TCTAGGGTTTTGGGAGTTGG
15Ap02	AAAATGTTGCTGCCGTTTTC	GGCCTTGCTTTCCTTTCTTT
15Ju14	GAGAAGAAGACGCGTTGAGG	CCTCTCTCCCCAAAACAATG
15Jy02	AGCAAAGGAAGGGAAGGAAG	CCCTGCATCAAAACAAACAA
15Ju10	AGGCAGGGGTGGAGAAAG	TTGGGGTTTAAATCATTAGGG
(4)	TGGTAGGAAATCACATTCAC	CGACAAATGAGATGGGAAG
RACE and qPCR	15Nv01_R	AACAGCTTTCTTCGCCTCCA
	15Nv05_F	AAGGAACTTTGGAGATTCATCAGG
	15Nv06_R	CGAAACCACAATCTCAAACCTTAC
	Ubiquitin_F	AACCAGCTGAGGCCCAAGA
	Ubiquitin_R	ACGATTGATTTAACCAGTCCATGA
pLGG:LGG^{Koshi}	LG-F-2550_inF	CTTCAGTTTAGCTTCATTTACCCATCATTCCCTCC
	LG-R-628	CATGCTCATAGCTTTTCTCC
	LG-F-642	AAAGCTATGAGCATGAAACA
	LG-R+836	GAATTCCTCGAAACGACAAC
	LG-F+822	CGTTTCGAGGAATTCAGAAT
	LG-R+3224	CATGGTTAGTTGTGGTAAGAG
	LG-F+3210	CCACAACTAACCATGACATA
	LG-R+4836_inF	ACTGATAGTTTAATTTAGTTTTATAGATATTGTTG
	Cambia_F2336	AATTAAACTATCAGTGTTTG
	Cambia_R11303	GAAGCTAAACTGAAGGCGGG
Genome Editing Vector	LG-F604BbsI	AAACCGCCCGGGGATTCGGCTTCA
	LG-R623BbsI	GTTGTGAAGCCGAATCCCCGGGCG
pCam_CLGFP_LGG	Cambia_R1	GGTCAAGAGTCCCCCGTGTT
	NinLGG_Cambia_F2055	TACATGGGGCATTAAATTGGTGACCAGCTCGAATT
	NinCam_pEZT-CL_GFP-F	GGGGGACTCTTGACCATGGTGAGCAAGGGCGAGGA
	inCam_pEZT-CL_GFP-R	GGCGTCCGCCTCCATGGCCGCTGCCGCAGCGGCAG
	LG-F+470_ATG	ATGGAGGCGGACGCCGGGAA
	LG-R+2055_TAA	TTAATGCCCCATGTAACCTG
pCam_LGG_NLGFP	Cambia_R1	GGTCAAGAGTCCCCCGTGTT
	NinGFP_Cambia_F2055	GAGCTGTACAAGTAAATTGGTGACCAGCTCGAATT
	NinCam_LG-F+470_ATG	GGGGGACTCTTGACCATGGAGGCGGACGCCGGGAA
	inGFP_LG-R+2055	GGCAGCGGCAGCAGCATGCCCCATGTAACCTG
	pEZT-NL_GFP-F	GCTGCTGCCGCTGCCGCTGC
	pEZT-NL_GFP-R	TTACTTGTACAGCTCGTCCA

4-3 Results

4-3-1 *nDart1* insertions in *Lgg*

In the previous chapter, the segregation pattern of grain length using F2 and F3 indicated that long grain trait in *Lgg* is controlled by a single factor. To determine the causal gene, *nDart1*-specific TD was applied for identifying the insertions of *nDart1* in the mutant and eight F2 plants with large grain phenotypes. Then, the specific band for the corresponding insertion site is expected to be present only in large grain phenotype lanes but absent in normal grain size plants, as illustrated in Fig. 4-2. The candidate bands were cut (Fig.4-3), and the sequences of eluted DNA were subjected to BLASTn search to the reference genomic sequence of *Oryza sativa* subsp. *japonica* in Nucleotide collection database on the National Center for Biotechnology Information (NCBI) and the rice annotation project database (RAP-DB). Although some selected bands in the TD results were unable to sequence, further analysis of the sequenced fragments revealed that a total of 14 additional insertions in the mutant were detected using four restriction enzymes (Table 4-2). Among the insertions, two insertion markers, which were detected by different restriction enzymes, *Mse*I and *Cvi*AII, were identical on chromosome 2. These insertions were distributed on chromosomes 1, 2, 3, 5, 8, 9, and 11, suggesting that *nDart1* elements moved frequently.

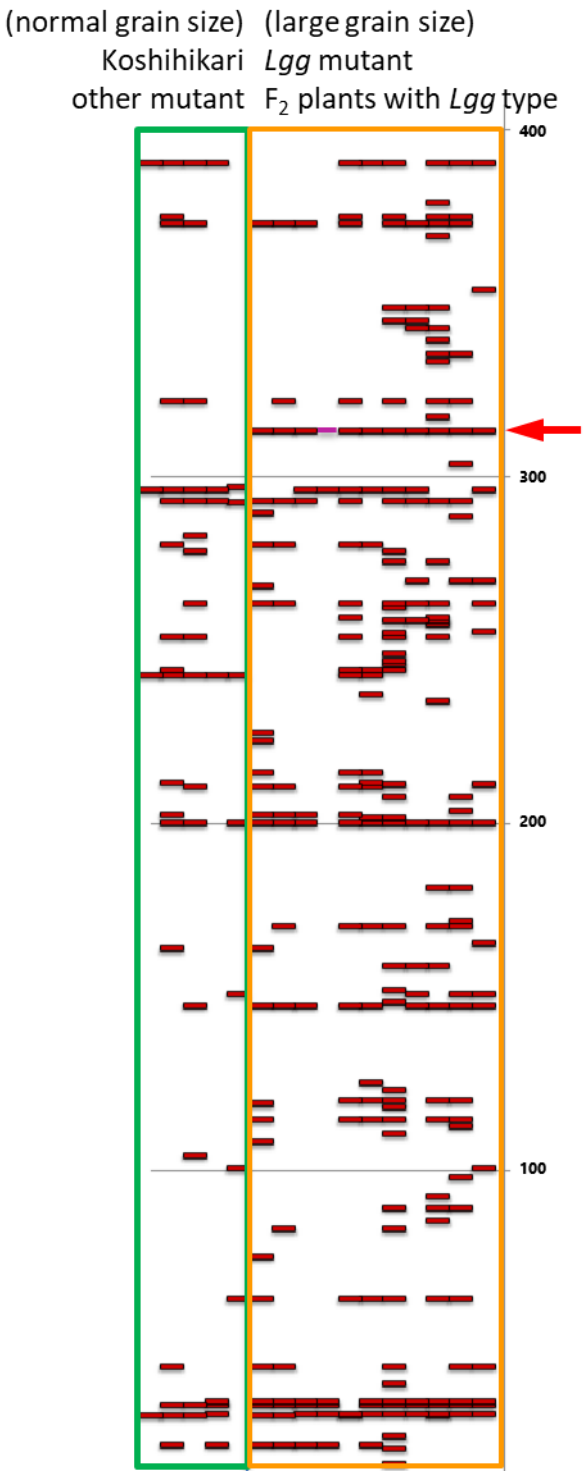

Fig. 4-2. The illustrated transposon display used in this study. Red arrow indicates the band co-segregated with the large grain phenotype.

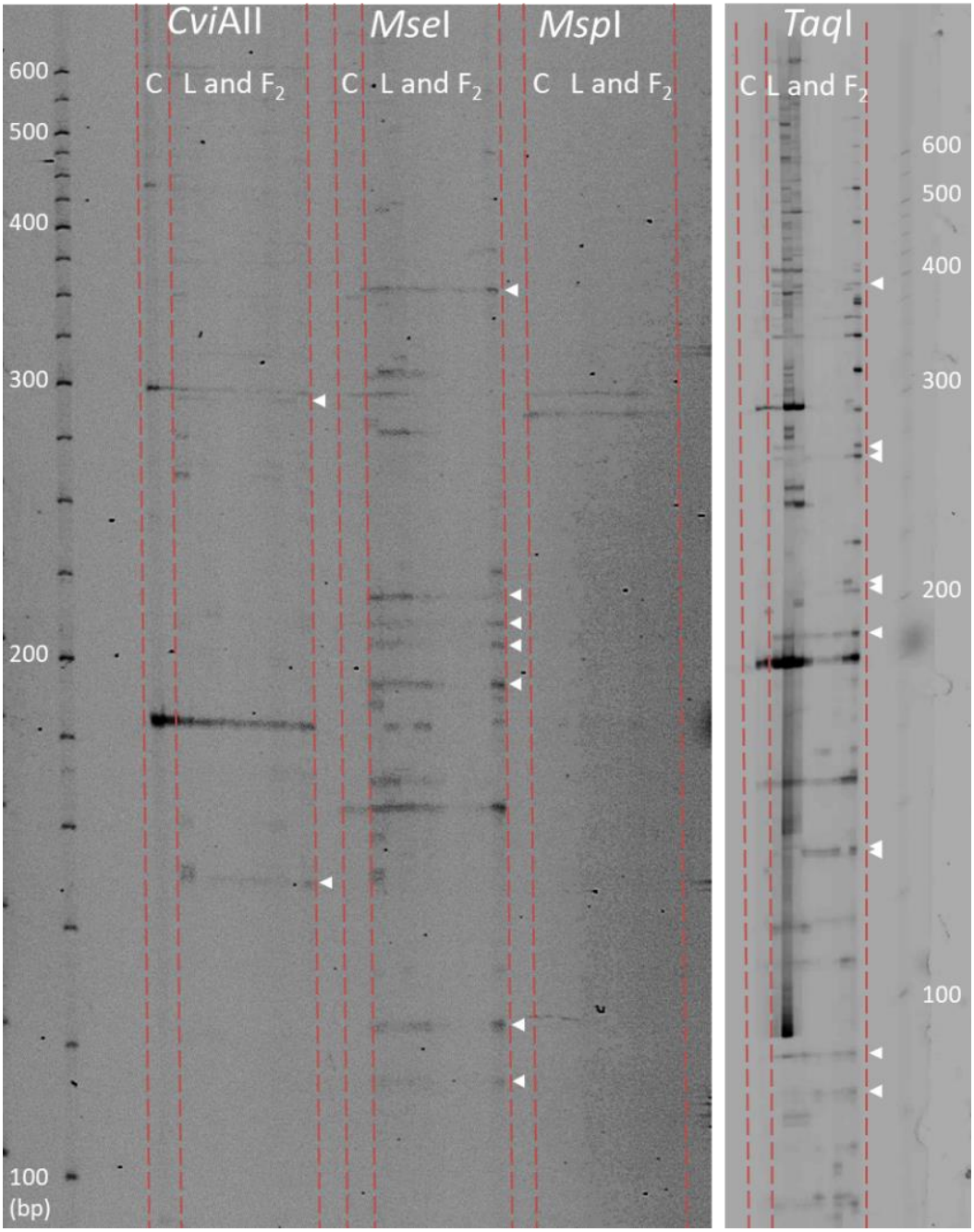

Fig. 4-3. Transposon display using *Cvi*AII, *Mse*I, *Msp*I, and *Taq*I in *Lgg* (L), F2 plants with large grain phenotype (F_2) and Koshihikari (C). Arrowheads indicate the fragments selected for sequencing.

Table 4-2 The insertion sites of *nDart1* detected in *Lgg*

Selected band	Enzyme	Chr.	Position†
15Ju05	*Taq*I	1	chr01:30,413,793..30,414,261
15Ju02	*Taq*I	3	chr03:35,689,039..35,689,500
15Ju06	*Taq*I	3	chr03:6,499,144..6,499,640
15Ju03	*Taq*I	8	chr08:26,537,174..26,537,647
15Ap05	*Taq*I	9	chr09:906,565..906940
15Ju15	*Mse*I	2	chr02:6,649,902..6,650,252
15Ju16	*Mse*I	2	chr02:7,034,724..7,035,119
15Ju18	*Mse*I	2	chr02:12,164,273..12,164,712
15Ju12	*Mse*I	3	chr03:31,460,099..31,460,468
15Ap02	MseI	5	chr05:22,353,860..22,354,221
15Ju14	*Mse*I	11	chr11:22,271,404..22,271,883
15Jy02	*Mse*I	11	chr11:25,192,996..25,193,462
15Ju10	*Cvi*AII	2	chr02:7,034,724..7,035,119
(4)	*Cvi*AII	3	chr03:1,078,201..1,078,588

†; According to IRGSP 1.0 (RAP-DB).

4-3-2 Identification and the transcripts of *LGG*

Cosegregations between the genotype of 13 insertion markers and large grain phenotype were examined in 40 F2 plants. Only one insertion marker, 15Jy02, on chromosome 11 was found to be cosegregated with large grain phenotype (Table 4-3). The *Lgg*-homozygous F2 plants with over 5.3 mm of grain length were homozygous in marker 15Jy02 (Table 4-3). Additionally, the F3 lines from 15Jy02-heterozygous F2 plants showed intermediate grain length and large standard errors (Fig. 4-4). On the other hand, F2 plants that marker types were Koshihikari and the mutant homozygous and their progenies had normal and long grain size, respectively, with small standard errors (Fig. 4-4). These results confirmed that 15Jy02 maker was cosegregated with large grain phenotype.

Subsequently, the PCR product for the marker 15Jy02 of the mutant was sequenced and aligned with the sequence of Koshihikari. The sequence analysis demonstrated that 500-bp of 5′ terminal region of *nDart1-3* was inserted in 5′ UTR of *Os11g0637700* and remaining 107-bp of 3′ terminal region was absent in the mutant. Moreover, 355-bp sequence followed by the

insertion was found to be deleted. Thus, this gene was presumed to be the candidate gene for the long grain phenotype. Hereafter, this gene, *Os11g0637700*, was designated as *LGG*.

Database searches revealed that *LGG* is annotated as an RNA-recognition motif (RRM) containing protein (Table 4-4) and contains four exons, and the coding region is from exon 2 to exon 3 (Fig. 4-6). To know the structure of the transcript, 5′ and 3′ RACE analyses were carried out and showed that the transcript initiation site (TIS) in the mutant was shifted to 30-bp upstream from the 5′ splicing site of exon 2, resulting in another extra translation start codon that located at 29-nt upstream of the original translation start codon that out of frame with original LGG protein (Fig. 4-6), while the 3′ region of the transcripts were identical between Koshihikari and *Lgg* (Fig. 4-6), suggesting that *Lgg* allele retained full coding region and splicing sites between intron 2 and intron 3.

Table 4-3 Homozygous *nDart1* inserted in F2 plants of the cross between *Lgg* and Koshihikari

Line	Grain length	Selected band													
		15Ju05	15Ju02	15Ju06	15Ju03	15Ap05	15Ju15	15Ju16	15Ju18	15Ju12	15Ap02	15Ju14	15Jy02	15Ju10	(4)
70-12	5.09 ± 0.31					++				++					++
70-13	5.36 ± 0.34	++			++	++						++	++		
70-14	5.13 ± 0.29				++	++	++	++	++		++			++	
70-15	5.14 ± 0.28						++	++	++	++				++	
70-16	5.16 ± 0.26					++					++				
70-17	4.96 ± 0.29	++													
70-18	4.93 ± 0.24		++				++	++						++	
70-10	5.37 ± 0.36										++		++		
70-22	5.13 ± 0.32				++										
70-23	5.34 ± 0.49	++		++	++		++	++					++	++	++
70-24	5.17 ± 0.30														++
70-25	5.36 ± 0.39						++	++	++			++	++	++	
70-26	5.38 ± 0.38											++	++		
70-27	4.94 ± 0.29			++						++					
70-28	5.15 ± 0.33		++			++			++	++	++				++
70-29	5.13 ± 0.30														
70-20	5.35 ± 0.44								++			++	++		
70-32	5.22 ± 0.26	++	++												
70-33	5.18 ± 0.24		++		++		++	++		++		++		++	
70-34	5.12 ± 0.24	++				++									
70-35	4.89 ± 0.23		++	++			++	++	++	++				++	
70-36	5.07 ± 0.26		++						++	++					
70-37	5.16 ± 0.30			++											
70-38	5.10 ± 0.29											++			++
70-39	5.06 ± 0.28	++							++		++				++
70-30	5.38 ± 0.35	++		++						++	++	++	++		++
70-42	5.19 ± 0.27									++					
70-43	5.35 ± 0.32		++									++	++		
70-44	5.42 ± 0.30		++			++	++	++	++	++	++		++	++	
70-45	5.11 ± 0.25				++						++				
70-46	5.12 ± 0.29			++		++		++			++			++	

++, homozygous genotype of *nDart1* insertion. Orange colored plant shows grain length over 5.3 mm with homozygous *Lgg*.

Table 4-3 Continued

Line	Grain length	Selected band													
		15Ju05	15Ju02	15Ju06	15Ju03	15Ap05	15Ju15	15Ju16	15Ju18	15Ju12	15Ap02	15Ju14	15Jy02	15Ju10	(4)
70-47	5.38 ± 0.35	++	++	++		++	++	++		++	++	++	++	++	++
70-48	5.42 ± 0.31	++				++							++		
70-49	5.40 ± 0.33											++	++		
70-40	5.15 ± 0.25				++										
70-52	5.14 ± 0.32		++	++						++		++			
70-53	5.12 ± 0.33	++	++							++		++			
70-54	5.37 ± 0.28	++											++		
70-55	5.00 ± 0.29			++	++	++			++						++
70-56	5.21 ± 0.29		++								++				

++, homozygous genotype of *nDart1* insertion. Orange colored plant shows grain length over 5.3 mm with homozygous *Lgg*.

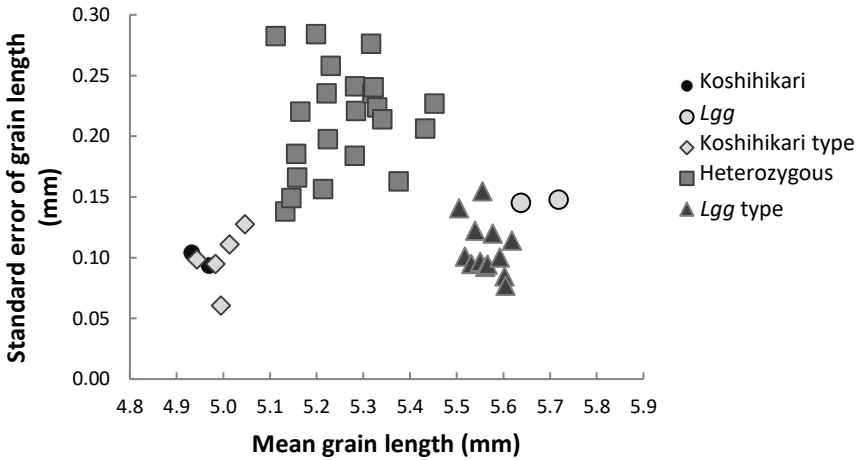

Fig. 4-4. Relationship between mean grain length and the standard error of grain length in F3 lines derived from the cross between *Lgg* and Koshihikari together with the genotype for *nDart1* insertion site of 15Jy02.

