A DISSERTATION FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

BY

BACKKI KIM

FEBRUARY, 2015

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

UNDER THE DIRECTION OF DR. HEE-JONG KOH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

BY
BACKKI KIM

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE
NOVEMBER, 2014

APPROVED AS QUALIFIED DISSERTATION OF BACKKI KIM
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS
FEBRUARY, 2015

CHAIRMAN _____
Suk-Ha Lee, Ph.D.

VICE-CHAIRMAN _____
Hee-Jong Koh, Ph.D.

MEMBER _____
Byon-Woo Lee, Ph.D.

MEMBER _____
Tae-Jin Yang, Ph.D.

MEMBER _____
Joohyun Lee, Ph.D.

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

BACKKI KIM

GENERAL ABSTRACT

Asian cultivated rice (*Oryza sativa* L.), divided into sub-species, *Oryza sativa. indica* and *Oryza sativa. japonica* has been cultivated in all over the world including Asia. Recent evidence suggests an ancient *indica* and *japonica* divided between 200,000 and 440,000 years ago based on nuclear genome sequence comparisons (Ma and Bennetzen, 2004; Tang et al., 2004; Vitte et al., 2004). During cultivation and domestication, these subspecies have developed the unique morphologies and characteristic agronomic traits. Many phenotypic differences are obvious between *O. sativa indica* and *japonica* and the beneficial traits from each subspecies are useful to develop a new varieties. However, various obstacles such as reproductive barriers, prevent gene exchange or gene flow between two subspecies. Although several studies have tried to explain the differences between *indica* and *japonica* at a certain developmental stage or molecular and

physiological level, data from these studies are quite limited to explain general differences between *indica* and *japonica*. Whole genome analysis to elucidate the differences between *indiica* and *japonica* genome will be useful to explain the genome structure that led to their distinct features.

In this study, we analyzed the nucleotide-level genome structure of Tongil rice which is a high-yielding rice variety derived from a three-way cross between *indica* and *japonica* and compared it to those of the parental varieties. Sequence data were obtained by whole-genome resequencing using the Illumina Hiseq. A total of 17.3 billion reads, 47× genome coverage, were generated for Tongil rice. Three parental varieties of Tongil rice, two *indica* types and one *japonica* type, were also sequenced at approximately 30x genome coverage. *Indica-japonica* genome composition was determined based on SNP data by comparing Tongil with three parental genome sequences using the sliding window approach. Analyses revealed that 91.8% of the Tongil genome originated from the *indica* parents and 7.9% from the *japonica* parent. Copy number of SSR motifs, ORF gene distribution throughout the whole genome, gene ontology (GO) annotation, and yield-related QTLs or genes variations were also comparatively analyzed between Tongil and parents using sequence-based tools. These results indicated that each genetic factor was transferred from parents into Tongil in proportion to the whole-genome composition.

The yellow colored pericarp and embryo lethal mutant was derived from chemical mutagenesis using N-methyl-N-nitrosourea (MNU) on a *japonica* rice cultivar, Hwacheong. In this study, we cloned the gene responsible for

yellow pericarp and lethal embryo phenotype using a map-based approach. Fine mapping revealed that the mutant gene, *yel* was located on the long arm of chromosome 2 and sequencing of candidate genes identified the gene responsible for the yellow pericarp and embryo lethal phenotype, *Os*COP1 (Constitutive Photomorphogenic 1), orthologs of *Arabidopsis* COP1 which encodes a protein of comprising RING-finger, coiled-coil, and WD40 domains. A 706-bp deletion in the first exon including 5′UTR and start codon of *yel* gene, LOC_Os02g53140 was found thus, it is conceivable that the yellow pericarp phenotype may have been the result of loss of function in rice *yel* gene. HPLC-MS analysis of the four candidate standards with extracts of embryo and endosperm of the mutant grain showed that extremely high level of orientin, luteolin-8-*C*-glucoside accumulated in mutant embryo and endosperm and was the major pigment both in embryo and endosperm. In addition, the antioxidant activities and total phenolic contents of embryo and endosperm of the mutant seed were significantly higher than those of wild-type. Fatty acid analysis showed that the content of saturated fatty acid was increased in mutant embryo on the other hand, the content of unsaturated fatty acid was decreased than that of wild-type. These results indicated that *yellow embryo lethal* (*yel*) gene might be involved in the flavonoids biosynthetic pathway and other ingredient characteristics of rice seed.

Keywords: Rice, Tongil Rice, NGS, Genome structure, Yellow pericarp, Embryo lethal Mutant, Orientin, Antioxidant capacity. *Indica-japonica*

Student Number: 2010-30291

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

COP1	Constitutive photomorphogenic 1
NGS	Next generation sequencing
GO	Gene ontology
ORF	Open reading frame
QTL	Quantitative trait loci
SNP	Single nucleotide polymorphism
LD	Linkage disequilibrium
WCV	Wide compatibility variety
MNU	N-methyl-N-nitrosourea
UHPLC	Ultra-high performance liquid chromatography
DPPH	2, 2 – diphenyl – 1 – picrylhydrazyl
FRAP	Ferric reducing antioxidant power
TEAC	Trolox equivalent antioxidant capacity
RNAi	RNA interference

GENERAL INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important staple food and provides over 21 percent of the caloric needs for more than half of the world's population, and also contributes about 30 percent of Asia's caloric supply (IRRI Statistics, 2009), specially up to 76 percent of the caloric intake of the population of southeast Asia (Fitzgerald et al., 2009; Khush, 1997). Asian cultivated rice divided into two sub-species of *O. sativa. indica* and *japonica*. Recent evidences suggest an ancient *indica*/*japonica* diverged between 200,000 and 440,000 years ago based on nuclear genome sequence comparisons and between 86,000 and 200,000 years ago based on chloroplast sequences, respectively prior to the domestication of rice (Ma and Bennetzen, 2004; Tang et al., 2004; Vitte et al., 2004). In terms of morphology and living environments, *indica* varieties exhibit greater plant heights, longer leaves, and heat tolerance but are sensitive to low temperatures thus, *indica* varieties are cultivated at low latitudes and in humid regions on the other hand, *japonica* varieties have lower plant heights and shorter leaves than that of *indica* but are tolerant to low temperatures therefore, *japonica* varieties are more suitable for high latitudes and for the lower latitudes of high altitude cultivation (Dan et al., 2014). Due to the deep genetic structure that evolved during its domestication and adaptation, current *oryza. sativa* cultivars and landraces can be sub-divided in more detail into five genetically differentiated groups: *indica*, aus, aromatic, temperate *japonica*, and tropical *japonica* (Garris et al., 2005).

