A DISSERTATION FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

Identification of a Novel Gene for Floral Organ Number and Haplotype Analysis in Myanmar Rice Landraces

February, 2023

Graduate School of Agriculture and Life Sciences
Seoul National University
Crop Science and Biotechnology Major

Phyu Phyu MAUNG

Identification of a Novel Gene for Floral Organ Number and Haplotype Analysis in Myanmar Rice Landraces

Under the direction of Dr. Hee-Jong Koh

Submitting a Ph.D. Dissertation of Agriculture

November 2022

Graduate School of Agriculture and Life Sciences
Seoul National University
Crop Science and Biotechnology Major

Phyu Phyu MAUNG

Confirming the Ph.D. Dissertation written by
Phyu Phyu MAUNG
December 2022

Chair _____
Suk-Ha Lee, Ph.D.

Vice Chair _____
Hee-Jong Koh, Ph.D.

Examiner _____
Nam-Chon Paek, Ph.D.

Examiner _____
Tae-Jin Yang, Ph.D.

Examiner _____
Soon-Wook Kwon, Ph.D.

General abstract

Approximately 50% of all calories consumed by the entire human population are directly derived from rice. Floral organ number is crucial for successful seed setting and mature grain development. In the first chapter of this report, the novel gene responsible for floral organ number was identified. In the floral organ number mutant, 37% of the spikelets showed an increase in the number of floral organs, especially stamens and pistils. We confirmed Os08g0299000 as the novel gene regulating floral organ number changes and was designated as the floral organ number 7 (*fon7*) gene. A single nucleotide polymorphism (G to A) at the intron splicing donor site of the *FON7* causes the skipping of whole exon 6 in the mutant, resulting in the deletion of 144 nt. The function of the *fon7* gene was confirmed by using T-DNA tagging lines. These results provide valuable insight into the mechanism of floral organ differentiation and formation in rice.

In the second chapter, the haplotype variation and new non-functional haplotypes of heading date and grain size-related genes were detected in 46 Myanmar rice landraces. Because heading date and grain size are the important determinants for yield and key targets for genetic improvement. Haplotype-based breeding and haplotype-assisted genomic selection are efficient improvement strategies to deploy superior haplotypes to hasten breeding progress. In this study, the combinations of functional and non-functional haplotypes of heading date genes were investigated. Additionally, we identified the *GS5* gene as a major gene underlying grain length variation in indica sub-species of Myanmar landraces.

Keywords: Rice, floral organ number, heading date, grain size, exon skipping, haplotypes

Student Number: 2019-35644

CONTENTS

CHAPTER II

Haplotype Analysis in Myanmar Rice Landraces

LIST OF TABLES

LIST OF FIGUREURES

LIST OF FIGUREURES

LIST OF ABBREVIATIONS

FON7	Floral organ number 7
dCAPS	Derived cleaved amplified polymorphic sequence
qRT-PCR	Quantitative real- time PCR
SNP	Single nucleotide polymorphism
Memo 1	Mediator of ErbB2-driven cell motility protein
CTAB	Cetyl-trimethyl ammonium bromide
GS5	Grain size5
NF	Non-functional haplotype
F	Functional haplotype
PCA	Principal component analysis
IND	Indica
JAP	Japonica
ARO	Aroma
AUS	Aus

CHAPTER I

Identification of a Novel Gene for Floral Organ Number in Rice

Abstract

Floral organ number is crucial for successful seed setting and mature grain development. Although some genes and signaling pathways controlling floral organ number have been studied, the underlying mechanism is complicated and requires further investigation. In this study, a floral organ number mutant was generated by the ethyl methanesulfonate treatment of the Korean *japonica* rice cultivar Ilpum. In the floral organ number mutant, 37% of the spikelets showed an increase in the number of floral organs, especially stamens and pistils. Histological analysis revealed that the number of ovaries was determined by the number of stigmas; spikelets with two or three stigmas contained only one ovary, whereas spikelets with four stigmas possessed two ovaries. The floral organ number mutant showed pleiotropic phenotypes including multiple grains, early flowering, short plant height, and reduced tiller number compared with the wild-type. Genetic and MutMap analyses revealed that floral organ number is controlled by a single recessive gene located between the 8.0 and 20.0 Mb region on chromosome 8. Calculation of SNP-index confirmed *Os08g0299000* as the candidate gene regulating floral organ number, which was designated as *FLORAL ORGAN NUMBER7* (*FON7*). A single nucleotide polymorphism (G to A) was discovered at the intron splicing donor site of *FON7*, which caused the skipping of the entire sixth exon in the mutant, resulting in the deletion of 144 bp. Furthermore, the T-DNA-tagged line displayed the same floral organ number phenotype as the *fon7* mutant. These results provide valuable insight into the mechanism of floral organ differentiation and formation in rice.

Keywords: Floral organ number, multiple pistils, multiple grains, exon skipping, rice

Introduction

The floral organs of spikelets, especially stamens and pistils, are fundamentally important for the fertilization process, and control normal seed setting and mature grain development [1]. Homeotic transformation of floral organs alters the floral organ number or leads to the complete loss of floral organs, causing abnormal seed formation or no seed setting in rice [2]. Therefore, understanding the factors affecting floral organ number at the molecular level is crucial for normal seed development and for gaining insight into yield improvement.

A milestone understanding of the detailed molecular mechanism and the ABC model of floral organ patterning was first developed in eudicots, namely, *Arabidopsis thaliana* and *Anthirrhinum majus* [3]. Subsequently, the regulation of D- and E-class genes in floral organ development was identified, and the ABC model was updated to the ABCDE model [4]. In rice, a modified version of the ABCDE model was established based on the identification of numerous genes regulating floral and spikelet development [4-9]. In this model, the A-class genes (*OsMADS14*, *OsMADS15*, *OsMADS18*, and *OsMADS20*) regulate the formation of lemma and palea [7]; B-class genes (*OsMADS2*, *OsMADS4*, and *OsMADS16*) and C-class genes (*OsMADS3*) regulate stamen identity [4, 10]; C-class genes (*OsMADS58*) determine pistil identity [11]; D-class genes (*OsMADS13*, and *OsMADS21*) regulate the specification of the ovule development [12, 13]; E-class genes (*OsMADS1*-

LHS1, *OsMADS5*, *OsMADS6*, *OsMADS7*, *OsMADS8*, and *OsMADS34*) specify the identities of stamens, pistils, and ovary [10, 13, 14].

Floral organ specification is determined not only by the ABCDE model, but also by the CLAVATA (CLV)–WUSHEL (WUS) signal transduction pathway. Mutations in *CLV* genes in *Arabidopsis* cause the progressive enlargement of floral meristem, resulting in flowers with extra sepals, petals, stamens, and ovaries [15]. The CLV–WUS module regulates floral organ number through the negative regulation of stem cell accumulation and positive regulation of floral meristem [16]. Rice *FLORAL ORGAN NUMBER1* (*FON1*) and *FON2/FON4* genes encode *Arabidopsis* CLV1 and CLV3 homologs, respectively [17-19]. Thus, the FON1–FON2 signaling pathway in rice corresponds to the CLV1–CLV3 signaling system in *Arabidopsis*. Additionally, the rice CLV–WUS pathway has been proven to regulate floral meristem and floral organ number in other crop plants, such as Maize, Brassica, Tomato [16]. The *FON2/FON4* genes act in parallel with other floral homeotic regulators such as, *OsMADS16*, *OsMADS58*, *OsMADS13*, and *OsMADS1* [20]. FON1 is required for the activation of FON2/FON4 protein function, whereas the other CLV3/EMBRYO SURROUNDING REGION (CLE) homolog, FON2 SPARE1 (FOS1), can substitute for FON2 activity without requiring FON1 [21]. Two other CLE homologs, FON2-LIKE CLE PROTEIN1 (*FCP1*) and *FCP2*, negatively

regulate vegetative stem cell activity and promote leaf initiation by repressing the expression of *WUSCHEL-RELATED HOMEOBOX4* (*WOX4*) [16, 22].

In addition, *SUPERWOMAN1 (SPW1*), *DROOPING LEAF* (*DL*), *abnormal floral organ (afo), TONGARI BOUSHI1 (TOB1)/YABBY5*, retrotransposon *Tos17, osmads1-z, ABERRANT PANICLE ORGANIZATION 1* (*APO1*, ortholog of *UFO*), and *ABERRANT PANICLE ORGANIZATION 2* (*APO2*, ortholog of *LFY*) also control floral organ identification in rice [14, 23-27]. Recently, studies on the interaction among *FON4*, *APO1*, and C- and D-class genes suggested a regulatory module that fine-tunes floret patterning and floral organ determinacy in rice [10]. Thus, accumulating evidence indicates that floral organ development is a multi-step process and involves numerous genes in a spatiotemporally regulated manner [4]. Despite these findings, our understanding of the floral organ regulatory pathway remains limited. Identification of additional floral organ genes is required to attain a clear understanding of the molecular mechanism of floral organ and spikelet development.

In this study, we characterized a chemically mutagenized *japonica* rice mutant exhibiting increased floral organ (stamen and pistil) numbers. The gene underlining floral organ number (*FON7*) was identified by MutMap analysis, and its function was confirmed using a T-DNA-tagged line.

Materials and Methods

Plant materials

The mutant was identified from the ethyl methanesulfonate (EMS)-induced mutational library of a Korean japonica rice cultivar, Ilpum. F_1 and F_2 populations derived from crosses between mutant and wild type were used for the genetic analysis and mapping of *fon7*. All plants were grown under normal conditions in the experimental paddy field of the Seoul National University, Suwon, Korea.

Agronomic characterization of the mutant

Agronomic traits such as plant height, tiller number, panicle length, and internode length were measured from five biological replicates of wild-type and mutant plants, and statistically analyzed using SPSS version 25. During flowering, five panicles of wild-type and mutant plants were selected randomly, and the components of each floret were investigated. Photographs of spikelets were also captured under a microscope (Olympus OX 31).

Histological analysis

The freshly collected spikelets were fixed in formalin-acetic acid-alcohol (FAA; 5% formaldehyde, 5% acetic acid, and 45% ethanol), and stored at 4˚C. The fixed spikelets were dehydrated in a graded ethanol series from 65% to 100%. Then, the samples were infiltrated with xylene substitute for 2 h, dipped in paraffin for 3 h, and subsequently embedded in new paraffin.

The paraffin-embedded samples were sectioned into 8-μm thick slices using the MICROM HM 325 Rotary microtome. These slides were deparaffinized in xylene, and stained with 0.05% toluidine blue. Photographs were captured under the Olympus CX 31 light microscope.

Genetic analysis of the *fon7* mutant

A total of 179 F2 plants derived from crosses between mutant and wild-type (Ilpum) plants were subjected to genetic analysis; F2 plants exhibiting increased floral organ numbers, short plant height, early flowering, and multiple pistils were considered as mutants. The segregation ratios of F2 populations were analyzed by the chi-square ($\chi 2$) test in SPSS version 25.

Identification of the *fon7* gene using MutMap

DNA was extracted from young and healthy leaf samples of the wild type and F_2-generation mutants (early flowering and short plant height) using the cetyltrimethylammonium bromide (CTAB) method. The genomic DNA samples of 12 F_2 mutants were combined in equal amounts, and the bulked DNA was sequenced on the Illumina NovaSeq 6000 system. at the National Instrumentation Center for Environmental Management (NICEM) of Seoul National University (NICEM, Seoul, Korea). The resequencing data of Ilpum was used for MutMap analysis (version 2.3.2) [28]. Subsequently, the SNP-index plot was generated using SNPs with a minimum of 0.4 SNP -index within 3-Mb windows.

SNP genotyping using the dCAPS marker

The candidate SNPs with SNP-index values > 0.9 were targeted for designing allele-specific markers. The allele-specific primers were designed using Primer 3 Input version 0.4.0 (https://bioinfo.ut.ee). A total of 24 genotypes (12 wild type and 12 mutant) were selected for SNP genotyping. The dCAPS finder 2.0 (http://helix.wustl.edu/dcaps/ dcaps.html) was used to design the derived cleaved amplified polymorphic sequence (dCAPS) markers for the selected SNPs. The PCR products were digested with appropriate restriction enzymes for 2 h, and analyzed by electrophoresis on 3% agarose gel. The sizes of wild-type and mutant PCR products were visualized under the ImageQuant LAS 4000 mini biomolecular imager.

Identification of exon skipping in *fon7* mutant

To identify nucleotide changes and splicing patterns in the candidate gene, total RNA was isolated from wild-type and mutant leaves using the GeneAll Hybrid-RTM kit (GENEALL Bio, South Korea), and then treated with RNase-free Recombinant DNase I (Takara Bio, Japan) to eliminate genomic DNA contamination. Total RNA was used for cDNA synthesis using an M-MLV reverse transcriptase kit (Promega, Madison, WI, USA), and cDNA was amplified using candidate gene-specific primers. PCR products were purified using the DNA purification kit (Inclone, Korea), and subjected to Sanger sequencing with both forward and reverse primers. Nucleotide changes were detected by aligning the wild-type and mutant

cDNA sequences using the Codon Code aligner software (Codon Code Corporation, USA). Candidate Protein deleterious SNP was predicted by performing Protein Blast candidate gene with the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

RNA extraction and qRT-PCR analysis

Roots, stems, leaves, spikelets, and seeds were sampled from the mutant plants and the T-DNA-tagged line as well as from the corresponding wild types (Ilpum and Dongjin, respectively) in three biological replicates, with each replicate containing three technical repeats. Total RNA was extracted from the harvested plant samples using the Takara kit (TaKaRa Bio, Kusatsu, Japan), and first-strand cDNA synthesis was carried out using oligo (dT) primers and M-MLV reverse transcriptase (Promega). Quantitative real-time PCR (qRT-PCR) was performed using sequence-specific primers and SYBR Premix ExTaq (TaKaRa) on the CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA), according to the manufacturer's instructions. Expression levels of *FON7* were normalized relative to those of Actin, a housekeeping gene, and relative gene expression levels were calculated using the ΔΔCt method.

Validation of the mutation causing altered floral organ number phenotype

PCR-based genotyping of T-DNA insertion mutant plants was conducted using the T-DNA-specific left border primer in combination with gene-specific primers. Then, homozygous T-DNA insertion mutants were selected by loading the amplification products on 1% agarose gel. The phenotypic traits and relative *fon7* expression levels of homozygous T-DNA mutants were analyzed and compared with those of the wild type.

Table 1-1. Details of primers used for candidate gene identification.

Primer name	Application	Forward primer (5'→'3')	Reverse primer (5'→3')	Note
MP 1	mpeh gene sequencing	GCGAACAACACTGGAAAACA	TATTTCTGAGGCCGAGAACG	
MP 3	mpeh gene sequencing	GCTCTTAGCGGGAGTGAGTG	GGAACGCTGCTATTGCGTAT	
MP 4	mpeh gene sequencing	TTCAGGGGTACTTTCTTTTCTGA	CGGCAAGAGAACTCTAAAGATTG	
MP 5	mpeh gene sequencing	CGTCCAACATGCTAGATGTGAT	TGCATATAAGCTCCCCCTGA	
MP 4- F	segregation test	TTCTGCCACTGGGGAACCCG	CGGGTTCCCCAGTGGCAGAA	dCAPS (MspI)
MP 5- F	segregation test	GTCATCATTGGACCAGACAA	TTGTCTGGTCCAATGATGAC	dCAPS
cDNA 1-F	cDNA synthesis	CATCGAAACCGCTCTTCTTC	ATGGATTGCAAACCGATCAT	for whole gene
cDNA 2-F	cDNA synthesis	TTGTTGGTGCCCTTAACTCC	TGCAGGTACTGCTTGAATGC	for only exon 6
RT-PCR-MP1	mRNA expression	GCCACCCCATCAGCGTTTTC	CTCACGCTGCTGTCCCTCAT	
RT-PCR-MP2	mRNA expression	ATGAGGGACAGCAGCGTGAG	GGGTTGTACAAGTGATCCGTGC	

Table 1-2. Details of primers used in this study for T-DNA tagging line.

Primer name	Application	Forward primer (5'→'3')	Reverse primer (5'→3')
1C03.1	T-DNA line genotyping	TGGACAGACCCCTGTCTAGATT	GCAAGCAGATTGTTGAAGCA
1C03N	T-DNA line genotyping	GGGGATTTGCCAGTAGACC	AAAGGCTATGCATTGCTACTGA
1C03.3	T-DNA line genotyping	TCAGACAGAGCCATTTGTGC	GAAGTTGCTCAAAAGGGCTG
1C03.4	T-DNA line genotyping	GAAGTTGCTCAAAAGGGCTG	TCAGACAGAGCCATTTGTGC
3C00	T-DNA line genotyping	GATCGTCGATACCGCATTTT	AGGGGGCAGATTTAGGAAGA
3C00-2	T-DNA line genotyping	AACGACTAGGGATGGCAATG	CGACACACCCGGAAAAATAC
2A	T-DNA line genotyping	GGATTCAGCGCAGGCTGT	CTCCCCACTCCCGAAAAGT
LB- 2707	T-DNA line genotyping	GGTGAATGGCATCGTTTGAA	
LB- 2717	T-DNA line genotyping	ATGGCAGTGAATTAACATAGC	
LB- 2715	T-DNA line genotyping	TTGGGGATCCTCTAGAGTCGAG	
RB	T-DNA line genotyping	AACGCTGATCAATTCCACAG	

Results

Floral organ morphogenesis of the mutant

The typical rice floret is composed of one pistil (that produces one ovary), six stamens, two lodicules, one palea, and one lemma. Compared with the wild type, the mutant spikelets showed no obvious differences in the outer whorl floral organ number (lemma, palea, and lodicules) but exhibited an increased number of stamens and pistils. In the *fon7* mutant, approximately 37% of the florets showed abnormal floral organ number (Table 1-1), including up to nine stamens or up to two pistils and four stigmas (Figure 1j–1o). The correlation between stigma number and ovary number per spikelet was investigated through histological analysis of the transverse section of wild type and mutant spikelets. Interestingly, spikelets with two or three stigmas contained only one ovary, while those with four stigmas showed two ovaries (Figure 2b–d). The two ovaries contained within a spikelet were sometimes attached and sometimes located separately (Figure 2c, 2d, 2g, and 2h). If pollination and fertilization were successful, the two ovaries developed as two seeds enwrapped by one glume (Figure 1c). These two seeds showed normal germination.

Table 1-3. Agronomic traits of the wild type and *fon7* mutant.

Genotype	Traits[a]						
	Normal florets (%)	**Abnormal florets (%)**	**No. of days to heading**	**Plant height (cm)**	**Tiller number**	**Panicle length (cm)**	**No. of spikelets per panicle**
Wild type	100	0	108	98.9 ± 2.6	11.5 ± 1.6	23.5 ± 0.8	123.3 ± 7.8
fon 6	62.98	37.02	95	75.6 ± 3.6**	8.2 ± 1.7**	20.3 ± 1.0**	93.9 ± 8.6**

aData represent mean ± standard deviation (SD; n = 5). Asterisks indicate significant differences (**P < 0.01; unpaired t-test).

Table 1-4. Floral and plant morphology of the wild type and *fon7* mutant.

	No. of spikelets with multiple stamens				**No. of spikelets with multiple stigmas**		
	Six stamens	Seven stamens	Eight stamens	Nine stamens	Two stigmas	Three stigmas	Four stigmas
Wild type	100	0	0	0	100	0	0
Mutant	90.95 ± 18	8.33 ± 1.3	0.57 ± 0.7	0.14 ± 0.4	70.64 ± 18.60	25 ± 3.7	5.36 ± 2.3

Data represent mean ± SD. Five panicles of five plants of each genotype were randomly selected, and the numbers of stamens and stigmas in each spikelet were counted.

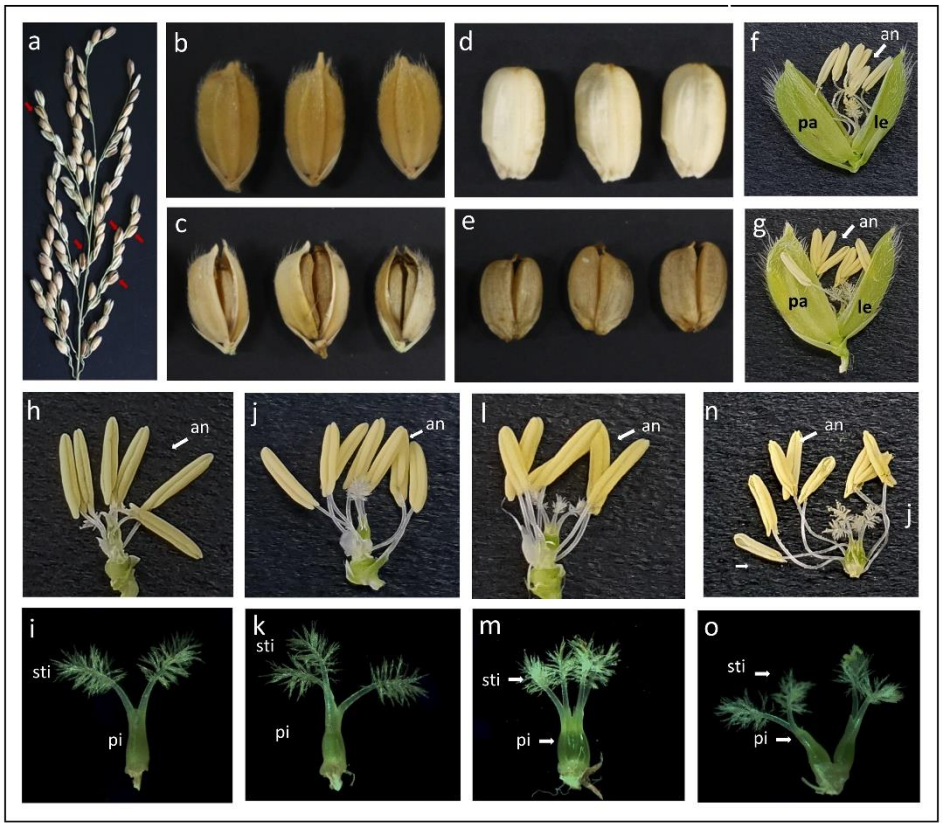


Figure 1-1. Grain and floral organ phenotypes of the wild type and *fon7* mutant. (a) Occurrence of multiple pistil grains (red arrows) in the *fon7* mutant. (b, c) Wild type (b) and fon 6 mutant (c) grains. (d, e) Wild type (d) and *fon7* mutant (e) grains with hull removed. (f, g) Wild type (f) and *fon7* mutant (g) spikelets. (h) Wild type spikelet with lemma and palea removed. Six stamens and two stigmas are visible. (i) Wild type pistil with two stigmas. (j) *fon7* spikelet with lemma and palea removed, showing seven stamens and three stigmas. (k) *fon7* pistil with three stigmas. (l) *fon7* spikelet with lemma and palea removed, showing six stamens and four stigmas. (m) *fon7* pistil containing four stigmas, with two ovaries attached to each other. (n) *fon7* spikelet with lemma and palea removed, showing eight stamens and four stigmas. (o) *fon7* pistil containing four stigmas, with two ovaries separated from each other. All spikelets in five panicles of five randomly selected plants of each genotype were examined, and representative images are shown. an, anther; pi, pistil; sti, stigma.