Fig. 4-5. The structure alteration in OSJNBa0095P22. (a) Illustration of the insertion site, which was co-segregated with large grain phenotype. Blue arrows indicate the primer positions. Dashed line represents 355-bp deletion; pink box with black triangle represents 500-bp truncated *nDart1-3*. (b) Alignment of *nDart1* insertion region of OSJNBa0095P22, Koshihikari, and *Lgg*. Pink and white boxes indicate the insertion of truncated *nDart1-3* and deletion, respectively. Black underline is the 19-bp terminal inverted repeat. Blue arrows indicate the primer positions.

Table 4-4 *LGG* (*Os11g0637700*) information from online databases

Database	Locus ID	Description
RAP-DB	Os11g0637700	Nucleotide-binding, alpha-beta plait domain containing protein
MSU Osa1	LOC_Os11g41890	RNA recognition motif containing protein, putative, expressed
NCBI	LOC4350986	Heterogeneous nuclear ribonucleoprotein 1 (hnRNP1)

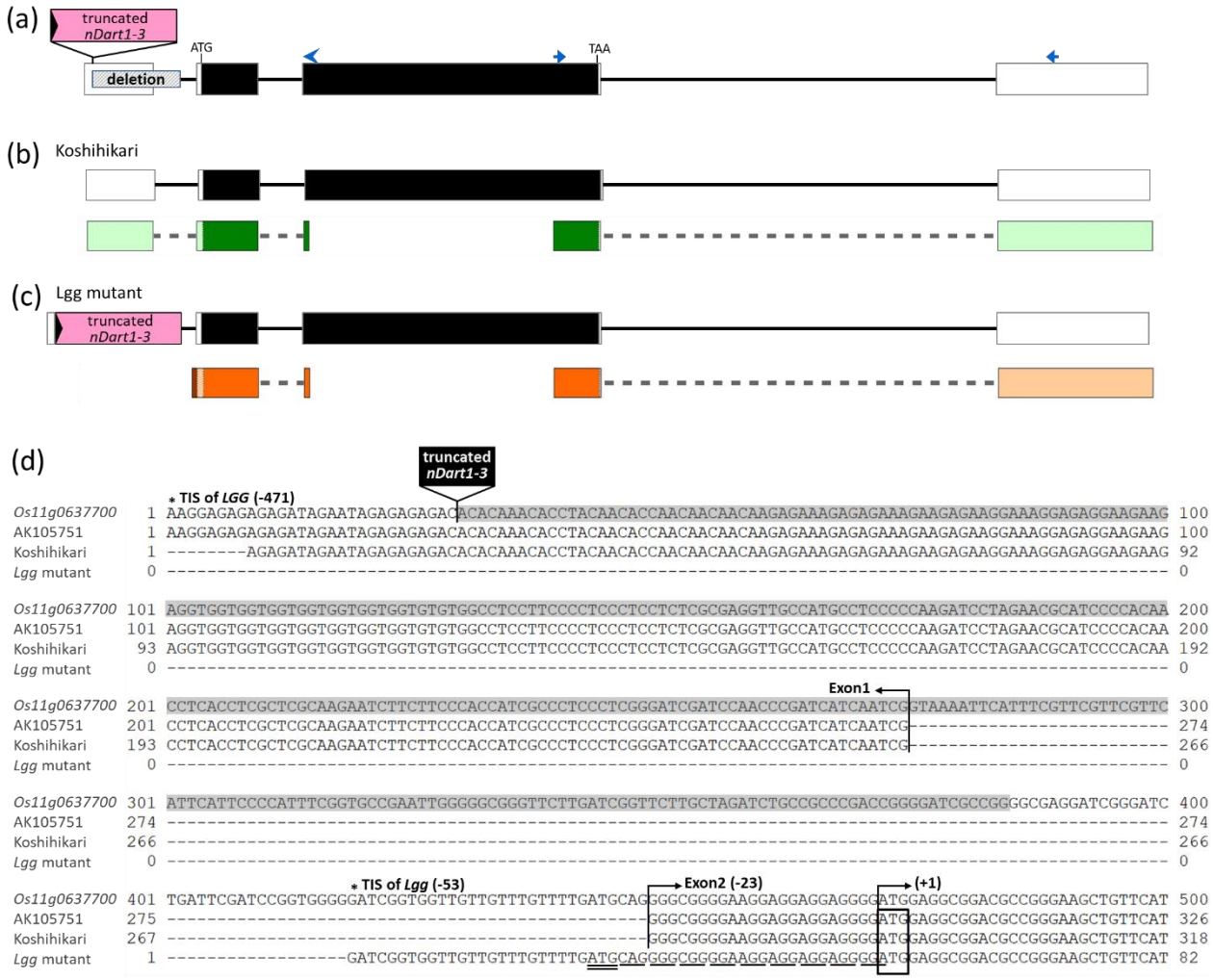

Fig. 4-6. Diagram of *LGG* structure and the sequence alignment. (a) The diagram of *LGG* with the mutation. Blue arrowhead and arrows indicate the primer positions for 5′ RACE and qRT-PCR, respectively. (b) The diagram of the genomic sequence and RACE results in Koshihikari (c) The diagram of genomic sequence and RACE results in *Lgg*. (d) Transcript sequences of 5′ region of *LGG* in Koshihikari and *Lgg* together with genomic sequence (*Os11g0637700)* of Nipponbare. 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. *; Transcription initiation site (TIS).

4-3-3 Transformation assay

To verify that LGG is responsible for the long grain phenotype, CRISPR/Cas9-mediated knockout (genome editing, GE) and overexpression (OE) lines were produced by transforming a CRISPR/Cas9 construct targeting DNA sequence corresponding to the RRM region of LGG (Fig. 4-7), and an *LGG* construct with its WT promoter (pLGG:LGG^{Koshi}; OE; Fig. 4-8) into cv. Nipponbare (NP), respectively, because we could not obtain any transformants in the Koshihikari genetic background. The sequences of the genome-editing target region (Fig. 4-7) were amplified and sequenced in T1 and T2 generation of GE lines. The sequence showed that all GE lines had 1-, 4-, 5-, or 7-bp deletion in target sequences (Fig. 4-9), suggesting that the frameshift occurred in the coding region of *LGG*.

Fig. 4-10 showed the plant and panicle morphology of T2 generation of GE and OE lines together with Nipponbare. The spikelet lengths were significantly increased in two GE lines and significantly reduced in the OE lines compared to that of Nipponbare (Fig. 4-11 and Table 4-5). Examination of cell composition of the spikelet hull was done by cryomicrotomic sectioning. The cell length of the lemmas was similar among Nipponbare, GE, and OE lines (Fig. 4-12), suggesting that the cell number is altered for the different spikelet length in the GE and OE lines.

Besides, the culm length was significantly decreased in GE and OE lines, and the fertility was drastically reduced in the OE lines relative to those in Nipponbare. The other traits, panicle length, panicle number, and spikelet number were similar among Nipponbare, GE, and OE lines (Table 4-5).

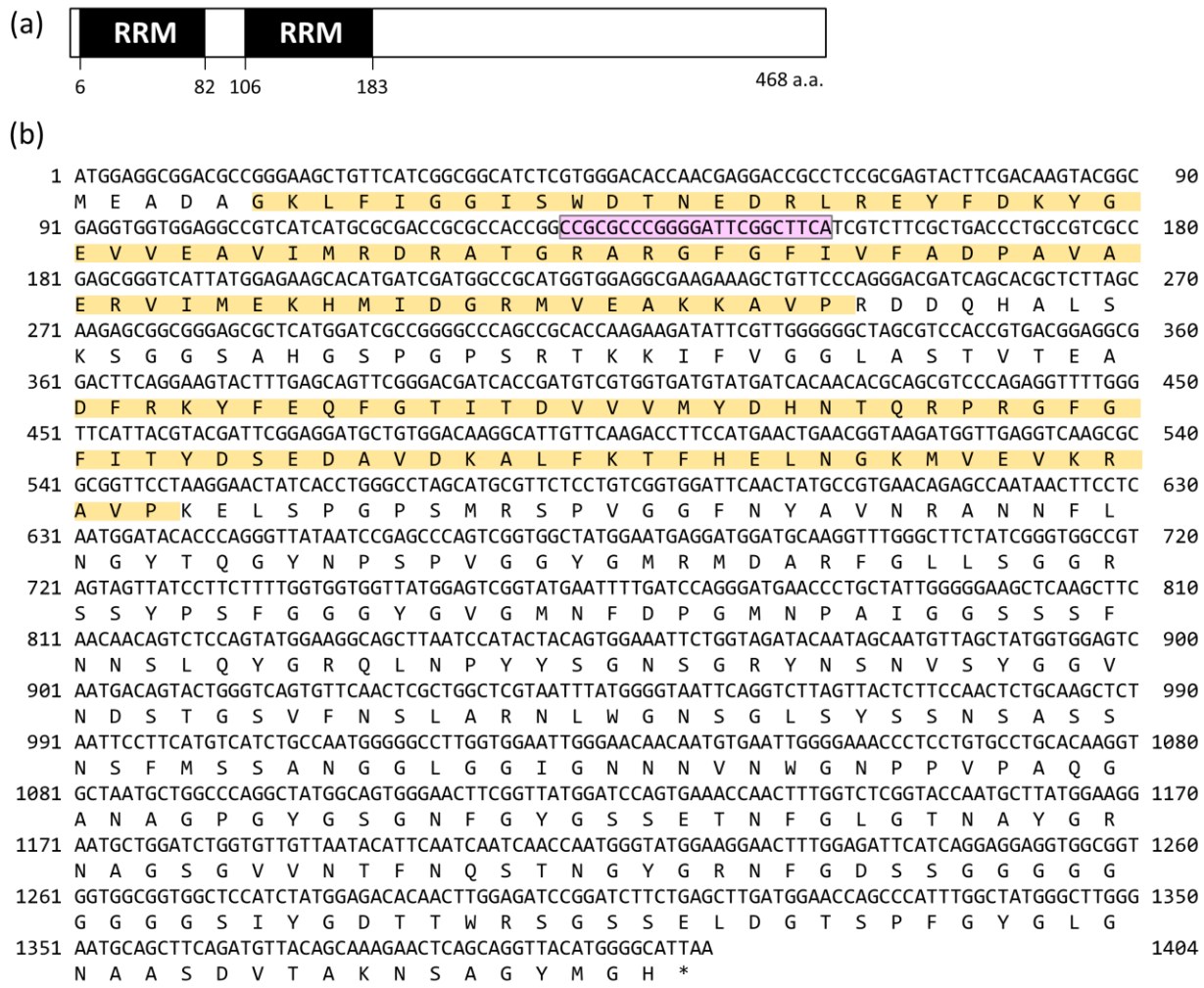

Fig. 4-7. LGG sequence. (a) Diagram of the amino acid motif in the candidate gene. RRM; RNA-recognition motif. (b) Nucleotide and deduced amino acid sequences of LGG. Yellow boxes indicate RRMs. Pink box indicates genome editing target region.

Fig. 4-8. Diagram of the construct for overexpression (OE) lines

```
Nipponbare   1 CCGCGCCCGGGGATTCGGCTTCA 23

GE10-4-3_1   1 CCGCGCC-GGGGATTCGGCTTCA 22
GE10-4-3_2   1 CCGCGC----GGATTCGGCTTCA 19

GE10-4-4_1   1 CCGCGCC-----ATTCGGCTTCA 18
GE10-4-4_2   1 CCGCGC-------TTCGGCTTCA 16

GE10-4-6     1 CCGCGC-------TTCGGCTTCA 16
```

Fig. 4-9. Genomic sequences of GE plants and Nipponbare in genomic editing target region.

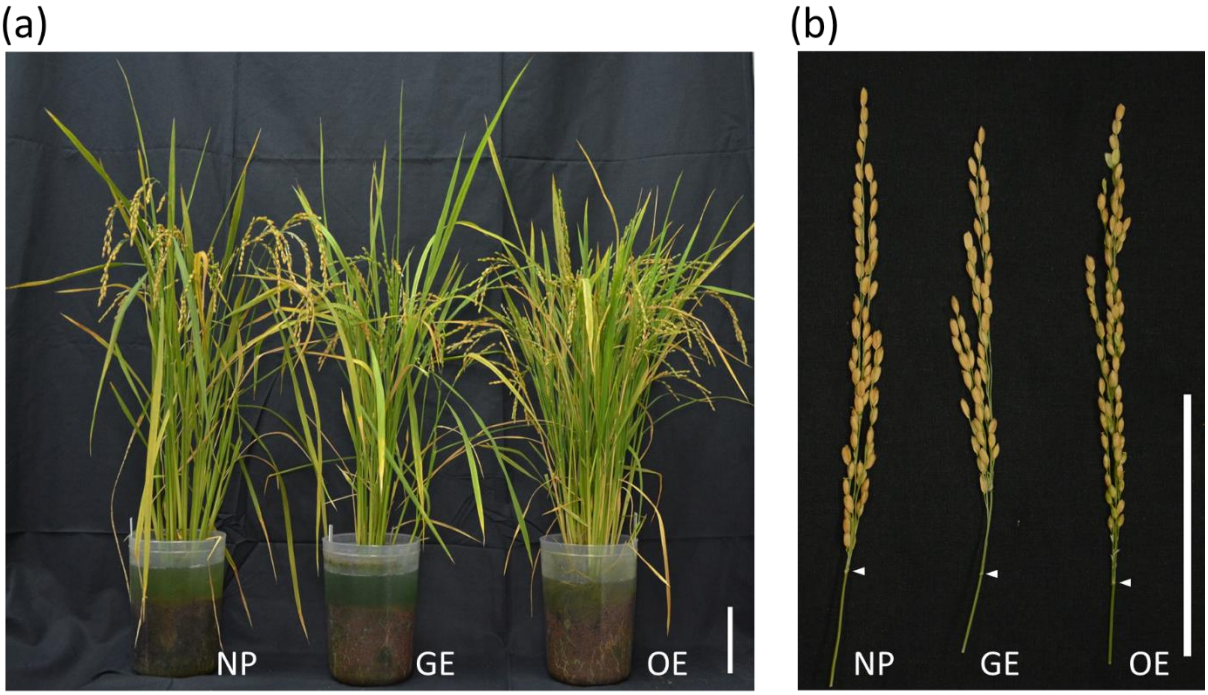

Fig. 4-10. Phenotypes of the plants (a) and the panicles (b) of Nipponbare (NP), the CRISPR/Cas9-mediated knockout (GE), and overexpression (OE) lines. Arrowheads indicate the panicle base. Bar = 10 cm.

Table 4-5 Agronomic trait of NP, GE, and OE lines

Line	Spikelet length (mm)	Panicle length (cm)	Culm length (cm)	No. panicles per plant	No. spikelets per panicle	Fertility (%)
NP	6.231 ± 0.126 [b] (6.215-6.239)	17.9 ± 1.5 [a] (16.5-19.5)	55.7 ± 1.0 [a] (54.7-56.7)	15.7 ± 3.8 [a] (13-20)	43.7 ± 12.2 [a] (33-57)	90.0 ± 1.9 [ab] (87.8-91.2)
GE	6.528 ± 0.191 [a] (6.361-6.644)	16.4 ± 1.0 [a] (15.5-17.5)	45.0 ± 6.6 [b] (38.3-51.5)	23.0 ± 8.0 [a] (15-31)	47.0 ± 3.6 [a] (43-50)	93.5 ± 2.7 [a] (90.7-96.0)
OE	5.767 ± 0.178 [c] (5.642-5.877)	16.9 ± 1.8 [a] (15.2-18.7)	46.3 ± 4.0 [b] (42.8-50.6)	20.7 ± 9.0 [a] (15-31)	60.3 ± 15.5 [a] (45-76)	67.8 ± 21.1 [b] (44.4-84.5)

Values are mean ± SD (n =3). Maximum and minimum values are in parentheses. Different letters indicate the significance level at 5% by Fisher's LSD test. The correlation coefficient between culm length and panicle length was estimated to be positively significant in NP, GE and OE ($r = 0.752$, $P < 0.05$).

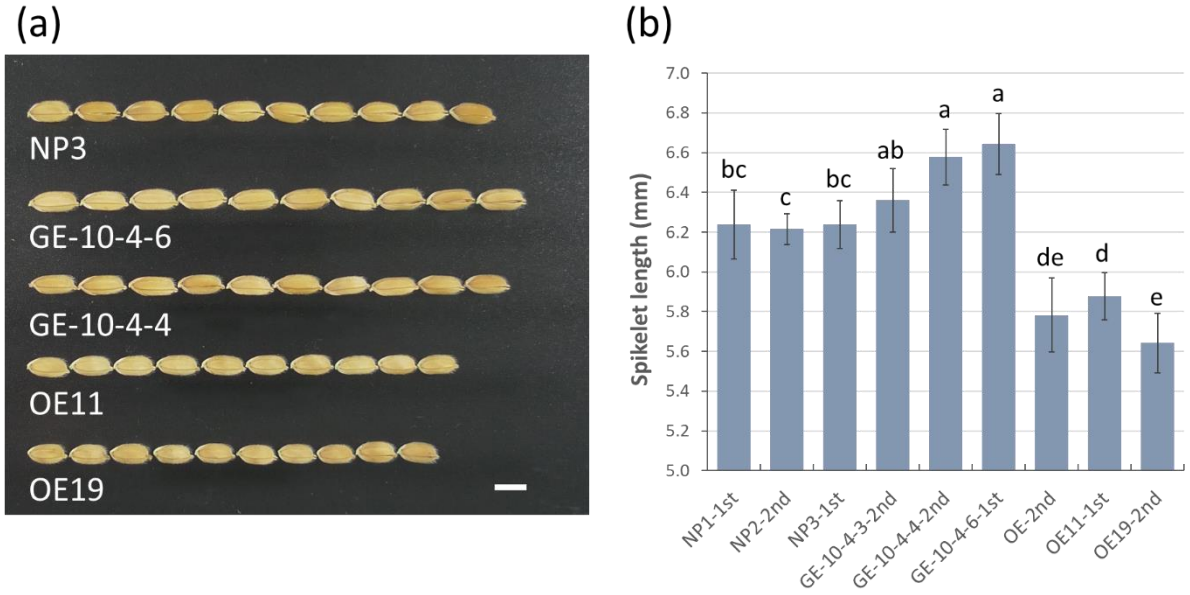

Fig. 4-11. The spikelets morphology in NP, GE, and OE lines. (a) The spikelets. Bar = 5 mm. (b) Mean spikelet hull length. Different letters indicate the significance level at 5% by Fisher's LSD test. n = 10.