Since the green revolution, many countries overcame food shortage and have developed high yielding rice varieties however, the productivity and yield of rice are still important factors in rice cultivation because demand from population growth and the needs of biofuels exceeds global food supply (Brown and Funk, 2008). In addition, presently consumptions of functional health promoting rice, such as black and red rice are rapidly growing due to their healthy functional food ingredients (Kim et al., 2008). Phytochemicals such as anthocyanins and flavonoids, which are found in a variety of crops including fruits and vegetables (Abdel-Aal el et al., 2006) and have been recognized as health-promoting functional food ingredients due to their antioxidant activity (Nam et al., 2006), anticancer (Zhao et al., 2004), hypoglycemic and anti-inflammatory effects (Tsuda et al., 2003). Rice also contains a various naturally occurring compounds in bran, thus rice grain is a valuable source to provide health promoting ingredients. To meet the demands deriving from rapid population growth, lack of planting area, worldwide climate change, and the industrial trend for human health or biofortification continuous efforts to increase rice production by using the genetic and genomic improvement technologies will be of great importance (Takeda and Matsuoka, 2008).

Rice, the first genome sequenced crop plant is an excellent model plant that has been comprehensively studied in functional genomics. In addition, rice has a relatively small genome size compared to other crops. Draft genome sequences of two rice subspecies, *O. sativa ssp. japonica* (cv. Nipponbare) and *O. sativa ssp. indica* (cv. 93-11) were reported (Goff et al.,

2002; Yu et al., 2002), and subsequently the International Rice Genome Sequencing Project completed the final sequence of the entire rice genome of Nipponbare (Matsumoto et al., 2005). These achievements have provided us with a vast amount of information on the rice genome and genome structure and also allowed us to perform detailed genetic analysis, functional studies, and association studies (Miura et al., 2011). Using this genome information, researchers have succeeded in isolating and identifying many important genes and QTLs, which related to useful agronomic traits and rice yield potential. In addition, next generation sequencing technologies have allowed genome-wide genetic variation genotyping and understanding the genetic basis of phenotype variation, and have facilitated reduced laborious works and saving time and increased resolution of genetic mapping. With the Nipponbare sequence as a reference, genome re-sequencing of a large number of rice accessions (The 3,000 rice genomes project, 2014) has led to the discovery of millions of SNPs and insertion/deletion sites (indels), enabling genome-wide association studies (GWAS) to identify agronomically important genes in rice (Huang et al., 2012c; Xu et al., 2012).

LITERATURE REVIEW

Development of "Tongil" and Tongil-type rice

Tongil rice is the first rice variety developed from a hybridization between *indica* and *japonica* in South Korea. It was developed from the progeny of IR667, which was derived from three-way cross (IR8//Yukara/Taichung Native 1) and released to the farmers in 1972. In addition, Tongil is the first Korean rice variety introduced semi-dwarf gene (*sd1*) which is well known as the "Green Revolution gene" and contributed to the dramatic increase in rice production in the 1960s and 1970s, specially, in Asia. Genetic basis of the short-stature of Tongil and Tongil-type rice that led the Green Revolution of South Korea was derived from the *sd1* gene of *indica* rice, IR8 and Taichung Native 1 (TN1). Tongil has short-statured in their plant architecture with erect leaves, high yield potential, and tolerance to nitrogen application and lodging.

The development of "Tongil" rice variety raised the milled rice productivity to 5.13 MT/ha, which was about 30 percent higher than that of the leading *japonica* rice variety at that time. Subsequently, 25 Tongil-type varieties were released by 1977, the area of cultivating Tongil-type rice including Tongil rice was rapidly increased. The area reached to 76.2 percent of the total rice farming areas in 1977, leading to the Korean Green Revolution and self-sufficiency in rice (Chung and Heu, 1980). During the

1970s and 1980s, total 40 varieties of Tongil-type rice were developed and cultivated. The productivity of Tongil-type varieties constantly increased to 5.76 MT/ha in 1976. The national average yield of milled rice in farmers' fields dramatically increased to 4.93 MT/ha, in 1977 as compared to *japonica* varieties that yielded 3.37~4.69 MT/ha (NARI, 2012). Total rice production reached 4.67 million MT in 1976, and 5.21 million MT in 1977 and 6.01 million MT in 1978, respectively (NARI, 2012). For the first time, rice production exceeded the demands and achieved self-sufficiency in rice production and the Green Revolution in South Korea. The success of Tongil and Tongil-type rice varieties was not only a new milestone for future rice breeding but also a practical opportunity to utilize *indica* germplasm in temperate region (NARI, 2012).

Next generation sequencing

The advent of the next generation sequencing technology changed the paradigm of genome research and have provided immeasurable opportunity for the genotyping and genomic study and also greatly enhanced rice functional genomics and molecular breeding studies (Gao et al., 2012; Xie et al., 2010). NGS techniques not only increase sequencing throughput but also allow sequencing a large number of samples using a multiplexed sequencing strategy, simultaneously (Craig et al., 2008; Cronn et al., 2008). These recent technical advances facilitated the development of a sequencing-based high-throughput genotyping that has advantages

of time and cost effectiveness, dense marker coverage, high mapping accuracy and resolution, and more comparable genome and genetic maps among mapping populations and organisms (Huang et al., 2009). Massive parallel sequencing technology has proven revolutionary, shifting the paradigm of genomics to address biological questions at a genome-wide scale (Koboldt et al., 2013). Three main NGS platforms, Illumina/Solexa, Roche/454 and ABI/SOLiD sequencing, which are known as high-throughput sequencing, can generate large amounts of sequence data in a single run (Ansorge, 2009).

NGS technology has been widely used in plant breeding area, such as whole genome sequencing of varieties and population re-sequencing, developing new and numerous molecular markers, SNP-based or InDel markers in a number of plant species, and constructing saturated high resolution genetic and physical maps for association study. The whole genome sequence of a crop species contains important information regarding of the origin, evolution, development, and physiology of that species. In addition, the systematic study of genome sequences can allow exploration of the gene sequences of the species and provide a method for identifying the genes that allowed the species to adapt to a specific environment. Whole genome re-sequencing is a useful method for studying genome to have a reference genome sequence such as Arabidopsis and rice. Besides re-sequencing, to obtain the genome sequence of a crop, the whole-genome de novo sequencing method, which involves the first genome sequence from the species which do not have

reference sequence can be used and subsequently can use bioinformatic tools to assemble the sequences and obtain the genomic map for that species (Gao et al., 2012). The new generation sequencing technologies have been provided and readily adapted to high-quality genome researches, and facilitated the sequencing of abundant diverse germplasm resources of crops.