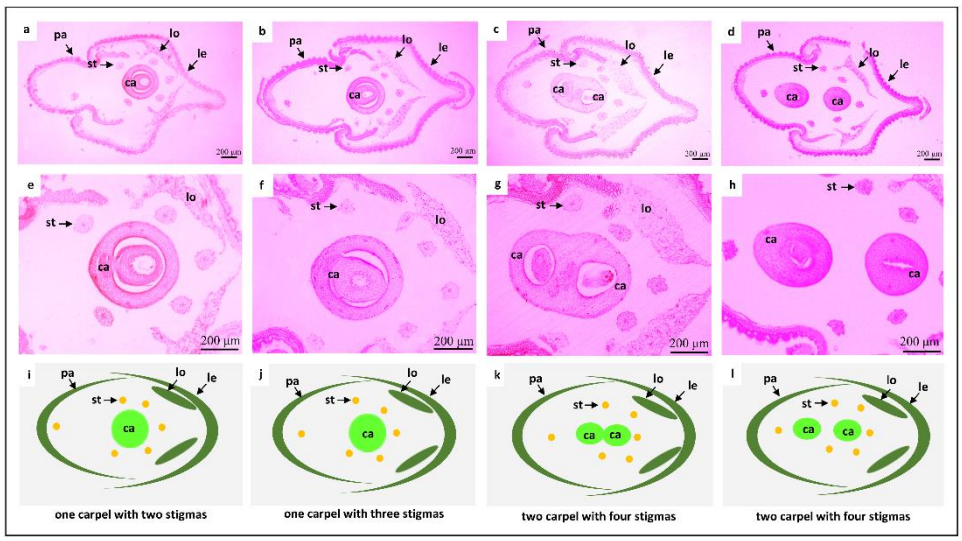


Figure 1-2. Microscopic analyses of paraffin sections of wild-type and *fon7* mutant spikelets located at the basal position of the corresponding spikelets. (a) Transverse section of wild-type spikelet. (b–d) Transverse sections of *fon7* spikelets formed from flowers with three stigmas and one ovary (b), four stigmas and two fused ovaries (c), and four stigmas and two separated ovaries (d). (e–h) Transverse sections of the ovary in upper spikelets. (i–l) Sketches depicting the upper photos of paraffin sections. ca, carpel; le, lemma; lo, lodicule; ov, ovary; pa, palea; st, stamen.

Phenotypic characterization of the *fon7* mutant

Mutants obtained by the ethyl methanesulfonate (EMS) treatment of Ilpum were grown together with the wild type (Ilpum) at the experimental Farm Seoul National University (Suwon, Korea) under normal cultivation conditions. Weak vigor and short height of mutant seedlings were apparent at the transplanting stage. Additionally, the mutant flowered approximately 2 weeks earlier than the wild type (Table 1), and showed significant reduction in tiller number per plant, plant height, and internode length compared with the wild type (Table 1 and Figure1-3). This indicated that loss of *fon7* function reduced plant height in rice by shortening the internode length. Consistent with reduced plant height, the mutant exhibited shorter panicles and fewer spikelets per panicle (Table 1).

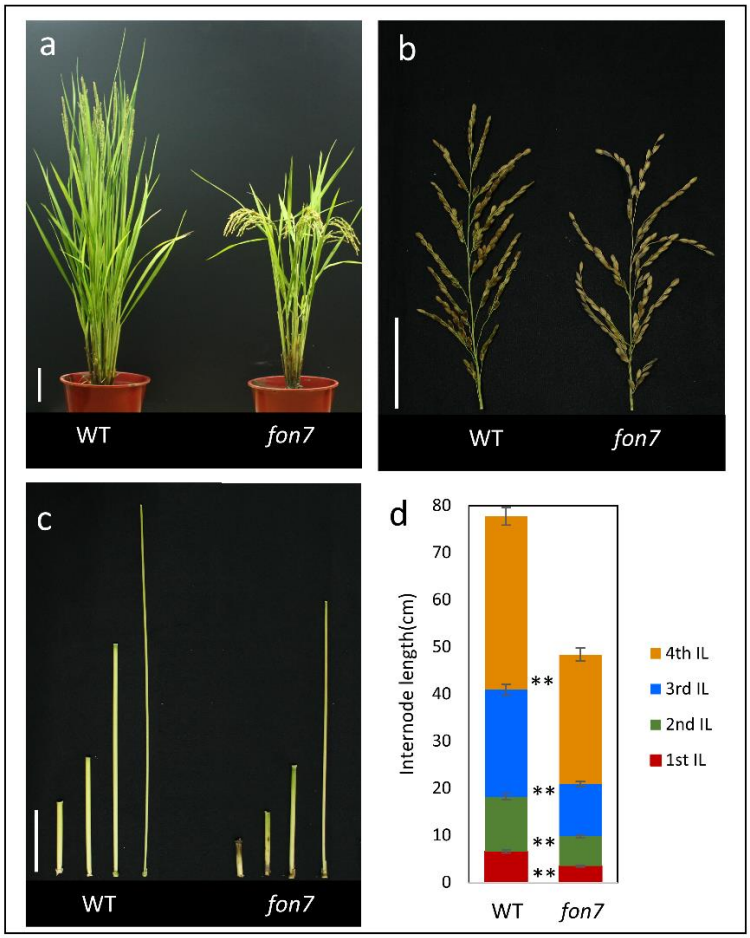


Figure 1-3. Comparison of plant morphology between the wild type and *fon7* mutant. (a) Internode length comparison between the wild type and *fon7* mutant (b) morphology of internode length (c) plant morphology of wild type and *fon7* mutant (d) panicle morphology of wild type and *fon7* mutant.

Genetic analysis of the *fon7* mutant

The F_1 and F_2 progenies derived from crosses between the wild-type (Ilpum) and mutant plants were phenotypically evaluated at the heading stage. All F_1 plants showed wild-type phenotypes, suggesting that the mutant trait was recessive. In the F_2 population, 137 out of a total of 179 plants exhibited wild-type phenotypes; however, phenotypes of the remaining 42 plants were similar to those of their mutant parents, which were characterized by increased stamen and pistil numbers, early flowering, short plant height, and multiple pistil grains. Additionally, the chi-square test revealed the wild type: mutant segregation ratio of 3:1 (χ 2 0.225 < χ2 0.05(1) = 3.841), indicating that floral organ number is controlled by a single recessive gene.

Identification of the gene controlling floral organ number in rice

Causal candidate SNPs responsible for the changes in floral organ number were predicted using MutMap analysis. SNP plots were generated by calculating the SNP-index of each SNP from the Illumina data of bulked F_2 mutant DNA. Among all 12 chromosomes of rice, the average SNP peak index was detected in the 8–20 Mb region of chromosome 8, which was selected as the candidate region (Fig 1-4a). A total of 97 SNPs and indels were detected in the candidate regions of chromosome 8. Among these polymorphisms, three SNPs designated as SNP-1 (*Os08g0223833*; a frameshift variant), SNP-2, (*Os08g0299000*; a splice donor variant), and SNP-3 (*Os08g0408200*; a missense variant) with SNP-index > 0.9 were found

in coding regions or at a splice site. SNP-2 (*Os08g0299000*), a G-to-A polymorphism at the 12,175,170 bp position, was used to develop dCAPS markers (Fig 1-4 b–d). Genotyping the F_2 population using SNP-2 dCAPS marker revealed a complete co-segregation with altered floral organ number phenotype, thus confirming that *Os08g0299000* controls the changes in floral organ number in rice (Fig 1-4 e).

To clarify whether the SNP-2 in *Os08g0299000* is naturally found in the rice germplasm, the nucleotide sequence flanking the 12,175,170 bp position on chromosome 8 was compared among the sequencing data of 4,726 cultivated rice accessions available at the RiceVarMap v2.0 website. No SNP was detected near this region in rice accessions, suggesting that the SNP responsible for the *fon7* mutant phenotype does not represent natural variation in the rice germplasm (data not shown). Therefore, we identified *Os08g0299000* as the candidate gene controlling floral organ number in rice. Because the G-to-A SNP in the intron 6 (splice donor site) of *FON7* was associated with increasing floral organ number (stamens and stigmas) in rice, we determined the full-length *FON7* cDNA sequence in the wild-type and *fon7* mutant to examine potential differences in the deduced amino acid sequence of *FON7* between the two genotypes. The full-length *FON7* cDNA sequence was 1,019 bp in the wild-type, as predicted, but was shorter in length in the *fon7* mutant (Fig 1-5 a, b). Sanger sequencing revealed that *FON7* cDNA carried a 144 bp deletion in the *fon7* mutant compared with the wild-

type. Sequence comparison revealed that full-length *FON7* cDNA contains eight exons in the wild-type but lacks the entire exon 6 in the mutant (Fig 1-5 c). Consequently, the FON7 protein is predicted to contain 298 amino acids in the wild-type but only 250 amino acids in the mutant. Overall, this result suggests that G-to -A mutation at the intron donor site of *FON7* fails to splice and leads to exon skipping in the *fon7* mutant.

Table 1-5. Summary of candidate SNPs (SNP-index = ~1) on chromosome 8.

Name	Position (bp)	SNP-index	Annotation	RAP locus ID	Description
SNP-1	7510150	0.938	Frameshift variant	Os08g0223833	Similar to H0702G05.10 protein
SNP-2	12175170	1	splice_donor_variant	Os08g0299000	Mediator of ErbB2-driven cell motility (Memo), related domain containing protein
SNP-3	19526883	0.9	missense_ variant	Os08g0408200	WD40 subfamily protein, Salt stress

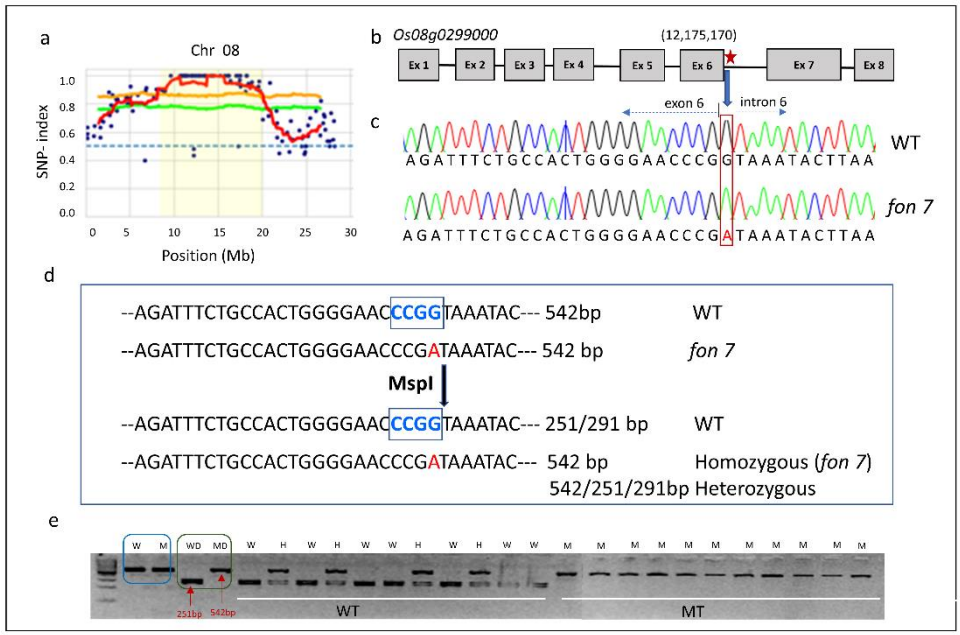


Figure 1-4. SNP-index graph and development of dCAPS molecular marker. (a) SNP-index graph generated by the MutMap analysis of the *fon7* gene. Green and orange lines indicate 95% and 99% confidence intervals, respectively. Red regression lines were obtained by averaging SNP indices from a moving window of five consecutive SNPs and shifting the window one SNP at a time. Blue dots represent SNP-index values at the SNP position. Pale-yellow shaded area on chromosome 8 indicates the region corresponding to the candidate gene controlling floral organ number in rice. (b) Gene structure of *Os08g0299000*. Gray boxes represent exons; black lines represent introns; red star indicates the SNP position. (c) Sequence comparison between the wild type and *fon7* mutant. Red color indicates the SNP (G ->A). (d) Development of the dCAPS marker for *fon7*. The sequence within the gray box indicates the mismatched base, which was used for developing the dCAPS marker. (e) Co-segregation of the SNP with the floral organ number phenotype of F2 plants derived from the *fon7* mutant wild type (Ilpum) cross. PCR products were digested with MspI, and then separated on agarose gel. WT, wild type; *fon7*, mutant; WD, wild type DNA fragment digested with MspI; MD, mutant DNA fragment digested with MspI.

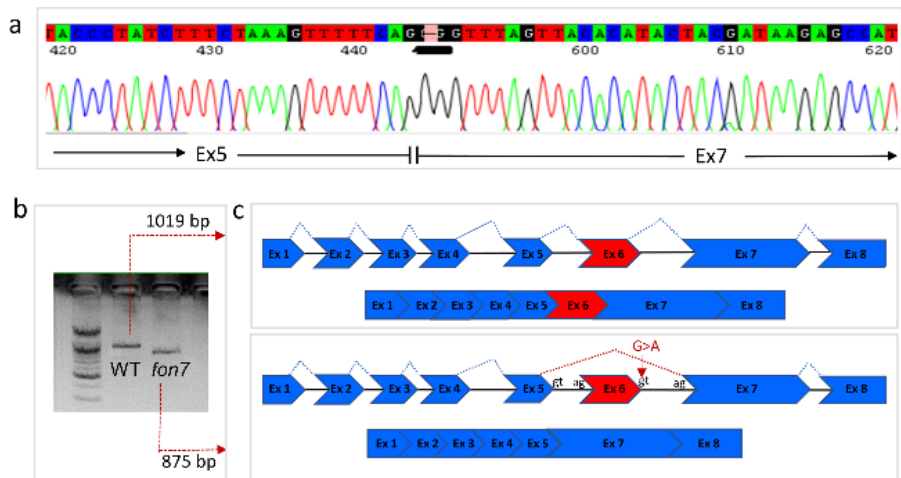


Figure 1-5. Exon skipping in the *fon7* gene. (a) cDNA sequence analysis of the *fon7* locus in the mutant. (b) Gel electrophoresis of RT-PCR products of the *fon7* gene in the wild type and *fon7* mutant. (c) Gene structure of Os08g0299000 in the wild type and mutant. Blue boxes represent exons, and gray lines represent introns. The mutation occurred in intron 6, changing G to A (splice donor variant), which caused the deletion of exon 6 (exon skipping).

Expression analysis of *FON7*

The spatial expression pattern of *FON7* was determined by qRT-PCR to further understand the gene function. The mutant showed variable expression levels of *FON7* in vegetative tissues with higher expression in the leaf and lower expression in the root than the wild type. However, in reproductive tissues, especially spikelets and seeds, the mutant showed significantly lower level of *FON7* expression than the wild type (Figure 1-6).

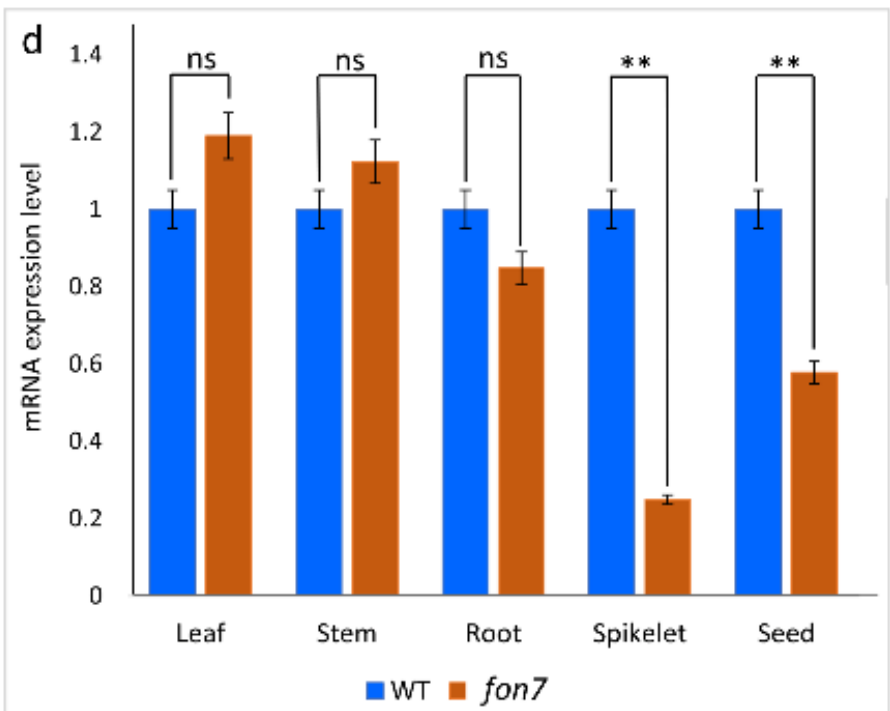


Figure 1-6. Relative expression level of Os08g0299000 in the wild type (Ilpum) and *fon7* mutant as examined by real-time quantitative PCR (qRT-PCR). WT, wild type; *fon7*, mutant.

Validation of the mutation causing altered floral organ number

The role of *Os08g0299000* in floral organ number determination was verified using a T-DNA-tagged line (PFG_3C-00521). T-DNA tagged plants were isolated from wild-type plants using HPT primer. Then, heterozygous and homozygous T-DNA plants were identified using T-DNA left border primer in combination with gene-specific primers (table 1-2) flanking the T-DNA insertion site. T-DNA was inserted in the promoter region of the *FON7* gene (Figure 1-7 a), and plants carrying homozygous T-DNA insertion showed an increased number of floral organs (stamens and pistils) and multiple pistil grains, consistent with the traits of the *fon7* mutant (Figure 1-7e–g). Interestingly, the homozygous T-DNA insertion plants also exhibited early flowering, lower tiller number, and shorter plant height than the wild type. These results indicated that the mutant phenotype is caused by the non-functional mutation of *FON7*. Compared with the wild type, the relative expression levels of *FON7* were significantly lower in the spikelet and seed but higher in the leaf, stem, and root of T-DNA plants (Figure 1-7 i).

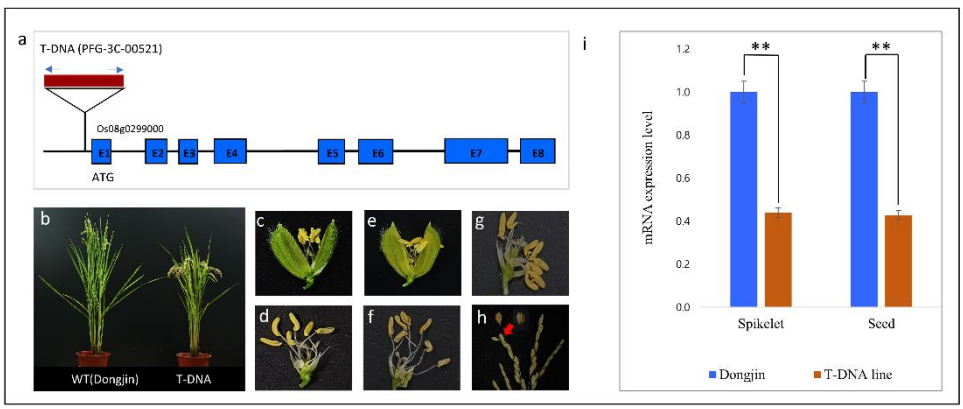


Figure 1-7. Morphological and *FON7* expression phenotypes of the T-DNA-tagged line. (a) T-DNA insertion position in the *FON7* gene model. (b) Comparison of plant morphology between the wild type and T-DNA-tagged line. (c) Wild type spikelet. (d) Wild type spikelet with lemma and palea removed, showing six stamens and two stigmas. (e) Spikelet of the T-DNA plant. (f, g) Spikelet of the T-DNA plant with lemma and palea removed, showing seven stamens and two stigmas (f) or eight stamens and four stigmas (g). (h) Occurrence of multiple pistils in the T-DNA plant. WT, wild type (Dongjin); T-DNA plant, PFG_3C-00521. Red arrowheads indicate seeds containing multiple pistils. (i) Relative expression level of *Os08g0299000* in the T-DNA plants and its corresponding wild type (Dongjin) as examined by qRT-PCR.

FON7 gene encodes Memo 1

Database analysis suggested that *FON7* is a member of the mediator of ErbB2-driven cell motility protein (Memo 1) family. HER2/ErbB2 is a member of the epidermal growth factor (EGF/ErbB) family of receptor tyrosine kinases, and ErbB signaling regulates cell growth, differentiation, and migration. Additionally, the ErbB proteins in mice is essential for vitality and loss of those proteins causes embryonic lethal, with defects found in multiple organ systems [29]. Homologs of Memo protein were found in 62 genes of 47 organisms including human, animals and plants. Among them 80% of memo protein occurred in plant species. Multiple sequence alignment indicated that rice *FON7* shared high sequence identity with its homologs in corn (Zm00001doaa850, 86%), wheat (TraesCS7D02G522800, 83%), tobacco (LOC107809489, 73%), tomato (Solyc08g029110.3, 72%), and Arabidopsis (AT2G25280, 69%) (Figure 1-7). The Memo 1 protein is a newly found plant signaling protein with unknown function, and it's signaling pathway needs to be investigated.

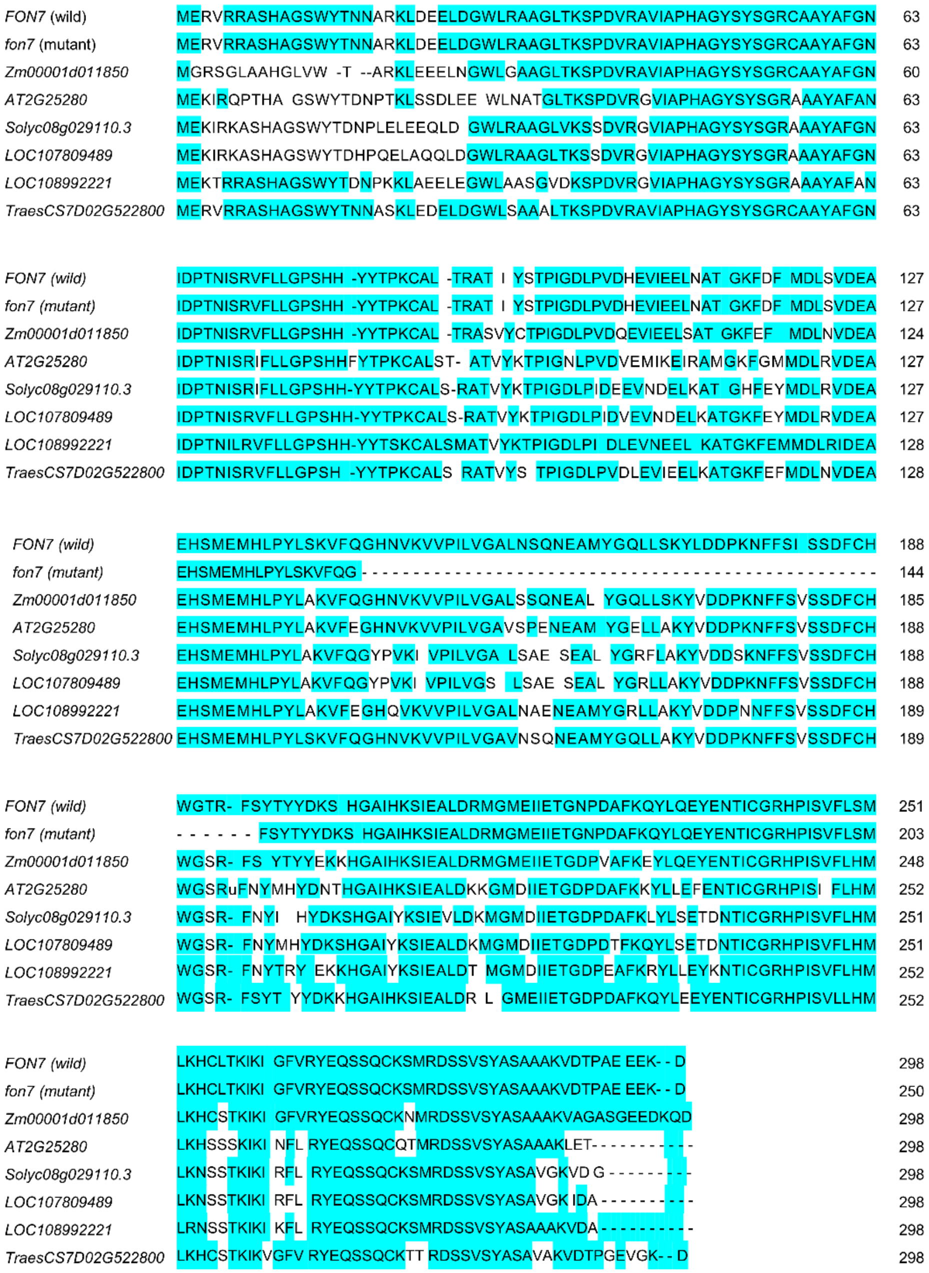


Figure 1-8. Alignment of *FON7* homologs. Amino acid sequence alignment of *FON7*. Zm00001d011850, *Zea mays*; AT2G25280, *Arabidopsis thaliana*; Solyc08g029110.3, *Solanum lycopersicum*; LOC107809489, *Nicotiana tabacum*; TraesCS7D02G522800, *Triticum aestivum*.