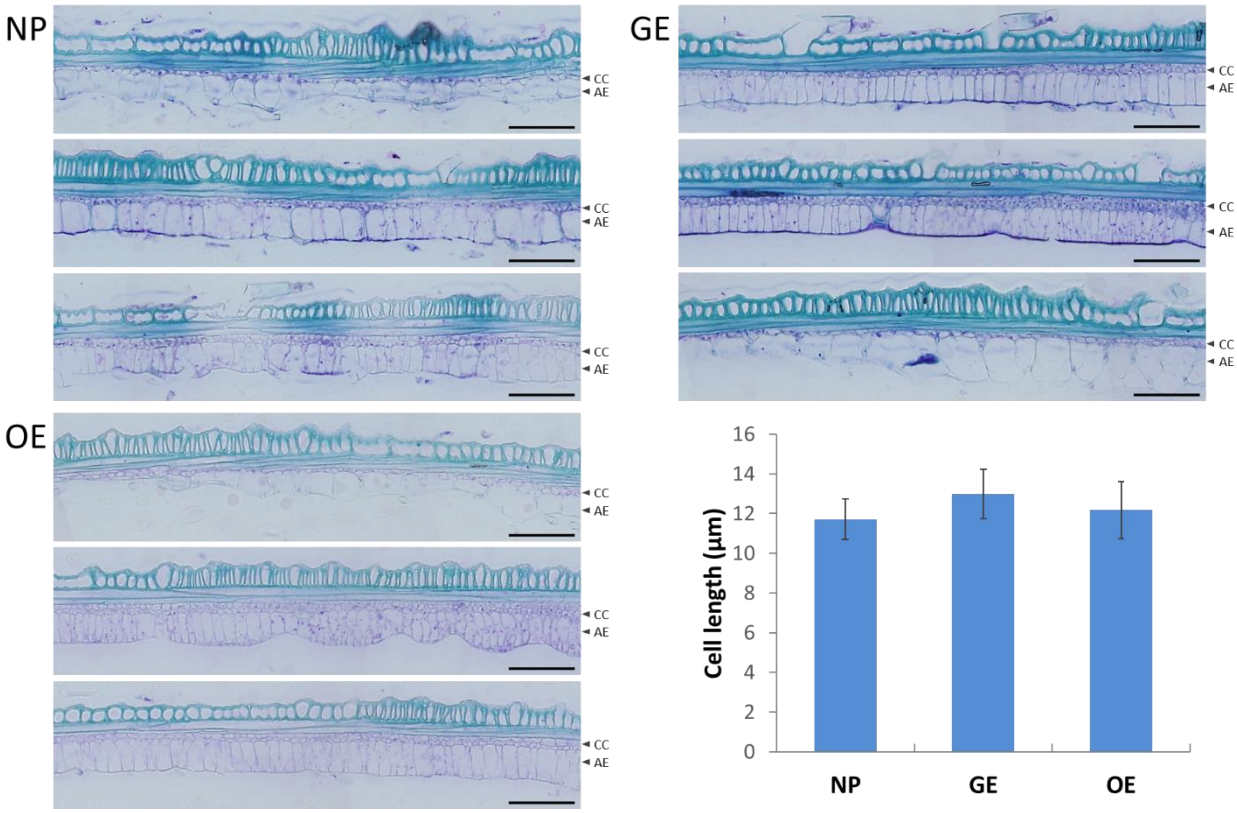

Fig. 4-12. Longitudinal sections and cell length of the lemma of NP, GE, and OE, respectively. AE; Adaxial epidermis. CC; Chlorophyll-contained parenchyma cell. Bar = 100 μm.

4-4 Discussion

TD is an effective method for detecting transposon insertions. Our TD analysis using four restriction enzymes indicated that there are 13 additional insertions located across the genome in *Lgg* (Table 4-2). Tsugane et al. (2006) mentioned that *nDart1* is the most active among *nDart1*-related elements, and *nDart1*-related elements tend to insert into intragenic regions (Takagi et al., 2010). This result suggests that *nDart1* moves frequently under the presence of active *aDart1*. With the advantage of *nDart1* tagging method, using a small segregating population, in this case, a total 85 plants, is sufficient for TD and progeny test, instead of using thousands of plants from a segregating population for fine mapping of the gene in map-based cloning method. In this study, TD analysis using ten large grain plants from the F2 could easily detect the candidate insertion of *nDart1*, and the cosegregated insertion marker was found more efficiently than by map-based cloning.

After sequencing the genomic region of the causal insertion, it was revealed that the truncated 500-bp *nDatr1-3* (107-bp deletion in 3′ terminal) was inserted in the 5′ UTR of *LGG* and 355-bp of the genomic region were deleted (Fig. 4-5), resulting in a shift of the TIS in the mutant. This deletion event is not likely 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). The deletion event is probably induced by microhomology (CGG)-mediated homologous recombination between *nDart1-3* and the first intron of *Os11G0637700* (Fig. 4-13).

To validate the function of LGG in regulating grain length, *LGG* knockout, and overexpressed transgenic plants were produced. It was confirmed that LGG controls grain length through

altering the cell number in the spikelet hull (Fig. 4-12), while the ploidy levels of 1-mm young panicle in Nipponbare, GE, and OE lines were found to be unchanged (Fig. 4-14). These results suggest that the regulatory mechanism for spikelet hull development could be associated with cell cycle-related pathway.

There was no significant difference in the number of spikelets per panicle among Nipponbare, GE, and OE lines, whereas the spikelet number had a significant negative correlation to grain length in the F2 population from the cross between *Lgg* and Koshihikari. These results suggest that there might be another factor that had linked to *LGG* affecting spikelet number. The grain fertility of GE lines was at the same level as Nipponbare. However, OE lines showed decreased fertility to Nipponbare, suggesting that the function of *LGG* might have an effect on grain fertility.

To date, the role of RNA-binding proteins (RBPs) in grain size regulation has not been reported. RBPs participate in RNA metabolism and influence plant growth and the response to stresses (Lorković, 2009; Ambrosone et al., 2012). There are only a few reports about the involvement of RBPs in rice development, for example, *MEIOSIS ARRESTED AT LEPTOTENE1* (*MEL1*) and *MEL2* (Nonomura et al., 2007, 2011), the components of exon junction complex (EJC) (Gong and He, 2014; Huang et al., 2016), *LAGGING GROWTH AND DEVELOPMENT 1* (*LGD1*) (Thangasamy et al., 2012), and RBP-P (Tian et al., 2018). Therefore, this study could shed new insight into the regulatory mechanism of grain size in rice.

Fig. 4-13. The relationship between *nDart1-3* insertion-deletion in the 5′ UTR of *LGG* in the *Lgg* mutant. Blue GGC: microhomology for putative homologous recombination. Blue line surrounded sequence: *nDart1-3* inserted in 5′ UTR of *LGG*. Green line surrounded sequence: deletion sequence. Underlined sequence: target site duplication.

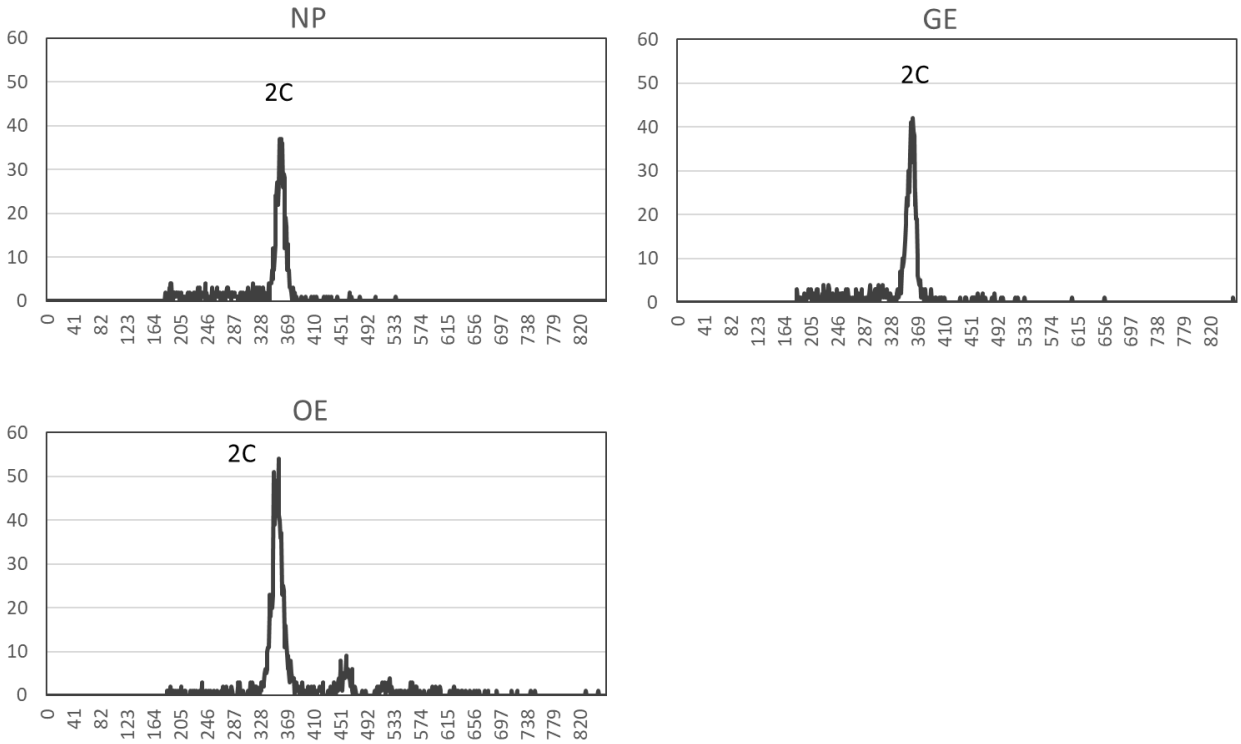

Fig. 4-14. The ploidy level of 1-mm young panicle in NP, GE, and OE. 2C; Unreplicated.

Chapter 5 Functional analysis of LGG

5-1 Introduction

RBPs play a crucial role in post-transcriptional gene regulation and RNA metabolisms, such as mRNA processing, export, localization, translation, and degradation. Most of RBPs contain one or more conserved domains with the common RNA binding domains including recognition RNA motif (RRM), K homology (KH) domain, zinc finger domain, DEAD/DEAH box (Asp-Glu-Ala-Asp/Asp-Glu-Ala-His), Pumilio/FBF domain, double-stranded RNA binding domain (DS-RBD), Piwi/Argonaute/Zwille (PAZ) domain, etc. (Ambrosone et al., 2012; Fedoroff, 2002). RRM and KH domains are conserved through evolution and have been found in various species from bacteria to human (Burd and Dreyfuss, 1994). RRM contains two consensus sequences, ribonucleoprotein (RNP) domains 1 and 2, arranged in $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ order and folds into an αβ sandwich structure. RRM provides high RNA-binding affinity and specificity. Two or more RRMs intensely increase the affinity. Moreover, the proteins with two RRMs have also been found to be involved in protein-protein interaction. Consequently, the RRMs contribute to multiple functions in RNA regulation (Maris et al., 2005).

RBPs are also involved in development, stress response and genome organization in plants (Ambrosone et al., 2012; Fedoroff, 2002; Lorković, 2009). For example, FCA (Macknight et al., 1997), FPA (Schomburg et al., 2001), FLK (Lim et al., 2004) and AtGRP7 (Streitner et al., 2008) regulate floral transition and induction in *Arabidopsis thaliana*. MEL1 and MEL2 regulate the premeiotic cell cycle transition for gametes formation in rice (Nonomura et al., 2007; Nonomura et al., 2011). The components of EJC regulate growth and development at both vegetative and reproductive stages in rice (Gong and He, 2014; Huang et al., 2016). The

cellular structures and organelles such as cytoskeleton, chloroplasts, and mitochondria are also associated with the function of RPBs in plants (Tian and Okita, 2014; Ichinose and Sugita, 2016; Nawaz and Kang, 2017).

The functions of RBPs in rice remains largely unknown. In the previous chapter, using *Lgg* of *nDart1*-tagged lines we identified a putative RBP regulating grain size. To reveal the function of LGG, analyses of *LGG* expression and transcriptome profiles of Koshihikari, *Lgg*, Nipponbare, and the transgenic plants using microarray and RNA-Seq were carried out.

5-2 Materials and Methods

5-2-1 Plant materials

Koshihikari, *Lgg*, Nipponbare, CRISPR/Cas9-mediated knockout line, and overexpression lines were used, as described in Chapter 1 and Chapter 4.

5-2-2 Quantitative reverse transcription PCR

RNAs were reverse transcripted to cDNA using PrimerScript™ RT reagent Kit (Perfect Real Time) (TaKaRa), and qPCR was performed using SYBR® Premix Ex TaqTM (Perfect Real Time) (TaKaRa) by LightCycler® 2.0 (Roche) with three biological replicates and three technical replicates. The ubiquitin was used as the internal control, and the primers, 15Nv05_F and 15Nv06_R, used for qRT-PCT are listed in Table 4-1, and the PCR conditions were shown below;

Denature	PCR	Melting	Cooling
1 cycle	40 cycles	0.1°C/sec.	
95°C 10 sec.	95°C 5 sec. 55°C 15 sec. 72°C 20 sec.	60°C to 95°C 15 sec.	40°C 30 sec.

5-2-3 Protein extraction and SDS-PAGE

Five mature grains of Koshihikari and the Lgg mutant were crushed with mortar and pestle and 30 mg of the powder was stirred with a vortex mixer for 20 min in a buffer containing 1 mL SDS-Urea solution, 4 M Urea, 2% sodium laurel sulfate, 5% 2-mercaptoethanol, 10% glycerin and 50 mM Tris-HCl buffer, pH 6.8. After centrifugation at 7,000 g for 10 min, 10 µL of the supernatant was subjected to SDS-PAGE. SDS-PAGE was carried out according to Laemmli (1970) on a 12% acrylamide gel. After electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250 (CBB-R250).

5-2-4 Vector construction and transformation

The sub-cellular localization vector was derived from pCAMBIA1305.1. p35S:GFP-LGG-CL and p35S:LGG-GFP-NL were assembled from LGG cDNA (AK105751) and EGFP fragments (Fig. 5-1 and Table 4-1), respectively. *Agrobacterium*-mediated transformations were performed, as described in Chapter 4. Rice calli were observed and photos were taken by confocal microscopy (A1, Nikon). DAPI (4′,6-Diamidino-2-phenylindole) is used for nuclear staining.

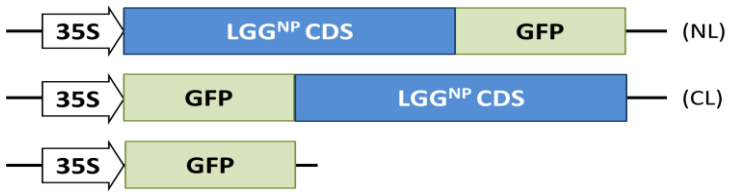

Fig. 5-1. Diagram of the constructs for sub-cellular localization.

5-2-5 Phylogenetic analysis

Amino acid sequences of 102 LGG orthologs from 17 species were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/) with the category of the ORTHOLOGY: K14411 and the non-redundant protein sequences database on the National Center for Biotechnology Information (NCBI) website

(https://www.ncbi.nlm.nih.gov/) by BLAST using the full length of amino acid of LGG. The homologs of LGG were listed in Table 5-1. Sequence alignment and the Neighbor-Joining Phylogenetic tree was constructed using the Poisson correction method with MEGA6.0 (Tamura et al., 2013).

5-2-6 Microarray and RNA-Seq analysis

RNA extraction was performed as described in Chapter 4, and RNA integrity and quality were checked by 2100 Bioanalyzer (Agilent Technologies). Total RNAs were extracted from 1- and 2-mm long young panicles of Koshihikari and the Lgg mutant, respectively. RNA samples with three biological replicates were labeled and hybridized on rice 44k microarray chips (Agilent Technologies), and the data were analyzed using GeneSpring (Agilent Technologies) with 200-row signal and 3-fold-change cut-off. For RNA-Seq analysis, RNA samples were isolated from 1-mm long young panicles of Nipponbare, GE, and OE with three biological replicates. The RNA samples were sequenced using Illumina Hi-Seq 4000 platform generating 100-bp paired-end reads (BGI, Japan). Data sets of Next-generation sequencing (NGS) RNA sequences were analyzed using the protocol as described previously (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 (Os-Nipponbare-Reference-IRGSP-1.0, https://rapdb.dna.affrc.go.jp) and determine the splicing sites of transcripts. For assembling the reads to the genome, Cufflinks version 2.2.1 (http://cole-trapnell-lab.github.io/cufflinks/cuffmerge/) and as reference genome annotation, the Ensemble Plants *O. sativa japonica* transcriptomes (https://plants.ensembl.org/Oryza_sativa/Info/Index) were used. Then, Cuffmerge was used for assembling the transcripts from individual reads. Finally, Cuffdiff calculated the 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 was used g:profiler (https://biit.cs.ut.ee/gprofiler/).

5-3 Results

5-3-1 The expression and subcellular localization of LGG

To examine the expression levels of *LGG*, qRT-PCR was carried out using the leaves and roots of 2-week-old seedlings and four different sizes (0.6, 1, 2, and 3 mm) of young panicles. It has been reported that the beginning of spikelet hull primordia differentiation occurs when the young panicle reaching to 1.5 mm (Hoshikawa, 1989). *LGG* expression was examined in all tested tissues (Fig. 5-2 and 5-3). *LGG* transcripts in the leaves were about twice higher than that in the roots (Fig. 5-2). The 0.6-mm young panicle tissue had the highest *LGG* mRNAs level among different sizes of young panicles in Koshihikari (Fig. 5-3). By contrast, *Lgg* mRNAs in the mutant were significantly lower (36.2% to 4.5%) than those in Koshihikari (Fig. 5-2 and 5-3). These results suggest that the mutation negatively affects *LGG* gene expression levels.

LGG contains two RRMs at the N-terminal region and is annotated as a putative RBP. To investigate the molecular function of LGG, sub-cellular localization of LGG was analyzed using two kinds of LGG and green fluorescent protein (GFP) fused constructs, 35S:LGG^{NP}-GFP (NL) and 35S:GFP-LGG^{NP} (CL) (Fig. 5-1). GFP fluorescent was observed in nuclei and co-localized with DAPI signals in both NL and CL expressing calli (Fig. 5-4), suggesting that LGG localizes in the nucleus.

To analyze the protein level in the grain, total protein was extracted from the brown rice of Koshihikari and *Lgg* and subjected to SDS-PAGE analysis. As shown in Fig. 5-5, the amounts of total protein were similar between Koshihikari and the mutant.