SNP genotyping and application

Advanced genome sequencing technologies provide unprecedented opportunities to characterize individual genomic differences and identify variations. Specifically, whole genome sequencing using next generation sequencing technologies is widely used due to the moderate costs, manageable data amounts and straightforward interpretation of analysis results (Pabinger et al., 2014). For these reasons, next generation sequencing is now available even for single laboratory. Recently, several re-sequencing projects have conducted and provided a great deal of information about rice genomic structure and genetic diversity (Huang et al., 2010; Huang et al., 2012c; Xu et al., 2012). These sequencing projects found millions of polymorphisms along the entire rice genome, including InDels and several millions of SNPs. The recent achievements in rice genomics research is based on establishing a high-throughput genotyping system. Next generation sequencing and array-based SNP detection are two major high-throughput variation detection platforms. This high-

throughput genotyping system is important for gene identification and molecular breeding. Furthermore, various SNP based platforms have become attractive tools for genotyping and the high-throughput genotyping technologies have been developed rapidly. Recently, a high-resolution 44K Affymetrix custom array has been designed for rice SNP genotyping (McCouch et al., 2010; Zhao et al., 2011). These SNP genotyping assays were considered critically important for associating phenotype–genotype in rice (Tung et al., 2010). High-throughput genotyping platforms play important roles in genes and QTLs cloning, and GWAS analysis. By 2012, over 800 genes had been cloned that are responsible for useful agronomic traits such as yield, plant architecture, grain quality, stresses resistances, and nutrient-use efficiency (Chen et al., 2013; Jiang et al., 2012). The traditional method for gene identification is based on genetic mapping of naturally or artificially occurring mutations and the traits for natural variations. Newly developed genomics-based strategies such as re-sequencing have greatly improved the resolution and accuracy of genetic mapping (Wang et al., 2011; Yu et al., 2011). Recently, a method called MutMap was used to rapidly isolate genes by crossing the mutant to the wild-type line followed by whole genome re-sequencing of the bulked DNA from a segregating population (F_2 population) of plants showing the mutant phenotype (Abe et al., 2012). MutMap-Gap approach also reported for isolating genes by whole genome re-sequencing of bulked DNA of mutant F_2 progeny combined with de novo assembly of gap regions in rice (Takagi et al., 2013). With the development of high-

throughput genomics based platforms, GWAS has been more frequently used in rice (Huang et al., 2010; Huang et al., 2012c; Rafalski, 2010; Zhao et al., 2011). A diversity panel which were selected to dissect the genetic diversity of domesticated rice, consisting of 413 *O. sativa* accessions, was recently genotyped with 44K SNPs and used for the GWAS analysis of aluminum tolerance and other complex traits (Famoso et al., 2011). The new SNP-based and genomic based mapping approaches will show advantages in detecting power, resolution, and time compared to marker-assisted map-based cloning or linkage mapping of QTLs.

Indica-japonica hybridization in rice

Asian cultivated rice (*Oryza sativa* L.) was domesticated approximately 7000–8000 years before the present in several regions such as India, the Yangtze River area in China, the southern Himalayas, and coastal swamp habitats in Southeast Asia (Gross and Zhao, 2014). Since the beginning of domestication and followed by its lasting cultivation, Asian cultivated rice has experienced genetic differentiation, adapting to different ecological and environmental conditions under both natural and human selection. Abundant genetic diversity was generated by genetic differentiation in rice, such as *indica* and *japonica* ecotypes (Vaughan et al., 2008), lowland and upland ecotypes, and non-glutinous and glutinous grain quality types (Olsen et al., 2006). The most momentous genetic differentiation in rice was *indica–japonica* differentiation, as a result, Asian cultivated rice evolved

into two major ecotypes. The *indica* rice is mainly found in tropical and subtropical area, with either low latitudes or altitudes, whereas the *japonica* rice is mostly found in temperate regions with high latitudes. The *japonica* rice can be cultivated in high mountainous areas in some rice planting regions of low latitude. In consequence of adaptation to various ecological conditions, *indica* and *japonica* rice has been diverged according to plant morphology, agronomical characteristics, and physiological–biochemical features (Vaughan et al., 2008). Usually, *indica* and *japonica* rice varieties are distantly related in terms of genetic background, therefore, the inter-subspecific hybridization between *indica* and *japonica* rice will result in remarkable genetic recombination and variation. The research of the mechanism of genetic differentiation between *indica* and *japonica* rice provides critical information for better understanding of adaptive evolution in rice under the different environments.

Notably, the hybridization between *indica* and *japonica* rice varieties generate strong hybrid vigor (Khush, 2001). This characteristic enables to utilize heterosis from inter-subspecies hybridization between *indica* and *japonica* by breeders. The magnitude of hybrid vigor depends on the genetic diversity between the two parents of the hybrids. The greater the genetic difference between the parents, the higher the heterosis (Khush, 1996). The inter-subspecific hybridization between *indica* and *japonica* has strong heterosis compared to intra-subspecies hybridization between *japonica* and *japonica* or *indica* and *indica* (Jiang J, 2002). Thus, the *indica*–*japonica* hybrid rice may has great yield potential in rice. The utilization of

inter-subspecific hybridization between *indica* and *japonica* is the most feasible approach for high yielding rice. Hybrid rice between *indica* and *japonica* varieties have a yield increase of approximately 25 to 30%. However, the F_1 hybrids often have a low seed setting rate due to the reproductive isolation of inter-subspecies (Li et al., 1997; Ouyang et al., 2010). Over the past decades, several studies have conducted regarding the fertility or low seed setting rate in the F_2 progenies (Chen et al., 2008a; Mizuta et al., 2010; Song et al., 2005). The reasons for this phenomenon include pollen sterility, embryo sac abortion or incompatibility. Since the interaction of genes which involved in formation of zygote in *indica-japonica* hybridization is complex, reproductive isolation is still a major obstacle to breed inter-subspecific hybrid rice.