Discussion

Rice spikelets are the ultimate sink organs, and produce seeds as the final products. Female reproductive organs (pistils) and male reproductive organs (stamens) play crucial roles in the fertilization process, and the number of these organs affect successful seed setting and normal mature grain development [2]. The mechanisms regulating floral organ numbers remain poorly understood, and more floral organ genes need to be identified. In this study, we isolated a new candidate gene *FON7* for floral organ number which was named after five genes (*FON1*, *FON 2/ FON4*, *FON3*, and *FON5*) previously reported influencing floral organ identity. Like the *fon7* mutant, the *fon1*, *fon2*, and *fon5* mutants were also generated by chemical mutagenesis, whereas *fon3* was a spontaneous mutant. Mutations in the corresponding genes affects floral meristem size and homeotic conversion [30-32]. The *fon5 and fon7* mutants showed similar phenotypes, i.e., increased stamen and pistil numbers and homeotic conversion of stamens and pistils [32]. In *fon1* and *fon2* mutants, the lodicule number was altered, and homeotic conversion was limited to lodicules and stamens [31]. However, the multi-grain 1 mutant (*mg1),* a novel allele of *fon1*, showed an increasing number of stamens and pistils, extra lemma-like organs, and extra palea-like organs, and homeotic conversion only in the outer whorl [33]. This strong homeotic conversion effect observed in the *fon1* mutant was similar to that

observed in the *fon3* mutant, which showed the alteration of nearly all floral organ numbers and noticeable changes in panicle morphology [31].

The result of agronomic trait analysis indicated that the early flowering phenotype of the *fon7* mutant was similar to that of the *fon5* mutant. A previous report suggested that early flowering resulted from the ectopic expression of *OsMADS1*, which encodes a positive regulator of flowering [32, 34]. The *fon7* mutant showed a considerable reduction in tiller number and plant height. These phenotypes were similar to those of floral organ number mutants *fon1-3* and *fon 1-4*, in which the reduced tiller number was caused by increased auxin production from the enlarged shoot apical meristem (SAM) and enhanced apical dominance [17]. Apical dominance is enhanced by the production of plant hormones such as auxin, which inhibit the growth of the axillary buds, thus reducing the tiller number [35].

FON7 encoded the MEMO (mediator of ErBb2-driven cell motility) protein and MEMO-like protein family contains members of all branches of life, including Archaea, Bacteria and Eukaryota. Recent evidence suggests that MEMO1 integrates with a broader range of cell signaling pathways; RTK (Receptor Tyrosine Kinase) signaling, steroid signaling, and G-protein coupled receptor signaling. Traditionally, RTK signaling using MEMO1 as a mediator leads to the activation of ERK, AKT and PLCγ pathways. A major outcome of MEMO1 mediated signaling is cell migration which is an essential event during organismal development, adult homeostasis (e.g.,

immune cells, wound healing, etc.) or during pathogenesis (e.g., tumor metastasis) [36]. Memo was identified as an essential mediator of tumor cell motility induced by receptor tyrosine kinase activation. MEMO1 (Q9Y316-MEMO1_HUMAN) was originally discovered through its ability to bind one of these 'docking' sites on the erythroblastic oncogene B (ERBB2) receptor within a human epithelial breast cancer cell Rice spikelets are the ultimate sink organs, and produce seeds as the final products. Female reproductive organs (pistils) and male reproductive organs (stamens) play crucial roles in the fertilization process, and the number of these organs affects success rate of seed setting and mature grain development [2]. In this study, we isolated a novel floral organ number regulatory gene, *FON7*, which was named after six genes (*FON1*, *FON2/FON4*, *FON3*, *FON5*, and *FON6*) previously reported to influence floral organ identity. The first abnormal floral organ number phenotype was reported in *multiple pistil-1* (*mp1*) and *mp2* mutants, which exhibit floral homeotic transformation [6]. Of the six genes previously reported to regulate floral organ number, only three genes (*FON1*, *FON2/FON4*, *FON3*) have been cloned and functionally characterized to date. While the approximate chromosomal positions of *FON5* and *FON6* have been determined, their exact gene-related information and biological functions remain unknown. The *fon1*, *fon2*, *fon5*, and *fon7* mutants were generated by chemical mutagenesis, whereas *fon3* was found to exist as a spontaneous mutant in the Gu-Guang-Huang farm line [17, 29-32]. The *fon5 and fon7*

mutants exhibit similar phenotypes, i.e., increased stamen and pistil numbers and homeotic conversion of stamens and pistils (Fig 1j–o) [32]. In *fon1* and *fon2* mutants, the lodicule number is altered, and homeotic conversion is limited to lodicules and stamens [31]. However, the *multi-grain 1* (*mg1*) mutant, a novel allele of *fon1*, shows an increasing number of stamens and pistils, extra lemma-like and extra palea-like organs, and changing spikelet meristem determinacy [33]. The *fon3* mutant exhibits strong homeotic conversion, which is characterized by the alteration of nearly all floral organ numbers and noticeable changes in panicle morphology [31].

The *fon7* mutant identified in the current study showed a considerable reduction in tiller number and plant height compared with the wild-type (Table 1 and S1 Fig). These phenotypes were similar to those of floral organ number mutants *fon1-3* and *fon1-4*, in which the reduced tiller number was caused by increased auxin production from the enlarged shoot apical meristem (SAM) and enhanced apical dominance [17]. Apical dominance is enhanced by the production of plant hormones such as auxin, which inhibit the growth of axillary buds, thus reducing the tiller number [35]. Among all *fon* mutants identified to date, only the *fon5* mutant exhibited the early flowering trait of *fon7* (Table 1) [32].

FON1 encodes the CLV1 receptor kinase, and *FON2/FON4* encodes the CLV3 homolog of Arabidopsis. Thus, the *FON1* and *FON2/FON4* genes regulate floral organ number through the CLV–WUS signaling pathway. Our

results indicated that *FON7* encodes the MEMO1 protein, which is also involved in the regulation of floral organ number [37]. A major outcome of MEMO1-mediated signaling is cell migration, which is an essential event during organismal development, adult homeostasis (e.g., cellular immunity, wound healing, etc.), and pathogenesis (e.g., tumor metastasis) [36]. MEMO was reported as an oxidase, and was shown to sustain NOX-mediated O_2^- production and to increase localized ROS abundance [38]. MEMO1 contributes to the overall redox state of the cell through oxidize Ras Homolog Family Member A (Rho A), possibly interacting with other key redox regulators, and creates a localized oxidized environment conducive for signaling and migratory purposes [36, 38]. According to a recent study, Rho-like small G proteins such as RAC/ROPs act as switches of multi-functional signaling that affects leaf epidermal cell morphogenesis, polarized cell growth, and hormone [39]. In rice, OsRopGEF7B regulates floral organ development, and loss of OsRopGEF7B increases the number of floral organs in the inner whorl (stamen and ovary), leading to abnormal lemma and ectopic lodicule growth, and eventually reducing seed setting [40]. In this study, the loss of MEMO1 protein led to increased inner whorl floral organ number (stamen and ovary) and reduced floret fertility. Detailed functional analysis of the rice MEMO1 protein is needed to understand its role in floral organ number determination.

In the *fon7* mutant, 37% of the florets showed an increasing number of floral organs, including up to nine stamens and four stigmas. In addition, the *fon7* mutant showed the multiple-grain phenotype (Table 1). Multiple grains were not formed by all abnormal florets but only by florets containing four stigmas. The occurrence of multiple grains, also known as twin grain, multi-grain, and polycarpellary grain, has also been reported in the other *fon* mutants identified previously, including *fon1* (*mg1*), *fon2*/*fon4*, and *fon3*. Recently, the *twin grain1* (*tg1*) gene (allelic to *fon2/fon4)* was introgressed into the cytoplasmic male sterile (CMS) line, and a new CMS line was established with enhanced glume opening, stigma exsertion, high outcrossing rate, and high hybrid seed yield. Ye et al. (2017) suggested that floral organ number genes show great potential for increasing hybrid seed yield, and the floral organ mutant could serve as a valuable germplasm for CMS hybrid rice breeding [41]. This implies that the *fon7* mutant could be used to improve CMS lines for increasing the production of hybrid seeds.

CHAPTER II

Haplotype Analysis

in Myanmar Rice Landraces

Abstract

Heading date and grain size are the important determinants for yield and key target for genetic improvement. Several key genes associated with heading date and grain size related traits have been functionally characterized in the past. Haplotype-based breeding, haplotype-assisted genomic selection is one of the efficient improvement strategies to deploy superior haplotypes to hasten breeding progress. In this study, haplotype variation and new non-functional haplotypes of heading date and grain size related genes were detected in 46 Myanmar rice landraces. In addition, combination of functional and non-functional allele of heading date genes was also investigated. In japonica sub-species, the combination of NFF (non-functional *HD1*, functional *HD3A*, and functional *RFT1*) combination and NFW (non-functional *HD1*, functional *HD3A*, and weak function *RFT1*) combination indicated the delayed days to heading in short-day condition. For seven grain-size related genes, high nucleotide diversity was detected in *GS3* whereas lowest nucleotide diversity was observed in *GL3* of Myanmar landraces. *GS5* gene underlies as a major gene underlying grain length variation in indica sub-species of Myanmar landraces.

Keywords: Haplotype, heading date, grain size, functional and non-functional haplotypes, rice

Introduction

Rice is the most important calorie source in Asia, and its domestication began at least began 8000–9000 years ago in the Yangtze Valley [42]. Through the domestication and continuous selection of traditional farmers, wild rice transformed into today cultivated rice germplasm with desirable traits [43]. Among these traits, the heading date is one of the crucial traits for rice diversification and domestication [44].

Heading date is an important determinant for rice yield and is controlled by multiple quantitative trait loci (QTLs). Among them, a major quantitative trait locus (QTL), *Heading date 3a* (*Hd3a*) and *RICE FLOWERING LOCUS T1* (*RFT1*), the closest paralog of *Hd3a*, also functions as a mobile flowering signal that works mainly under long day conditions (LDs). Both of the "florigen" genes are regulated by *Hd1*, *Early heading date 1* (*Ehd1*), and *Days to heading 2* (*DTH2*) [45]. *Heading date 1* (*Hd1*) (counterpart of Arabidopsis CONSTANS, CO) (14) promotes *Hd3a* expression under SDs but suppresses it under LDs, whereas *DTH2*, another CO like protein, induces flowering by promoting the expression of *Hd3a* and *RFT1* under LDs. *Ehd1*, encoding a B-type response regulator, up-regulates both florigen gene expression under SDs and LDs [46]. Further, dozens of upstream negative regulators of *Ehd1* have been identified. Grain number, plant height, and heading date 7 (*Ghd7*), encoding a CO, CO-like, TOC1 (CCT) domain protein, and *Days to heading 8* (*DTH8*), encoding a putative

HAP3 subunit of the CCAAT box-binding transcription factor, act as LD-specific repressors of *Ehd1*. On the other hand, several positive *Ehd1* regulators have also been cloned by isolating mutants with extreme late flowering. Therefore, it appears that *Ehd1* is a pivotal convergence point that integrates multiple signaling pathways, including "SD-activation pathway," "LD-repression pathway," and "LD-activation pathway," to regulate the flowering time of rice under diverse environmental conditions [47].

Grain size is also the important factor determining yield and quality in rice. A large number of genes are involved in the regulation of grain size parameters such as grain length and grain width. Grain shape traits identified as quantitative trait loci (QTLs) controlled by variation in naturally occurring genes [48]. Some grain shape genes in rice are also correlated with grain weight and yield. Many major genes and QTLs have been reported for grain shape traits in rice [49]. The characterized grain size QTL were *GS3*, *GW3.1*, *GL3.1/GL3*, *GL 3.2*, *GL 3.3*, *qGL3*, *TGW3/GL3.3*, *qGL7*, *qGL7.2*, *GLW7/OsSPL13*, *SLG7*, *GW2*, *GIF1*, *GS5*, *GW6a*, *TGW6*, *OsSPL16/GW7*, *GW8/OsSPL16*, and *GL7/GW7* [50-56].

Haplotype characterization approach is a powerful tool that help in understanding a particular germplasm and identify and validate trait SNPs for applied breeding purposes and proven to be vital in understanding the trait genetic architecture in a particular germplasm [57]. For the heading date traits, the investigation of the presence of non-functional allele proved that more

than one-third of the world's rice accessions carry non-functional haplotypes of *hd1* and the combination of non-functional alleles of *Ghd7*, *Ghd8* and *Hd1* involve the distribution of rice cultivars to higher latitudes. Loss of function of some heading date genes were highly associated with the expansion and adaptation of rice throughout the world [58, 59].

Many haplotype characterization studies of heading date related genes, key genes related with yield and grain size were conducted in many countries many areas. Myanmar is known as one of the gene centers for rice, with high diversity of both wild and cultivated rice being well conserved compared with the surrounding countries [60, 61]. However, haplotype diversity analysis of Myanmar landraces still needed to study. In this study, the identification of sub-species of Myanmar rice landraces and the haplotypes analysis of heading date genes were conducted. Functional and non-functional haplotypes of heading date related genes and grain size related genes were manipulated. Then, the distribution of those haplotypes was estimated.

Materials and methods

Plant materials

A set of 46 Myanmar rice (O. sativa L.) accessions was collected from the National GeneBank of the Rural Development Administration (RDA-Genebank, Republic of Korea). Landraces' Information, including variety name, collected region, longitude and latitude origin and subpopulation identity, is listed in Table 1. All accessions were planted in the experimental farm of Seoul National University, Suwon, Korea. Five plants in the middle of each line were selected to investigate the values of agronomic traits. The selected phenotypic traits included days to heading, plant height, and spikelets per panicle were used as the phenotypic data.

Phenotypic traits statistical analysis

The mean value, the distribution, correlation of phenotypic traits, and hierarchical cluster analysis are calculated using IBM SPSS statistics 22.0 software.

Table 2-1. Geographical origin of Myanmar landraces used in this study

No.	Accession	Code	Collection Site	No.	Accession	Code	Collection Site
1	AYEPYAUNG	B1	Mon state	24	GAW AHTUN	B24	Bago division
2	IN SITT	B2	Kachin state	25	GWA NGASEIN (GWA)	B25	Rakhine division
3	LETER	B3	Shan state	26	MA WAINE OHN	B26	Mandalay division
4	TAUNG YAR SABA	B4	Magway division	27	TOEPWAGALAY(TOEPWA)	B27	Un
5	SIN E KAYI-2	B5	Chin state	28	YA THAY SAN	B28	Magway division
6	ATHIN BAYKYAUNG	B6	Yangon Division	29	GAUK YA	B29	Rakhine division
7	PALEPYU	B7	Sagaing Division	30	HMWAY SHAY	B30	Un
8	NGA YWAE NU	B8	Un	31	KHAUK CHAL	B31	Shan state
9	BU KYUT	B9	Shan state	32	MANAW THU KHA	B32	Un
10	KAUT MAE	B10	Shan state	33	KHAO PHA LUM	B33	Shan state
11	KHAO HLAING	B11	Shan state	34	SKIN THU KHA	B34	Bago/Mandalay
12	KHAUK PHI PHAN	B12	Shan state	35	TAUNG PYONE	B35	Un
13	KHAUT SANN	B13	Yangon Division	36	SA BA NI	B36	Mandalay division
14	LONG PHYU	B14	Shan state	37	PAW SAN	B37	Chin state
15	MOAT SOE MA KYWE KYAY	B15	Sagaing Division	38	SAWBWA (SAWPWAR)	B38	Ayeyawady division
16	MYAY SEIN	B16	Shan state	39	DAWEBYAN	B39	Kachin state
17	NGA SI	B17	Shan state	40	SITPWA	B40	Ayeyawady division
18	VE TA THE	B18	Shan state	41	YODANYA	B41	Magway division
19	KAUK YA	B19	Kachin state	42	PAW HTUN	B42	Un
20	KAUK HNYIN	B20	Kachin state	43	PYO NE THAR	B43	Chin state
21	DIGA	B21	Ayewaryady Division	44	KHAO SINE (KHAO SAING)	B44	Shan state
22	INHTAW KAGYI	B22	Kachin state	45	MU SOE MA KYWAY KYAY	B45	Bago division
23	BYAT KYAR	B23	Mon state	46	MOTENAUK (UPLAND)	B46	Kachin state

DNA extraction, genome sequencing and alignment with reference genome

Fresh leaves were sampled from field-grown plants and total genomic DNA was extracted using the cetyl-trimethyl ammonium bromide (CTAB) method. Qualified DNA samples were used for whole-genome resequencing of the collected rice landraces with an average coverage of approximately 16X on the Illumina HiSeq. 2000 Sequencing Systems Platform. Nipponbare genome was used as the reference genome and the genome sequence of 25 check varieties were downloaded from the 3K Rice Genome to identify the subpopulation of the Myanmar rice landraces.

NGS analysis and genotyping

Whole genome sequencing data were generated on the Illumina HiSeq x system to generate 2x 150 bp paired-end reads with a sequencing depth of > 10 x per sample. Raw reads were processed to remove adaptors and low-quality bases using Trimmomatic v0.38 with the parameters ILLUMIMACLIP:2:30:10 SLIDEWINDOW:4:15 MINLEN:50. Read were aligned to the rice reference genome (Nipponbare, IRGSP v1.0) and 25 check varieties selected from 3K rice genome using the BWA v0.7.17 MEM algorithm with the default parameters. Aligned reads were sorted using samtools v 1.9 and duplicated were removed using Picard v2.20.2. Nucleotide variants were called by the HaplotypeCaller function of GATK v4.1.2 with the parameters-max-missing 0.95 -minQ 30- minDP 5. In addition, nucleotide

variants with proportions of heterozygous genotypes of >0.05 were filtered using the vc. getHetcount command in GATK v4.12. (tiller no. paper,CMB)

Nucleotide diversity, Haplotype network and their distribution

Nucleotide diversity, including the average number of pairwise nucleotide differences per sites (π) and watterson's estimator derived from the total number of segregating sites (θ_w) were analyzed using DnaSP 5.0 (Librado and Rozas 2009). The haplotype network for each gene was generated using PopART v1.7. First, we used a python script to make FASTA data from the vcf file. Then, FASTA data alignment and transformation to nex format was performed using MEGA7. DnaSP v6 was employed for haplotype analysis.

For some heading date genes, the haplotypes were divided into functional and non-functional groups. The distribution of those functional and non-functional haplotypes was visualized using PopART v1.7.

Table 2-2. Heading date related genes for haplotype analysis in this study

No.	Gene ID	Gene Symbol	Gene Ontology
1	*Os06g0275000*	*Hd 1/ SE1*	Positive Regulation of Short-Day Photoperiodism, Negative Regulation of Long-Day Photoperiodism
2	*Os08g0174500*	*HD5/DTH8*	Negative Regulation of Short-Day Photoperiodism
3	*Os07g0695100*	*HD2/DTH7*	Negative Regulation of Long-Day Photoperiodism
4	*Os06g0157700*	*HD3A*	Positive Regulation of Short-Day Photoperiodism
5	*Os06g0157500*	*RFT1*	Regulation of Long-Day Photoperiodism
6	*Os07g0261200*	*Ghd7*	Negative Regulation of Long-Day Photoperiodism
7	*Os10g0463400*	*Ehd1/Ef1*	Floral inducer to promote short-day (SD) flowering pathway
8	*Os06g0160400*	*HGW*	Photoperiodism
9	*Os03g0762000*	*HD6*	Negative Regulation of Long-Day Photoperiodism

Table 2-3. Yield related genes for haplotype analysis in this study

No.	Gene ID	Gene Symbol	Gene Ontology
1	*Os01g0883800*	*SD1*	internode patterning, response to deep water
2	*Os06g0610300*	*MOC1*	regulation of meristem growth, root development
3	*Os08g0509600*	*IPA1*	inflorescence development, leaf development, negative regulation of seed germination
4	*Os11g0235200*	*SP1*	panicle type, panicle length, grain number, inflorescence branching
5	*Os06g0677000*	*DEP3*	grain yield, grain number, inflorescence branching
6	*Os01g0887700*	*PHD1*	regulation of transcription, transcription cofactor activity, metal ion binding, histone binding
7	*Os02g0674800*	*ROC5*	regulation of cell division
8	*Os03g0449200*	*AGO7*	organ morphogenesis, maintenance of shoot apical meristem identity

Table 2-4. Grain size related genes for haplotype analysis in this study

No.	Gene ID	Gene Symbol	Gene Ontology
1	*Os03g0646900*	*GL3.1*	grain length, leaf lamina joint bending, grain weight, grain size
2	*Os03g0417700*	*GL3.2*	monooxygenase activity, regulation of seed growth, seed size, grain yield, grain size, grain length
3	*Os03g0841800*	*GL3.3*	protein serine/threonine kinase activity, seed development, seed size, grain weight, grain length, grain size, seed development trait,
4	*Os12g0528400*	*GS3*	glutathione synthase activity, cytosol
5	*Os05g0158500*	*GS5*	serine-type carboxypeptidase activity, proteolysis, grain size, grain weight, grain width, grain length, seed maturation
6	*Os07g0603300*	*GL7/GW7*	unidimensional cell growth, plasma membrane, inflorescence development, seed development, grain size, grain length, grain shape, yield trait
7	*Os05g0187500*	*GW5*	amino sugar metabolic process, nucleotide-sugar metabolic process, response to heat, inflorescence development, regulation of cell proliferation, seed development, grain width, grain weight, grain size

Results

Phenotypic diversity and correlation analysis

The frequency distribution of three agronomic traits in 46 Myanmar landraces are presented in Figure 2-1. 30% of total landraces (14 landraces) had been observed the short plant height. Whereas, 10 % of landraces have been observed tall plant height which are susceptible to lodging. For heading date, nearly half of the landraces (52%) have been observed early heading date. The landraces possessing early heading date were indicated high spikelets per panicle.

According to correlation analysis, heading date was negatively correlated with all traits expect plant height. Whereas positive correlation was observed between 5 traits, panicle length, primary branch number, secondary branch number, filled grains and spikelet per panicle. (Figure 2-1). Panicle length was positively correlated with primary branch number, secondary branch number, filled grains and spikelet per panicle respectively.

Two distinct groups were obtained from clustering result based on the 15 phenotypic traits (Figure 2-1). Most of the Indica sub-population were clustered together with CK1 (IR64) in cluster I and japonica landraces were grouped into cluster II along with CK2 (Nipponbare) and CK3 (Malaki Sinaguing).