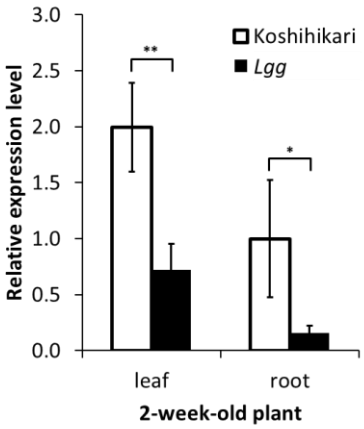

Fig. 5-2. Expression of *Os11g0637700* in leaf and root of the 2-week-old plant in Koshihikari and *Lgg*. Data are given as mean ± SD (n=3). * and ** indicate $P < 0.05$ and $P < 0.01$ by Student's *t*-test, respectively. *Ubiquitin* is used as internal control.

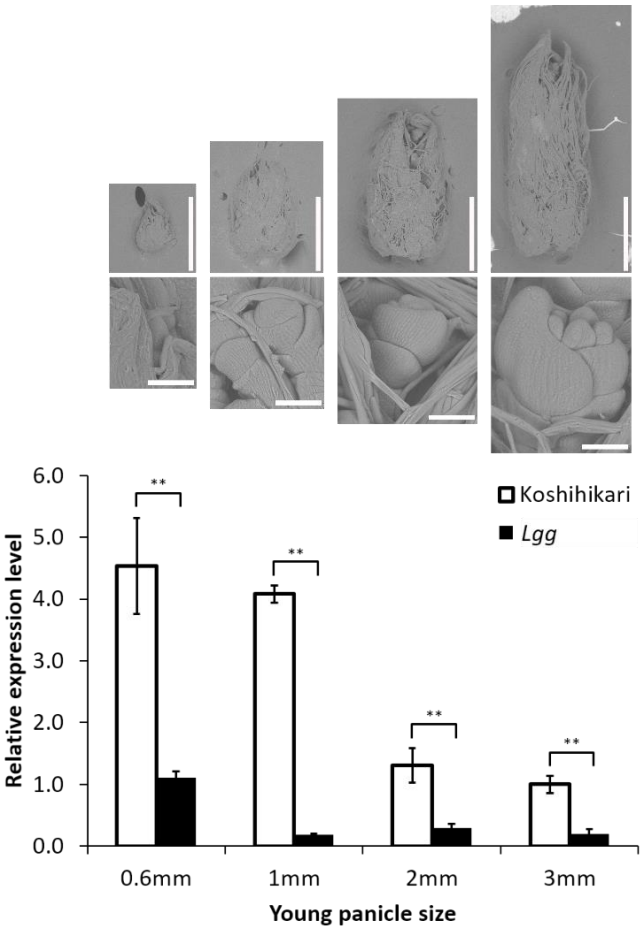

Fig. 5-3. The morphology of young panicles and spikelet differentiated and expression of *Os11g0637700* in 0.6-, 1-, 2-, and 3-mm long young panicle in Koshihikari and *Lgg*. Vertical bar = 1 mm. Horizontal bar = 100 μm. Data are given as mean ± SD (n=3). * and ** indicate $P < 0.05$ and $P < 0.01$ by Student's *t*-test, respectively. *Ubiquitin* is used as internal control.

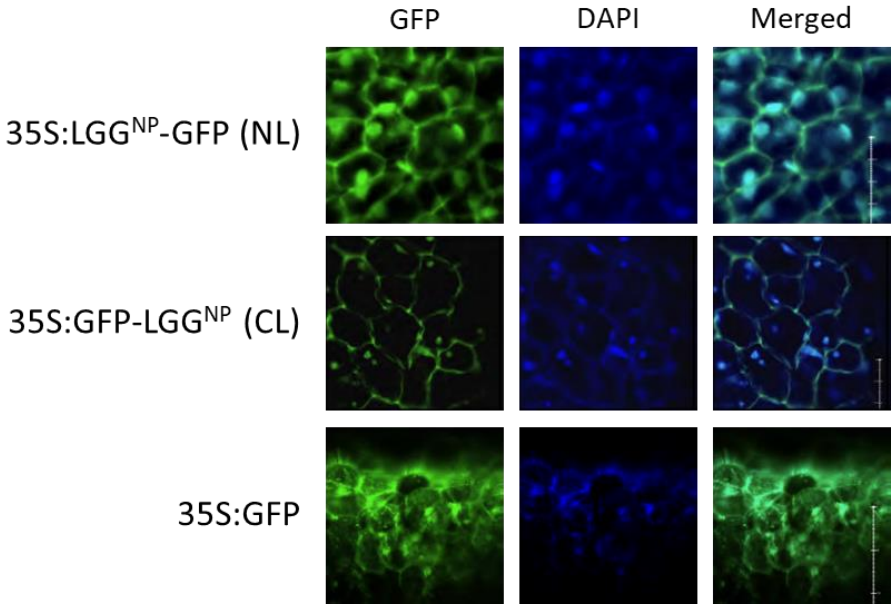

Fig. 5-4. Sub-cellular localization of LGG. Two types of fused proteins, 35S:LGG^{NP}- GFP (NL) and 35S:GFP-LGG^{NP} (CL) are expressed in rice calli. DAPI was used for nuclear staining. Bar = 20 μm.

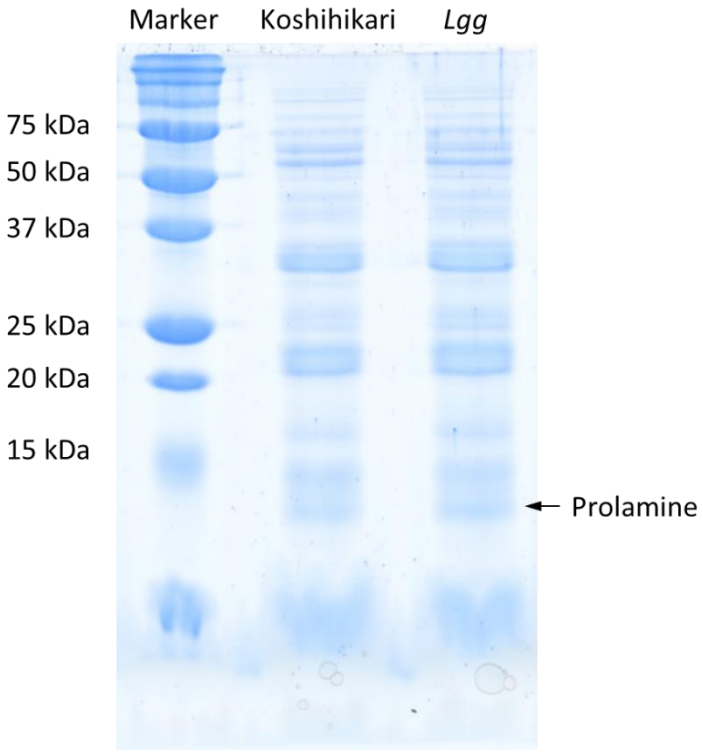

Fig. 5-5. SDS-PAGE analysis of seed protein extracted from whole mature grain in Koshihikari and *Lgg*.

5-3-2 Phylogenetic analysis of LGG homologs

From the database search, there are six homologs in the rice genome (KEGG ORTHOLOGY: K14411), including LGG itself (Fig. 5-6). Homologs of LGG were present across plants, fungi, and animals, and the number of LGG homologs in plant species is more than in animals (Table 5-1). The phylogenetic tree was built based on the full-length amino acid sequences of 102 homologs from 17 species. As shown in Fig. 5-7, the phylogenetic analysis demonstrated that the homologs from animals, dicot plants and monocot plants are separated into different clades, and LGG was grouped into a clade far related to the animal ones.

5-3-3 The transcriptome profiles of Koshihikari and *Lgg* using microarray

To find the possible regulatory pathway of LGG for grain length, microarray analysis was performed using 1- and 2-mm young panicles (1mm YP and 2mm YP) of Koshihikari and *Lgg* with three biological repeats. There were 19 down-regulated and seven up-regulated genes in 1mm YP (Table 5-2), and 37 down-regulated and two up-regulated genes in 2mm YP of the mutant (Table 5-3). The reduced signals of *LGG* (*Os11g0637700*) were also detected in both 1mmYP and 2mmYP. The extremely down-regulated gene (*Os04g0518000*) in both 1mm YP and 2mm YP was the gene annotated as *ADENOSINE KINASE 2* on chromosome 4. It is possible that a deletion had occurred at the *ADENOSINE KINASE 2* because it could not be amplified by specific primers on the genomic region in the mutant. Moreover, many of the down-regulated genes are the heat-shock proteins and uncharacterized genes, and interestingly, five of the seven up-regulated genes in 1mm YP were MADS-box genes that are related to floral identity regulation. To validate the expression, eight down-regulated and four up-regulated genes were selected for qRT-PCR. Except for *Os04g0352400* and *Os11g0506800*, others down-regulated and up-regulated genes did not show any significant difference between Koshihikari and the mutant (Fig. 5-8 and 5-9).

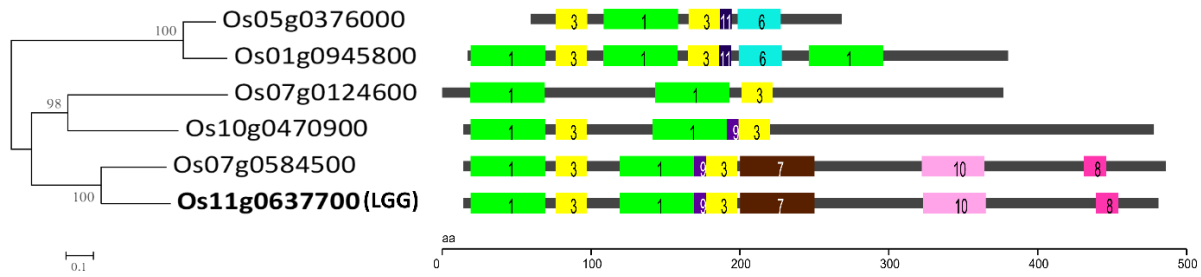

Fig. 5-6. The phylogeny of LGG homologs in rice. Motifs were generated using SALAD database (Mihara et al., 2010).

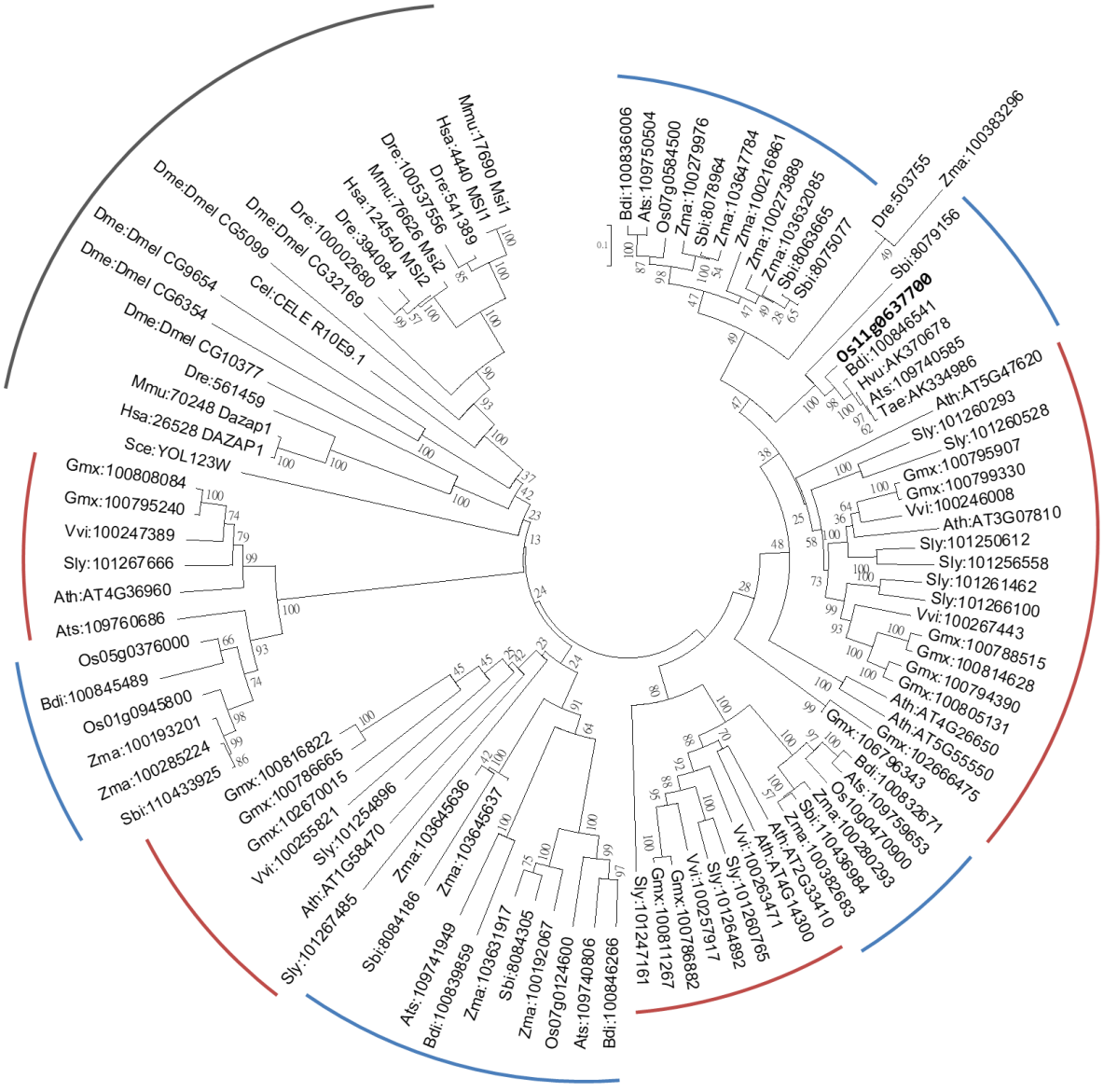

Fig. 5-7. Phylogenetic tree of 102 homologs of LGG in 17 species. Blue, red and black lines indicate monocotyledon, dicotyledon and animal clades, respectively.

Table 5-1 102 of LGG homologs from 17 species

Species	Identity	Similarity	Description
Arabidopsis thaliana (thale cress) (8)	46.7	66.3	>ath:AT1G58470 RBP1
	40.5	54.8	>ath:AT2G33410 RNA recognition motif-containing protein
	57.9	73.1	>ath:AT3G07810 RNA recognition motif-containing protein
	55.9	76.7	>ath:AT4G14300 heterogeneous nuclear ribonucleoprotein 1
	51.5	64.9	>ath:AT4G26650 RNA recognition motif-containing protein
	38	63.1	>ath:AT4G36960 RNA recognition motif-containing protein
	54.4	69.1	>ath:AT5G47620 RNA recognition motif-containing protein
	52.1	66.5	>ath:AT5G55550 RNA recognition motif-containing protein
Aegilops tauschii (wheat D†) (6)	91.2	94.9	>Ats:109740585
	41.6	57.4	>Ats:109740806
	39.3	54.2	>Ats:109741949
	60.5	73.7	>Ats:109750504
	39.7	53.1	>Ats:109759653
	34.4	58.3	>Ats:109760686
Brachypodium distachyon (6)	40.2	53.3	>bdi:100832671 heterogeneous nuclear ribonucleoprotein 1-like
	61.3	74.6	>bdi:100836006 heterogeneous nuclear ribonucleoprotein 1
	39.3	58.8	>bdi:100839859 RNA-binding protein 1-like
	35.5	58.4	>bdi:100845489 heterogeneous nuclear ribonucleoprotein A2 homolog 2
	42.3	57.7	>bdi:100846266 heterogeneous nuclear ribonucleoprotein 1-like
	91.2	94.9	>bdi:100846541 heterogeneous nuclear ribonucleoprotein 1-like
Caenorhabditis elegans (nematode)	47.3	67.9	>cel:CELE_R10E9.1 msi-1
Drosophila melanogaster (fruit fly) (5)	40.7	62.8	>dme:Dmel_CG10377 Hrb27C
	48.9	66.8	>dme:Dmel_CG32169 Rbp6
	41.4	63	>dme:Dmel_CG5099 msi
	39.5	61	>dme:Dmel_CG6354 Rb97D
	38.5	55.5	>dme:Dmel_CG9654 Rbp4
Danio rerio (zebrafish) (6)	44.1	63.4	>dre:100002680 msi2a
	45.1	65.2	>dre:100537556 RNA-binding protein Musashi homolog 1-like
	44.6	63.4	>dre:394084 msi2b
	45.2	69.9	>dre:503755 cb676
	45.8	65.6	>dre:541389 msi1
	46.7	68.5	>dre:561459 dazap1
Glycine max (soybean) (15)	47.9	65.4	>gmx:100786665 RNA-binding protein 1-like
	40.8	54.8	>gmx:100786882 heterogeneous nuclear ribonucleoprotein 1-like
	58.4	73.4	>gmx:100788515 heterogeneous nuclear ribonucleoprotein 1
	61.1	77.1	>gmx:100794390 heterogeneous nuclear ribonucleoprotein 1-like
	38.4	60	>gmx:100795240 DAZ-associated protein 1
	63.8	76.8	>gmx:100795907 heterogeneous nuclear ribonucleoprotein 1-like
	63.3	75.9	>gmx:100799330 heterogeneous nuclear ribonucleoprotein 1-like
	61.8	77.1	>gmx:100805131 heterogeneous nuclear ribonucleoprotein 1-like
	39.2	59.8	>gmx:100808084 heterogeneous nuclear ribonucleoprotein 1-like
	41.1	55.3	>gmx:100811267 heterogeneous nuclear ribonucleoprotein 1-like
	58.8	73.8	>gmx:100814628 heterogeneous nuclear ribonucleoprotein 1
	49.8	66.7	>gmx:100816822 RNA-binding protein 1-like
	50.6	74.2	>gmx:102666475 heterogeneous nuclear ribonucleoprotein 1-like
	65.8	81	>gmx:102670015 RNA-binding protein 1-like
	60.3	82.2	>gmx:106796343 heterogeneous nuclear ribonucleoprotein 1-like
Homo sapiens (human) (3)	45	63.9	>hsa:124540 MSI2
	41.6	59.7	>hsa:26528 DAZAP1
	46.9	65.1	>hsa:4440 MSI1
Hordeum vulgare (barley)	91	94.7	>Hvu:AK370678
Mus musculus (mouse) (3)	46.9	64.6	>mmu:17690 Msi1
	41.6	60.2	>mmu:70248 Dazap1
	45	63.9	>mmu:76626 Msi2

†; names were according to KEGG database. Identity and similarity were calculated by EMBOSS Matcher (https://www.ebi.ac.uk/Tools/psa/emboss_matcher/).