Flavonoids in plant

Flavonoids, secondary metabolites widely accumulate in plants and involved in several aspects of development and defense mechanism. They are classified in six major subgroups: chalcones, flavones, flavonols, flavadiols, anthocyasnins, and proanthocyanidins (Winkel-Shirley, 2002) and are low molecular weight compounds composed of a three-ring structure with various substitutions. Most of flavonoids accumulate in all organs and tissues, at different stages of development, and depending on internal plant condition or external environmental conditions. Flavonoids are synthesized in the cytosol and are transported to the vacuole for storage.

They can also be found in cell walls, the nucleus, chloroplasts, and even in the extracellular space (Zhao and Dixon, 2010). They represent the color and flavor of leaves, fruits, flowers, and seeds and also contribute to plant adaptation to environmental conditions such as cold or UV stresses, and pathogen attacks (Hichri et al., 2011). In addition, plant-derived flavonoids are important health promoting sources for human and provides various nutrition. Furthermore, these compounds possess pharmaceutical properties extremely attractive for human health. Over several decades, the genes and enzymes involved in flavonoid biosynthetic pathway have been characterized in several plant species such as maize (*Zea mays*), snapdragon (*Antirrhinum majus*), petunia (*Petunia hybrida*), and arabidopsis (*Arabidopsis thaliana*) (Saito et al., 2013).

Although flavonoids normally accumulate in plants as *O*-glycosylated derivatives, several species, including major cereal crops, predominantly synthesized flavone-*C*-glycosides, which are stable to hydrolysis and biologically active in planta and as dietary components (Brazier-Hicks et al., 2009). *C*-Glycosides are formed in microbes, plants, and insects, where they serve as antibiotics, antioxidants, attractants, and feeding deterrents (Hultin, 2005). Despite the common secondary metabolites in major cereal crops and medicinal plants, *C*-Glycosylation has received little attention. The most abundant *C*-glycosylated natural products in plants are flavonoids, a large group of polyphenolic compounds. Flavonoids normally accumulate as the *O*-linked glycosidic conjugates in the vacuoles of plant tissues however, flavonoids also accumulate as the *C*-glycosides in at least 20 families of

angiosperms (Harborne, 1986). These derivatives are major secondary metabolites in maize, wheat, and rice. In these cereals, *C*-glycosides of the simple flavones apigenin and/or luteolin predominate with conjugation occurring singly or doubly at the *C*-8 and/or *C*-6 position (Brazier-Hicks et al., 2009). Among these, four flavonoid *C*-glycosyl compounds (orientin, isoorientin, vitexin, and isovitexin) were first isolated compounds from bamboo leaves but are also found in numerous plants (Dietrych-Szostak and Oleszek, 1999). However, the effects of flavonoid *C*-glycosyl compounds; orientin (Luteolin-8-*C*-glucoside), isoorientin (Luteolin-6-*C*-glucoside), vitexin (Apigenin-8-*C*-glucoside) and isovitexin (Apigenin-6-*C*-glucoside), were not well-known.

COP1 (constitutive photomorphogenic 1)

The COP1 (constitutive photomorphogenic 1) acts as an E3 ubiquitin ligase at the central of light signaling to repress light signaling by targeting photoreceptors and downstream transcription factors for ubiquitination and degradation and is one of the best characterized locus among COP (Constitutive Photomorphogenic) / DET (De-Etiolated) / FUS (Fusca) locus subsequently cloned for the first time (Deng et al., 1992; Yi and Deng, 2005). COP1 and DET1 were found among the members of a group of genes termed the Constitutive Photomorphogenic / De-Etiolated (COP/DET) (Schwechheimer and Deng, 2000). At the beginning of study, these loci were identified through mutant screening in Arabidopsis (*Arabidopsis*

thaliana) at the seedling stage that display light-grown phenotypes under dark condition (COP/DET) or seeds that accumulate high levels of anthocyanin (FUS). COP1 mediates the degradation of various photomorphogenesis-promoting transcription factors by the ubiquitin-proteasome system, with the other COP/DET/FUS proteins (Holm et al., 2002; Saijo et al., 2003). In plants, the function of COP1 is closely related to the light signaling pathway. COP1 acts as a central repressor in light signal transduction, where it promotes the ubiquitination and degradation of the positive regulators and is itself regulated by various photoreceptors such as phyA, phyB, cry1, cry2, and UVR8 (Yi and Deng, 2005). Besides seedling photomorphogenesis related to light signaling, research in the past few years has expanded the role of COP1 in other pathway. These pathway include flowering (Jang et al., 2008; Liu et al., 2008; Yu et al., 2008), circadian rhythm (Yu et al., 2008), UV-B signaling (Oravecz et al., 2006; Wu et al., 2012), stomatal opening and development (Kang et al., 2009), shade avoidance response (Crocco et al., 2010), crosstalk between light and brassinosteroid signaling (Luo et al., 2010), cold acclimation response (Catala et al., 2011) and light-induced root elongation (Dyachok et al., 2011) in Arabidopsis, and juvenile–adult phase duration change in rice (Tanaka et al., 2011). These studies indicated that COP1 is an important regulator and integrator in light signal transduction.

The COP1 protein has three conserved domains: a RING finger, coiled-coil and seven WD40 domains, in both higher plants and vertebrates. These domains intermediate COP1 with other proteins and its self-dimerization

(Hoecker and Quail, 2001; Suzuki et al., 2002). The various signaling molecules have been reported to directly interact with COP1 and control its activity. In addition, the activity and location of COP1 in plant cells is altered to cytoplasmic and nuclear partitioning according to dark or light conditions. Under dark condition, COP1 targets these transcription factors, including HY5 (Elongated Hypocotyl 5), LAF1 (Long After Far-red Light 1) and HFR1 (Long Hypocotyl in Far-Red) for ubiquitination and degradation, leading to suppression of photomorphogenesis (Jang et al., 2005; Seo et al., 2003). In general, protein ubiquitination requires a specific E3 ubiquitin ligase, which can be a single protein or a protein complex. An E3 typically functions by recruiting ubiquitin-conjugating enzymes (E2s) through a RING-finger motif and the substrate.