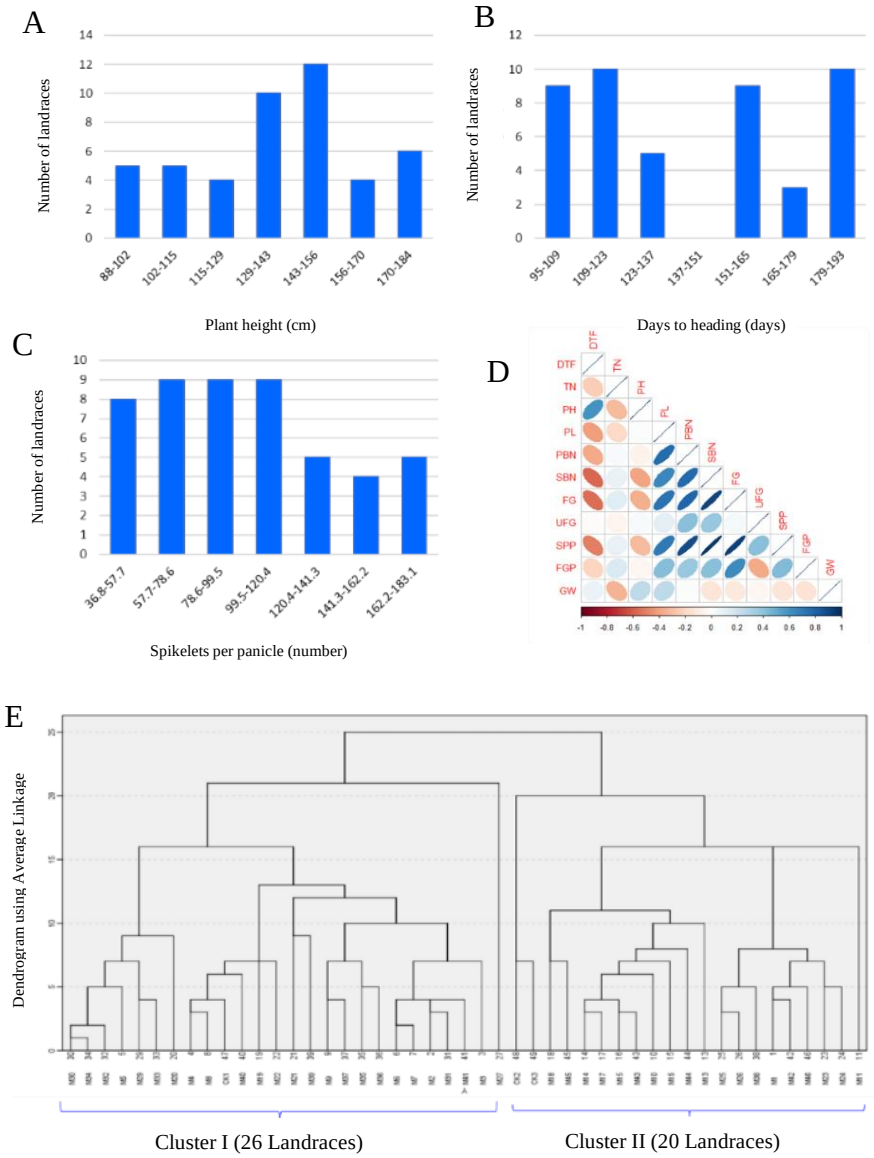


Figure 2-1. Phenotypic diversity and correlation analysis. (A-C) Frequency distribution of agronomic traits in the 46 Myanmar rice Landraces (D) Correlation Analysis of Agronomic Traits of 46 Myanmar Landraces (E) Cluster Dendrogram of 46 Myanmar Germplasm with three check varieties

PCA analysis and sub-species distribution in Myanmar

Based on the results of the principal component analysis, the first 14 components account for a total of 71.67% of the total variance (Table 6). The first two vectors accounted for 25.51% of the total variance, of which the first vector share was 18.15% and the second vector share was 7.35%. Based on the main coordinates of the diagram, the pattern of variation among individuals was plotted (Figure. 2-2B). In the diversity pattern diagram, 46 rice accessions were grouped into three major clusters (Figure 2-2C). Population I (indicated with red color) counted for 52% of total landraces (26 landraces) and all are Indica sub-species cultivars. The germplasm containing in this cluster group I showed the heading date ranges from 95 DAS to 195 DAS. Indica type landraces are distributed throughout the whole country and especially high distribution in the lowland area. Japonica populations were grouped together in cluster II (highlighted with purple color) which contained 41% of total accessions (19 landraces). Whereas, only one Aroma Germplasm (3% of total accessions) was occurred in cluster III (B22 germplasm) which possessed the early heading date phenotypes (90DAS). According to the distribution of indica-japonica sub-species in Myanmar, Japonica populations are dominant in hilly zone area. Interestingly, Aus sub-species germplasm has not occurred in these Myanmar landraces.

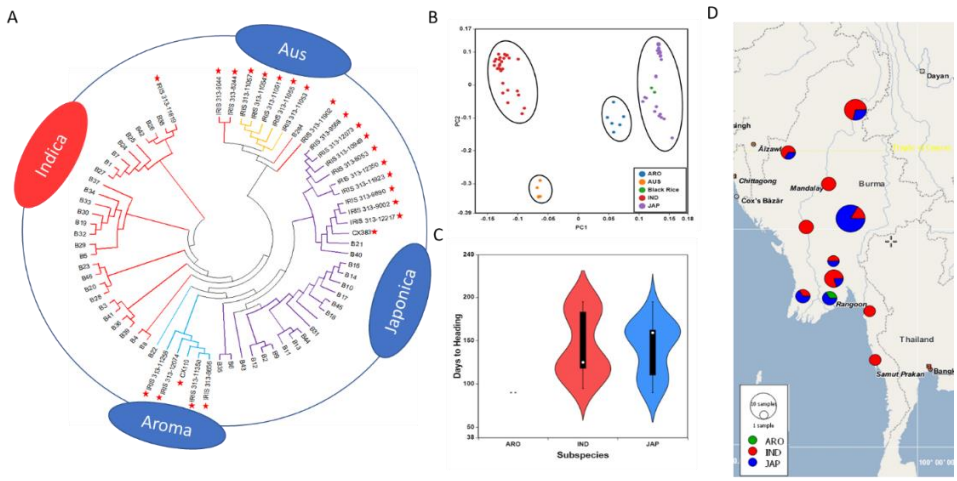


Figure 2-2. PCA analysis and distribution of sub-species in Myanmar (A) Neighbor joint tree, the check cultivars from the neighbor countries are highlighted with red star and the different sub-populations are highlighted with different colors. (B) Principal component analysis (C) Days to heading (days) associated with subspecies; *Aroma*, *Indica* and *Japonica* (D) Distribution of sub-population group of 46 Myanmar rice landraces. ARO- *Aroma*, IND- *Indica*, and JAP- *Japonica*.

Polymorphism and nucleotide diversity of genes associated with heading date

Nucleotide polymorphism of heading date genes were examined in the sequence of nine heading date related genes, (*Hd1*, *Hd5*, *Hd2*, *Hd3A*, *Rft1*, *Ghd7*, *Ehd1*, *HGW*, and *Hd6*). The aligned sequence length of *Hd1* was 1577 bp. A total of 19 nucleotide substitutions and 5 indels were detected with the average of 1 per 83 bp and 1 per 315 bp respectively. Number of indel sites and average indel length were 18 and 5.12 bp. The pairwise nucleotide diversity of overall coding sites were 0.49 respectively. In *HD3A*, total length was 847 bp and total of 26 nucleotide substitutions (1SNP per 33 bp) and 5 indel (1 indel per 169 bp) were detected. A total of 36 nucleotide substitution (1SNP per 24 bp) and 6 indels (1SNP per 144 bp) were detected across 866 bp of *RFT1*. In aligned 1009 bp of *Ghd7* sequence, a total of 21 SNPs (1 SNP per 48 bp) and 4 indels (indel per 252 bp) were detected. The aligned sequence of *Ehd1* was 1026 bp long. A total of 88 SNPs (average 1 SNP per every 12 bp) were detected and 7 indels (1 SNP per every 147). *HGW* sequence was aligned as length of 1070 bp, including 88 SNPs (1 SNP per 12 bp) and 19 indels (1 indel per 56 bp). Total 22 SNPs (1 SNP per 20) and 2 indels (1 indel per 384) were detected in 767 aligned based pairs of *HD6*. However, only one stop gain SNP substitution was detected in exon region.

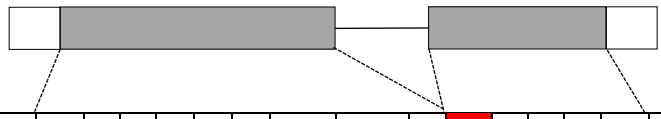
Table 2-5. Nucleotide diversity for heading date related genes

Gene Symbol	*Hd 1/ SE1*	*HD5/ DTH8*	*HD2/ DTH7*	*HD3A*	*RFT1*	*Ghd7*	*Ehd1/ Ef1*	*HGW*	*HD6*
Total length of coding site bp	1577	1608	3311	847	866	1009	1026	1070	767
total mutation	24	165	87	31	42	25	63	107	22
No. of nucleotide substitution	19	143	75	26	36	21	56	88	20
Frequency of nucleotide substitution per bp	0.012	0.089	0.023	0.031	0.042	0.021	0.055	0.082	0.026
No. of Indels	5	22	12	5	6	4	7	19	2
Frequency of Indels	0.003	0.014	0.004	0.006	0.007	0.004	0.007	0.018	0.003
total bp for 1 SNP	83	11	44	33	24	48	18	12	38
total bp for 1 Indel	315	73	276	169	144	252	147	56	384

Haplotype analysis and network of *Hd1*

Six haplotypes of *Hd1* were identified in 48 Myanmar rice landraces and 25 varieties of neighbor countries (Figure. 1). Hd1-H3 was the major haplotype and the frequency was 41%. Hd1-H3 haplotype was present in both indica and japonica sub-species. Distribution of this haplotype was found in China, Indonesia, Bangladesh and India. The Hd1-H5 was the second highest frequency (27%) and this type of haplotypes distributed in the India, Philippines and South-Korea. The rare haplotype (Hd1-H1 and Hd1-H4) occurred only in Myanmar landraces. Hd1-H1 haplotype group was present only in the indica sub-species whereas Hd1-H2 and Hd1-H4 detected only in the japonica sub-species. The haplotypes of the landraces were divided into two clades according to the functional and non-functional haplotypes. Three haplotypes showed non-functional haplotypes. The 2 bp deletion in the exon 2 leading to early stop of hd1 protein and non-functional of *hd1* was occurred in Hd1-H4 and Hd1-H5. In addition, Hd1-H6 was also observed as the non-functional haplotype which contained the 3 bp deletion and 1 bp deletion in the exon 1.

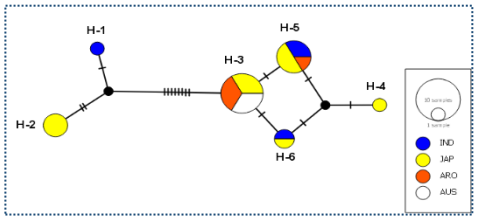
A.



Hd1 (*Os06g0275000*)

Hap group	FS	His106Tyr	Val76Ala	Lys67Asn	Glu66Asp	Asp151Asn	InDel	Del	Pro6Leu	FS (Type 7)		Arg299Ser	Gly387Ser	Distribution (%)	Occurrence
REF	T	C	A	C	C	G	A	C	G	CTT		A	G		
Hd1-H1	T	T	A	C	C	G	A	C	G	CTT		A	G	5	Myanmar
Hd1-H2	T	C	A	C	C	G	A	G	A	CTT		A	G	14	Myanmar, South-korea, Italy
Hd1-H3	TAG	C	G	A	A	A	ACAG	C	A	CTT		C	A	41	Myanmar, China, Indonesia, Bangladesh, India
Hd1-H4	T	C	G	A	A	A	*|*	1bp del	A	C		C	A	5	Myanmar
Hd1-H5	TAG	C	G	A	A	A	ACAG	C	A	C		C	A	27	Myanmar, India, Philippines, South-Korea
Hd1-H6	TAG	C	G	A	A	A	*|*	1bp del	A	CTT		C	A	9	Myanmar

B



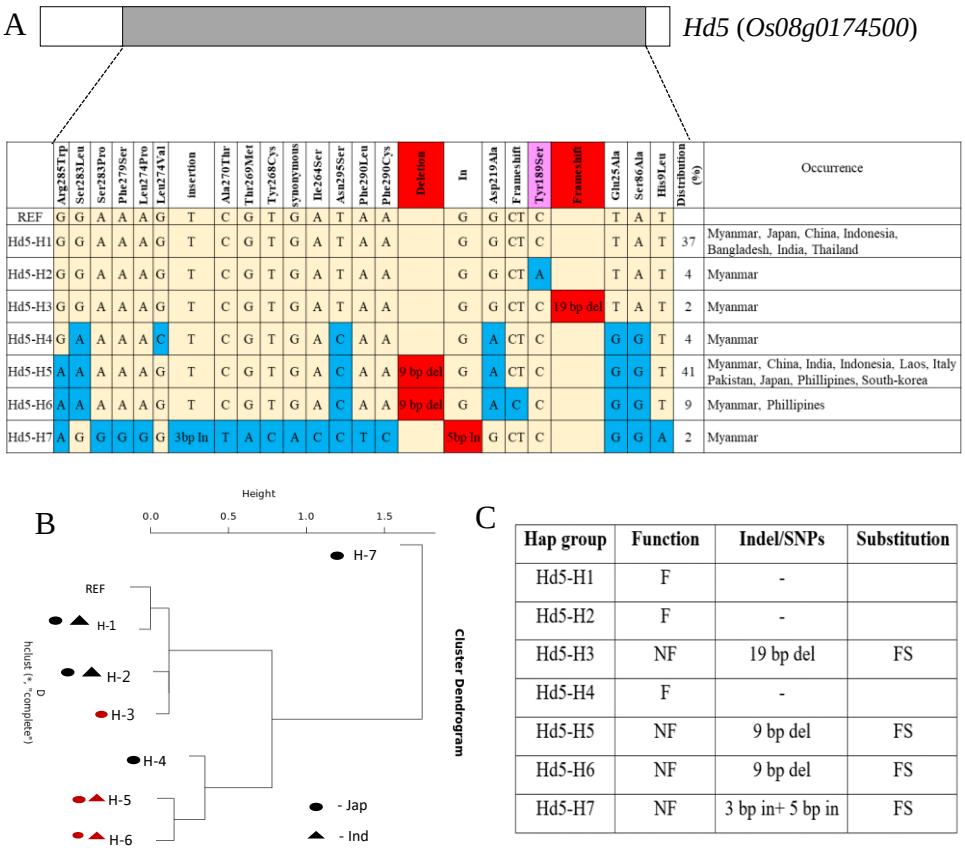
C

Hap group	Function	Indel/SNPs	Substitution
Hd1-H1	F	-	
Hd1-H2	F	-	
Hd1-H3	F	-	
Hd1-H4	NF	3 bp del+ 1bp del + 2 bp del	FS
Hd1-H5	NF	2bp del	FS
Hd1-H6	NF	3 bp del+ 1bp del	FS

Figure 2-3. Haplotype analysis and distribution of the *Hd1* gene. (A) The gene structure and haplotype analysis of the *Hd1* (*Os06g0275000*) gene in 76 germplasms. Type amino acid changes is shown in the first row. The distribution percent of haplotypes is presented in the last row. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non- functional (B) the haplotype network of six haplotypes of Hd1 genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS- (C) The information of function and non-functional haplotypes of *Hd1*. The red color shows the non-functional haplotypes and the green color indicates function haplotypes. F- functional haplotype and NF- non-functional haplotype.

Haplotype analysis of *Hd5*

Based on these nucleotide polymorphisms of 71 genotypes, seven haplotypes were detected in the *Hd5* gene region. Hd5-H5 was the most common predominant haplotype across 46 Myanmar rice landraces, found in 41 % of total landraces, of which about 47% (9 landraces) were of Indica sub population, and 42% were of Japonica sub population (8 landraces). The haplotype Hd5-H3 and Hd5-H4 occurred only in the japonica sub population. The rare haplotype Hd5-H3, Hd5-H6, and Hd5-H7 was distinguished from all other haplotypes by the Indel casual mutation allele in the second exon of Hd5 at 19 bp deletion (at the position of 4334626) and 5 bp insertion (at the position of 4333910) respectively. Based on the functional and non-functional haplotypes, we anticipated that Hd5-H3, Hd5-H5, Hd5-H6, and Hd5-H7 were non-functional haplotypes and occurred in 52 landraces. Other four haplotypes possessed functional haplotypes. In Myanmar, 28 % of landraces (n-13) carry non-functional haplotypes.



	Arg285Trp	Ser283Leu	Ser283Pro	Phe279Ser	Leu274Pro	Leu274Val	insertion	Ala270Thr	Thr269Met	Tyr268Cys	synonymous	Ile264Ser	Asn295Ser	Phe290Leu	Phe290Cys	Deletion	In	Asp219Ala	Frameshift	Tyr189Ser	Frameshift	Glu25Ala	Ser86Ala	His91Leu	Distribution (%)	Occurrence
REF	G	G	A	A	A	G	T	C	G	T	G	A	T	A	A		G	G	CT	C		T	A	T		
Hd5-H1	G	G	A	A	A	G	T	C	G	T	G	A	T	A	A		G	G	CT	C		T	A	T	37	Myanmar, Japan, China, Indonesia, Bangladesh, India, Thailand
Hd5-H2	G	G	A	A	A	G	T	C	G	T	G	A	T	A	A		G	G	CT	A		T	A	T	4	Myanmar
Hd5-H3	G	G	A	A	A	G	T	C	G	T	G	A	T	A	A		G	G	CT	C	19 bp del	T	A	T	2	Myanmar
Hd5-H4	G	A	A	A	A	C	T	C	G	T	G	A	C	A	A		G	A	CT	C		G	G	T	4	Myanmar
Hd5-H5	A	A	A	A	A	G	T	C	G	T	G	A	C	A	A	9 bp del	G	A	CT	C		G	G	T	41	Myanmar, China, India, Indonesia, Laos, Italy Pakistan, Japan, Phillipines, South-korea
Hd5-H6	A	A	A	A	A	G	T	C	G	T	G	A	C	A	A	9 bp del	G	A	C	C		G	G	T	9	Myanmar, Phillipines
Hd5-H7	A	G	G	G	G	G	3bp In	T	A	C	A	C	C	T	C		5bp In	G	CT	C		G	G	A	2	Myanmar

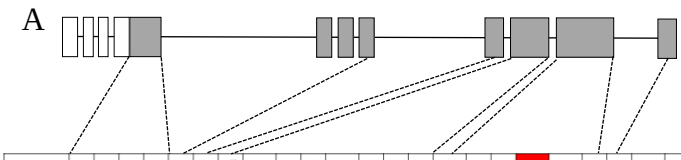
Hap group	Function	Indel/SNPs	Substitution
Hd5-H1	F	-	
Hd5-H2	F	-	
Hd5-H3	NF	19 bp del	FS
Hd5-H4	F	-	
Hd5-H5	NF	9 bp del	FS
Hd5-H6	NF	9 bp del	FS
Hd5-H7	NF	3 bp in+ 5 bp in	FS

Figure 2-4. Haplotype analysis and distribution of the *Hd5* gene. (A) The gene structure and haplotype analysis of the *Hd5* (*Os08g0174500*) gene in 76 germplasms. Type amino acid changes is shown in the first row. The distribution percent of haplotypes is presented in the last row. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non-functional (B) the haplotype network of six haplotypes of *Hd5* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS- (C) The information of function and non-functional haplotypes of *Hd5*. The red color shows the non-functional haplotypes. F- functional haplotype and NF- non-functional haplotype.

Haplotype analysis of *Hd2*

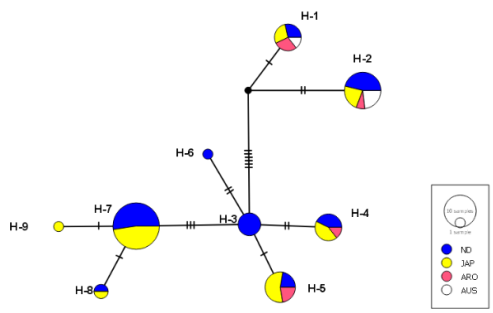
According to the results of haplotype analysis, 21 SNPs and indel mutations were observed and nine haplotypes were detected in the *Hd2* gene region. The most common predominant haplotype was detected in the Hd2-H7 haplotype group. The 32 % of total genotypes showed Hd2-H7 haplotype, of which 9 landraces were of Indica sub-species, and 8 landraces were of Japonica sub-species. The genotypes of Hd2-H7 haplotype group were located in Japan, Bangladesh, India and Thailand. While the rare haplotypes Hd2-H8 and Hd2-H9 was distinguished from all other haplotypes by the Indel casual mutation allele of 8 bp deletion and 2bp deletion respectively. Besides, Hd2-H3 and Hd2-H6 were present only in Myanmar. These haplotypes Hd2-H3 and Hd2-H6 occurred only in the indica sub -species whereas Hd2-H9 haplotype group was detected only in japonica sub-species. Based on the functional and non-functional haplotypes, we anticipated that Hd2-H8, and Hd2-H9 were non-functional haplotypes while other *Hd2* haplotype groups showed functional *Hd2*.

A



	Gly16Ala	Ile42Thr	Arg47Pro	Asp68Glu	Asn214Ser	Pro268Thr	splice region	InDel	Ala383Thr	synonymous	Gly420Asp	synonymous	synonymous	Asp436His	Met457Val	Leu462Pro	Ser466Ile	InDel	His583Asn	Gly656Glu	Arg742Thr	Distribution (%)	Occurrence
REF	G	T	G	T	A	C	G		G	A	G	C	G	G	A	T	G		C	G	G		
Hd2-H1	G	T	C	T	A	C	G		G	A	G	C	G	G	A	T	G		C	G	G	11	Myanmar, Japan, Bangladesh, India, Thailand
Hd2-H2	C	T	C	G	A	C	G		G	A	G	C	G	G	A	T	G		C	A	G	20	Myanmar, China, Bangladesh
Hd2-H3	G	C	C	G	G	C	G		A	A	G	T	G	G	G	C	G		C	G	G	8	Myanmar
Hd2-H4	G	C	C	G	G	C	G		A	G	G	T	C	G	G	C	G		C	G	G	11	Myanmar, South korea, Italy, Pakistan
Hd2-H5	G	C	C	G	G	C	G		A	A	G	T	G	G	G	C	G		A	G	G	14	Myanmar, India, Japan, Philippines, South-korea
Hd2-H6	G	C	C	G	G	A	G		A	A	G	T	G	G	G	C	G		C	G	C	2	Myanmar
Hd2-H7	G	C	C	G	G	C	A		A	A	A	T	G	G	G	C	T		C	G	G	32	Myanmar, China, Indonesia, Philippines, Laos
Hd2-H8	G	C	C	G	G	C	A		A	A	A	T	G	G	G	C	T	8 bp del	C	G	G	3	Myanmar, Sri-Lanka
Hd2-H9	G	C	C	G	G	C	A	2 bp del	A	A	A	T	G	G	G	C	T		C	G	G	2	Myanmar, Malaysia

B



C

Hap group	Function	Indel/SNPs	Substitution
Hd2- H1	F	-	
Hd2-H2	F	-	
Hd2-H3	F	-	
Hd2-H4	F	-	
Hd2-H5	F	-	
Hd2-H6	F	-	
Hd2-H7	F	-	
Hd2-H8	NF	8 bp del	FS
Hd2-H9	NF	2 bp del	FS

Figure 2-5. Haplotype analysis and distribution of the *Hd2* gene. (A) The gene structure and haplotype analysis of the *Hd2* (Os07g0695100) gene in 76 germplasms. Types amino acid changes are shown in the first row. The distribution percent of haplotypes is presented in the last row. The cream-color cells indicate the original nucleotides, the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non-functional (B) the haplotype network of six haplotypes of Hd1 genes. IND- Indica, JAP- Japonica, ARO- Aroma and AUS- (C) The information of function and non-functional haplotypes of *Hd2*. The red color shows the non-functional haplotypes and the green color indicates function haplotypes. F- functional haplotype and NF- non-functional haplotype.

Geographical distribution of *Hd1*, *Hd5* and *Hd2* of 3K rice

The distribution of functional and non-functional haplotypes of *Hd1*, *Hd5*, and *Hd2* in 3K rice genome project are presented in figure 2-3. The non-functional haplotypes of *hd1* widely spread nearly in all country. Whereas, the 3K rice genomes indicated that non-functional *Hd5* haplotypes were highly distributed in Europe, South America, and China. For *Hd2* haplotypes, functional haplotypes were highly distributed throughout all countries and some non- functional haplotypes were distributed in Africa and America.