Table 5-1 Continued

Species	Identity	Similarity	Description
Oryza sativa japonica (Japanese rice†) (6)	36	59.9	>dosa:Os01g0945800
	34.1	57.9	>dosa:Os05g0376000
	42.9	59.5	>dosa:Os07g0124600
	60.5	74.1	>dosa:Os07g0584500
	38.4	52.2	>dosa:Os10g0470900
	100	100	>dosa:Os11g0637700
Sorghum bicolor (sorghum) (8)	37	60.2	>sbi:110433925
	41.2	54.4	>sbi:110436984
	61.9	75.2	>sbi:8063665
	62.1	75.2	>sbi:8075077
	61.5	75	>sbi:8078964
	68.3	78	>sbi:8079156
	40	56.2	>sbi:8084186
	42.9	59	>sbi:8084305
Saccharomyces cerevisiae (budding yeast)	40.1	58.6	>sce:YOL123W HRP1
Solanum lycopersicum (tomato) (12)	56.2	74.2	>sly:101247161 heterogeneous nuclear ribonucleoprotein 1-like
	62.3	74.1	>sly:101250612 heterogeneous nuclear ribonucleoprotein 1
	44.2	61.1	>sly:101254896 RNA-binding protein 1
	57.8	71.2	>sly:101256558 heterogeneous nuclear ribonucleoprotein 1-like
	56.9	70	>sly:101260293 heterogeneous nuclear ribonucleoprotein 1
	53.1	66.4	>sly:101260528 heterogeneous nuclear ribonucleoprotein 1
	41.5	56.5	>sly:101260765 heterogeneous nuclear ribonucleoprotein 1
	60.3	72.9	>sly:101261462 heterogeneous nuclear ribonucleoprotein 1-like
	41.3	56.6	>sly:101264892 heterogeneous nuclear ribonucleoprotein 1-like
	59.5	74.6	>sly:101266100 heterogeneous nuclear ribonucleoprotein 1-like
	42.9	67	>sly:101267485 RNA-binding protein 1
	38	58	>sly:101267666 heterogeneous nuclear ribonucleoprotein A0
Triticum aestivum (wheat)	91	94.9	>Tae:AK334986
Vitis vinifera (wine grape) (6)	64.7	78.6	>vvi:100246008 heterogeneous nuclear ribonucleoprotein 1
	38.9	58.5	>vvi:100247389 RNA-binding protein Musashi homolog 2
	51.9	66.5	>vvi:100255821 RNA-binding protein Musashi homolog 2-like
	44.7	58.7	>vvi:100257917 heterogeneous nuclear ribonucleoprotein 1
	42.7	56.5	>vvi:100263471 heterogeneous nuclear ribonucleoprotein 1-like
	63.4	77	>vvi:100267443 heterogeneous nuclear ribonucleoprotein 1
Zea mays (maize) (14)	42	56.6	>zma:100192067 GRMZM2G042343; Ribonucleoprotein like protein
	38.1	61.9	>zma:100193201 GRMZM2G060977; heterogeneous nuclear ribonucleoprotein 27C
	57.8	71.8	>zma:100216861 pco066270, GRMZM2G162954
	60.8	74.5	>zma:100273889
	60	73.1	>zma:100279976 pco083367(678), GRMZM2G022313
	41.5	53.6	>zma:100280293 pco101966, GRMZM2G167505
	37.6	60.8	>zma:100285224 GRMZM2G030902; heterogeneous nuclear ribonucleoprotein 27C
	41.9	55.1	>zma:100382683 uncharacterized LOC100382683
	48.5	64.3	>zma:100383296 GRMZM2G346701; uncharacterized LOC100383296
	43.5	60.3	>zma:103631917 GRMZM2G006071; DAZ-associated protein 1-like
	61.4	74.5	>zma:103632085 GRMZM2G013065; heterogeneous nuclear ribonucleoprotein 1-like
	38.1	55.6	>zma:103645636 GRMZM2G456186; heterogeneous nuclear ribonucleoprotein 1-like
	39	54.7	>zma:103645637 GRMZM2G430461; heterogeneous nuclear ribonucleoprotein 1-like
	60.9	73.9	>zma:103647784 GRMZM2G014400; heterogeneous nuclear ribonucleoprotein 1

†; names were according to KEGG database. Identity and similarity were calculated by EMBOSS Matcher (https://www.ebi.ac.uk/Tools/psa/emboss_matcher/).

The qRT-PCR results confirmed that *Os04g0352400* showed reduced expression in 1mm YP of the mutant. This gene encodes peptidyl-prolyl cis-trans isomerase with FKBP-type domain, but the function remains unknown. The signal of *Os11g0506800* was decreased in microarray analysis, but the qRT-PCR result showed significantly increased expression level in 1mm YP of the mutant. However, the expressions of *Os11g0506800* were low in both Koshihikari and *Lgg*. The reason for the different results between microarray and qRT-PCR is unknown.

Table 5-2 Down-regulated genes expressed in young panicles of *Lgg* by microarray

Locus ID	Fold-change		Description
	1mm	2mm	
Os04g0518000	-196.68	-180.83	Similar to Adenosine kinase (Fragment). Adenosine kinase 2
Os03g0133100 †	-15.25	-9.23	Hypothetical protein. Uncharacterized
Os11g0506800 †	-12.87	-33.62	Uncharacterized; OsBAG5
Os03g0267000	-6.36	-7.14	Low molecular mass heat shock protein Oshsp18.0. 18.1 kDa class I heat shock protein; OSHSP18.0
Os06g0592500	-5.03	-3.24	Similar to Ethylene-responsive transcriptional coactivator. multiprotein-bridging factor 1c
Os06g0728700 †	-4.96	-3.55	Homeodomain-like containing protein. protein REVEILLE 1
Os11g0637700	-4.67	-3.11	Nucleotide-binding, alpha-beta plait domain containing protein. heterogeneous nuclear ribonucleoprotein 1
Os02g0225300	-4.04	-4.29	Conserved hypothetical protein. Uncharacterized
Os04g0352400 †	-3.77	-3.68	Peptidyl-prolyl cis-trans isomerase, FKBP-type domain containing protein. OsFKBP62b, OsFKBP65
Os02g0758000	-3.09	-5	Similar to Low molecular weight heat shock protein precursor (Mitochondrial small heat shock protein 22). 24.1 kDa heat shock protein, mitochondrial; OsHsp24.1
Os02g0731300	-3.03	-4.18	Non-protein coding transcript.
Os01g0880200 †	-4.69	--	Similar to secondary cell wall-related glycosyltransferase family 8. UDP-glucuronate:xylan alpha-glucuronosyltransferase 1
Os02g0606200	-4.64	--	Zinc finger, B-box domain containing protein. B-box zinc finger protein 25
Os05g0519700	-3.81	--	Similar to Heat shock protein 101. chaperone protein ClpB1; Oshsp101
Os01g0136100	-3.67	--	16.9 kDa class I heat shock protein 1.
Os02g0782500	-3.64	--	Similar to 17.5 kDa class II heat shock protein. 18.6 kDa class III heat shock protein
Os01g0264000 †	-3.55	--	Zinc finger, Dof-type family protein. cyclic dof factor 2
Os03g0305600	-3.41	--	Mitochondrial import inner membrane translocase, subunit Tim17/22 family protein. outer envelope pore protein 16-2, chloroplastic
Os08g0157600	-3	--	Oryza sativa late elongated hypocotyl., Circadian clock., Myb transcription factor. protein LHY; OsCCA1, COC1
Os11g0454200	--	-7.53	Dehydrin RAB 16B. dehydrin Rab16B; OsLEA28, OsRAB16D, OsRAB16b
Os01g0733200	--	-5.83	Similar to Heat shock transcription factor 29 (Fragment). heat stress transcription factor C-1b; HSF3/11
Os01g0822800	--	-5.32	Similar to RING-H2 finger protein ATL3C. RING-H2 finger protein ATL74
Os01g0102900	--	-4.67	Light regulated Lir1 family protein. light-regulated protein
Os08g0546800	--	-4.56	Similar to Heat stress transcription factor B-2b. heat stress transcription factor B-2b; HSF2, HSF21
Os04g0568700	--	-4.52	Similar to Heat stress transcription factor Spl7 (Heat shock factor RHSF10). heat stress transcription factor B-2a; OsHsf-14/1
Os06g0317200	--	-4.36	Glycine-rich protein (Fragment). keratin-associated protein 21-1
Os01g0971800 †	--	-4.3	Similar to Two-component response regulator ARR11 (Receiver-like protein 3). transcription factor PCL1
Os03g0265900	--	-4.19	Conserved hypothetical protein. Uncharacterized

Os07g0489800	--	-3.88	Beta-grasp fold, ferredoxin-type domain containing protein. photosynthetic NDH subunit of subcomplex B 3, chloroplastic
Os03g0304800	--	-3.71	Lg106-like family protein. Uncharacterized
Os12g0561900	--	-3.7	Conserved hypothetical protein. UDP-glycosyltransferase 91C1
Os02g0181300 †	--	-3.44	Similar to WRKY transcription factor. probable WRKY transcription factor 40; wrky38; OsWRKY71
Os09g0325700	--	-3.42	Similar to Protein phpsphatase 2C (PP2C) (EC 3.1.3.16). probable protein phosphatase 2C 68; OsPP2C1, OsPP108
Os11g0454300	--	-3.37	Similar to Water-stress inducible protein RAB21. water stress-inducible protein Rab21; OsRab16A, OsLEA29
Os03g0237500	--	-3.36	Protein of unknown function DUF775 family protein. protein OPI10 homolog
Os06g0142200	--	-3.34	Early nodulin. early nodulin-93; OsENOD93a
Os03g0184100	--	-3.33	Hypothetical protein. Uncharacterized
Os03g0586500	--	-3.27	Conserved hypothetical protein. Uncharacterized
Os05g0494600	--	-3.27	Conserved hypothetical protein. EID1-like F-box protein 3
Os09g0437500	--	-3.2	Dormancyauxin associated family protein. auxin-repressed 12.5 kDa protein
Os03g0276500	--	-3.18	Heat shock cognate 70 kDa protein 2; OsHsp71.1
Os04g0583200	--	-3.18	Conserved hypothetical protein.
Os03g0322900	--	-3.1	late embryogenesis abundant protein 1; OsLEA17/OsLEA3-2
Os10g0159300	--	-3.09	Conserved hypothetical protein.
Os08g0460000	--	-3.04	Similar to Germin-like protein 1 precursor. germin-like protein 8-14; OsGLP1, GER1, GER5, GLP110, OsGER1, OsGER5

†; selected for qRT-PCR. --; indicates fold-change values below 3. Fold-change value is shown by the highest value of each gene.

Table 5-3 Up-regulated genes expressed in young panicles of *Lgg* mutant by microarray

Locus ID	Fold-change		Description
	1mm	2mm	
Os09g0507200 †	6.03	--	MADS box protein. MADS-box transcription factor 8; OsMADS8/24
Os01g0201700 †	4.97	--	Similar to MADS box protein. MADS-box transcription factor 3; RICE AGAMOUS1(RAG1)
Os04g0580700	4.78	--	MADS box transcription factor MADS17. MADS-box transcription factor 17
Os01g0530100	4.34	--	Conserved hypothetical protein.
Os02g0682200	3.41	--	Similar to MADS box protein. MADS-box transcription factor 6; MOSAIC FLORAL ORGANS 1 (MFO1)
Os07g0160100 †	3.39	--	Protein YABBY 1; FILAMENTOUS FLOWER 1 (FIL1); dropping leaf
Os08g0531700 †	3.31	--	MADS-box transcription factor 7; OsMADS7/45, AGAMOUS-like 6 (AGL6)
Os01g0106400	--	5.53	Similar to Isoflavone reductase homolog IRL (EC 1.3.1.-). isoflavone reductase homolog IRL
Os07g0484900	--	3.08	Protein of unknown function DUF1210 family protein.

†; selected for qRT-PCR. --; indicates fold-change values below 3. Fold-change value is shown by the highest value of each gene.

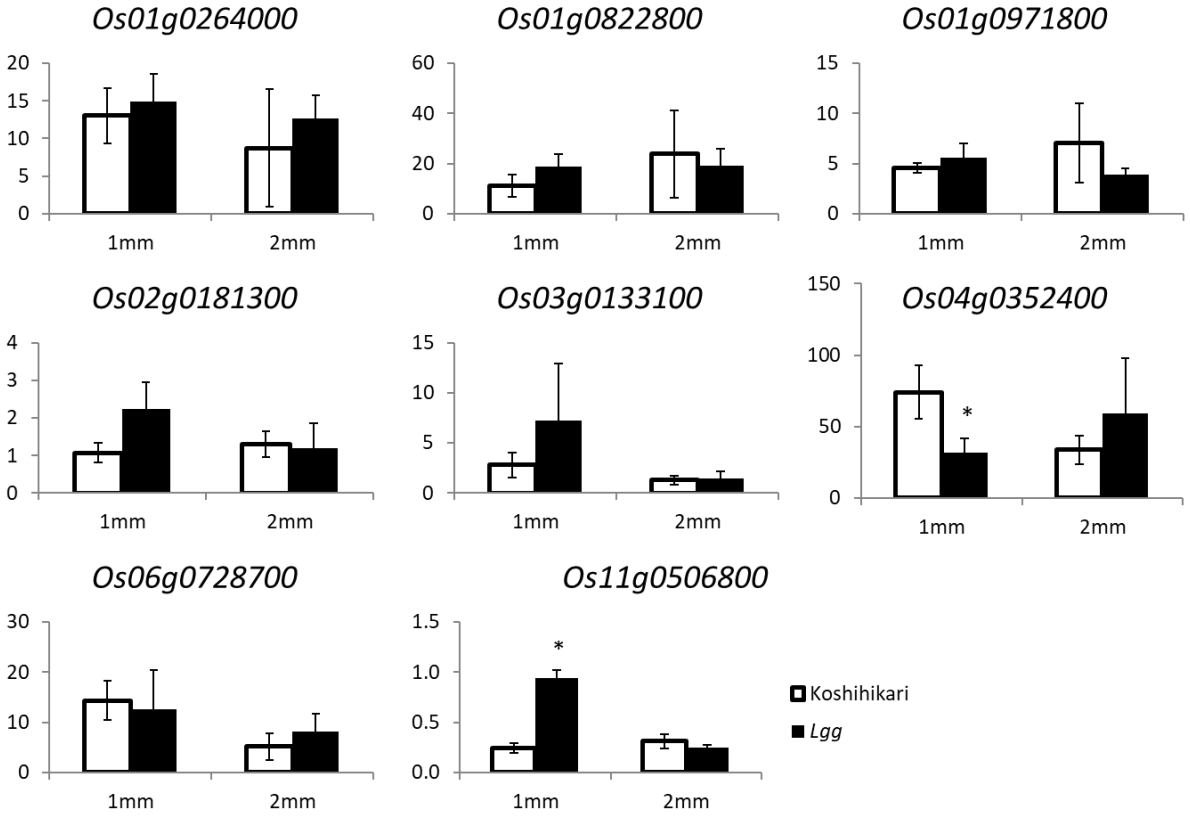

Fig. 5-8. Expression level of down-regulated genes selected from microarray analysis in 1- and 2-mm young panicle.

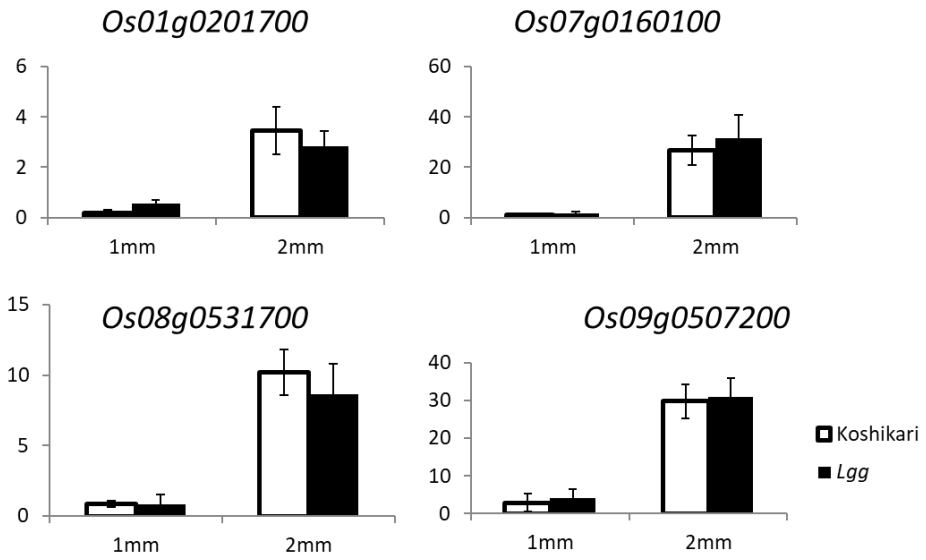

Fig. 5-9. Expression level of up-regulated genes selected from microarray analysis in 1- and 2-mm young panicle.

5-3-4 The transcriptome analysis in Nipponbare, GE and OE by RNA-Seq

RNA-Seq analysis was performed on 1-mm-long young panicles of Nipponbare (NP), CRISPR/Cas9-mediated knockout (GE), and overexpression (OE) plants. The 1-mm-long panicles were selected for analysis because *LGG* expression was the highest in 0.6-mm-long young panicles in Koshihikari, and LGG is expected to regulate downstream genes at later development stages. The mRNA levels of *LGG* were similar between NP and GE plants but significantly increased in OE plants. Moreover, as shown in Fig. 5-10 and 5-11, the expression of 477 genes in GE plants were reduced compared to NP, whereas 351 genes of GE plants 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 and cell proliferation genes, were found to be reduced in GE plants (Fig. 5-10). On the other hand, GE plants showed increased expressions of genes related to transcription, transcription factor activity and response to abiotic stimulus (Fig. 5-11). Furthermore, we examined the expression levels of the cell cycle, cell division, and cell proliferation genes in GE, OE, and NP plants, because LGG is suggested to regulate longitudinal cell number in the spikelet hull, and cell number is controlled by these genes. We found that 38 of 46 cell cycle-related genes showed lower expression and that eight genes showed higher expression levels in GE plants than OE and NP plants (Fig, 5-12). In particular, the order of expression levels of *CYCD3-2* (*Cyclin-D3-2*) and *Os04g0488100* (*OsRad21-2*) were GE < NP < OE and GE > NP > OE, respectively. Expression levels of half of the cell division and cell proliferation-related genes were reduced in GE plants compared to OE and NP plants. *Os05g0389000* (*OsAP2*/*ERF142*/*SMOS1*) showed the highest expression among GE, OE and NP with the expression level order, GE > NP > OE. Furthermore, *EL2* and *SMR*, cyclin-dependent kinase (CDK) inhibitors, showed increased expression in GE compared to OE and NP plants (Fig, 5-13).