In contrast to the typical roles in the photomorphogenesis, COP1 and HY5 are involved in long-wavelength UV-B induced photomorphogenesis. This response is initiated by absorbing UV-B through its internal chromophore tryptophan residues by the UV-B receptor, UVR8 (UV Resistance Locus 8) (Christie et al., 2012; Wu et al., 2012). UVR8 then monomerizes to interact with the UV-B inducible protein, COP1 for downstream signaling (Favory et al., 2009; Huang et al., 2012b). The physical changes of this process include anthocyanin accumulation, and tolerance against damaging UV-B. The loss of function mutant of either COP1 or HY5 showed the decreased activation of UV-B–responsive genes, impaired photomorphogenesis, and defective UV-B acclimation in previous studies (Brown et al., 2005; Oravecz et al., 2006).

CHAPTER I

Genome Analysis of a Cultivar 'Tongil' Rice

ABSTRACT

Tongil (IR667-98-1-2) rice, developed in 1972, is a high-yielding rice variety derived from a three-way cross between *indica* and *japonica*. Tongil contributed to the self-sufficiency of rice in Korea, during a period known as the 'Korean Green Revolution'. In this study, we analyzed the nucleotide-level genome structure of Tongil rice and compared it to those of the parental varieties. Sequence data were obtained by whole-genome resequencing using the Illumina Genome Analyzer. A total of 17.3 billion reads, 47X genome coverage, were generated for Tongil rice. Three parental accessions of Tongil rice, two *indica* types and one *japonica* type, were also sequenced at approximately 30x genome coverage. A total of 2,149,991 SNPs were detected between Tongil and Nipponbare; the average SNP frequency of Tongil was 5.77 per kb. Genome composition was determined based on SNP data by comparing Tongil with three parental genome sequences using the sliding window approach. Analyses revealed that 91.8% of the Tongil genome originated from the *indica* parents and 7.9% from the *japonica* parent. Copy number of SSR motifs, ORF gene distribution throughout the whole genome, gene ontology (GO) annotation, and some yield-related QTLs or genes variation were also comparatively analyzed

between Tongil and parents using sequence-based tools. Each genetic factor was transferred from the parents into Tongil rice in amounts that were in proportion to the whole genome composition and these analyses support the finding that the Tongil genome is primarily made up of the *indica* genome.

Keywords: Tongil rice, three-way cross, next-generation sequencing, SEG map*, indica*/*japonica* hybridization

INTRODUCTION

Rice (*Oryza sativa* L.) is a staple food for more than half of the world's population, providing about 19 percent of the world's and 29 percent of Asia's caloric supply (IRRI, 2009). Although demands on the nutritional and industrial functionality of rice are increasing, especially to improve human health and quality of life, improving the yield potential of rice is still a major challenge for rice breeders, who must address the rapid growth of the world population along with dramatic reductions in the amount of cultivated land (Khush, 1999), as well as environmental challenges(Nelson and International Food Policy Research Institute., 2009). Asian varieties of cultivated rice include two major subspecies, *O. sativa indica* and *O. s. japonica,* which are differentiated based on morphological and physiological characteristics and geographical distribution (Morishima and Oka, 1981; Sano and Morishima, 1992). *Oryza sativa. indica* cultivars have higher genetic diversity (Lu et al., 2002), a broader cultivation range, and stronger resistance to prominent diseases and insect pests compared to *Oryza sativa. japonica* cultivars (Chung and Heu, 1991). Inter-subspecific hybridization between *indica* and *japonica* rice cultivars may enrich allelic variation and facilitate hybrid vigor by creating new genetic recombination (Cheng et al., 2007). In spite of these advantages, the introduction of desirable *indica* traits into the *japonica* variety has not been successful due to reproductive barriers and incorporation of undesirable characteristics, such as low eating quality for people who prefer the taste of *japonica* rice (Chung and Heu, 1991).

Tongil rice (IR667-98-1-2) is the first semi-dwarf variety obtained by a three-way cross of *indica*/*japonica* varieties as part of a collaborative research project between the International Rice Research Institute (IRRI) and the government of South Korea. The development of Tongil rice resulted in a significant yield increase from 4 to 5 t ha^{-1}, corresponding to a 30% yield increase relative to the leading *japonica* varieties grown in Korea (Chung and Heu, 1980). After the introduction of Tongil rice in 1972, Korean rice production significantly increased and the South Korean government announced the achievement of rice self-sufficiency (the so-called 'Green Revolution') in 1977. However, the genome characterization and structure of Tongil rice have never been analyzed.

Rice is a useful model crop for studying genome structure due to its relatively small genome. Furthermore, its genetic and physical data have been extensively analyzed by the International Rice Genome Sequencing Project (IRGSP) (Matsumoto et al., 2005). The recent improvement of next-generation sequencing (NGS) technology has enabled high-throughput genotyping and elucidation of genome structures of various rice cultivars (Huang et al., 2009; Huang et al., 2012a). Most sequence-based rice genome analyses are based on DNA polymorphisms, single nucleotide polymorphisms (SNPs) and insertion-deletions (InDels). SNP detection is the first step for comparing DNA variation and is an effective tool to elucidate genome structure and composition (Chen et al., 2014; Feltus et al., 2004; McNally et al., 2009; Shen et al., 2010).

In this study, we sequenced the whole genomes of Tongil rice (*Oryza*

sativa L.) and its parental varieties to analyze the genome structure of Tongil in detail and to identify regions of the *indica* and *japonica* parental genomes that introgressed in the Tongil genome. In addition, we analyzed previously reported yield-related genes (*Gn1a*, *Ghd7*, *sd1*, *GS3* and *qSW5*), SSRs, GO annotation, and other genetic characteristics of the Tongil genome.

MATERIALS AND METHODS

Plant Materials

Plant lines subjected to whole-genome resequencing in the present study included Tongil (SNU accession no. 260697) and its parental lines: Yukara, an early maturing temperate *japonica* cultivar (RDA-Genebank Information Center accession no. IT004665); Taichung native 1 (TN1), the first semi-dwarf *indica* variety with high adaptability (RDA-Genebank Information Center accession no. IT004120); and IR8, an improved high-yielding semi-dwarf variety developed at the International Rice Research Institute (IRRI, IRTP 195). The Tongil variety was developed through a three-way cross, IR8//Yukara/TN1. With generation advancement after the cross, the most promising line, IR667-98-1-2, was selected and released to farmers in Korea under the name 'Tongil' (Chung and Heu, 1991).