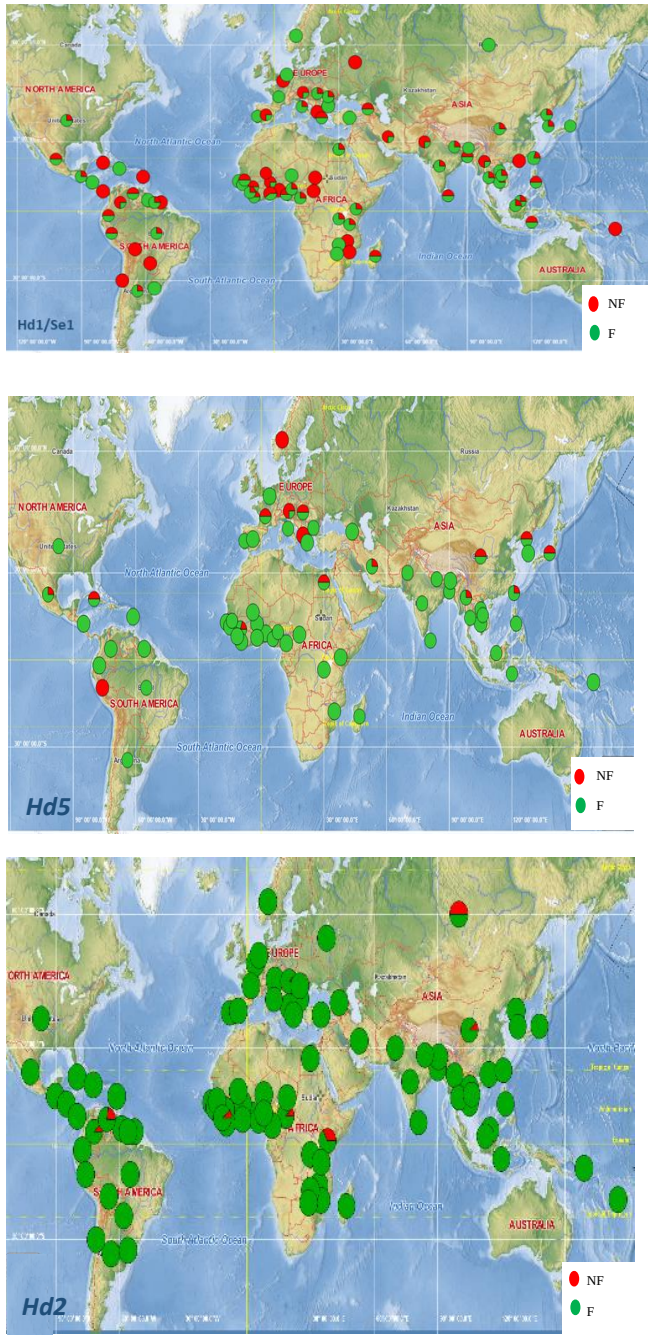


Figure 2-6 The distribution of function and non-functional haplotypes of *Hd1*, *Hd5* and *Hd2* in 3K rice germplasms. The red color shows the non-functional haplotypes and the green color indicates function haplotypes. F- functional haplotype and NF- non-functional haplotype.

Geographical distribution of functional and non-functional haplotypes in long day suppressor heading date genes

In Myanmar rice germplasms, non-functional haplotypes were present in heading date genes. Non-functional haplotypes were detected clearly in three haplotypes (*Hd1*-H4, *Hd1*-H5, and *Hd1*-H6) of *Hd1*, four haplotypes (*Hd5*-H3, *Hd5*-H5, and *Hd5*-H6) of *Hd5*, and two haplotypes (Hd2-H8 and Hd2-H8) of *Hd2*. The genetic effects and geographical distribution of the genotypes at the level of three-gene combinations among known functional alleles were estimated. A total of six combinations were found among 71 genotypes (table.2-3).

The combinations of FFF (Functional haplotypes of *Hd1*, *Hd5* and *Hd2*) and NFF (Non-functional haplotypes of *Hd1*, functional haplotypes of *Hd5* and *Hd2*) were the dominant combinations in Myanmar germplasms. Whereas the FFN combination (Functional haplotypes of *Hd1*, *Hd5*, and non-functional haplotypes *Hd2*) was not occurred in Myanmar germplasms used in this study. In combination of NFN (non-functional haplotypes *Hd1*, functional haplotypes *Hd5*, and functional haplotypes *Hd2*), *japonica* and *indica* subpopulation involved nearly the same distribution in Myanmar. But the high distribution of japonica was in NNF combination (Non-functional haplotypes *Hd1*, non-functional haplotypes *Hd5*, and functional haplotypes *Hd2*). The NFN combinations (Non-functional haplotypes *Hd1*, functional

haplotypes *Hd5*, and non-functional haplotypes *Hd2*) was observed in only one japonica sub-species with early heading date.

The geographical distribution study of those three gene combinations investigated that the main combination of FNF was widely distributed in many countries, China, Vietnam, Laos, Malaysia, Philippines, and Japan. Then the FFF combination was also found in China, Cambodia, India, Indonesia, Philippines and Japan. In the contrary, the combination NFN occur only in Myanmar and the FFN combination was only in Sri-Lanka. The favorable early heading date traits was examined in the NFN combination of *japonica* sub-species. The other combination was under the wide range of heading date.

Table 2-6. Heading date associated with all of the detected three-gene combinations of *Hd1*, *Hd5* and *Hd2* in rice (*Oryza sativa*)

Combinations			Sub-species			Distribution %
Hd1	***Hd5/ DTH8***	***Hd2/DTH7***	**IND**	**JAP**	**ARO**	
F	F	F	8	5	1	30.4
F	N	F	6	0	0	13.0
N	F	F	9	6	0	32.6
N	F	N	0	1	0	2.2
N	N	F	4	6	0	21.7

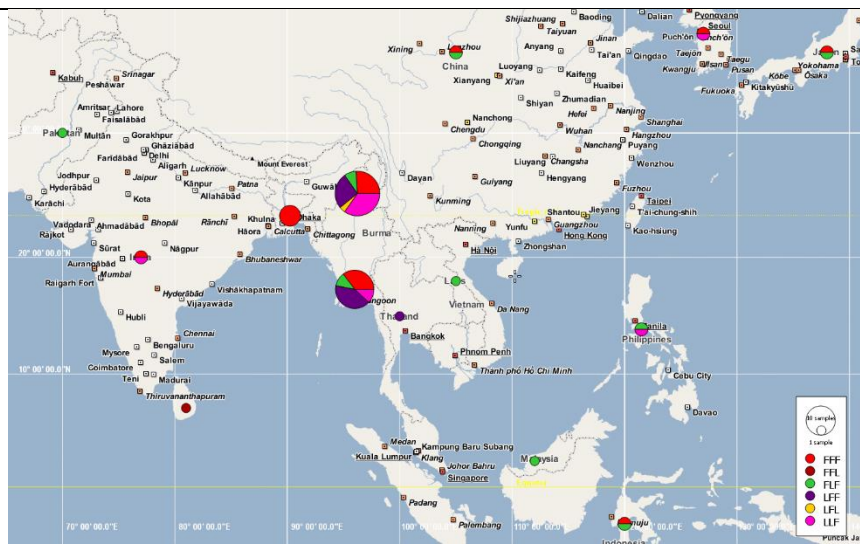


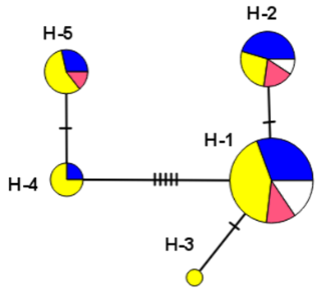
Figure 2-7. Geographical distribution of functional and non-functional haplotypes of the three gene combinations in rice (*Oryza sativa*). The different colors represent different combinations of *Hd1*, *Hd5* and *Hd2*. F, Functional haplotypes of *Hd1* (Hd1-H1, Hd1-H2, Hd1-H3, and Hd1-H6), *Hd5* (Hd5-H1, Hd5-H2, Hd5-H4, and Hd5-H7), and *Hd2* (Hd2-H1, Hd2-H2, Hd2-H3, Hd2-H4, Hd2-H5, Hd2-H6, Hd2-H7, and Hd5-H9); N, Non-functional haplotypes of *Hd1*(Hd1-H4, Hd1-H5, and Hd1-H6), *Hd5* (Hd5-H3, Hd5-H5, and Hd5-H6), and *Hd2* (Hd2-H8, and Hd2-H9)

Haplotype analysis of short-day promoter heading date genes *RFT1*, and *Hd3A*

Two florigen genes RFT1 and Hd3A are generated haplotype analysis to determine the haplotype variation in 71 genotypes. Among the five-haplotype groups of Rft1, Rft1-H1 was the most common SNP haplotype (nearly 53% of total distribution), of which all sub-species, indica, japonica, aroma and aus were present. Rare haplotype group Rft1-H3 was occurred only in Myanmar, of which all landraces belonged to the Japonica sub-species. (Fig2-8A-B). Rft1-H5 haplotype group showed the nucleotide substitution of G to A which causes the loss of function of Rft1.

In Hd3A haplotype analysis, there was no nucleotide substitution and indel mutation in 71 genotypes. Five haplotype groups are classified according to nucleotide changes in the promoter region. Hd3A-H4 haplotype group with 11 bp insertion in promoter region caused the weakness in the Hd3A function. 13% of total genotypes showed this 11 bp insertion mutation, of which both indica and japonica sub-species were observed. Hd3A-H1 was observed as the most common haplotype group in the panel and which are similar to reference sequence and occurred in all sub-species group. The other four haplotype groups also showed indel mutation of 1 bp insertion at 294963 and 2 bp insertion at 2940963 position.

	Val31Ala	synonymous	Pro95Ser	Glu105Lys	FS	synonymous	Ser144Asn	Thr160Ala	Distribution (%)	Occurrence
REF	T	T	C	G	CG	C	G	A		
Rft1-H1	T	T	C	G	CG	C	G	A	53	Myanmar, Bangladesh, China, India, Japan,Thailand, Philippines,South-Korea
Rft1-H2	T	T	T	G	CG	C	G	A	22	Myanmar, Bangladesh, Indonesia, Japan
Rft1-H3	T	T	C	G	C	C	G	A	2	Myanmar
Rft1-H4	C	C	C	G	CG	T	A	G	8	Myanmar
Rft1-H5	C	C	C	A	CG	T	A	G	14	Myanmar, China, South-korea, Laos, Italy, Pakistan Malaysia



	2940395	2940414	2940434	2940442	2940962	2940963	2941042	2941269	2941346	2941526	2941551	2941599	2941629	2941633	2941669	2941680	2941682	2941700	2941965	2942052	Distribution (%)	Occurrence
Ref	G	CT	A	A	A	C	C	T	T	C	C	G	C	T	T	C	C	C	A	C		
Hd3A-H1	G	CT	A	A	A	C	C	T	T	C	C	G	C	T	T	C	C	C	A	C	69	Myanmar, China, Indonesia, Bangladesh, India, Thailand, Philippines, South-Korea
Hd3A-H2	G	CT	A	T	A	2bp In	T	T	T	C	C	A	G	T	G	C	T	T	A	T	6	Myanmar
Hd3A-H3	G	C	A	T	1bp In	C	C	T	T	T	C	A	G	T	G	T	T	T	A	T	8	Myanmar, Sri-Lanka
Hd3A-H4	G	C	A	T	1bp In	C	C	T	T	T	11bp In	A	G	T	G	T	T	T	A	T	13	Myanmar, South-korea, Italy, Japan
Hd3A-H5	T	CT	G	T	A	2bp In	T	C	G	C	C	A	G	A	G	C	T	T	T	T	4	Myanmar

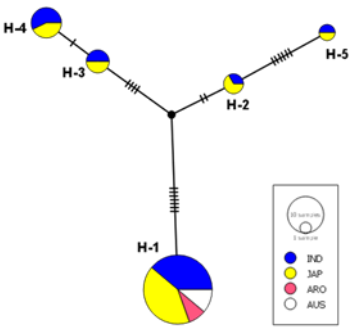


Figure 2-8. Haplotype analysis of the and *Rft1* gene, and *Hd3A* in 71 rice germplasms

Combinations of short-day promoter genes (*Hd1, Hd3A and RFT1)* and their association with heading date

Three heading genes (*Hd1* and *Hd3A*) which enhanced flowering in the short-day condition were classified functional and non- functional haplotype groups and *RFT1* haplotype groups were grouped into two types the normal function and weak-function according to the indel mutation in the promoter region. Then, these functional and non-functional haplotypes were analyzed to determine the combination of those in Myanmar landraces. Eight combinations were present among 46 Myanmar landraces. Among 46 germplasms, nearly 26% of total showed the FFF (three functional haplotypes for *Hd1*, *Hd3A* and *RFT1*). These FFF combination was present in both indica and japonica sub-species. Whereas, FFW (functional *HD1*, functional *HD3A*, and weak function *RFT1*) combination, FNF (functional *HD1*, non-functional *HD3A*, and weak function *RFT1*) combination, FNW (functional *HD1*, non-functional *HD3A*, and weak function *RFT1*) combination, NNF (non-functional *HD1*, non-functional *HD3A*, and weak function *RFT1*) combination, and NNW (non-functional *HD1*, non-functional *HD3A*, and weak function *RFT1*) combination were present only in indica sub-species. NFW (non-functional *HD1*, functional *HD3A*, and weak function *RFT1*) combination was observed only in japonica sub-species. Then, the association of these combinations with days to heading was calculated. In japonica sub-species, the combination of NFF (non-functional *HD1*, functional *HD3A*, and

functional *RFT1*) combination and NFW (non-functional *HD1*, functional *HD3A*, and weak function *RFT1*) combination indicated the delayed days to heading in short-day condition. We can conclude that these three genes (Hd1, Hd3A and RFT1) involved as major gene in heading date variation of Myanmar landraces and involvement of non-functional allele of one of those three genes causes delayed heading in short-day environment.

Table 2-7. Combinations of short-day promoter genes (*Hd1, Hd3A and RFT1)* and their association with heading date

Combinations			Sub-species			Distribu-tion%	Occurrence
Hd1	*Hd3A*	*RFT1*	*IND*	*JAP*	*ARO*		
F	F	F	7	3	1	26.2	Myanmar, China, Japan, Indonesia, India, Bangladesh, Philippines
F	F	W	2	0	0	4.8	Myanmar, Japan
F	N	F	4	0	0	9.5	Myanmar, Sri-Lanka
F	N	W	1	0	0	2.4	Myanmar, China, Indonesia, South-korea, Laos, Italy, Pakistan, Malaysia
N	F	F	10	11	0	50	Myanmar, India, Thailand, Philippines, South-korea
N	F	W	0	1	0	2.4	Myanmar
N	N	F	1	0	0	2.4	Myanmar
N	N	W	1	0	0	4.9	Myanmar

F- functional haplotype, N - non-functional haplotype, W- weak function haplotype of heading date genes

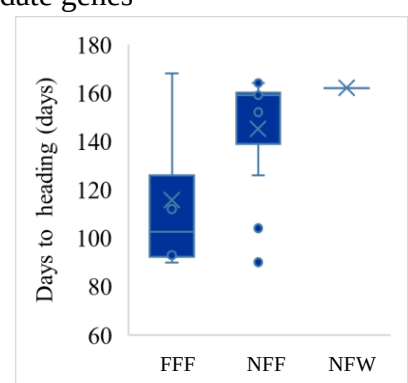


Figure 2-9. Combination of three short-day promoter genes (*Hd1, Hd3A, RFT1* associated with days to heading in SD

Polymorphism and nucleotide diversity of genes associated with yield related genes

Seven yield related genes (*SD1*, *MOC1*, *IPA1*, *SP1*, *DEP3*, *PHD1*, *ROC5* and *AGO7*) were selected for nucleotide polymorphism analysis and haplotype analysis. In *SD1*, total length was 1232 bp and total of 19 nucleotide substitutions (1SNP per 65 bp) and 4 indel (1 indel per 308 bp) were observed. Whereas, a total of 7 nucleotide substitution (1SNP per 30 bp) and 2 indels (1SNP per 313 bp) were detected across 626 bp of *MOC1*. In aligned 1624 bp of *IPA1* sequence, a total of 54 SNPs (1 SNP per 30 bp) and 11 indels (indel per 148 bp) were detected. The aligned sequence length of *SP1* was 2508 bp. A total of 132 nucleotide substitutions and 24 indels were detected with the average of 1 per 19 bp and 1 per 105 bp respectively. In the aligned sequence 1213 bp of *DEP3*, a total of 9 SNPs (average 1 SNP per every 137 bp) were detected and there was no indel. *PHD1* sequence was aligned as length of 1297 bp, including 44 SNPs (1 SNP per 29 bp) and 11 indels (1 indel per 100 bp). For entire sequence length 3435 bp of *ROC5* gene, the 23 SNPs and 3 indels were detected with average 1 SNP per 149 bp and 1 indel per 1145 bp. Total 29 SNPs (1 SNP per 117) and 2 indels (1 indel per 1695) were detected in 3390 aligned based pairs of *AGO7*. For yield related genes, the high frequency of nucleotide substitution was detected in *SP1* gene and high indel was in *PHD1* gene. Where *ROC 5* showed the lowest nucleotide substitution.

Table 2-8. Nucleotide diversity for yield component related genes

Gene Symbol	*SD1*	*MOC1*	*IPA1*	*SP1*	*DEP3*	*PHD1*	*ROC5*	*AGO7*
Total length of coding site bp	1232	626	1624	2508	1237	1297	3435	3390
total mutation	23	9	65	156	9	57	26	31
No. of nucleotide substitution	19	7	54	132	9	44	23	29
Frequency of nucleotide substitution per bp	0.015	0.011	0.033	0.053	0.007	0.034	0.007	0.009
No. of Indels	4	2	11	24	0	13	3	2
Frequency of Indels	0.003	0.003	0.007	0.010	0.000	0.010	0.001	0.001
total bp for 1 SNP	65	89	30	19	137	29	149	117
total bp for 1 Indel	308	313	148	105	0	100	1145	1695

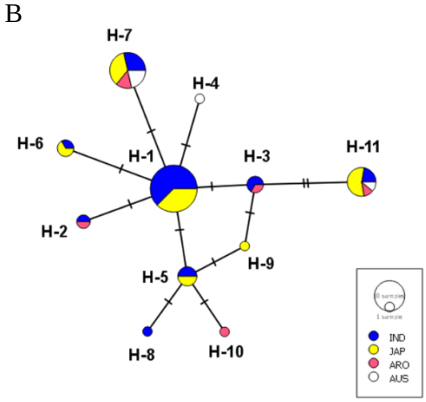
Haplotype analysis of *PHD1* (spikelet per panicle related genes)

The analyzed total 71 landraces contained 11 haplotypes according to the detected 10 SNPs and 3 InDel in coding region of *PHD1*. Haplotypes PHD1–H1 was represented mainly in all accessions (38% of distribution) and occurred in both indica and japonica sub-species. This major haplotype group (PHD1-H1) was widely occurred in China, South-korea, Laos, Sri-Lanka, Japan, Philippines and Malaysia regions (Figure 2-11 A). Rare haplotypes PHD1–H2, PHD1–H6, PHD1–H8, and PHD1–H9 were only detected in Myanmar. Rare haplotypes PHD1-H8 was only detected in indica sub-species whereas PHD1-H9 was occurred only in japonica sub-species (Figure 2-11B). Three indel mutation sites was noted in PHD1 sequence region of 71 landraces, in which 6 bp insertion at 14997854 position, 3 bp del at 14998081 position and 52 insertions at 14998100 positions.

According to phenotypic association of PHD1 haplotype group, PHD1-H2 group showed the lowest spikelet per panicle (40 spikelets). The haplotype group PHD1-H5, PHD1-H8, and PHD1-H9 which contained 6 bp insertion at 1499854 position exhibited the high spikelet per panicle (159, 137, and 109 respectively).

	14993220	14993246	14993394	14993514	14996635	14997670	14997854	14997938	14998012	14998081	14998100	Distribution	Occurrence
REF	T	A	T	T	G	T	G	G	C		G		
H-1	T	A	T	T	G	T	G	G	C		G	38	Myanmar, China, South-korea, Laos, Sri-Lanka, Japan, Philippines, Malaysia
H-2	A	A	T	T	G	T	G	G	C		G	3	Myanmar,
H-3	T	A	T	T	A	T	G	G	C		G	5	Myanmar, Pakistan
H-4	T	A	T	T	G	T	G	G	T		G	2	Bangladesh
H-5	T	A	T	T	G	T	6bp In	G	C		G	6	Myanmar, Thailand
H-6	T	A	T	C	G	T	G	G	C		G	5	Myanmar,
H-7	T	A	TA	T	G	T	G	G	C		G	22	Myanmar, China, Indonesia, India, Philippines
H-8	T	A	T	T	G	T	6bp In	A	C		G	2	Myanmar
H-9	T	A	T	T	A	T	6bp In	G	C	3bp del	G	2	Myanmar
H-10	T	A	T	T	G	T	6bp In	G	C	N	52 bp In	2	Japan
H-11	T	G	T	T	A	A	G	G	C	3bp del	G	14	Myanmar, Bangladesh, India, South-korea

B



C

Hap Group	Spikelet per panicle	
	Range	Mean value
PHD1-H1	70-184	104
PHD1-H2	40	40
PHD1-H3	78-168	123
PHD1-H4	-	-
PHD1-H5	138-173	159
PHD1-H6	69-143	103
PHD1-H7	48-106	80
PHD1-H8	137	137
PHD1-H9	108	108
PHD1-H10	-	-
PHD1-H11	62-128	89

Figure 2-11. Haplotype analysis and network of the *PHD1* gene. (A) The haplotype analysis of the *PHD1* (*Os01g0887700*) gene in 71 germplasms. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non-functional. (B) the haplotype network of six haplotypes of *PHD1* genes. IND-Indica, JAP-Japonica, ARO- Aroma and AUS- (C) The association of grain length trait with *PHD1* haplotype groups.

Polymorphism and nucleotide diversity of grain size related genes

Nucleotide polymorphism of grain genes were examined in the coding sequence of 7 genes, (*GL3.1*, *GL3.2*, *GL 3.2*, *GS3*, *GS5*, *GL7* and *GW5*). *GL3.1* sequence was aligned as length of 3631 bp, including 5 SNPs (1 SNP per 762 bp) and no indel was detected. Total 6 SNPs (1 SNP per 354 bp) and only one indels were detected in 1769 aligned based pairs of *GL3.2*. The aligned sequence length of *GL3.3* was 1980 bp. A total of 21 nucleotide substitutions and 4 indels were detected with the average of 1 per 94 bp and 1 per 495 bp respectively. In *GS3*, total length was 3485 bp and total of 76 nucleotide substitutions (1SNP per 60 bp) and 18 indel (1 indel per 194 bp) were detected. A total of 26 nucleotide substitution (1SNP per 120 bp) and 6 indels (1SNP per 521 bp) were detected across 3128 bp of *RFT1*. In aligned 1413 bp of *GW5* sequence, a total of 5 SNPs (1 SNP per 471 bp) and 2 indels (indel per 471 bp) were detected. Among these seven grain-size related genes, high nucleotide diversity was detected in *GS3* and lowest nucleotide diversity was observed in *GL3*.

Table 2-9. Nucleotide diversity for grain size related genes

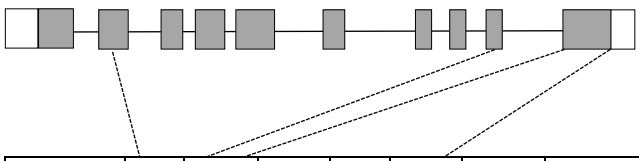
Gene Symbol	*GL3.1*	*GL3.2*	*GL3.3*	*GS3*	*GS5*	*GL7/GW7*	*GW5*
Total length of coding site bp	3631	1769	1980	3485	1822	3128	1413
total mutation	5	6	25	76	5	32	5
No. of nucleotide substitution	5	5	21	58	2	26	3
Frequency of nucleotide substitution per bp	0.001	0.003	0.011	0.017	0.001	0.008	0.002
No. of Indels	0	1	4	18	3	6	2
Frequency of Indels	0.000	0.001	0.002	0.005	0.002	0.002	0.001
total bp for 1 SNP	726	354	94	60	911	120	471
total bp for 1 Indel	0	1769	495	194	607	521	707

Haplotype analysis of *GS5* grain size related genes

Among the seven grain-size related genes, *GS5* showed highly associated with the grain size traits and three haplotype groups were observed in *GS5* sequence comparison of 71 landraces. GS5-Hap1 was present only in *indica* sub-species (1.4% distribution), all of which were previously classified as long-grain germplasm. Major haplotype group was observed in the GS5-H2 group which composed of 72.9% distribution of all genotypes (Figure 2- A).