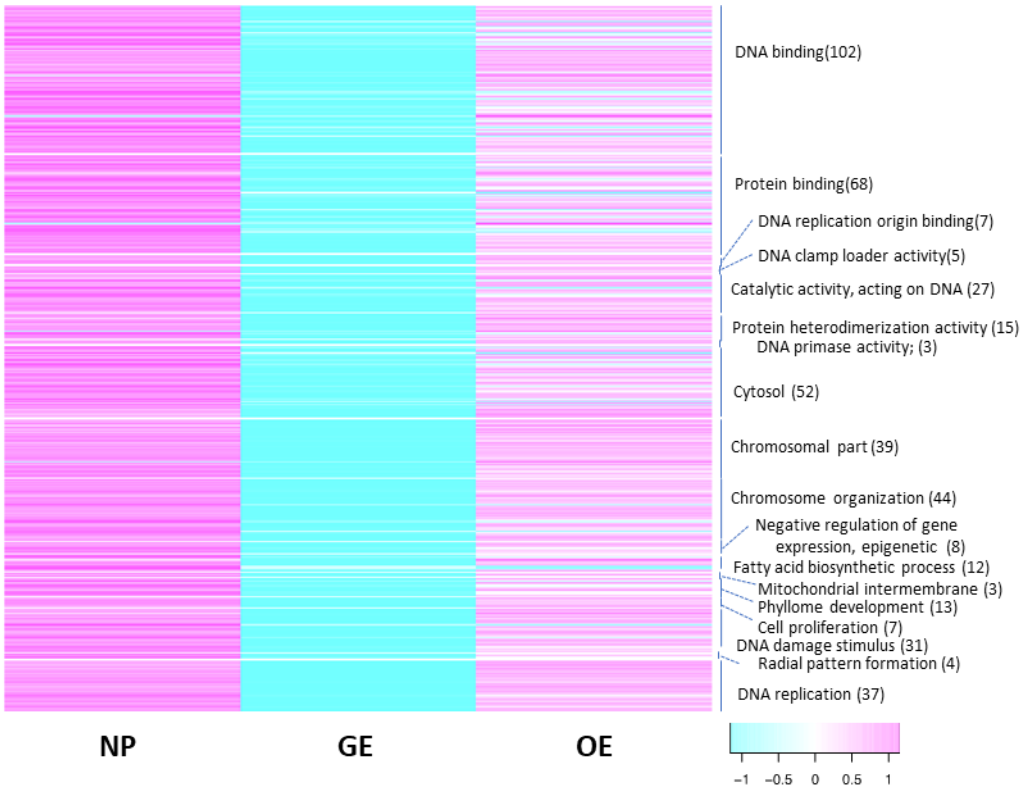

Fig.5-10. Profile of down-regulated genes in 1-mm-long young panicles of NP, GE, and OE.

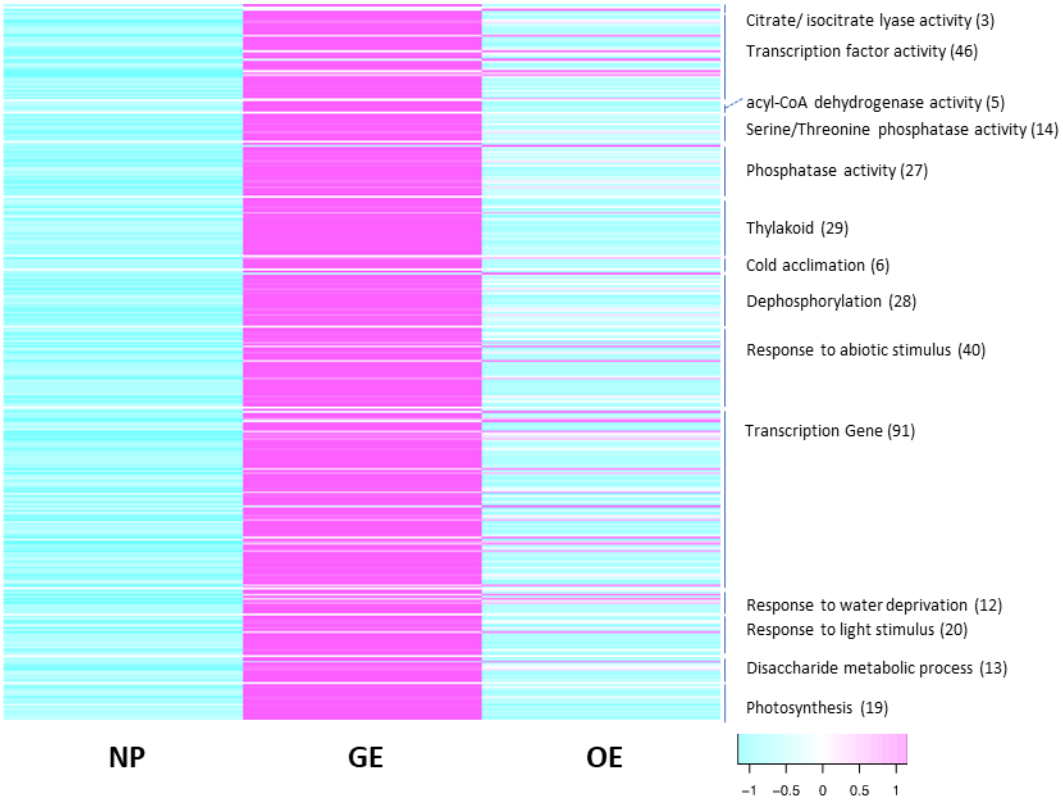

Fig. 5-11. Profile of up-regulated genes in 1-mm-long young panicles of NP, GE, and OE.

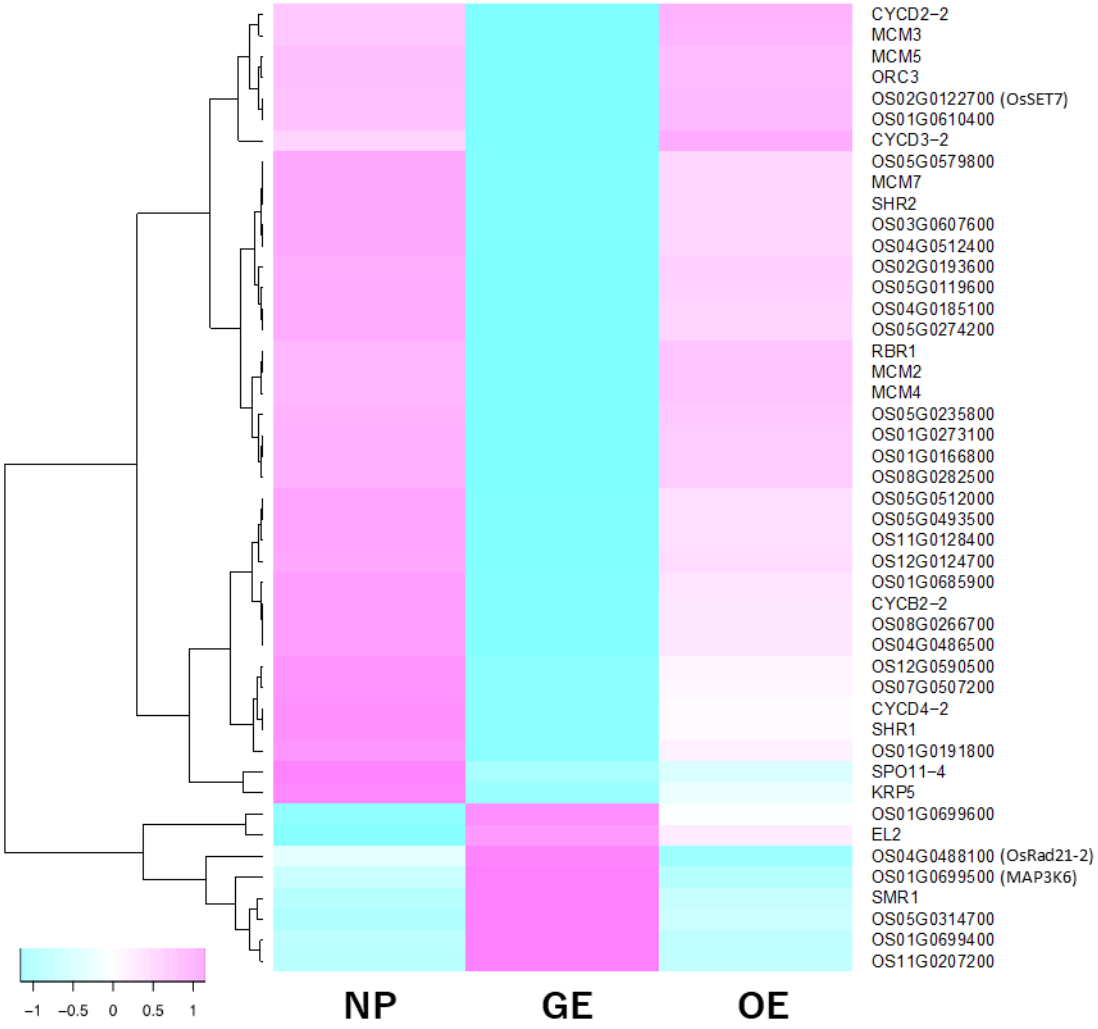

Fig. 5-12. Differential expression of cell cycle-related gene in GE, OE and NP using RNA-Seq.

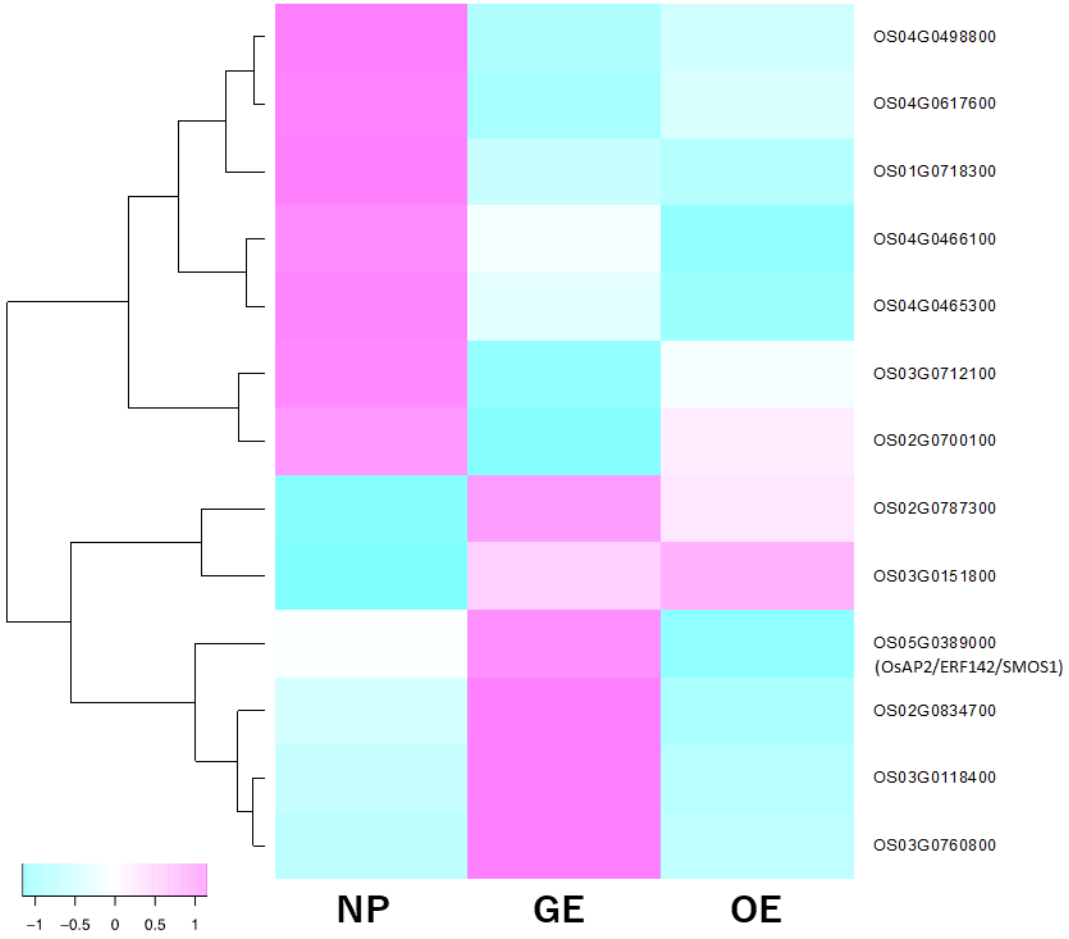

Fig. 5-13. Differential expression of cell division and cell proliferation-related gene in GE, OE and NP using RNA-Seq.

5-4 Discussion

LGG is the causal gene for grain length in *Lgg* found in *nDart1*-tagged lines and annotated as a putative RBP. The expressions of *LGG* was detected in root, leaf, and young panicles, and the expression showed the highest level in 0.6-mm-long young panicle and decreased along with young panicle growth (Fig. 5-2 and 5-3). RBPs function in post-transcription regulation based on their characteristics and RBPs are predicted to be localized in the nucleus and cytoplasm (Keene, 2007). Although the results of sub-cellular localization analysis showed the nuclear localization of LGG in rice calli (Fig. 5-4), further detailed sub-cellular localization analysis in young panicles is needed.

Phylogenetic analysis indicated that Os11g0637700 (LGG) and AtRBP1, which was first identified in Arabidopsis and had a role in cell proliferation in meristematic tissues (Suzuki et al., 2000), belong to the Musashi family (Fig. 5-7). *Musashi1* (*Msi1*) that regulated asymmetric differentiation of sensory organ precursor cells was discovered in *Drosophila* (Nakamura et al., 1994), and the homologs exist in fungi, animals, and plants (Fig. 5-7). This suggests that the genes belonging to the Musashi family might play a fundamental role in cell proliferation and differentiation.

In animals, *Musashi1* and *Musashi2* were reported to target *Numb* in Notch signaling, as reviewed by Fox et al. (2015), but the homologs of *Numb* and the Notch signaling have not been reported in plants. Plants contain more LGG homologs than animals, and LGG was grouped into a monocot-specific clade in the phylogenetic tree and distantly related to *Drosophila* Msi1 and human Msi1 (Fig. 5-7). Moreover, *LGG* was highly expressed in the spikelet hull primordium (Fig. 5-3). The spikelet hull is a specific characteristic of 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, LGG is likely to function in the spikelet hull differentiation in Gramineae.

It is likely that LGG controls a key player in a cell cycle-related gene network. Through RNA-Seq analysis of 1-mm-long young panicles of GE, OE, and Nipponbare plants, the expression levels of many cell cycle-related genes were found to be significantly reduced in GE lines (Fig. 5-12). One was *CYCD3-2*, differentially expressed in GE < NP < OE order. CYCD3 was reported to regulate cell proliferation and secondary growth in the cambium of *Arabidopsis* (Collins et al., 2015). CDK inhibitor genes, *EL2* and *SMR5*, were highly expressed in GE compared to OE and Nipponbare (Fig. 5-12). These genes are responsible for the control of cell cycle by the antagonistic interaction between the CYC-CDK complex and the inhibitors, EL2 (Peres et al., 2007) and SMR (Komaki and Sugimoto, 2012). Since CYCD3 is the target of SMR (Hamdoun et al., 2016), it is plausible that antagonistic expression of *SMR1* and *CYCD3-2* were observed in GE and OE.

On the other hand, the expression levels of some cell cycle-related genes, e.g. *Os01g0699500* (*MAP3K6*), *Os04g0488100* (*Rad21-2*), and *Os05g0389000* (*OsAP2*/*ERF142*/*SOMS1*) were significantly increased in GE plants compared to OE plants and Nipponbare (Fig. 5-12 and 5-13). Rad21/Rec8-like proteins are also related to cell division (da Costa-Nunes et al., 2006; Gong et al., 2011), and OsRAD21-2 controls cell division and growth in actively dividing tissues, including premeiotic flowers, stem intercalary meristems and leaf meristems (Gong et al., 2011). Aya et al. (2014) reported that *SMOS1* functions as an auxin-dependent regulator of cell expansion. On the other hand, *OsLG3*, encoding APETALA2 (AP2)/ethylene-responsive

element binding protein 125 (EREBP 125), functions as a positive regulator of grain length, by increasing the expression of cell cycle-related genes, such as *CDKA1*, *CDC3*, *MCM3*, and *CYCA2.1* (Yu et al., 2017b). However, the reason for reduced expression levels of many cell cycle-related genes and some cell division/cell proliferation genes in GE plants remain unknown.

Considering that AK105751 (Os11G0637700) is the key RBP for mRNA localization of prolamines at the mid-stage of seed development in rice and was closely related to AK067725 (Os07g0584500) (Crofts et al., 2010; Yang et al., 2014), *Lgg* was expected to show lower prolamines content in the mature grains. Prolamine is a storage protein with the molecular weight range from 10 to18 kDa and predominantly 13-15 kDa in rice (Li and Okita, 1993). Although the expression of *LGG* in seeds was not tested in this study, there were no obvious differences in the seed protein accumulations, including prolamines, between Koshihikari and the mutant in SDS-PAGE analysis (Fig. 5-5). Thus, it is unlikely that LGG is related to prolamines accumulation in the rice endosperm.

Chapter 6 General Discussion and Conclusion

Promoting grain yield is one of the main objectives of rice breeding. Grain weight is largely determined by grain size, therefore investigating the regulatory mechanisms of grain size can help us to control grain size and to improve rice yield. Many genes involving in the regulation of grain size have been identified and they function in different pathways (Li et al., 2018). However, the molecular regulatory network of grain size remains to be elucidated. To discover novel genes and define the gene function, one direct approach is to utilize transposon-tagged mutants. The active DNA transposon, *aDart* and its non-autonomous elements, *aDart/nDart* tagging system was established by Tsugane et al. (2006). The *nDart1*-tagged lines are very advantageous for gene functional analysis and breeding in rice (Maekawa et al., 2011). The aim of this study is to characterize *Lgg* mutant found from *nDart1*-tagged lines of Koshihikari as well as to investigate the gene implicated in grain size and to explore the novel regulatory pathway controlling grain size.

Lgg exhibited longer and heavier grains weight than those of Koshihikari. The mutated gene, *LGG* (*Os11g0637700*), was identified through the transposon display analysis and the cosegregation between grain-length phenotypes in F2 and F3 (Fig. 4-4; Table 4-3). *LGG* encodes a putative RBP with two RRMs (Fig. 4-7) and located at the long arm of chromosome 11. *Lgg* allele contains 500-bp truncated *nDart1-3* insertion and 355-bp deletion at 5′ UTR (Fig. 4-5), resulting in transcript initiation site shifted (Fig. 4-6) and reduced expression level (Fig. 5-2 and 5-3). A similar trend was also observed in other *nDart1*-insertion mutants. For example, the expression levels of *TAW1* (Yoshida et al., 2013) and *APO1* (Ikeda-Kawakatsu et al., 2009) were altered, while the transcription initiation sites of *OsClp5* (Tsugane et al., 2006) and *SWL1* (Hayashi-Tsugane et al., 2014) were shifted. These results indicate that *aDart*/*nDart* tagging

system is effective for gene tagging and functional study.