Whole- genome DNA sequencing

Four rice varieties were sequenced: Tongil and its parental varieties, Yukara, IR8, and TN1. Whole-genome shotgun sequencing of the four rice genomes was performed using the Illumina/Solexa GAII system. DNA sequencing, including construction of shotgun DNA libraries, was performed according to the methods recommended by the manufacturer (Illumina, San Diego, CA, USA). Briefly, whole-genome DNA shotgun paired-end sequencing libraries were generated by fragmentation of DNA

into 500-bp segments using a Covaris DNA shearing machine (Covaris, CA, US), followed by ligation of paired-end adapters ligation of 53 and 68 bp for sequencing on the FlowCell, size selection of the adapter-ligated fragments within the desired size range (500–600 bp), and PCR enrichment using complete primer constructs required for binding and clustering on the FlowCell. Illumina GAII sequencing was performed by identifying the emission color of single-base extensions on the FlowCell.

DNA variation

Illumina whole-genome shotgun 100-bp paired-end DNA sequencing data were filtered to obtain high-quality sequence data and to map reads to the Nipponbare reference genome sequence, which as downloaded from NCBI. Briefly, high-quality sequence with at least QC20-justified phred quality score was mapped to the reference Nipponbare sequence using CLC NGS Cell software (http://www.clcbio.com). The DNA sequence variation DB was converted to text format, including DNA variation based on the reference position, for the analysis of genome structure.

SNP calling – probabilities

Genotype calling to identify regions originating from the *japonica* and *indica* genomes was performed using the sliding-window approach suggested by Huang et al. (Huang et al., 2009). In each window, the proportion of SNPs originating from each parent was examined for

genotype calling. Huang et al. determined optimum window size by calculating the probability of finding a specific number of *japonica* SNPs in a window based on SNP error rates. Recent improvements in sequencing technology, however, resulted in fewer errors in SNP identification. Thus, the method suggested by Huang et al. (Huang et al., 2009) was not directly applicable in this study. Even with a window size of 2, for example, calling accuracy could reach 99.99%. Instead of calculating this probability, the optimum window size was determined iteratively by comparing the portion of *japonica* SNPs (*O*) and the portion of the genome originating from *japonica* (*P*). Tongil was resequenced to obtain SNPs originating from its parents and to calculate the percentage of *japonica* SNPs in each chromosome. SEG-Map software (Huang et al., 2009) was also used for genotype calling on each chromosome. Because the optimum window size was unknown, a range of window sizes from 1 to 199 was used. Then, the Nash-Sutcliffe efficiency (*E*) between *O* and *P* was calculated as follows:

$$E = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - O_m)^2},$$

Here, an individual chromosome is denoted by *i*. The average percentage of *japonica* SNPs on each chromosome is denoted by O_m. The optimal window size was defined as that with a maximum value of *E*; values of *E* ranged from -29 to 0.963. This maximum value of E occurred with a window size of 9. The percentage of *indica* SNPs was at its second highest (0.966) with a window size of 9. At a window size of 10, the *E* value dropped

rapidly for *japonica* SNPs (0.037) and *indica* SNPs (-0.018). Thus, a window size of 9 was selected as the optimum for data analysis (Figure 1-3).

Parental genome composition of Tongil

We compared DNA variation between the parental and Tongil genomes. Genomic regions originating from the *japonica* (Yukara) and *indica* (TN1 or IR8) parents were identified by comparing the Tongil genome sequence to parental sequences. Estimated *indica* and *japonica* regions in the Tongil genome sequence were calculated based on the methods of Zhao et al (Zhao et al., 2010a).

Gene ontology and classification

Annotated Nipponbare reference genes were classified based on parental origin in the Tongil genome and assigned to the three main GO-term categories (cellular component, molecular function, and biological process) using BLAST2GO software (www.blast2go.com) (Conesa et al., 2005).

Simple sequence repeats (SSRs)

SSR loci were searched using SSR search software (Initiative, 2000) and classified with respect to their parental origin.

Accession codes

Raw sequence data obtained in our study have been submitted to the NCBI Short Read Archive with the following accession numbers: Tongil [SRA: SRR923809, SRA: SRR923810], IR8 [SRA: SRR921498], TN1 [SRA: SRR921505], and Yukara [SRA: SRR925387].

RESULTS

Genome structure of Tongil

The whole genomes of Tongil and its three parental varieties, Yukara, IR8, and TN1 (Taichung Native 1) (Figure 1-1), were sequenced on the Illumina-GAII platform. A large number of short reads were mapped onto the reference Nipponbare genome and then assembled into a consensus sequence. A total of 199,543,820 reads of the Tongil genome, corresponding to 17,339,883,560 bp (17.3 Gb), were generated, representing a 47-fold sequence depth and covering 88.8% of the Nipponbare pseudomolecules (Table 1-1 and Table 1-2). We detected a total of 2,149,991 SNPs between Tongil and Nipponbare sequences. The two *indica* parents of Tongil, IR8 and TN1, had 6.22 and 6.04 SNPs per kb, respectively, whereas the *japonica* parent of Tongil, Yukara, had only 0.49 SNP per kb (Table 1-3). Using the SNP data sets from Tongil and its parents, we defined the genomic origins of regions of the Tongil genome by SNP calling (Figure 1-3 and Table 1-4 and Table 1-5; see also the SNP calling section in the Materials and Methods), and then performed a SEG-Map analysis (Zhao et al., 2010b) of Tongil (Figure 1-2). The whole genome of Tongil consisted of an average contribution of 91.8% from *indica*, 7.9% from *japonica*, and 0.3% unknown (i.e., not defined as *indica* or *japonica* regions) (Figure 1-2 and Table 1-4). The contribution of *indica* to the Tongil genome varied across chromosomes, from 74% (Chr. 2) to 100% (Chr. 12). A relatively high proportion of the *japonica* genome was found on chromosomes 1, 2, and 3, whereas the *japonica* sequences were barely