All indica, japonica, aroma and aus sup-species involved in GS5-H2 group. GS5-Hap3 was occurred only in Myanmar which showed high difference in nucleotide sequence from other genotypes. Four indel mutations were detected in GS5-H3 Haplotype group, including 2 bp del at 3442782 position of exon 2, 6 bp del at 3443491 position, 6 bp insertion at 3443576 and 3 bp insertion at 3443653 position of exon 1 respectively. A comparison of grain length in landraces carrying GS5-H2 and GS5-H3 was highly significant ($p < 0.01$) in *indica* sub-species from GS5-H1 group. The GS5-H1 group had an average grain length of 11.6 mm, while the GS5-H2 group had an average length of 9.9 mm and GS5-H2 group 10 mm respectively. Whereas, the average grain length of GS5-H1 and GS5-H3 in japonica sub-species were the same (9.5 mm) (Fig 2- C). Thus, we concluded that the *GS5* gene underlies as a major gene underlying grain length variation in indica sub-species of Myanmar landraces.

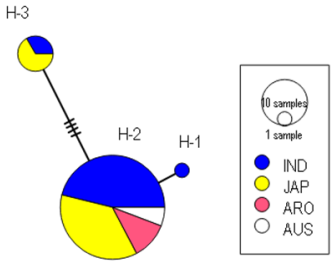
A



GS5 (*Os05g0158500*)

Hap Group	3439806	3442782	3443491	3443576	3443653	Distribution (%)	Occurrence
REF	T	TGG		T	A		
GS5-H1	T	TGG		T	A	1.4	Myanmar
GS5-H2	T	TGG		T	3 bp in	72.9	Myanmar, China, Japan, Indonesia, Bangladesh, Thailand, Philippines, Malaysia, South-Korea
GS5-H3	G	2 bp del	6 bp del	6 bp in	3 bp in	8.6	Myanmar

B



C

Sub-species	Hap Group	Grain length(mm)	Grain width(mm)	Grain length-to-width ratio
Indica	GS5-H1	11.6	3.8	3.1
	GS5-H2	9.9	3.5	2.9
	GS5-H3	10.0	3.9	2.6
Japonica	GS5-H1	-	-	-
	GS5-H2	9.5	4.0	2.4
	GS5-H3	9.5	4.2	2.3

Figure 2-10. Haplotype analysis and network of the *GS5* gene. (A) The gene structure and haplotype analysis of the *GS5* (*Os05g0158500*) gene in 71 germplasms. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of Hd1 genes. IND- *Indica*, JAP-*Japonica*, ARO- *Aroma* and AUS- *aus* (C) The association of grain length trait with *GS5* haplotype groups.

Discussion

Nucleotide polymorphism and haplotype analysis of heading date related genes

The highly diverse patterns of nucleotide polymorphisms were detected in the genes related to heading dates. Nucleotide substitutions of *Hd1* were most frequently observed in coding sequence of with 1 SNP for every 83 bp. In contrary, only one SNP, causing premature stop codon formation, was detected in coding site of *Hd6*. Frequency of indels was highest from *Hd5* with 1 indel per 123 bp. Among these, 2 SNPs and 3 indels were nonsynonymous causing the changes of protein.

Hd1 has a primary function for promoting the heading date through enhancing the expression of *Hd3A* and *RFT1* and mutations in either *Hd1* or *Ehd1* cause delayed flowering under SD [46]. On the contrary, *hd1* mutant alleles promote flowering under long day (LD) conditions, indicating that *Hd1* can either promote or repress flowering, depending on the photoperiod [59]. Previous research suggests that there is a correlation between genetic diversity of long-day suppressor genes and heading date, and that this diversity was critical for the expansion of cultivated rice production to higher latitudes [62]. Genetic diversity of *Hd1* was an important reason for flowering date diversity in cultivated rice. They showed that a loss of function of *Hd1* occurred 8,000 to 10,000 years ago and played an important role in the adaptation of rice crops to higher latitudes and in the development of various

growing strategies [45]. High natural genetic variation is associated to the *Hd1* locus contributed to diversification of flowering time in many varieties and adaptation to northern regions [63]. Several European varieties bear mutant alleles of *hd1* and genetic evidence also showed that such mutations can increase *Ehd1* transcription and promote flowering. Seven non-functional *hd1* haplotypes in 3K rice genomes had been reported and those non-functional allele haplotypes occurred in subtropical and tropical regions, which indicates human selection [64]. In Myanmar landraces, only three types of non-functional haplotypes distributed about 32% of *japonica* and *indica* sub-species. In addition, some non-functional mutation resulted from frame shift mutations or create premature stop codons was identified.

Functional *DTH8(Hd5)* alleles are more sensitive to photoperiod, whereas those with non-functional *DTH8(Hd5)* alleles are less sensitive to photoperiod. A high degree of polymorphism in the *DTH8 (Hd5)* sequences and different types of non-functional alleles were examined in many studied. In this paper, we also identified three different types of non- functional alleles in which one new non-functional allele containing 5 bp insertion (Fig 2-).

Genotyping of the world rice collection revealed that the widely distribution of japonica cultivars harboring non-functional PRR37-2a(*Hd2*) and functional *Ghd7-2* alleles were occurred across northern Asia, and *PRR37* plays a major role in rice adaptation to high-latitude regions by reducing photoperiod sensitivity and causing early flowering (Wei et al., 2009). In our

study, non-functional of *Hd2* was detected about 3% distribution in both *japonica* and *indica* sub-species. *Hd2* sequence of 71 genotypes also showed the non-functional alleles due to the deletion of 2 bp and 8 bp in the coding region of *Hd2*. Yamamoto (2000) indicated that *Hd2* is epistatic to the *Hd6* in LD-dependent flowering pathway and indica subspecies contained non-functional *hd2* and functional *Hd6* whereas japonica sub-species involved functional *Hd2* and non-functional *hd6* [65]. However, in our studied detected that all Myanmar landraces carried non-functional *Hd6* allele.

In long day condition, the *Ghd7* in functional *DTH8* background required for the expression of *Hd1* floral repressor activity [66]. The heading repression in LD by *Hd1* also depends on *DTH8* in functional *Ghd7* background [67]. Gao et al. (2014) showed that different haplotype combinations of the long-day suppressor genes *DTH7*, *DTH8,* and *Ghd7* were closely related to heading date diversity in cultivated rice. The combination of non-functional alleles of *Ghd7*, *Ghd8* and *Hd1* was one of the factors for expansion of rice cultivars to the higher latitudes. In cultivated rice of northern China mainly carried the combination of non- functional allele in *DTH7*, *DTH8*, and *Ghd7* [68]. Zhang (2015) reported that ancestral Asian cultivars possessed a combination of SSF, strong alleles of *Ghd7*, *Ghd8* and functional *Hd1* [47]. In our research finding, six haplotype combinations of *Hd1*, *Hd5* and *Hd2* were detected and only 30% of total genotypes carried the combination of three functional allele of *Hd1*, *Hd5* and *Hd2*. In *japonica* sub-

species, combination of LFL (non-functional *Hd1*, Function *Hd5* and non-functional *Hd2*) associated with earlier heading date.

Expression of *Hd3a* and *RFT1* is induced by *HEADING DATE 1 (Hd1)* and *EARLY HEADING DATE 1 (Ehd1*) and the genetic architecture of flowering in rice depends on a regulatory network controlling expression of two proteins, encoded by *HEADING DATE 3a (Hd3a)* and *RICE FLOWERING LOCUS T 1 (RFT1)* [69]. In our studies, the combination of those three flowering promoter genes (*Hd1*, *RFT1* and *Hd3A*) in Myanmar landraces are determined. Nearly 74% of 46 landraces carry at least one or more weak and non-functional allele of these three genes *Hd1*, *Hd3A* and *RFT1*. We observed that these three genes (*Hd1*, *Hd3A* and *RFT1*) involved as major gene in heading date variation of Myanmar landraces and loss of function of one of those three genes causes delayed heading in short-day environment.

Nucleotide diversity and haplotype analysis of grain size related genes

Rice yield potential is partially affected by grain size and weight and large number of genes regulate the grain size parameters. And larger number of functional and non-functional mutation in the grain size related genes were investigated. *GS3* encoding a heterotrimeric G protein involved in the regulation grain size in rice and the wild-type *GS3* protein produces medium grain. Loss-of-function *GS3* allele results in long grain, whereas truncated forms lacking the C-terminus produce very short grain [70]. A causal C to A

mutation in the second exon of the *GS3* gene which creates a prematurely truncated GS3 protein and enhanced the grain length in rice [71]. In the US rice germplasms, GS3 is a major grain length gene, and qGL7.1 is a novel locus on chr7 that significantly affects grain shape [56].

In addition, *GRAIN WIDTH and WEIGHT2* (*GW2*) gene, which encodes a RING-type E3 ubiquitin ligase, and two alleles of *GW2* gene which distinguished by a single nucleotide polymorphism (SNP) in exon 4 resulted in a premature translational termination of *GW2* protein. This mutation was reported to be responsible for the enhanced grain width and weight [72].

Here, we investigated seven core grain size-related functional genes (*GL3.1*, *GL3.2*, *GL3.3*, *GS3*, *GS5*, *GL7*, and *GW5*) and observed a wide phenotypic variation for five agronomic traits (grain length, grain width and grain length–width ratio,) in 46 Myanmar rice landraces. We investigated that high nucleotide polymorphism of *GS3* in Myanmar landraces which showing total of 76 nucleotide substitutions (1 SNP per 60 bp) and 18 indel (1 indel per 194 bp) mutation. *GS5* showed highly associated with the grain size traits and three haplotype groups were observed in *GS5* sequence comparison of 71 landraces. A comparison of grain length in landraces carrying GS5-H2 and GS5-H3 was highly significant ($p < 0.01$) in indica sub-species from GS5-H1 group. We found that the *GS5* gene underlies as a major gene underlying grain length variation in indica sub-species of Myanmar landraces.

References

1. Das SP, Jasrotia RS, Deb D, Iquebal MA, Jaiswal S, Dey N. Genomic analysis of polycarpellary rice (Oryza sativa L.) through whole genome resequencing. Journal of Plant Biochemistry and Biotechnology. 2020;30(2):364-72. doi: 10.1007/s13562-020-00602-8.
2. Tanaka W, Toriba T, Hirano H-Y. Flower Development in Rice. The Molecular Genetics of Floral Transition and Flower Development. Advances in Botanical Research2014. p. 221-62.
3. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. Nature. 1991;353(6339):31-7. doi: 10.1038/353031a0. PubMed PMID: 1715520.
4. Ali Z, Raza Q, Atif RM, Aslam U, Ajmal M, Chung G. Genetic and Molecular Control of Floral Organ Identity in Cereals. Int J Mol Sci. 2019;20(11). Epub 20190604. doi: 10.3390/ijms20112743. PubMed PMID: 31167420; PubMed Central PMCID: PMCPMC6600504.
5. Callens C, Tucker MR, Zhang DB, Wilson ZA. Dissecting the role of MADS-box genes in monocot floral development and diversity. Journal of Experimental Botany. 2018;69(10):2435-59. doi: 10.1093/jxb/ery086. PubMed PMID: WOS:000431251900002.
6. Nagasawa N, Miyoshi M, Kitano H, Satoh H, Nagato Y. Mutations associated with floral organ number in rice. Planta. 1996;198(4):627-33. PubMed PMID: WOS:A1996UB45100018.
7. Wu F, Shi XW, Lin XL, Liu Y, Chong K, Theissen G, et al. The ABCs of flower development: mutational analysis of AP1/FUL-like genes in rice provides evidence for a homeotic (A)-function in grasses. Plant Journal. 2017;89(2):310-24. doi: 10.1111/tpj.13386. PubMed PMID: WOS:000395810600010.
8. Bommert P, Satoh-Nagasawa N, Jackson D, Hirano HY. Genetics and evolution of inflorescence and flower development in grasses. Plant and Cell Physiology. 2005;46(1):69-78. doi: 10.1093/pcp/pci504. PubMed PMID: WOS:000226807600007.
9. Ciaffi M, Paolacci AR, Tanzarella OA, Porceddu E. Molecular aspects of flower development in grasses. Sexual Plant Reproduction. 2011;24(4):247-82. doi: 10.1007/s00497-011-0175-y. PubMed PMID: WOS:000300109300001.
10. Shen C, Li G, Dreni L, Zhang D. Molecular Control of Carpel Development in the Grass Family. Front Plant Sci. 2021;12:635500. Epub 20210216. doi: 10.3389/fpls.2021.635500. PubMed PMID: 33664762; PubMed Central PMCID: PMCPMC7921308.
11. Yamaguchi T, Lee DY, Miyao A, Hirochika H, An GH, Hirano HY. Functional

diversification of the two C-class MADS box genes OSMADS3 and OSMADS58 in Oryza sativa. Plant Cell. 2006;18(1):15-28. doi: 10.1105/tpc.105.037200. PubMed PMID: WOS:000234300500002.

12. Dreni L, Jacchia S, Fornara F, Fornari M, Ouwerkerk PBF, An GH, et al. The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. Plant Journal. 2007;52(4):690-9. doi: 10.1111/j.1365-313X.2007.03272.x. PubMed PMID: WOS:000250730600007.
13. Dreni L, Pilatone A, Yun DP, Erreni S, Pajoro A, Caporali E, et al. Functional Analysis of All AGAMOUS Subfamily Members in Rice Reveals Their Roles in Reproductive Organ Identity Determination and Meristem Determinacy. Plant Cell. 2011;23(8):2850-63. doi: 10.1105/tpc.111.087007. PubMed PMID: WOS:000295254700008.
14. Jeon JS, Jang S, Lee S, Nam J, Kim C, Lee SH, et al. leafy hull sterile1 is a homeotic mutation in a rice MADS box gene affecting rice flower development. Plant Cell. 2000;12(6):871-84. doi: DOI 10.1105/tpc.12.6.871. PubMed PMID: WOS:000087989600003.
15. Clark SE, Running MP, Meyerowitz EM. Clavata1, a Regulator of Meristem and Flower Development in Arabidopsis. Development. 1993;119(2):397-418. PubMed PMID: WOS:A1993MB89200008.
16. Fletcher JC. The CLV-WUS Stem Cell Signaling Pathway: A Roadmap to Crop Yield Optimization. Plants (Basel). 2018;7(4). Epub 20181019. doi: 10.3390/plants7040087. PubMed PMID: 30347700; PubMed Central PMCID: PMCPMC6313860.
17. Suzaki T, Sato M, Ashikari M, Miyoshi M, Nagato Y, Hirano HY. The gene FLORAL ORGAN NUMBER1 regulates floral meristem size in rice and encodes a leucine-rich repeat receptor kinase orthologous to Arabidopsis CLAVATA1. Development. 2004;131(22):5649-57. doi: 10.1242/dev.01441. PubMed PMID: 15509765.
18. Suzaki T, Toriba T, Fujimoto M, Tsutsumi N, Kitano H, Hirano HY. Conservation and diversification of meristem maintenance mechanism in Oryza sativa: Function of the FLORAL ORGAN NUMBER2 gene. Plant Cell Physiol. 2006;47(12):1591-602. Epub 20061020. doi: 10.1093/pcp/pcl025. PubMed PMID: 17056620.
19. Chu H, Qian Q, Liang W, Yin C, Tan H, Yao X, et al. The floral organ number4 gene encoding a putative ortholog of Arabidopsis CLAVATA3 regulates apical meristem size in rice. Plant Physiol. 2006;142(3):1039-52. Epub 20060929. doi: 10.1104/pp.106.086736. PubMed PMID: 17012407; PubMed Central PMCID: PMCPMC1630730.
20. Xu W, Tao J, Chen M, Dreni L, Luo Z, Hu Y, et al. Interactions between FLORAL ORGAN NUMBER4 and floral homeotic genes in regulating rice flower development. J Exp Bot. 2017;68(3):483-98. doi: 10.1093/jxb/erw459. PubMed PMID: 28204535; PubMed Central PMCID: PMCPMC6055531.

21. Suzaki T, Ohneda M, Toriba T, Yoshida A, Hirano HY. FON2 SPARE1 redundantly regulates floral meristem maintenance with FLORAL ORGAN NUMBER2 in rice. PLoS Genet. 2009;5(10):e1000693. Epub 20091016. doi: 10.1371/journal.pgen.1000693. PubMed PMID: 19834537; PubMed Central PMCID: PMCPMC2752996.
22. Ohmori Y, Tanaka W, Kojima M, Sakakibara H, Hirano HY. WUSCHEL-RELATED HOMEOBOX4 Is Involved in Meristem Maintenance and Is Negatively Regulated by the CLE Gene FCP1 in Rice. Plant Cell. 2013;25(1):229-41. doi: 10.1105/tpc.112.103432. PubMed PMID: WOS:000315572400019.
23. Nagasawa N, Miyoshi M, Sano Y, Satoh H, Hirano H, Sakai H, et al. SUPERWOMAN1 and DROOPING LEAF genes control floral organ identity in rice. Development. 2003;130(4):705-18. doi: 10.1242/dev.00294. PubMed PMID: WOS:000181190500007.
24. Wang KJ, Tang D, Hong LL, Xu WY, Huang J, Li M, et al. DEP and AFO Regulate Reproductive Habit in Rice. Plos Genetics. 2010;6(1). doi: ARTN e1000818 10.1371/journal.pgen.1000818. PubMed PMID: WOS:000274194300026.
25. Toriba T, Harada K, Takamura A, Nakamura H, Ichikawa H, Suzaki T, et al. Molecular characterization the YABBY gene family in Oryza sativa and expression analysis of OsYABBY1. Mol Genet Genomics. 2007;277(5):457-68. doi: 10.1007/s00438-006-0202-0. PubMed PMID: WOS:000245888100001.
26. Ikeda K, Ito M, NagasawaO N, Kyozuka J, Nagato Y. Rice ABERRANT PANICLE ORGANIZATION 1, encoding an F-box protein, regulates meristem fate. Plant Journal. 2007;51(6):1030-40. doi: 10.1111/j.1365-313X.2007.03200.x. PubMed PMID: WOS:000249424200008.
27. Kater MM, Dreni L, Colombo L. Functional conservation of MADS-box factors controlling floral organ identity in rice and Arabidopsis. J Exp Bot. 2006;57(13):3433-44. Epub 20060912. doi: 10.1093/jxb/erl097. PubMed PMID: 16968881.
28. Jin Y, Luo Q, Tong H, Wang A, Cheng Z, Tang J, et al. An AT-hook gene is required for palea formation and floral organ number control in rice. Dev Biol. 2011;359(2):277-88. Epub 20110907. doi: 10.1016/j.ydbio.2011.08.023. PubMed PMID: 21924254.
29. Ren D, Xu Q, Qiu Z, Cui Y, Zhou T, Zeng D, et al. FON4 prevents the multi-floret spikelet in rice. Plant Biotechnol J. 2019;17(6):1007-9. Epub 20190206. doi: 10.1111/pbi.13083. PubMed PMID: 30677211; PubMed Central PMCID: PMCPMC6524161.
30. Moon S, Jung KH, Lee DE, Lee DY, Lee J, An K, et al. The rice FON1 gene controls vegetative and reproductive development by regulating shoot apical meristem size. Mol Cells. 2006;21(1):147-52. PubMed PMID:

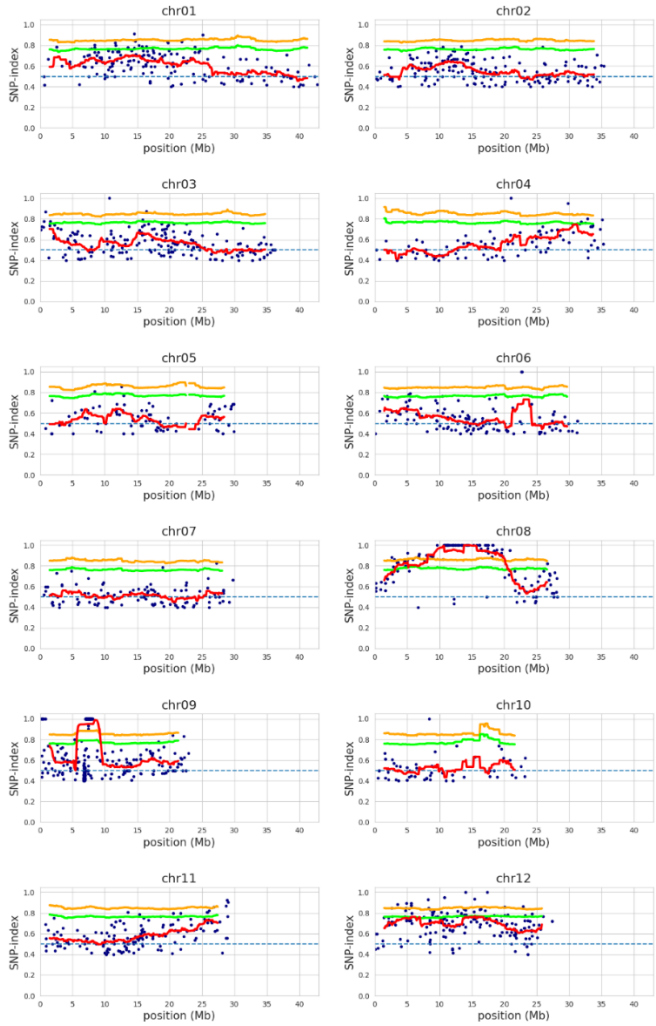
WOS:000235685000018.
31. Jiang L, Qian Q, Mao L, Zhou QY, Zhai WX. Characterization of the rice floral organ number mutant fon3. J Integr Plant Biol. 2005;47(1):100-6. doi: DOI 10.1111/j.1744-7909.2005.00017.x. PubMed PMID: WOS:000227090100012.
32. Zhang XQ, Zou JS, Zhu HT, Li XY, Zeng RZ. [Genetic analysis and gene mapping of an early flowering and multi-ovary mutant in rice (Oryza sativa L.).]. Yi Chuan. 2008;30(10):1349-55. doi: 10.3724/sp.j.1005.2008.01349. PubMed PMID: 18930897.
33. Zhang T, You J, Zeng XQ, Yu GL, Zhang Y, Li YD, et al. Gene mapping and candidate gene analysis of multi-grains 1 (Mg1) in rice. Crop Science. 2020;60(1):238-48. doi: 10.1002/csc2.20100. PubMed PMID: WOS:000525798500019.
34. Li Y, Xu P, Zhang H, Peng H, Zhang Q, Wang X, et al. Characterization and Identification of a Novel Mutant fon(t) on Floral Organ Number and Floral Organ Identity in Rice. Journal of Genetics and Genomics. 2007;34(8):730-7. doi: 10.1016/s1673-8527(07)60082-4.
35. Kariali E, Mohapatra PK. Hormonal regulation of tiller dynamics in differentially-tillering rice cultivars. Plant Growth Regul. 2007;53(3):215-23. doi: 10.1007/s10725-007-9221-z. PubMed PMID: WOS:000250381300008.
36. Schotanus MD, Van Otterloo E. Finding MEMO-Emerging Evidence for MEMO1 ' s Function in Development and Disease. Genes-Basel. 2020;11(11). doi: ARTN 1316
10.3390/genes11111316. PubMed PMID: WOS:000593329500001.
37. Marone R, Hess D, Dankort D, Muller WJ, Hynes NE, Badache A. Memo mediates ErbB2-driven cell motility. Nat Cell Biol. 2004;6(6):515-22. doi: 10.1038/ncb1134. PubMed PMID: WOS:000221759400011.
38. MacDonald G, Nalvarte I, Smirnova T, Vecchi M, Aceto N, Doelemeyer A, et al. Memo Is a Copper-Dependent Redox Protein with an Essential Role in Migration and Metastasis (vol 7, er4, 2014). Sci Signal. 2014;7(333). doi: 10.1126/scisignal.2005661. PubMed PMID: WOS:000338828900005.
39. Mathur J, Hulskamp M. Signal transduction: Rho-like proteins in plants. Curr Biol. 2002;12(15):R526-R8. doi: Pii S0960-9822(02)01029-1
Doi 10.1016/S0960-9822(02)01029-1. PubMed PMID: WOS:000177358100009.
40. Huang J, Liu H, Berberich T, Liu Y, Tao LZ, Liu T. Guanine Nucleotide Exchange Factor 7B (RopGEF7B) is involved in floral organ development in Oryza sativa. Rice (N Y). 2018;11(1):42. Epub 20180730. doi: 10.1186/s12284-018-0235-0. PubMed PMID: 30062598; PubMed Central PMCID: PMCPMC6066601.
41. Ye S, Yang W, Zhai R, Lu Y, Wang J, Zhang X. Mapping and application of the twin-grain1 gene in rice. Planta. 2017;245(4):707-16. Epub 20161220. doi: 10.1007/s00425-016-2627-y. PubMed PMID: 27999987.
42. Gao LZ, Liu YL, Zhang D, Li W, Gao J, Liu Y, et al. Evolution of Oryza

chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019;2. doi: ARTN 278 10.1038/s42003-019-0531-2. PubMed PMID: WOS:000477966700003.
43. Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. Genetic structure and diversity in Oryza sativa L. Genetics. 2005;169(3):1631-8. doi: 10.1534/genetics.104.035642. PubMed PMID: WOS:000228450800036.
44. Izawa T. Adaptation of flowering-time by natural and artificial selection in Arabidopsis and rice. Journal of Experimental Botany. 2007;58(12):3091-7. doi: 10.1093/jxb/erm159. PubMed PMID: WOS:000250677300003.
45. Takahashi Y, Teshima KM, Yokoi S, Innan H, Shimamoto K. Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. P Natl Acad Sci USA. 2009;106(11):4555-60. doi: 10.1073/pnas.0812092106. PubMed PMID: WOS:000264278800089.
46. Doi K, Izawa T, Fuse T, Yamanouchi U, Kubo T, Shimatani Z, et al. Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. Genes Dev. 2004;18(8):926-36. Epub 20040412. doi: 10.1101/gad.1189604. PubMed PMID: 15078816; PubMed Central PMCID: PMCPMC395851.
47. Zhang J, Zhou X, Yan W, Zhang Z, Lu L, Han Z, et al. Combinations of the Ghd7, Ghd8 and Hd1 genes largely define the ecogeographical adaptation and yield potential of cultivated rice. New Phytol. 2015;208(4):1056-66. Epub 20150706. doi: 10.1111/nph.13538. PubMed PMID: 26147403.
48. Lee CM, Park J, Kim B, Seo J, Lee G, Jang S, et al. Influence of Multi-Gene Allele Combinations on Grain Size of Rice and Development of a Regression Equation Model to Predict Grain Parameters. Rice (N Y). 2015;8(1):33. Epub 20151030. doi: 10.1186/s12284-015-0066-1. PubMed PMID: 26519289; PubMed Central PMCID: PMCPMC4627975.
49. Sun S, Wang L, Mao H, Shao L, Li X, Xiao J, et al. A G-protein pathway determines grain size in rice. Nat Commun. 2018;9(1):851. Epub 20180227. doi: 10.1038/s41467-018-03141-y. PubMed PMID: 29487318; PubMed Central PMCID: PMCPMC5829277.
50. Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, et al. The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. Cell Res. 2012;22(12):1666-80. Epub 20121113. doi: 10.1038/cr.2012.151. PubMed PMID: 23147796; PubMed Central PMCID: PMCPMC3515756.
51. Hu Z, Lu SJ, Wang MJ, He H, Sun L, Wang H, et al. A Novel QTL qTGW3 Encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that Interacts with OsARF4 to Negatively Regulate Grain Size and Weight in Rice. Mol Plant. 2018;11(5):736-49. Epub 20180320. doi: 10.1016/j.molp.2018.03.005. PubMed PMID: 29567449.
52. Bai X, Luo L, Yan W, Kovi MR, Zhan W, Xing Y. 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. 2010;11:16. Epub 20100226. doi: 10.1186/1471-2156-11-16. PubMed PMID: 20184774; PubMed Central PMCID: PMCPMC2846863.
53. Shao G, Tang S, Luo J, Jiao G, Wei X, Tang A, et al. Mapping of qGL7-2, a grain length QTL on chromosome 7 of rice. J Genet Genomics. 2010;37(8):523-31. doi: 10.1016/S1673-8527(09)60072-2. PubMed PMID: 20816385.
54. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, et al. OsSPL13 controls grain size in cultivated rice. Nat Genet. 2016;48(4):447-56. Epub 20160307. doi: 10.1038/ng.3518. PubMed PMID: 26950093.
55. Zhou Y, Miao J, Gu H, Peng X, Leburu M, Yuan F, et al. Natural Variations in SLG7 Regulate Grain Shape in Rice. Genetics. 2015;201(4):1591-9. Epub 20151004. doi: 10.1534/genetics.115.181115. PubMed PMID: 26434724; PubMed Central PMCID: PMCPMC4676533.
56. Angira B, Cerioli T, Famoso AN. Discovery and Validation of Grain Shape Loci in US Rice Germplasm Through Haplotype Characterization. Front Genet. 2022;13. doi: ARTN 923078
57. Brinton J, Ramirez-Gonzalez RH, Simmonds J, Wingen L, Orford S, Griffiths S, et al. A haplotype-led approach to increase the precision of wheat breeding. Commun Biol. 2020;3(1). doi: ARTN 712
58. Zong WB, Ren D, Huang MH, Sun KL, Feng JL, Zhao J, et al. Strong photoperiod sensitivity is controlled by cooperation and competition among Hd1, Ghd7 and DTH8 in rice heading. New Phytologist. 2021;229(3):1635-49. doi: 10.1111/nph.16946. PubMed PMID: WOS:000580564300001.
59. Kim SR, Torollo G, Yoon MR, Kwak J, Lee CK, Prahalada GD, et al. Loss-of-Function Alleles of Heading Date 1 (Hd1) Are Associated With Adaptation of Temperate Japonica Rice Plants to the Tropical Region. Frontiers in Plant Science. 2018;9. doi: ARTN 1827
60. Okoshi M, Nishikawa T, Akagi H, Fujimura T. Genetic diversity of cultivated rice (Oryza sativa L.) and wild rice (Oryza rufipogon Griff.) in Asia, especially in Myanmar, as revealed by organelle markers. Genet Resour Crop Ev. 2018;65(3):713-26. doi: 10.1007/s10722-017-0566-5. PubMed PMID: WOS:000425544300002.
61. Shishido R, Akimoto M, Htut T, Sato T, Ishii T. Assessment of genetic diversity and genetic structure of wild rice populations in Myanmar. Breeding Sci. 2019;69(3):471-7. doi: 10.1270/jsbbs.18165. PubMed PMID: WOS:000486384000012.
62. Wei FJ, Tsai YC, Wu HP, Huang LT, Chen YC, Chen YF, et al. Both Hd1 and Ehd1 are important for artificial selection of flowering time in cultivated rice. Plant Sci. 2016;242:187-94. doi: 10.1016/j.plantsci.2015.09.005. PubMed PMID: WOS:000367106100018.

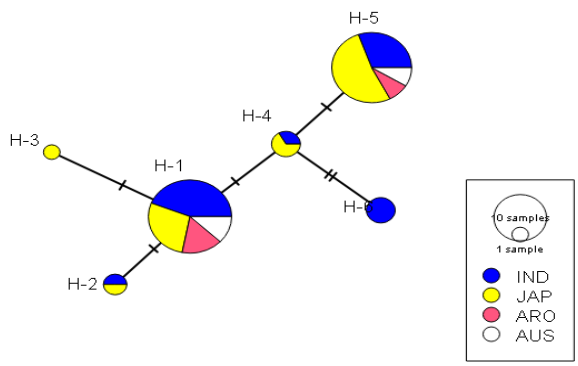
63. Naranjo L, Talon M, Domingo C. Diversity of floral regulatory genes of japonica rice cultivated at northern latitudes. Bmc Genomics. 2014;15. doi: Artn 101 10.1186/1471-2164-15-101. PubMed PMID: WOS:000331117600001.
64. Wu CC, Wei FJ, Chiou WY, Tsai YC, Wu HP, Gotarkar D, et al. Studies of rice Hd1 haplotypes worldwide reveal adaptation of flowering time to different environments. Plos One. 2020;15(9). doi: ARTN e0239028 10.1371/journal.pone.0239028. PubMed PMID: WOS:000573848800044.
65. Yamamoto T, Lin HX, Sasaki T, Yano M. Identification of heading date quantitative trait locus Hd6 and characterization of its epistatic interactions with Hd2 in rice using advanced backcross progeny. Genetics. 2000;154(2):885-91. PubMed PMID: WOS:000085178700034.
66. Nemoto Y, Nonoue Y, Yano M, Izawa T. Hd1,a CONSTANS ortholog in rice, functions as an Ehd1 repressor through interaction with monocot-specific CCT-domain protein Ghd7. Plant Journal. 2016;86(3):221-33. doi: 10.1111/tpj.13168. PubMed PMID: WOS:000378717500002.
67. Du AP, Tian W, Wei MH, Yan W, He H, Zhou D, et al. The DTH8-Hd1 Module Mediates Day-Length-Dependent Regulation of Rice Flowering. Molecular Plant. 2017;10(7):948-61. doi: 10.1016/j.molp.2017.05.006. PubMed PMID: WOS:000404889800006.
68. Gao H, Jin MN, Zheng XM, Chen J, Yuan DY, Xin YY, et al. Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice (vol 111, pg 16337, 2014). P Natl Acad Sci USA. 2014;111(51):18399-. doi: 10.1073/pnas.1422341112. PubMed PMID: WOS:000346767200075.
69. Tsuji H, Taoka K, Shimamoto K. Florigen in rice: complex gene network for florigen transcription, florigen activation complex, and multiple functions. Curr Opin Plant Biol. 2013;16(2):228-35. doi: 10.1016/j.pbi.2013.01.005. PubMed PMID: WOS:000320681100014.
70. Mao HL, Sun SY, Yao JL, Wang CR, Yu SB, Xu CG, et al. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. P Natl Acad Sci USA. 2010;107(45):19579-84. doi: 10.1073/pnas.1014419107. PubMed PMID: WOS:000283997800081.
71. Fan CC, Yu SB, Wang CR, Xing YZ. A causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. Theoretical and Applied Genetics. 2009;118(3):465-72. doi: 10.1007/s00122-008-0913-1. PubMed PMID: WOS:000262577300007.
72. Rasheed H, Fiaz S, Khan MA, Mehmood S, Ullah F, Saeed S, et al. Characterization of functional genes GS3 and GW2 and their effect on the grain size of various landraces of rice (Oryza sativa). Mol Biol Rep. 2022;49(6):5397-403. doi: 10.1007/s11033-022-07119-4. PubMed PMID: WOS:000742320900005.

Appendix



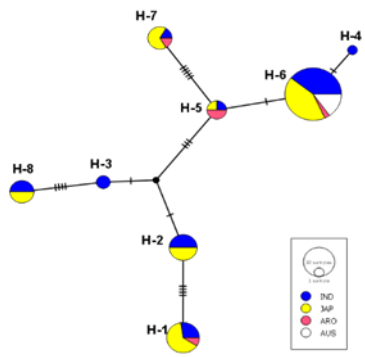
Appendix Figure1-1. Application of MutMap to *fon7* mutants. Single-nucleotide polymorphism (SNP)-index plots of twelve chromosomes generated by the MutMap analysis, showing a genomic region with the highest SNP-index peak harboring the candidate mutation. Green and orange lines indicate 95% and 99% confidence intervals, respectively. Red regression lines were obtained by averaging SNP indices from a moving window of five consecutive SNPs and shifting the window one SNP at a time. Blue dots represent SNP-index values at the SNP position. Y-axis shows SNP-index values ranging from 0–1, and X-axis indicates the SNP position (Mb).

	Pro233Ala	Asp174Val	Ala169Pro	Asp125Tyr	Glu122Gly	Ser80Pro	Distribution (%)	Occurrence
REF	G	T	C	C	G	A		
Ghd7-H1	G	T	C	C	G	A	44	Myanmar, China, Bangladesh, India, Philippines, Malaysia
Ghd7-H2	G	T	C	C	G	G	4	Indonesia
Ghd7-H3	G	T	C	A	G	A	2	Myanmar
Ghd7-H4	G	A	C	C	G	A	5	Myanmar
Ghd7-H5	C	A	C	C	G	A	40	Myanmar, Japan, Bangladesh, Thailand, Italy, South-korea
Ghd7-H6	G	A	G	C	C	A	5	Myanmar



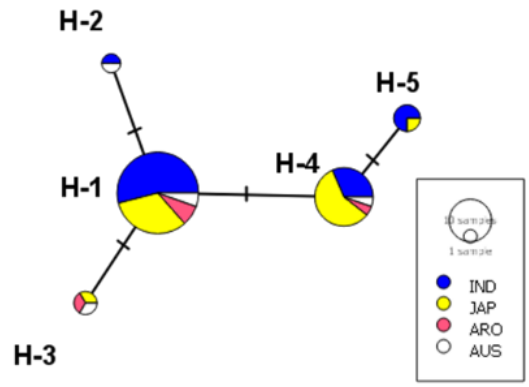
Appendix Figure 2-1. Haplotype analysis and network of the *Ghd7* gene.
(A) The haplotype analysis of the *Ghd7* (*Os07g0261200*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *Ghd7* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus-sup-species.

	Glu313Lys	Lys291Met	Ala265Val	His263Gln	syn	Asp195Asn	syn	Ile152Thr	syn		Gly65Arg	Glu63Val	Leu35Met	Arg33Gln	Glu21Lys	Met19Ile	Ala17Ala	Distribution (%)	Occurrence
REF	C	T	G	G	T	C	C	A	T	T	C	T	A	C	C	C	C		
Ehd1-H1	C	T	G	G	T	C	C	A	T	T	C	T	A	C	C	C	C	15	Myanmar, India, Philippines, Thailand
Ehd1-H2	C	A	G	G	T	C	G	A	T	TA	C	A	A	C	C	C	C	11	Myanmar
Ehd1-H3	C	A	G	G	T	T	G	A	T	TA	T	A	A	C	C	C	C	3	Myanmar
Ehd1-H4	C	A	G	G	T	C	G	A	T	T	T	A	T	T	T	C	T	1	Myanmar
Ehd1-H5	C	A	G	G	T	C	G	A	T	TA	T	A	T	C	T	C	T	6	Myanmar, Laos, Pakistan
Ehd1-H6	C	A	G	G	T	C	G	A	T	TA	T	A	T	T	T	C	T	46	Myanmar, China, Indonesia, Italy, Japan, Philippines, South-Korea, Sri-Lanka
Ehd1-H7	C	A	G	T	T	C	A	G	C	TA	C	A	T	C	T	C	T	8	Myanmar, India, Malaysia, South-Korea
Ehd1-H8	T	A	A	G	C	T	G	A	T	TA	T	A	A	C	C	T	C	8	Myanmar



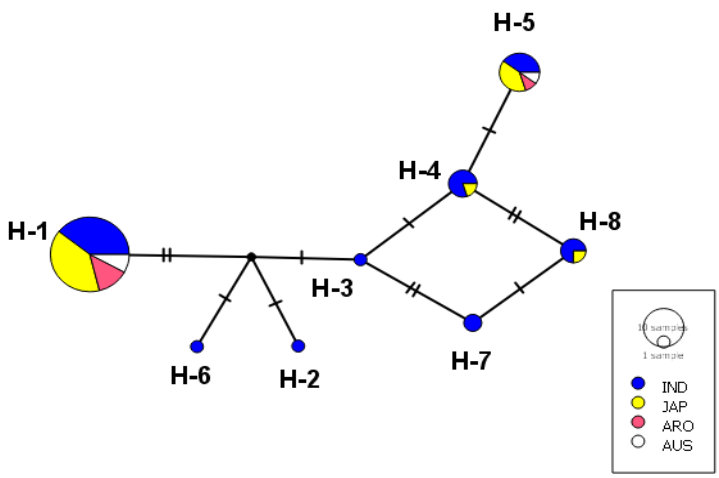
Appendix Figure 2-2. Haplotype analysis and network of the *Ehd1* gene.
(A) The haplotype analysis of the *Ehd1* (*Os10g0463400*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *Ehd1* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus-sup-species.

	38383221	38383286	38385168	38385199	Distribution (%)	Occurrence
REF	C	C	C	G		
SD1-H-1	C	C	C	G	57	Myanmar, China, Indonesia, Bangladesh, India, Philippines, Laos, Sri-Lanka, Pakistan, Malaysis,
SD1-H-2	C	C	A	G	3	Bangladesh,Indonesia
SD1-H-3	C	C	C	A	5	Myanmar, Japan, Bangladesh
SD1-H-4	T	C	C	G	29	Myanmar, Bangladesh, Thailand, Italy, South-korea
SD1-H-5	T	T	C	G	6	Myanmar



Appendix Figure 2-3. Haplotype analysis and network of the *SD1* gene. (A) The haplotype analysis of the *SD1* (*Os01g0883800*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *SD1* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus-sup-species.

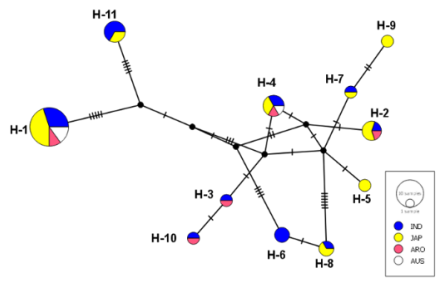
	24312528	24312543	24312552	24312568	24312578	24312654	24312668	24312972	24312990	Distribution (%)	Occurrence
Ref	C	C	A	C		A	A		G		
H-1	C	C	A	C		A	A		G	61	Myanmar, China, Indonesia, Philippines, South-korea, Laos, Sri-Lanka, Pakistan, Japan. Malaysia
H-2	T	C	A	T		G	A		G	2	Myanmar
H-3	T	C	G	T		A	A		G	2	Myanmar
H-4	T	C	G	T	7bp del	A	A		G	8	Myanmar
H-5	T	C	G	T	7bp del	A	A	3bp del	G	16	Myanmar, Japan, Bangladesh, Italy
H-6	T	T	A	T	7bp del	G	A		G	2	China
H-7	T	C	G	T		A	T		C	3	Myanmar
H-8	T	C	G	T	7bp del	A	T		C	6	Myanmar, Thailand



Appendix Figure 2-4. Haplotype analysis and network of the *MOC1* gene. (A) The haplotype analysis of the *MOC1* (*Os06g0610300*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *MOC1* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus-sup-species.

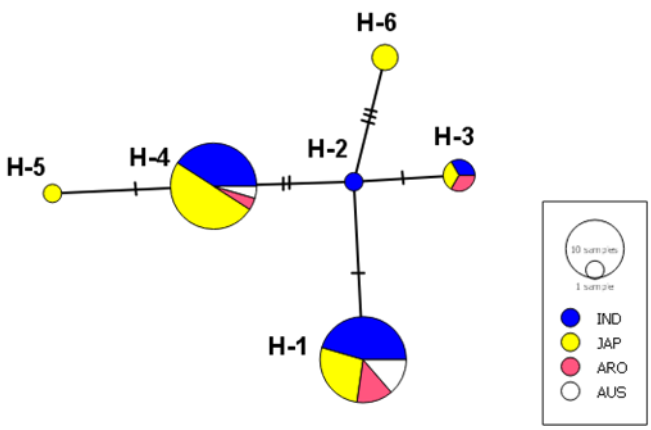
	14012532	14012580	14012648	14016854	14017023	14017070	14017424	14017477	14017499	14017572	14017574	14017636	14017761	14017791	14018179	14018273	14018354	14018591	14018608	14018645	14018685	14018821	14019098	14019119	14019421
REF	A	T	T	C	G	C	C	G	G	T	T	C	G	T	C	C	G	A	G	C	C	C	G	G	
H-1	A	T	T	C	G	C	C	G	G	T	T	C	G	T	C	C	G	A	G	C	C	C	G	G	
H-2	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	A	G	C	C	T	G	G	20 bp del
H-3	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	T	G	T	C	C	G	G	
H-4	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	T	G	C	C	T	G	G	
H-5	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	T	G	C	C	C	5bp	G	20 bp del
H-6	A	G	A	C	A	T	CT	C	A	T	A	T	G	C	C	C	G	A	G	C	C	C	G	G	20 bp del
H-7	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	T	T	C	C	C	G	G	20 bp del
H-8	A	G	A	C	A	T	CT	C	A	T	A	T	G	C	C	C	G	A	G	C	C	C	5bp In	G	20 bp del
H-9	T	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	G	T	T	C	C	C	G	C	20 bp del
H-10	A	G	A	C	G	C	C	C	G	C	A	T	G	C	T	C	A	T	G	T	C	C	G	G	
H-11	A	G	A	CT	G	C	CT	C	G	T	A	T	A	C	T	T	G	A	G	C	T	C	G	G	

	Distri-bution %	Occurrence
H-1	38	Myanmar, China, Indonesia, Bangladesh, India, Phillippines, South-korea, Japan, Malaysis
H-2	9	Myanmar, South-korea
H-3	4	Myanmar, Bangladesh
H-4	11	Myanmar, Bangladesh, Japan
H-5	4	Myanmar
H-6	6	Myanmar
H-7	4	Myanmar, Italy
H-8	4	Myanmar, Thailand
H-9	4	Myanmar
H-10	4	Myanmar
H-11	11	Myanmar



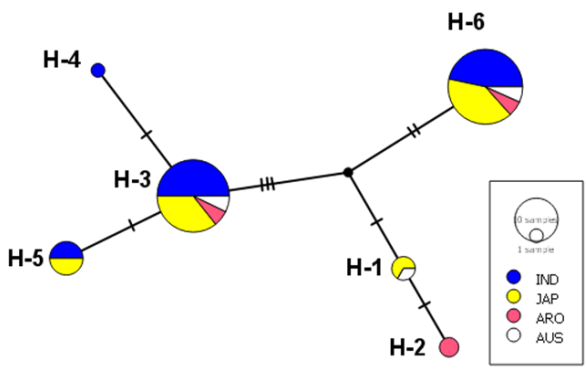
Appendix Figure 2-5. Haplotype analysis and network of the *IPA1* gene. (A) The haplotype analysis of the *IPA1* (*Os08g0509600*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *IPA1* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus-sup-species.