Grain size is limited by spikelet hull size, and the organ size is determined by the cell division and cell expansion. Segami et al. (2017) reported that known grain length-related genes were categorized into two types based on gene functions: for cell length (*GL7*/*GW7*, *D2*, *D11*, *D61*, *BRD1*, *SRS1*, *SRS3*, *SRS5* and *GLW7*) and for cell division (*GS3*, *GW2*, *qSW5*/*GW5*, *GS5*, *GW8*, *GL3.1*, *TGW6*, *D1*, *SG1* and *TUD1*). While *GS5* (Li et al. 2011) and *GW8* (Wang et al. 2012) are positive regulators for the cell cycle, *GW2* is a negative regulator of cell division (Song et al. 2007). *Lgg* and GE lines showing longer spikelet hull had more cells in the longitudinal direction, and OE lines had decreased cell number (Table 2-3 and Fig. 4-12), suggesting that *LGG* negatively regulates the spikelet hull size through altering of the cell number.

However, RNA-Seq analysis revealed that many cell cycle-related genes were down- and up-regulated in 1-mm-long young panicles of GE and OE lines, respectively (Fig. 5-12 and 5-13). 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. Loss of function of *TGW6* gene for endosperm length and grain filling in rice leads to long grain through reduced expressions of core cell cycle genes including *CycB2;2* (Ishimaru et al., 2013). In this study, through RNA-Seq analysis of GE lines, expressions of several cell cycle genes including *CYCB2-2* were also reduced (Fig. 5-12). Since *CycB2;2* works with the highest level at the G2-M transition and during M phase (de Veylder et al., 2007), it is likely that cell division in the spikelet hull primordium might be delayed with lower expressions of crucial cell cycle genes including *CYCB2-2* in GE lines, resulting in increased cell number of the spikelet hull. However, further investigation could elucidate how LGG played a role in cell proliferation and promotion of the

expressions of several AP2-type genes, although the expressions of core cell-cycle genes were reduced in the spikelet differentiation stage.

LGG localizes at the nucleus of the rice calli (Fig. 5-4) with consistent with RNA-binding characteristic of LGG. Previously, it was reported that LGG (AK105751) localizes at the nucleus and co-localizes with microtubules (Crofts et al., 2010). Furthermore, LGG protein is suggested to form a complex with other proteins (Yang et al., 2014). LGG protein starts to accumulate in seeds at seven days after flowering and play a role in the localization of prolamine mRNA (Crofts et al., 2010). Our results showed that prolamine content in seeds is similar between Koshihikari and *Lgg* (Fig. 5-5) but the grain quality has not been tested in *Lgg*. Thus, it will be interesting to examine whether *LGG* has pleiotropy effects on grain quality. On the other hand, *LGG* transcripts expressed in root, leaf, and young panicles (Fig. 5-2 and 5-3), however, the protein level of LGG before flowering remains unknown. To more understand LGG function, it is necessary to investigate LGG protein levels in different tissue.

In conclusion, the results in the present study showed that *LGG* encodes a putative RBP belonging to Musashi family and it is involved in regulation of grain size. LGG might regulate downstream genes required for cell cycle and cell division in spikelet hull primordium.

Prospects

Grain size is an important factor of the yield in rice. There are more than 50 QTLs and genes related to grain size were identified and involved in several regulatory pathways. However, the post-transcriptional regulation of grain size is rarely discussed. This study reveals that the RBP modulates the grain size in rice. From the results of the microarray and the RNA-Seq analysis, this regulatory mechanism was presumed to be different from the phytohormone regulatory network.

The function of LGG possibly affects the expression of the cell cycle-related genes responsible for the cell number in the spikelet hull, resulting in regulation of the grain size. Based on this notion, combining or pyramiding the genes of the phytohormone-related and cell size regulator could improve rice yield.

References

Abe Y, Mieda K, Ando T, Kono I, Yano M, Kitano H, Iwasaki Y (2010) The *SMALL AND ROUND SEED1* (*SRS1*/*DEP2*) gene is involved in the regulation of seed size in rice. Genes Genet Syst 85: 327–339

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

Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. Proc Natl Acad Sci USA 96: 10284–10289

Aya K, Hobo T, Sato-Izawa K, Ueguchi-Tanaka M, Kitano H, 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

Bai X, Huang Y, Hu Y, Liu H, Zhang B, Smaczniak C, Hu G, Han Z, Xing Y (2017) Duplication of an upstream silencer of *FZP* increases grain yield in rice. Nat Plants 3: 885–893

Bai X, Luo L, Yan W, Kovi MR, Zhan W, Xing Y (2010) Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus *qGL7*. BMC Genet 11: 16

Barabaschi D, Tondelli A, Desiderio F, Volante A, Vaccino P, Valè G, Cattivelli L (2016) Next generation breeding. Plant Sci 242: 3–13

Burd CG, Dreyfuss G (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science 265: 615–621

Chakravorty D, Trusov Y, Zhang W, Acharya BR, Sheahan MB, McCurdy DW, Assmann SM, Botella JR (2011) An atypical heterotrimeric G-protein γ-subunit is involved in guard cell K^+-channel regulation and morphological development in *Arabidopsis thaliana*. Plant J 67: 840–851

Chandraratna MF (1964) Genetics and Breeding of Rice. Longmans, London

Che R, Tong H, Shi B, Liu Y, Fang S, Liu D, Xiao Y, Hu B, Liu L, Wang H, et al (2015) Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses. Nat Plants 2: 15195

Chen C, Begcy K, Liu K, Folsom JJ, Wang Z, Zhang C, Walia H (2016) Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. Plant Physiol 171: 606–622

Chen J, Gao H, Zheng X-M, Jin M, Weng J-F, Ma J, Ren Y, Zhou K, Wang Q, Wang J, et al (2015) An evolutionarily conserved gene, *FUWA*, plays a role in determining panicle architecture, grain shape and grain weight in rice. Plant J 83: 427–438

Chen Y, Lübberstedt T (2010) Molecular basis of trait correlations. Trends Plant Sci 15: 454–461

Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K (2013) The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. Plant Physiol 163: 1673–1685

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

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

Crofts AJ, Crofts N, Whitelegge JP, Okita TW (2010) Isolation and identification of cytoskeleton-associated prolamine mRNA binding proteins from developing rice seeds. Planta 231: 1261–1276

da Costa-Nunes JA, Bhatt AM, O'Shea S, West CE, Bray CM, Grossniklaus U, Dickinson HG (2006) Characterization of the three *Arabidopsis thaliana RAD21* cohesins reveals differential responses to ionizing radiation. J Exp Bot 57: 971–983

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

Deng ZY, Liu LT, Li T, Yan S, Kuang BJ, Huang SJ, Yan CJ, Wang T (2015) OsKinesin-13A Is an Active Microtubule Depolymerase Involved in Glume Length Regulation via Affecting Cell Elongation. Sci Rep 5: 9457

Duan P, Ni S, Wang J, Zhang B, Xu R, Wang Y, Chen H, Zhu X, Li Y (2015) Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. Nat Plants 2: 15203

Duan P, Rao Y, Zeng D, Yang Y, Xu R, Zhang B, Dong G, Qian Q, Li Y (2014) *SMALL GRAIN 1*, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. Plant J 77: 547–557

Duan P, Xu J, Zeng D, Zhang B, Geng M, Zhang G, Huang K, Huang L, Xu R, Ge S, et al (2017) Natural Variation in the Promoter of *GSE5* Contributes to Grain Size Diversity in Rice. Mol Plant 10: 685–694

Eun C-H, Takagi K, Park K-I, Maekawa M, Iida S, Tsugane K (2012) Activation and epigenetic regulation of DNA transposon *nDart1* in rice. Plant Cell Physiol 53: 857–868

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112: 1164–1171

Fedoroff NV (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, Zhang S, Zhang H, Yang C, Hu J, et al (2016) *SLG* controls grain size and leaf angle by modulating brassinosteroid homeostasis in rice. J Exp Bot 67: 4241–4253

Fitzgerald MA, McCouch SR, Hall RD (2009) Not just a grain of rice: the quest for quality. Trends Plant Sci 14: 133–139

Folsom JJ, Begcy K, Hao X, Wang D, Walia H (2014) Rice fertilization-Independent Endosperm1 regulates seed size under heat stress by controlling early endosperm development. Plant Physiol 165: 238–248

Fox RG, Park FD, Koechlein CS, Kritzik M, Reya T (2015) Musashi signaling in stem cells and cancer. Annu Rev Cell Dev Biol 31: 249–267

Fujisawa Y, Kato T, Ohki S, Ishikawa A, Kitano H, Sasaki T, Asahi T, Iwasaki Y (1999) Suppression of the heterotrimeric G protein causes abnormal morphology, including dwarfism, in rice. Proc Natl Acad Sci USA 96: 7575–7580

Gao F, Wang K, Liu Y, Chen Y, Chen P, Shi Z, Luo J, Jiang D, Fan F, Zhu Y, et al (2015) Blocking *miR396* increases rice yield by shaping inflorescence architecture. Nat Plants 2: 15196

Gong C, Li T, Li Q, Yan L, 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

Gong P, He C (2014) Uncovering Divergence of Rice Exon Junction Complex Core Heterodimer Gene Duplication Reveals Their Essential Role in Growth, Development, and Reproduction. Plant Physiol 165: 1047–1061

Guo L, Ma L, Jiang H, Zeng D, Hu J, Wu L, Gao Z, Zhang G, Qian Q (2009) Genetic analysis and fine mapping of two genes for grain shape and weight in rice. J Integr Plant Biol 51: 45–51

Guo T, Chen K, Dong N-Q, Shi C-L, Ye W-W, Gao J-P, Shan J-X, Lin H-X (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 JC, Kwon A, Lu H (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, Kobayashi H, Iida S, Tsugane K (2011) Examination of transpositional activity of *nDart1* at different stages of rice development. Genes Genet Syst 86: 215–219

Hayashi-Tsugane M, Takahara H, Ahmed N, Himi E, Takagi K, Iida S, Tsugane K, Maekawa M (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

Heang D, Sassa H (2012a) An atypical bHLH protein encoded by *POSITIVE REGULATOR OF GRAIN LENGTH 2* is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein APG. Breed Sci 62: 133–141

Heang D, Sassa H (2012b) Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. PLoS ONE 7: e31325

Heang D, Sassa H (2012c) Overexpression of a basic helix–loop–helix gene Antagonist of PGL1 (APG) decreases grain length of rice. Plant Biotechnol (Tokyo) 29: 65–69

Hirano K, Yoshida H, Aya K, Kawamura M, Hayashi M, Hobo T, Sato-Izawa K, Kitano H, Ueguchi-Tanaka M, Matsuoka M (2017) *SMALL ORGAN SIZE 1* and *SMALL ORGAN SIZE 2*/*DWARF AND LOW-TILLERING* Form a Complex to Integrate Auxin and Brassinosteroid Signaling in Rice. Mol Plant 10: 590–604

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783–7788

Hong Z, Ueguchi-Tanaka M, Fujioka S, Takatsuto S, Yoshida S, Hasegawa Y, Ashikari M, Kitano H, Matsuoka M (2005) The Rice brassinosteroid-deficient *dwarf2* mutant, defective in the rice homolog of Arabidopsis *DIMINUTO*/*DWARF1*, is rescued by the endogenously accumulated alternative bioactive brassinosteroid, dolichosterone. Plant Cell 17: 2243–2254

Hong Z, Ueguchi-Tanaka M, Umemura K, Uozu S, Fujioka S, Takatsuto S, Yoshida S, Ashikari M, Kitano H, Matsuoka M (2003) A rice brassinosteroid-deficient mutant, *ebisu dwarf* (*d2*), is caused by a loss of function of a new member of cytochrome P450. Plant Cell 15: 2900–2910

Hoshikawa K (1989) The growing rice plant: an anatomical monograph. Nobunkyo, Tokyo, Japan

Hsing Y-I, Chern C-G, Fan M-J, Lu P-C, Chen K-T, Lo S-F, Sun P-K, Ho S-L, Lee K-W, Wang Y-C, et al (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol Biol 63: 351–364

Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H, Shi Z, Pan J, Zhang D, Kang S, et al (2015) A Rare Allele of *GS2* Enhances Grain Size and Grain Yield in Rice. Mol Plant 8: 1455–1465

Hu Z, He H, Zhang S, Sun F, Xin X, Wang W, Qian X, Yang J, Luo X (2012) A Kelch motif-containing serine/threonine protein phosphatase determines the large grain QTL trait in rice. J Integr Plant Biol 54: 979–990

Huang C-K, Sie Y-S, Chen Y-F, Huang T-S, Lu C-A (2016) Two highly similar DEAD box proteins, OsRH2 and OsRH34, homologous to eukaryotic initiation factor 4AIII, play roles of the exon junction complex in regulating growth and development in rice. BMC Plant Biol 16: 84

Huang J, Zhang K, Shen Y, Huang Z, Li M, Tang D, Gu M, Cheng Z (2009a) Identification of a high frequency transposon induced by tissue culture, *nDaiZ*, a member of the *hAT* family in rice. Genomics 93: 274–281

Huang K, Wang D, Duan P, Zhang B, Xu R, Li N, Li Y (2017) *WIDE AND THICK GRAIN 1*,

which encodes an otubain-like protease with deubiquitination activity, influences grain size and shape in rice. Plant J 91: 849–860

Huang X, Kurata N, Wei X, Wang Z-X, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, et al (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490: 497–501

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009b) Natural variation at the *DEP1* locus enhances grain yield in rice. Nat Genet 41: 494–497

Ichinose M, Sugita M (2016) RNA Editing and Its Molecular Mechanism in Plant Organelles. Genes (Basel). doi: 10.3390/genes8010005

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

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436: 793–800

Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B-I, Onishi A, 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

Jagodzik P, Tajdel-Zielinska M, Ciesla A, Marczak M, Ludwikow A (2018) Mitogen-Activated Protein Kinase Cascades in Plant Hormone Signaling. Front Plant Sci 9: 1387

Jang S, An G, Li H-Y (2017) Rice Leaf Angle and Grain Size Are Affected by the *OsBUL1* Transcriptional Activator Complex. Plant Physiol 173: 688–702

Jiang Y, Bao L, Jeong S-Y, Kim S-K, Xu C, Li X, Zhang Q (2012) *XIAO* is involved in the control of organ size by contributing to the regulation of signaling and homeostasis of brassinosteroids and cell cycling in rice. Plant J 70: 398–408

Jin J, Hua L, Zhu Z, Tan L, Zhao X, Zhang W, Liu F, Fu Y, Cai H, Sun X, et al (2016) *GAD1* Encodes a Secreted Peptide That Regulates Grain Number, Grain Length, and Awn Development in Rice Domestication. Plant Cell 28: 2453–2463

Kato T, Segami S, Toriyama M, Kono I, Ando T, Yano M, Kitano H, Miura K, Iwasaki Y (2011) Detection of QTLs for grain length from large grain rice (*Oryza sativa* L.). Breed Sci 61: 269–274

Kawamoto T, 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

Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8: 533–543

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125: 1198–1205

Khush GS (1999) Green revolution: preparing for the 21st century. Genome 42: 646–655

Kikuchi K, Terauchi K, Wada M, Hirano H-Y (2003) The plant MITE *mPing* is mobilized in anther culture. Nature 421: 167–170

Kitagawa K, Kurinami S, Oki K, Abe Y, Ando T, Kono I, Yano M, Kitano H, Iwasaki Y (2010) A novel kinesin 13 protein regulating rice seed length. Plant Cell Physiol 51: 1315–1329

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

Komatsu M, Shimamoto K, Kyozuka J (2003) Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. Plant Cell 15: 1934–1944

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685

Leng Y, Yang Y, Ren D, Huang L, Dai L, Wang Y, Chen L, Tu Z, Gao Y, Li X, et al (2017) A Rice *PECTATE LYASE-LIKE* Gene Is Required for Plant Growth and Leaf Senescence. Plant Physiol 174: 1151–1166

Li D, Wang L, Wang M, Xu Y-Y, Luo W, Liu Y-J, Xu Z-H, Li J, Chong K (2009a) Engineering *OsBAK1* gene as a molecular tool to improve rice architecture for high yield. Plant Biotechnol J 7: 791–806

Li J, Chu H, Zhang Y, Mou T, Wu C, Zhang Q, Xu J (2012a) The rice *HGW* gene encodes a ubiquitin-associated (UBA) domain protein that regulates heading date and grain weight. PLoS ONE 7: e34231

Li J, Thomson M, McCouch SR (2004) Fine mapping of a grain-weight quantitative trait locus in the pericentromeric region of rice chromosome 3. Genetics 168: 2187–2195

Li N, Li Y (2015) Maternal control of seed size in plants. J Exp Bot 66: 1087–1097

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

Li S, Liu W, Zhang X, Liu Y, Li N, Li Y (2012b) Roles of the Arabidopsis G protein γ subunit *AGG3* and its rice homologs *GS3* and *DEP1* in seed and organ size control. Plant Signal Behav 7: 1357–1359

Li S, Qian Q, Fu Z, Zeng D, Meng X, Kyozuka J, Maekawa M, Zhu X, Zhang J, Li J, et al (2009b) *Short panicle1* encodes a putative PTR family transporter and determines rice panicle size. Plant J 58: 592–605

Li T, Jiang J, Zhang S, Shu H, Wang Y, Lai J, Du J, Yang C (2015) *OsAGSW1*, an ABC1-like kinase gene, is involved in the regulation of grain size and weight in rice. J Exp Bot 66: 5691–5701

Li X, Okita TW (1993) Accumulation of Prolamines and Glutelins during Rice Seed Development: a Quantitative Evaluation. Plant Cell Physiol 34: 385–390

Li X, Sun L, Tan L, Liu F, Zhu Z, Fu Y, Sun X, Sun X, Xie D, Sun C (2012c) *TH1*, a DUF640 domain-like gene controls lemma and palea development in rice. Plant Mol Biol 78:

351–359
Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, et al (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. Nat Genet 43: 1266–1269
Lim M-H, Kim J, Kim Y-S, Chung K-S, Seo Y-H, Lee I, Kim J, Hong CB, Kim H-J, Park C-M (2004) A new Arabidopsis gene, *FLK*, encodes an RNA binding protein with K homology motifs and regulates flowering time via *FLOWERING LOCUS C*. Plant Cell 16: 731–740
Lin Z, Zhang X, Yang X, Li G, Tang S, Wang S, Ding Y, Liu Z (2014) Proteomic analysis of proteins related to rice grain chalkiness using iTRAQ and a novel comparison system based on a notched-belly mutant with white-belly. BMC Plant Biol 14: 163
Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, Tian P, Cheng Z, Yu X, Zhou K, et al (2017) *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat Plants 3: 17043
Liu L, Tong H, Xiao Y, Che R, Xu F, Hu B, Liang C, Chu J, Li J, Chu C (2015a) Activation of *Big Grain1* significantly improves grain size by regulating auxin transport in rice. Proc Natl Acad Sci USA 112: 11102–11107
Liu S, Hua L, Dong S, Chen H, Zhu X, Jiang J, Zhang F, Li Y, Fang X, Chen F (2015b) *OsMAPK6*, a mitogen-activated protein kinase, influences rice grain size and biomass production. Plant J 84: 672–681
Lombardo F, Yoshida H (2015) Interpreting lemma and palea homologies: a point of view from rice floral mutants. Front Plant Sci 6: 61
Lorković ZJ (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. Trends Plant Sci 14: 229–236
Luo J, Liu H, Zhou T, Gu B, Huang X, Shangguan Y, Zhu J, Li Y, Zhao Y, Wang Y, et al (2013) *An-1* encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. Plant Cell 25: 3360–3376
Macknight R, Bancroft I, Page T, Lister C, Schmidt R, Love K, Westphal L, Murphy G, Sherson S, Cobbett C, et al (1997) *FCA*, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell 89: 737–745
Maekawa M, Tsugane K, Iida S (2011) Effective Contribution of the *nDart1* Transposon-Tagging System to Rice Functional Genomics. Advances in Genetics Research. Nova Science, New York, pp 259–272
Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc Natl Acad Sci USA 107: 19579–19584
Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272: 2118–

2131
McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, et al (2016) Open access resources for genome-wide association mapping in rice. Nat Commun 7: 10532
Mihara M, Itoh T, Izawa T (2010) SALAD database: a motif-based database of protein annotations for plant comparative genomics. Nucleic Acids Res 38: D835-842
Mikami M, Toki S, Endo M (2015) Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. Plant Mol Biol 88: 561–572
Moon S, Jung K-H, Lee D-E, Jiang W-Z, Koh HJ, Heu M-H, Lee DS, Suh HS, An G (2006) Identification of active transposon *dTok*, a member of the *hAT* family, in rice. Plant Cell Physiol 47: 1473–1483
Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M (2006) Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice. Plant Physiol 141: 924–931
Nagasawa N, Hibara K, Heppard EP, Vander Velden KA, Luck S, Beatty M, Nagato Y, Sakai H (2013) *GIANT EMBRYO* encodes CYP78A13, required for proper size balance between embryo and endosperm in rice. Plant J 75: 592–605
Nakagawa H, Tanaka A, Tanabata T, Ohtake M, Fujioka S, Nakamura H, Ichikawa H, Mori M (2012) *Short grain1* decreases organ elongation and brassinosteroid response in rice. Plant Physiol 158: 1208–1219
Nakamura M, Okano H, Blendy JA, Montell C (1994) *Musashi*, a neural RNA-binding protein required for Drosophila adult external sensory organ development. Neuron 13: 67–81
Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421: 170–172
Nawaz G, Kang H (2017) Chloroplast- or Mitochondria-Targeted DEAD-Box RNA Helicases Play Essential Roles in Organellar RNA Metabolism and Abiotic Stress Responses. Front Plant Sci 8: 871
Nishimura H, Ahmed N, Tsugane K, Iida S, Maekawa M (2008) Distribution and mapping of an active autonomous *aDart* element responsible for mobilizing nonautonomous *nDart1* transposons in cultivated rice varieties. Theor Appl Genet 116: 395–405
Nonomura K-I, Eiguchi M, Nakano M, Takashima K, Komeda N, Fukuchi S, Miyazaki S, Miyao A, Hirochika H, Kurata N (2011) A novel RNA-recognition-motif protein is required for premeiotic G1/S-phase transition in rice (*Oryza sativa* L.). PLoS Genet 7: e1001265
Nonomura K-I, Morohoshi A, Nakano M, Eiguchi M, Miyao A, Hirochika H, Kurata N (2007) A germ cell specific gene of the *ARGONAUTE* family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. Plant Cell 19:

2583–2594

Oh J-M, Balkunde S, Yang P, Yoon D-B, Ahn S-N (2011) Fine mapping of grain weight QTL, *tgw11* using near isogenic lines from a cross between *Oryza sativa* and *O. grandiglumis*. Genes Genom 33: 259–265

Omidbakhshfard MA, Proost S, Fujikura U, Mueller-Roeber B (2015) Growth-Regulating Factors (GRFs): A Small Transcription Factor Family with Important Functions in Plant Biology. Mol Plant 8: 998–1010

Peng P, Liu L, Fang J, Zhao J, Yuan S, Li X (2017) The rice *TRIANGULAR HULL1* protein acts as a transcriptional repressor in regulating lateral development of spikelet. Sci Rep 7: 13712

Peres A, Churchman ML, Hariharan S, Himanen K, Verkest A, Vandepoele K, Magyar Z, Hatzfeld Y, Van Der Schueren E, Beemster GTS, et al. (2007) Novel plant-specific cyclin-dependent kinase inhibitors induced by biotic and abiotic stresses. J Biol Chem 282: 25588–25596

Qi P, Lin Y-S, Song X-J, Shen J-B, Huang W, Shan J-X, Zhu M-Z, Jiang L, Gao J-P, Lin H-X (2012) The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. Cell Res 22: 1666–1680

Qiu X, Gong R, Tan Y, Yu S (2012) Mapping and characterization of the major quantitative trait locus *qSS7* associated with increased length and decreased width of rice seeds. Theor Appl Genet 125: 1717–1726

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE 8: e66428

Rice Annotation Project, Itoh T, Tanaka T, Barrero RA, Yamasaki C, Fujii Y, Hilton PB, Antonio BA, Aono H, Apweiler R, et al. (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. Genome Res 17: 175–183

Rogers K, Chen X (2013) Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell 25: 2383–2399

Schomburg FM, Patton DA, Meinke DW, Amasino RM (2001) *FPA*, a gene involved in floral induction in Arabidopsis, encodes a protein containing RNA-recognition motifs. Plant Cell 13: 1427–1436

Segami S, Kono I, Ando T, Yano M, Kitano H, Miura K, Iwasaki Y (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, Miura K (2017) Overexpression of *SRS5* improves grain size of brassinosteroid-related dwarf mutants in rice (*Oryza sativa* L.). Breed Sci 67: 393–397

Shao G, Wei X, Chen M, Tang S, Luo J, Jiao G, Xie L, Hu P (2012) Allelic variation for a candidate gene for *GS7*, responsible for grain shape in rice. Theor Appl Genet 125:

1303–1312

Shida T, Fukuda A, Saito T, Ito H, 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, Qian Q, Terada R, Johzuka-Hisatomi Y, Iida S, 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, Yano M (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, Zhou T, Lu T, Zhu J, Shangguan Y, et al (2016) *OsSPL13* controls grain size in cultivated rice. Nat Genet 48: 447–456

Singh R, Singh AK, Sharma TR, Singh A, Singh NK (2012) Fine mapping of grain length QTLs on chromosomes 1 and 7 in Basmati rice (*Oryza sativa* L.). J Plant Biochem Biotechnol 21: 157–166

Song X-J, Huang W, Shi M, Zhu M-Z, 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

Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, Segami S, Miura K, Ogawa D, Kamura T, et al (2015) Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. Proc Natl Acad Sci USA 112: 76–81

Sosso D, Luo D, Li Q-B, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, et al (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat Genet 47: 1489–1493

Streitner C, Danisman S, Wehrle F, Schöning JC, Alfano JR, Staiger D (2008) The small glycine-rich RNA binding protein *AtGRP7* promotes floral transition in *Arabidopsis thaliana*. Plant J 56: 239–250

Sun T, Li S, Ren H (2017) *OsFH15*, a class I formin, interacts with microfilaments and microtubules to regulate grain size via affecting cell expansion in rice. Sci Rep 7: 6538

Suzuki M, Kato A, 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, Iida S (2007) Transposon display for active DNA transposons in rice. Genes Genet Syst 82: 109–122

Takagi K, Maekawa M, Tsugane K, Iida S (2010) Transposition and target preferences of an active nonautonomous DNA transposon *nDart1* and its relatives belonging to the *hAT* superfamily in rice. Mol Genet Genomics 284: 343–355

Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram

B, Bustamante C, Yoshimura A, Doi K, et al (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics 182: 1323–1334

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30: 2725–2729

Tanabe S, Ashikari M, Fujioka S, Takatsuto S, Yoshida S, Yano M, Yoshimura A, Kitano H, Matsuoka M, Fujisawa Y, et al (2005) A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, *dwarf11*, with reduced seed length. Plant Cell 17: 776–790

Tanaka A, Nakagawa H, Tomita C, Shimatani Z, Ohtake M, Nomura T, Jiang C-J, Dubouzet JG, Kikuchi S, Sekimoto H, et al (2009) *BRASSINOSTEROID UPREGULATED1*, encoding a helix-loop-helix protein, is a novel gene involved in brassinosteroid signaling and controls bending of the lamina joint in rice. Plant Physiol 151: 669–680

Tanaka W, Toriba T, Ohmori Y, Yoshida A, Kawai A, Mayama-Tsuchida T, Ichikawa H, Mitsuda N, Ohme-Takagi M, Hirano H-Y (2012) The *YABBY* gene *TONGARI-BOUSHI1* is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. Plant Cell 24: 80–95

Tang W, Wu T, Ye J, Sun J, Jiang Y, Yu J, Tang J, Chen G, Wang C, Wan J (2016) SNP-based analysis of genetic diversity reveals important alleles associated with seed size in rice. BMC Plant Biol 16: 93

Tanksley SD (1993) Mapping polygenes. Annu Rev Genet 27: 205–233

Thangasamy S, Chen P-W, Lai M-H, Chen J, Jauh G-Y (2012) Rice *LGD1* containing RNA binding activity affects growth and development through alternative promoters. Plant J 71: 288–302

Tian L, Chou H-L, Zhang L, Hwang S-K, Starkenburg SR, Doroshenk KA, Kumamaru T, Okita TW (2018) RNA-Binding Protein RBP-P Is Required for Glutelin and Prolamine mRNA Localization in Rice Endosperm Cells. Plant Cell 30: 2529–2552

Tian L, Okita TW (2014) mRNA-based protein targeting to the endoplasmic reticulum and chloroplasts in plant cells. Curr Opin Plant Biol 22: 77–85

Tong H, Jin Y, Liu W, Li F, Fang J, Yin Y, Qian Q, Zhu L, Chu C (2009) *DWARF AND LOW-TILLERING*, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice. Plant J 58: 803–816

Tong H, Liu L, Jin Y, Du L, Yin Y, Qian Q, Zhu L, Chu C (2012) *DWARF AND LOW-TILLERING* acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate brassinosteroid responses in rice. Plant Cell 24: 2562–2577

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562–578

Trusov Y, Chakravorty D, Botella JR (2012) Diversity of heterotrimeric G-protein γ subunits

in plants. BMC Res Notes 5: 608

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

Utsunomiya Y, Samejima C, Takayanagi Y, Izawa Y, Yoshida T, Sawada Y, Fujisawa Y, Kato H, Iwasaki Y (2011) Suppression of the rice heterotrimeric G protein β-subunit gene, *RGB1*, causes dwarfism and browning of internodes and lamina joint regions. Plant J 67: 907–916

Van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D'hauw M, Van Montagu M, Gerats T (1998) Transposon Display identifies individual transposable elements in high copy number lines. Plant J 13: 121–129

Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, et al (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat Genet 40: 1370–1374

Wang M, Lu X, Xu G, Yin X, Cui Y, Huang L, Rocha PSCF, Xia X (2016) *OsSGL*, a novel pleiotropic stress-related gene enhances grain length and yield in rice. Sci Rep 6: 38157

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

Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C, et al (2015a) The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat Genet 47: 949–954

Wang S, Wu K, Qian Q, Liu Q, Li Q, Pan Y, Ye Y, Liu X, Wang J, Zhang J, et al (2017) Non-canonical regulation of SPL transcription factors by a human OTUB1-like deubiquitinase defines a new plant type rice associated with higher grain yield. Cell Res 27: 1142–1156

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

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

Wei F-J, Droc G, Guiderdoni E, Hsing Y-IC (2013) International Consortium of Rice Mutagenesis: resources and beyond. Rice (N Y) 6: 39

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

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8: 973–982

Wu T, Shen Y, Zheng M, Yang C, Chen Y, Feng Z, Liu X, Liu S, Chen Z, Lei C, et al (2014)

Gene *SGL*, encoding a kinesin-like protein with transactivation activity, is involved in grain length and plant height in rice. Plant Cell Rep 33: 235–244

Xiao Y, Liu D, Zhang G, Gao S, Liu L, Xu F, Che R, Wang Y, Tong H, Chu C (2018) *Big Grain3*, encoding a purine permease, regulates grain size via modulating cytokinin transport in rice. J Integr Plant Biol. doi: 10.1111/jipb.12727

Xie K, Wu C, Xiong L (2006) Genomic Organization, Differential Expression, and Interaction of SQUAMOSA Promoter-Binding-Like Transcription Factors and microRNA156 in Rice. Plant Physiol 142: 280–293

Xie X, Jin F, Song M-H, Suh J-P, Hwang H-G, Kim Y-G, McCouch SR, Ahn S-N (2008) Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* x *O. rufipogon* cross. Theor Appl Genet 116: 613–622

Xu C, Liu Y, Li Y, Xu X, Xu C, Li X, Xiao J, Zhang Q (2015) Differential expression of *GS5* regulates grain size in rice. J Exp Bot 66: 2611–2623

Xu R, Yu H, Wang J, Duan P, Zhang B, Li J, Li Y, Xu J, Lyu J, Li N, et al (2018) A mitogen-activated protein kinase phosphatase influences grain size and weight in rice. Plant J 95: 937–946

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

Yamamuro C, Ihara Y, Wu X, Noguchi T, Fujioka S, Takatsuto S, Ashikari M, Kitano H, Matsuoka M (2000) Loss of function of a rice *brassinosteroid insensitive1* homolog prevents internode elongation and bending of the lamina joint. Plant Cell 12: 1591–1606

Yang W, Gao M, Yin X, Liu J, Xu Y, Zeng L, Li Q, Zhang S, Wang J, Zhang X, et al (2013) Control of rice embryo development, shoot apical meristem maintenance, and grain yield by a novel cytochrome p450. Mol Plant 6: 1945–1960

Yang Y, Crofts AJ, Crofts N, Okita TW (2014) Multiple RNA binding protein complexes interact with the rice prolamine RNA cis-localization zipcode sequences. Plant Physiol 164: 1271–1282

Yoshida A, Sasao M, Yasuno N, Takagi K, Daimon Y, Chen R, Yamazaki R, Tokunaga H, Kitaguchi Y, Sato Y, 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

Yoshida S (1981) Fundamentals of Rice Crop Science. International Rice Research Institute, Los Baños, Laguna, Philippines

Yu H, Ruan B, Wang Z, Ren D, Zhang Y, Leng Y, Zeng D, Hu J, Zhang G, Zhu L, et al (2017a) Fine Mapping of a Novel *defective glume 1* (*dg1*) Mutant, Which Affects Vegetative and Spikelet Development in Rice. Front Plant Sci 8: 486

Yu J, Xiong H, Zhu X, Zhang H, Li H, Miao J, Wang W, Tang Z, Zhang Z, Yao G, et al (2017b) *OsLG3* contributing to rice grain length and yield was mined by Ho-LAMap.

BMC Biol 15: 28

Yuan H, Fan S, Huang J, Zhan S, Wang S, Gao P, Chen W, Tu B, Ma B, Wang Y, et al (2017) *08SG2*/*OsBAK1* regulates grain size and number, and functions differently in Indica and Japonica backgrounds in rice. Rice (N Y) 10: 25

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

Zhang W, Sun P, He Q, Shu F, Wang J, Deng H (2013a) Fine mapping of *GS2*, a dominant gene for big grain rice. Crop J 1: 160–165

Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, et al (2012) Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. Proc Natl Acad Sci USA 109: 21534–21539

Zhang Y-C, Yu Y, Wang C-Y, Li Z-Y, Liu Q, Xu J, Liao J-Y, Wang X-J, Qu L-H, Chen F, et al (2013b) Overexpression of microRNA *OsmiR397* improves rice yield by increasing grain size and promoting panicle branching. Nat Biotechnol 31: 848–852

Zhao J, Wu C, Yuan S, Yin L, Sun W, Zhao Q, Zhao B, Li X (2013) Kinase activity of *OsBRI1* is essential for brassinosteroids to regulate rice growth and development. Plant Sci 199–200: 113–120

Zhao Y, Wen H, Teotia S, Du Y, Zhang J, Li J, Sun H, Tang G, Peng T, Zhao Q (2017) Suppression of *microRNA159* impacts multiple agronomic traits in rice (Oryza sativa L.). BMC Plant Biol 17: 215

Zheng J, Zhang Y, Wang C (2015) Molecular functions of genes related to grain shape in rice. Breed Sci 65: 120–126

Zhou Y, Miao J, Gu H, Peng X, Leburu M, Yuan F, Gu H, Gao Y, Tao Y, Zhu J, et al (2015) Natural Variations in *SLG7* Regulate Grain Shape in Rice. Genetics 201: 1591–1599

Zhou Y, Zhu J, Li Z, Yi C, Liu J, Zhang H, Tang S, Gu M, Liang G (2009) Deletion in a quantitative trait gene *qPE9-1* associated with panicle erectness improves plant architecture during rice domestication. Genetics 183: 315–324

Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, Zhu J, Lucas WJ, Wang Z, Zhang D (2015) Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein. Plant J 82: 570–581

Zuo J, Li J (2014) Molecular genetic dissection of quantitative trait loci regulating rice grain size. Annu Rev Genet 48: 99–118

Acknowledgments

I would like to express my sincere gratitude to my advisor, Professor Masahiko Maekawa, for his notable guidance and patience during my Ph. D course and for giving me the opportunity to study in Japan. I am deeply grateful for his generosity and support in academics and beyond.

I would like to extend genuine thanks to Professor Kazuhiro Sato and Professor Wataru Sakamoto for their kind assistance in accomplishing this work and for their valuable advice in my study and research.

My sincere thanks to Dr. Himi, Dr. Rikiishi, Dr. Sugimoto, Dr. Ezaki and Dr. Utsugi for their advice, technical guidance and encouragement during my study. Many thanks to Mr. Hideki Nishimura and Ms. Yoko Kato for their help with my experiments and data collection.

Special thanks to Dr. Kazuo Tsugane and Dr. Mika Hayashi-Tsugane of National Institute for Basic Biology, for their collaboration, worthy advice and support, and Dr. Tadafumi Kawamoto of Tsurumi University for helping the cryomicrotomic sectioning.

I appreciate Dr. Emily Gichuhi, Tanaka and Tun, Institute of Plant Science and Resources staff and students for their precious friendship and selfless help.

I would like to acknowledge the Ohara Agriculture Scholarship and Kobayashi International Student Scholarship for the financial support during my study in Japan.

Finally, to my family and friends, without their continuous support and encouragement I would have not been able to come this far.