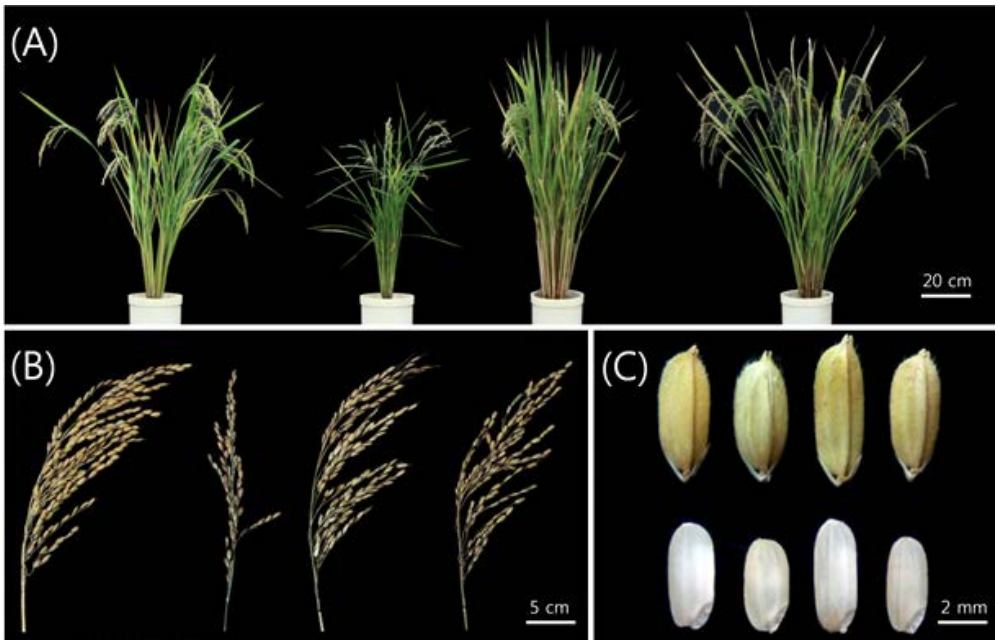


Figure 1-1. Morphological comparison of Tongil and its parental lines. From left to right: Tongil, Yukara, IR8, and TN1. (A) The plant architecture of Tongil, its *japonica* parent (Yukara), and its *indica* parent (IR8 and TN1). (B) The panicle phenotype of Tongil and its parents. (C) The brown rice shape and grain shape of Tongil and its parents. Scale bars are included in each panel.

Table 1-1. General sequencing statistics for Tongil and its parental genomes. Sequencing and mapping against Nipponbare reference genome.

Variety	Number of Reads	Total Read Length (bp)	Mapped Read Length (bp)	Sequencing Depth (×)	Coverage [a)] (%)	SNP Frequency (SNPs/kb)
Tongil	199,543,820	17,339,883,560	330,933,489	47	88.8	5.77
Yukara	114,615,268	12,429,060,750	345,058,384	34	92.6	0.49
IR8	109,304,614	11,790,909,253	327,065,806	32	87.7	6.22
TN1	105,708,026	11,299,286,038	326,132,058	30	87.5	6.04

[a)] Coverage to Nipponbare genome sequence

Table 1-2. Mapping coverage of Tongil rice and its three parents

Chromosome	Reference	Tongil		Yukara		IR8		TN1	
	Pseudomolecule	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)
1	45,038,604	38,574,603	89.2	39,521,561	91.4	37,917,584	87.7	37,633,042	87.0
2	36,792,247	32,784,930	91.2	33,736,987	93.9	32,281,381	89.8	32,249,409	89.8
3	37,312,367	33,412,277	91.9	34,084,555	93.7	33,168,624	91.2	33,108,547	91.0
4	36,060,865	31,735,280	89.3	33,210,463	93.5	31,468,452	88.6	31,193,791	87.8
5	30,073,438	27,991,645	93.5	28,624,984	95.6	27,856,337	93.0	27,770,415	92.7
6	32,124,789	27,784,237	89.0	28,323,987	90.7	27,516,550	88.1	27,400,223	87.7
7	30,357,780	26,328,583	88.6	27,297,454	91.8	25,826,093	86.9	25,896,533	87.1
8	28,530,027	25,808,735	90.7	27,150,402	95.5	25,591,577	90.0	25,618,079	90.1
9	23,895,721	19,770,440	85.9	21,074,657	91.6	19,774,924	86.0	19,460,581	84.6
10	23,703,430	20,191,910	88.0	21,233,194	92.5	19,872,700	86.6	19,739,332	86.0
11	31,219,694	22,752,159	78.7	24,726,750	85.5	22,345,311	77.3	22,512,524	77.8
12	27,679,166	23,798,690	86.4	26,073,390	94.7	23,446,273	85.2	23,549,582	85.5
Total or Ave.	382,788,128	330,933,489	88.8	345,058,384	92.6	327,065,806	87.7	326,132,058	87.5

Table 1-3. SNPs and SNP frequency of Tongil and its three parents

	Reference	Tongil		Yukara		IR8		TN1	
Chromosome	Pseudomolecule	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)
1	45,038,604	222,438	5.14	19,609	0.45	267,113	6.18	262,349	6.07
2	36,792,247	159,177	4.43	7,129	0.20	225,762	6.28	209,537	5.83
3	37,312,367	183,617	5.05	6,785	0.19	205,566	5.65	192,339	5.29
4	36,060,865	177,722	5.00	24,186	0.68	190,976	5.38	182,173	5.13
5	30,073,438	132,916	4.44	3,276	0.11	141,789	4.73	145,396	4.85
6	32,124,789	184,344	5.90	23,105	0.74	187,503	6.00	184,005	5.89
7	30,357,780	198,051	6.66	21,460	0.72	205,384	6.91	191,745	6.45
8	28,530,027	162,946	5.73	15,459	0.54	163,856	5.76	162,534	5.71
9	23,895,721	154,565	6.72	11,273	0.49	143,014	6.22	150,864	6.56
10	23,703,430	146,848	6.40	6,749	0.29	163,379	7.12	165,561	7.22
11	31,219,694	226,457	7.83	28,203	0.98	227,571	7.87	216,814	7.50
12	27,679,166	200,910	7.30	17,131	0.62	198,404	7.21	190,243	6.91
Total or Ave.	382,788,128	2,149,991	5.77	184,365	0.49	2,320,317	6.22	2,253,560	6.04