	28109248	28109610	28109645	28109722	28109741	28110428	28110429	28110814	28110885	**Distribution (%)**	**Occurrence**
REF	G	C	A	G	G	G	A	G	A		
H-1	G	C	A	G	G	G	A	G	A	43	Myanmar, China, Indonesia, Bangladesh, India, Philippines,South-korea
H-2	C	C	A	G	G	G	A	G	A	2	Myanmar
H-3	C	C	A	G	G	G	A	C	A	6	Myanmar, Japan
H-4	C	C	T	G	G	G	A	G	C	43	Myanmar, Bangladesh, South-korea
H-5	C	C	T	G	T	G	A	G	C	2	Myanmar
H-6	C	C	A	A	G	T	T	G	A	4	Myanmar, Thailand, Italy



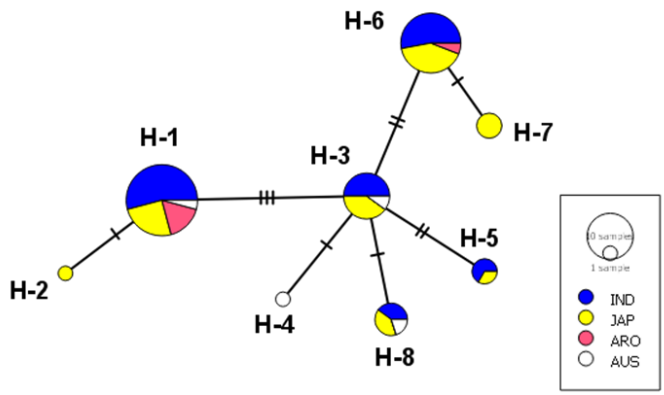
Appendix Figure 2-6. Haplotype analysis and network of the *DEP3* gene (panicle length related gene). (A) The haplotype analysis of the *DEP3* (*Os06g0677000*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *DEP3* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

	27487919	27488020	27489983	27490300	27492401	27493441	27494206	27494808	27494840	**Distribution (%)**	**Occurrence**
REF	A	T	T	T	G	A	T	G	T		
H-1	A	T	T	T	G	A	T	G	T	4	Myanmar, China, Indonesia, Bangladesh (not in Myanmar)
H-2	A	T	T	T	G	A	T	G	A	3	Myanmar,
H-3	T	T	T	C	G	G	C	G	T	40	Myanmar, China, Bangladesh, India, Indonesia, Philippines, South-korea, Laos, Sri-Lanka, Pakistan,Japan,Malaysia
H-4	T	T	T	C	A	G	C	G	T	1	Myanmar,
H-5	T	T	A	C	G	G	C	G	T	9	Myanmar,
H-6	T	C	T	T	G	A	T	A	T	43	Myanmar, Japan, Bangladesh, Thailand, Italy, South-korea



Appendix Figure 2-7. Haplotype analysis and network of the *ROC5* gene (yield related gene). (A) The haplotype analysis of the *ROC5* (*Os02g0674800*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *ROC5* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

	19244185	19244424	19246004	19246770	19246952	19247153	19247415	19247694	19247862	19247956	19248217	Distribution (%)	Occurrence
REF	T	T	C	G	C	T	C	G	G	T	C		
H-1	T	T	C	G	C	T	C	G	G	T	C	38	Myanmar, China, Bangladesh, India, Indonesia, Philippines, South-korea, Sri-Lanka, Pakistan,Japan, Malaysia
H-2	T	T	T	G	C	T	C	G	G	T	C	2	Myanmar,
H-3	T	T	C	G	C	T	C	A	A	C	C	16	Myanmar, Bangladesh, Laos, Italy
H-4	T	T	C	G	C	C	C	A	A	C	C	2	Myanmar,
H-5	T	T	C	G	T	T	C	A	A	C	A	5	Myanmar,
H-6	T	C	C	A	C	T	C	A	A	C	C	27	Myanmar, South-korea
H-7	T	C	C	A	C	T	T	A	A	C	C	5	Myanmar,
H-8	6 bp del	T	C	G	C	T	C	A	A	C	C	8	Myanmar, Bangladesh

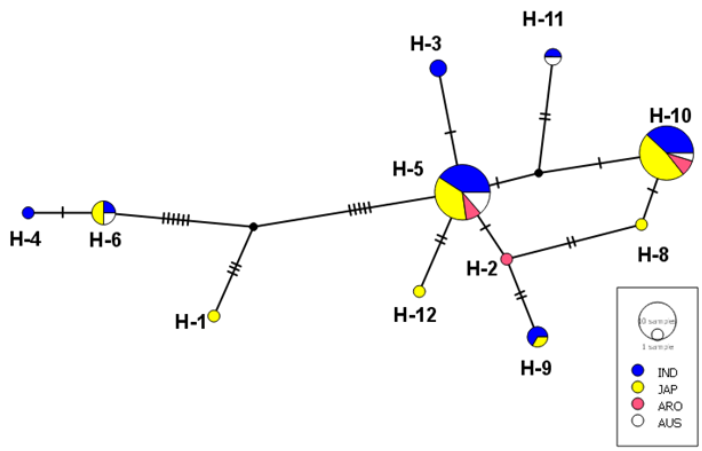


Appendix Figure 2-8. Haplotype analysis and network of the *AGO7* gene (yield related gene). (A) The haplotype analysis of the *AGO7* (*Os03g0449200*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *AGO7* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

Appendix Table 2-1: Haplotype analysis of the *SP1* gene (panicle length related gene) in 71 landraces.

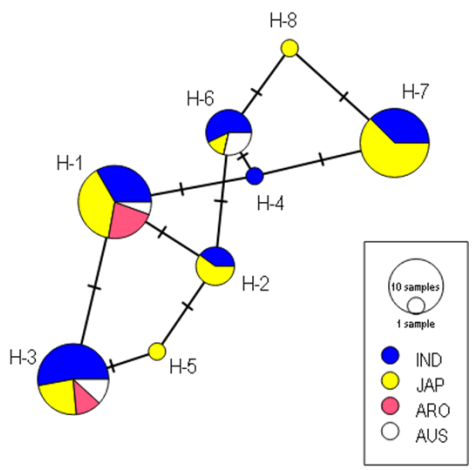
	7193407	7193484	7193503	7193590	7193603	7193611	7193656	7193663	7193683	7193688	7193692	7193906	7193921	7193923	7193929	7194004	7194052	7194133	7194160	7194214	7194301	7194359	7194365	7194368	7194580	7194951	7197002	7197007	7197016	7197023	7197046	7197173	7197180	7197250	Distribution (%)	Occurrence
REF	C		T	A	A	A		C	G	C	G		C	G	C	T	C		T	C	C	G	C	C	C		C	A	A		C	G	C	C		
H-1	C		T	A	A	A		C	G	C	G	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	C	2	Myanmar
H-2	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C		C	A	A		C	G	C	C	2	Pakista
H-3	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T		C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	C	3	Myanmar
H-4	C		T	A	A	A		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	G	C		C	G	A	9bp del	G	A	G	C	2	
H-5	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	C	36	Myanmar, China, Indonesia, Bangladesh, India, Philippines,South-korea, Japan,Malaysia
H-6	C		T	A	A	A		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	G	C	19 bp del	C	G	A	9bp del	G	A	G	C	7	Myanmar, Bangladesh, Italy
H-7	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	C	2	Myanmar
H-8	T	3bp del	C	C	T	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C		C	A	A		C	G	C	T	2	Myanmar
H-9	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	T		T	A	A		C	G	C	C	3	Myanmar
H-10	T	3bp del	C	C	T	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	T	34	Myanmar, Japan, Bangladesh, Thailand, South-korea
H-11	T	3bp del	C	C	A	G	8bp del	-	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	A	C	G	C	C	19 bp del	C	A	A		C	G	C	T	3	Myanmar, Bangladesh
H-12	T	3bp del	C	C	A	G		T	A	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	6bp in		C	G	C	C	2	Myanmar
H-13	T	3bp del	C	C	A	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	T		T	A	A		C	G	C	C	2	Myanmar
H-14	T	3bp del	C	C	T	G		T	G	G	A	7 bp del	A	A	G	G	T	7 bp del	C	C	G	C	G	C	C	19 bp del	C	A	A		C	G	C	T	2	Myanmar

The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides.



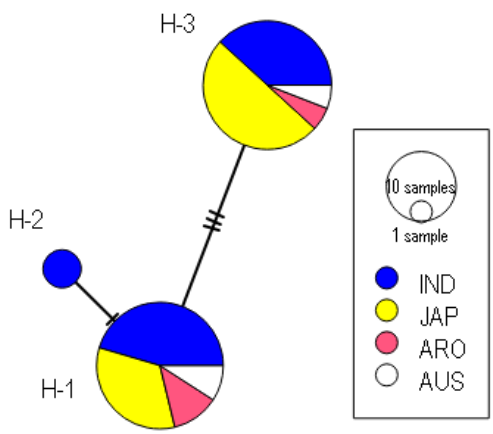
Appendix Figure 2-9. Haplotype network of the *SP1* gene (panicle length related gene). IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

	Ser141Cys	del	stop gain Cys55*	del	Distribution (%)	Occurrence
REF	G		G			
GS3-H1	G		G		25.4	Myanmar, China, Bangladesh, India, Philippines, South-korea, Pakistan, Japan
GS3-H2	G		G	14 bp del	7.0	Myanmar, Laos
GS3-H3	G		T		23.9	Myanmar, China, Japan, Indonesia, Bangladesh, Philippines, Malaysia
GS3-H4	G	3 bp del	G		1.4	Myanmar
GS3-H5	G		T	14 bp del	1.4	Myanmar
GS3-H6	G	3 bp del	G	14 bp del	9.9	Myanmar, Bangladesh
GS3-H7	C	3 bp del	G		22.5	Myanmar, Thailand, South-korea
GS3-H8	C	3 bp del	G	14 bp del	1.4	Italy



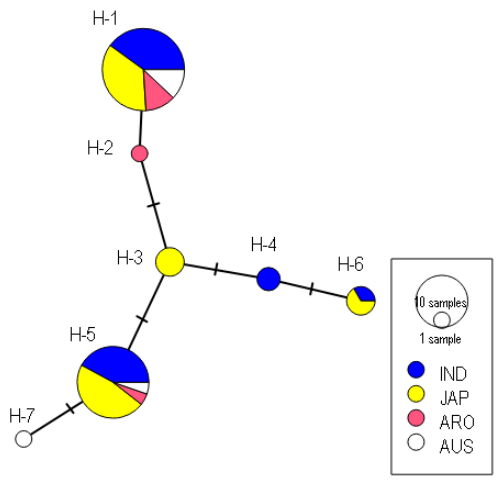
Appendix Figure 2-10. Haplotype network of the *GS3* gene (grain size related gene). (A) The haplotype analysis of the *GS3* (*Os12g0528400*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non- functional. (B) the haplotype network of six haplotypes of *GS3* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

	25043314	25043678	25046030	25049491	Distribution (%)	Occurrence
REF	G	C	A	G		
GL3.1-H1	G	C	A	G	47.9	Myanmar, China, Indonesia, Bangladesh, India, Phillipines, South-korea, Laos, Sri-Lanka, Pakistan, Malaysia,
GL3.1-H2	G	C	A	A	4.2	Myanmar, China
GL3.1-H3	A	T	G	G	47.9	Myanmar, Japan, Bangladesh, Thailand, Italy, South-korea



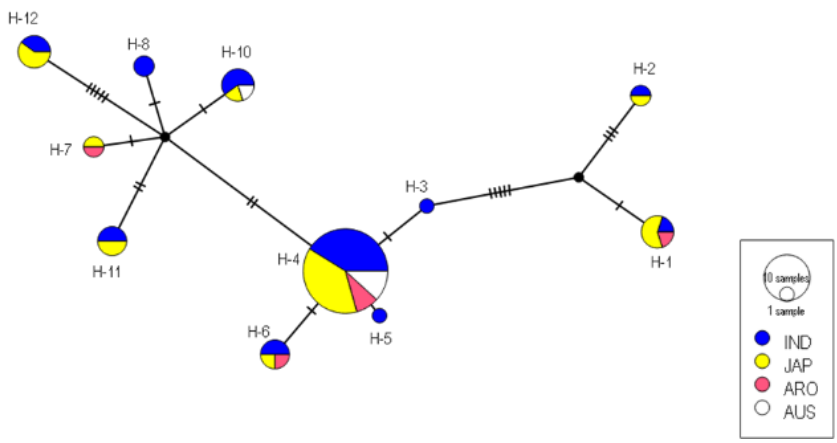
Appendix Figure 2-11. Haplotype network of the *GL3.1* gene (grain size related gene). (A) The haplotype analysis of the *GL3.1* (*Os03g0646900*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *GL3.1* genes. IND- Indica, JAP-Japonica, ARO-Aroma and AUS-Aus sup-species.

	stop loss	His492Asp	Arg27Trp	Gly358Glu	del	Distribution(%)	Occurrence
REF	T	G	C	C	GGGA		
GL3.2- H1	T	G	C	C	GGGA	35.2	Myanmar, China, Indonesia, Bangladesh, India, Indonesia, Philippines, South-korea, Italy, Malaysia
GL3.2-H2	T	C	C	C	GGGA	1.4	Pakistan
GL3.2-H3	T	C	C	T	GGGA	4.2	Myanmar, Laos
GL3.2-H4	T	C	T	T	GGGA	2.8	Myanmar
GL3.2-H5	C	C	C	T	GGGA	26.8	Myanmar, Bangladesh, Thailand
GL3.2-H6	T	C	T	T	3 bp del	4.2	Myanmar
GL3.2-H7	C	C	C	T	3 bp del	1.4	Bangladesh



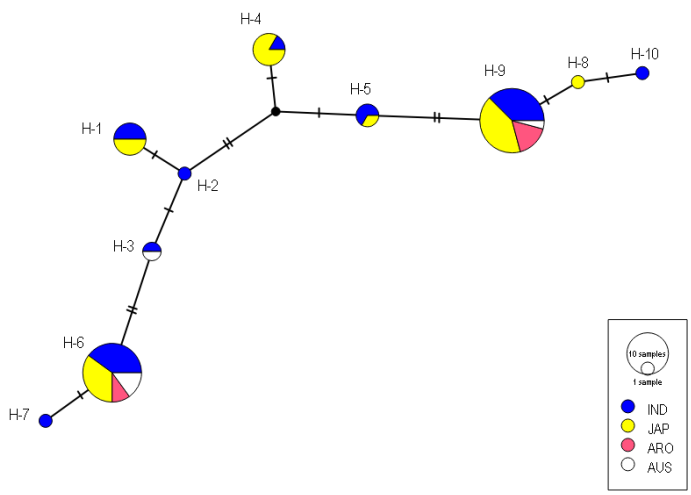
Appendix Figure 2-12. Haplotype network of the *GL3.2* gene (grain size related gene). (A) The haplotype analysis of the *GL3.2* (*Os03g0417700*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *GL3.2* genes. IND- Indica, JAP-Japonica, ARO-Aroma and AUS-Aus sup-species.

	synonymous	synonymous	synonymous	Gly31Val	synonymous	synonymous	*52G>A	*82C	His224Gln	stop gain	synonymous	synonymous	synonymous	synonymous	synonymous	Distribution %	Occurrence
REF	G	G	T	G	C	T	G	C	G	C	A	A	G	G	C		
GL3.3-H1	G	G	T	G	C	T	G	C	G	C	A	A	G	G	C	5.7	Myanmar, India, South-korea
GL3.3-H2	A	G	C	G	C	T	G	C	G	C	T	A	G	G	C	7.1	Myanmar
GL3.3-H3	G	G	T	G	T	T	G	C	G	C	T	A	G	G	C	12.9	Myanmar, China, Bangladesh, India, Indonesia, Philippines, Laos, Pakistan, Malaysia, South-Korea
GL3.3-H4	G	G	T	G	T	T	G	C	G	A	T	A	G	G	C	2.9	China
GL3.3-H5	G	G	T	G	T	T	A	C	G	C	T	A	G	G	C	1.4	Myanmar, Japan
GL3.3-H6	G	G	C	G	C	T	G	C	G	C	T	A	A	G	C	48.6	Myanmar
GL3.3-H7	G	G	C	G	C	A	G	C	G	C	T	A	G	G	C	1.4	Myanmar
GL3.3-H8	G	G	C	G	C	T	G	C	G	C	T	A	A	G	C	5.7	Myanmar
GL3.3-H9	G	G	C	G	C	T	G	C	G	C	T	G	G	G	C	1.4	Myanmar, Bangladesh, Italy
GL3.3-H10	G	G	C	T	C	T	G	C	G	C	T	A	G	G	C	2.9	Myanmar, Indonesia
GL3.3-H11	G	A	C	G	C	T	G	T	T	C	T	A	G	A	T	1.4	Myanmar



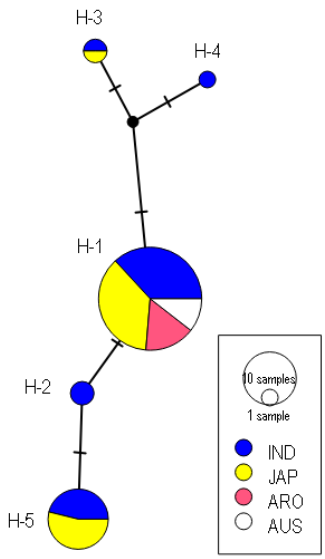
Appendix Figure 2-13. Haplotype network of the *GL3.3* gene (grain size related gene). (A) The haplotype analysis of the *GL3.3* (*Os03g0841800*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides, and the red color show SNP that causes non- functional (B) the haplotype network of six haplotypes of *GL3.3* genes. IND- Indica, JAP-Japonica, ARO- Aroma and AUS-Aus sup-species.

	24666092	24666135	24666398	24666566	24667068	24667175	24667282	24669157	24669200	24669225	24669233	24669274	Distribution (%)	Occurrence
REF	T	G	G	C	G	G	G	A	C	T				
GL7-H1	T	G	G	C	G	G	G	A	C	T			8.5	Myanmar, South-korea, Japan
GL7-H2	T	G	G	C	A	G	G	A	C	T			1.4	Myanmar
GL7-H3	T	G	G	C	A	G	G	A	C	T	11 bp del		2.8	Myanmar, Bangladesh
GL7-H4	T	G	G	C	A	G	G	C	C	C		18 bp del	8.5	Myanmar, South-korea
GL7-H5	T	G	G	C	A	G	G	C	A	T		18 bp del	4.2	Myanmar, Thailand
GL7-H6	C	G	A	C	A	G	G	A	C	T	11 bp del		28.2	Myanmar, China, Indonesia, Bangladesh, India, Phillipines, Malaysia
GL7-H7	C	G	A	C	A	G	A	A	C	T	11 bp del		1.4	Myanmar
GL7-H8	T	C	G	A	A	G	G	C	A	T			1.4	Myanmar
GL7-H9	T	C	G	A	A	G	G	C	A	T		18 bp del	33.8	Myanmar, Japan, Bangladesh, India, Italy, Pakistan
GL7-H10	T	C	G	A	A	A	G	C	A	T			1.4	Myanmar



Appendix Figure 2-14. Haplotype network of the *GL7* gene (grain size related gene). (A) The haplotype analysis of the *GL7* (*Os07g0603300*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *GL7* genes. IND- Indica, JAP-Japonica, ARO-Aroma and AUS-Aus sup-species.

	synonymous	synonymous	del	In	Ala404Gly	Distribution (%)	Occurrence
REF	G	T	TGCG		C		
GW5-H1	G	T	TGCG		C	54	Myanmar, Japan, China, Indonesia, Bangladesh, India, Indonesia, Thailand, Philippines, Malaysia, South-Korea
GW5-H2	A	T	TGCG		C	3	Myanmar
GW5-H3	G	C	3 bp del		C	1	Myanmar
GW5-H4	G	C	TGCG		G	1	Sri-Lanka
GW5-H5	A	T	TGCG	15 bp in	C	18	Myanmar



Appendix Figure 2-15. Haplotype network of the *GW5* gene (grain size related gene). (A) The haplotype analysis of the *GW5* (*Os05g0187500*) gene in 71 landraces. The cream-color cells indicate the original nucleotides and the blue-color cells describe the mutation nucleotides. (B) the haplotype network of six haplotypes of *GW5* genes. IND- Indica, JAP-Japonica, ARO-Aroma and AUS-Aus sup-species.

초록

벼의 화기변이체 유전자 동정 및 미얀마 재래종들의 유전적 다양성 분석

벼는 전 세계 사람들이 섭취하는 칼로리의 50 % 이상을 제공하는 주식 작물로서 매우 중요하다. 벼의 화기는 성공적인 종자 형성과 곡물 발달에 중요하기 때문에 화기의 유전 연구는 매우 중요한 주제 중 하나이다. 본 연구에서는 벼의 돌연변이를 통해 화기 변이체를 탐색하였고, 형태적 특성 조사 및 그 유전자 floral organ number 7 (fon7) 동정 실험을 수행하였다.

사용된 fon7 화기변이체는 자포니카형 품종인 일품벼에 Ethyl Methane Sulfonate (EMS)를 처리해서 확보하였다. 이 fon7 화기변이체에서는 영화 중 37%가 꽃 기관, 특히 수술과 암술의 수가 증가한 것으로 나타났다. 중 특히 자방이 복수로 존재하는 다배 현상이 관찰되어 일대잡종 종자생산에 이용가치가 있을 것으로 판단하였다. 뿐만 아니라, 화기변이체의 출수기는 2 주일 빨랐고, 식물체의 키가 작았으며(short plant height)와 분얼수가 적은(low tiller number) 표현형을 확인하였다. 일품벼와 fon 7 화기변이체 간의 F2 교배 집단은 3:1 분리비로 확인하며 화기변이체 표현형들이 단일 열성 유전자에 의해 조절되고 있음을 확인하였다. 유전자지도에 기초하고 MutMap 방법으로 염기서열을 분석한 결과 8 번 염색체의 8.0~20.0Mb 영역 사이에 후보 유전가 존재함을 알 수 있었다. dCAPs 마커로 분석하여 8 번 염색체에 위치한 유전자 Os08g0299000 에서 frameshift 돌연변이를 유발하는 하나의 염기서열 결손을 발견하였다. 그 유전자의 T-DNA 변이체를 평가한 결과 fon7 변이체와 동일한 화기형질을 보임으로써 Os08g0299000 가 fon7 유전자임을 확증하였다. 유전자 구조를

분석한 결과 fon7 변이체는 유전자의 intron splicing site 에서 G to A 변이가 발생하여 전사시에 6 번째 exon 이 결여됨으로써(exon skipping) 화기형질의 변이를 초래한다고 해석할 수 있었다. 위 결과는 벼의 화기형질 변이체와 관여 유전자를 동정함으로써 향후 일대잡종 종자 생산을 위한 소재 육성시 유용한 유전자와 육종 소재를 제공할 것으로 기대된다.

또한 벼의 출수기와 종자 크기는 수확량에 영향하기에 그 표현형을 개선하는 것은 중요하다. Haplotype-based breeding, haplotype-assisted genomic selection 방법으로 유용한 haplotypes 조합을 사용하면 원하는 품종을 빠르게 개발할 수 있다. 바람직한 haplotypes 을 찾기 위해 미얀마 재래종 46 개 accession 를 사용하여 연구를 수행하였다. 미얀마 재래종의 출수기와 종자 크기가 다양한 것을 확인하였고 출수기 관여 9 개 유전자, 수량 및 초형 관여 8 개 유전자, 그리고 종자 형질 관여 7 개 유전자들의 haplotype 을 분석하였다. 출수기 관여 9 개 유전자마다 다양한 haplotype 이 존재하였고, 그 haplotype 들의 조합에 의해 출수기의 조만성이 결정됨을 확인하였다. 종자형질 관여 7 개의 유전자 중 GS5 의 haplotype 변이는 미얀마 재래종들의 종자길이에 가장 크게 영향함을 분석하였다. 이러한 연구결과들은 미얀마 재래종의 주요 유전자 haplotype 분석을 통해 향후 미얀마 벼 품종 육성 방향을 제시하고 있어서 육종의 효율화에 기여할 것으로 기대된다.

주요단어: 벼, 화기변이체, 출수기, 종자 크기, exon skipping, haplotype
학번: 2019-35644

Acknowledgements

First and foremost, I would like to express my profound gratitude to my supervisor Prof. Hee-Jong Koh for his patient guidance, constructive suggestions and precious support throughout my Ph.D study and research. Besides, I would like to sincere thank Chairman of the examination committee Prof. Suk-Ha Lee, and the members of the examination committee, Prof. Nam-Chon Paek, Prof. Tae-Jin Yang and Prof. Soon-Wook Kwon for their valuable comments, questions and suggestions that enabled me to complete and widen my research.

My special thanks to Dr. Backki Kim for his very supportive guidance, suggestions, moral support, and unconditional valuable help during the study. I could not have reached this far without his great support and teaching. My sincere thanks also go to my Crop Molecular Breeding Laboratory (CMB) family, Dr. Hong Yeol Kim, Miss Kang Mi Kyong, Miss Eun-Byeol Koh, Mr. Zhuo Jin, Mr. Su Jang, Miss Yoon kyung Lee, Miss Ji-Young Nam, Miss Dasol Kim, Mr. Jin Woo Lee, Mr. Joenghwan Seo, and Mr. Seung Young Lee, for their tremendous support and kind assistance.

I would like to acknowledge the Korean Government Scholarship for giving me the chance to study in Korea and gain so many precious experiences.

I am especially grateful to Dr. Aung Htay Naing, Thant Zin Maung, May Zin Win, Khaing Khaing Htun, Su Thet Hnin, Kay Zin Kyaw, and Theint Thandar Latt for hanging around by my side when times get hard and supporting me in their own sweet ways. I also want to deliver a sincere thanks to all of my teachers, relatives, and other close friends for their great support.

Last but not least, I deep thank to my beloved parents (U Maung Maung San, Daw Yi Yi Win, daughter (Su Lei May Maung), dady (Htin Aung Kyaw), sisters (Zarchi Win, Aye Aye Aung), brother (Kyaw Thet Aung) and sister-in-law (Khin Zarni Cho) for their love, countless supporting and encouragement throughout my life.