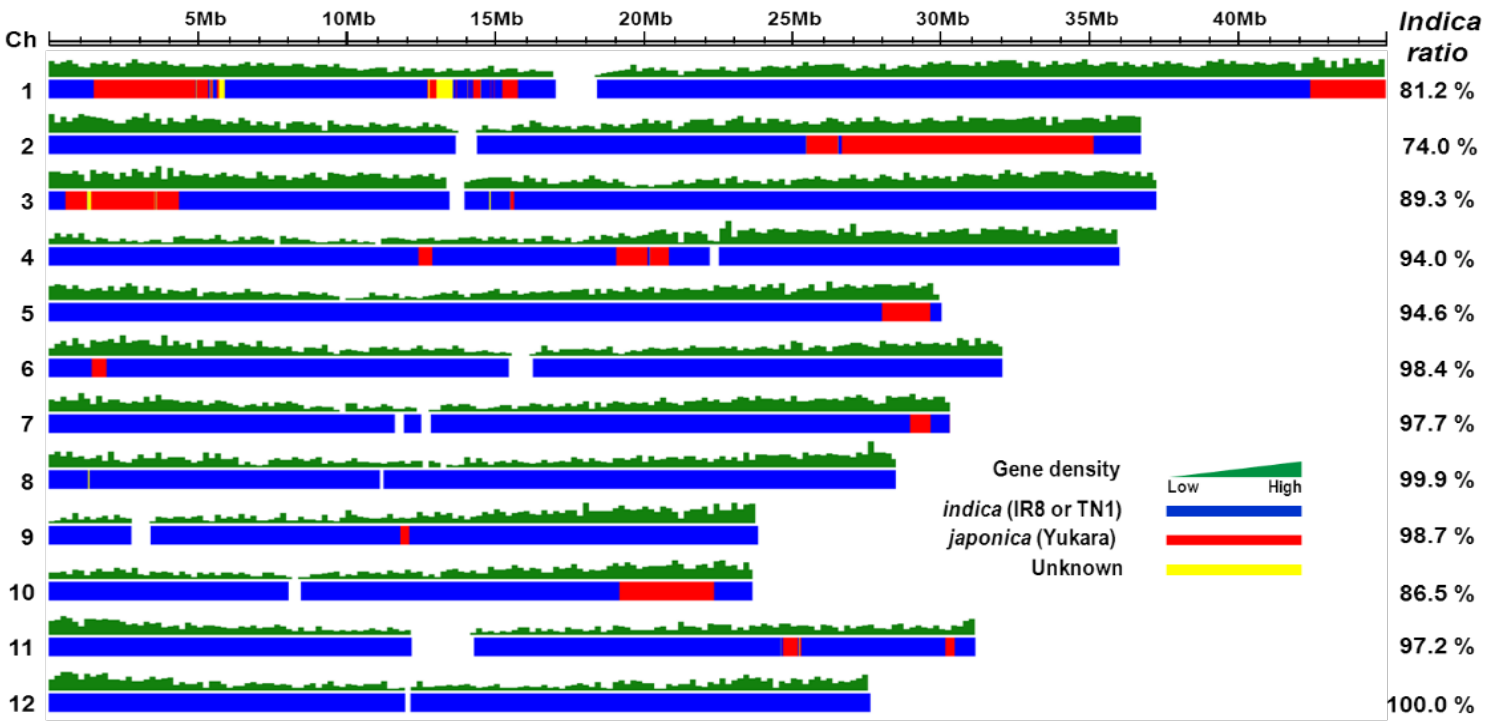


Figure 1-2. *Indica*/*japonica* genome organization on the 12 chromosomes of Tongil. Blue indicates the *indica* genome (TN1 and IR8); red indicates the *japonica* genome (Yukara); and yellow indicates a region from an unknown genome. The percentages indicate the proportion of *indica* contribution on each chromosome.

detectable on chromosomes 8 and 12. In addition, there were no differences in gene density between the *indica*- and *japonica*-derived genome regions of Tongil (Figure 1-2 and Table 1-6).

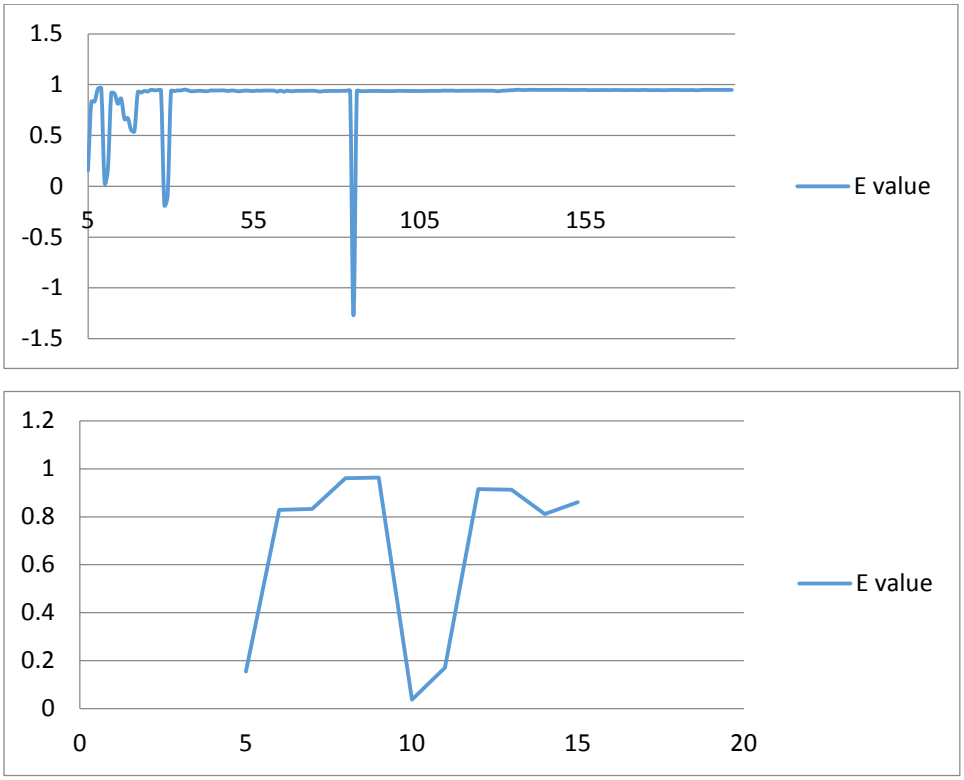


Figure 1-3. Determination of window size followed by *E*-value calculation

Table 1-4. Genome region definition by presence (O) or absence (X) of SNPs difference

Genome origin from / Define as	Presence of SNP or not		
	Yukara	IR8	TN1
IR8 / *indica*	O	X	O
TN1 / *indica*	O	O	X
IR8 and TN1 / *indica*	O	X	X
Yukara / *japonica*	X	O	O

Table 1-5. The ratio of *indica-japonica* specific SNPs determined by comparison of Tongil and its parental varieties.

Chromosome	*Indica* SNP	Ratio (%)	*Japonica* SNP	Ratio (%)	Total
1	129,021	82.1	28,045	17.9	157,066
2	96,765	74.8	32,628	25.2	129,393
3	115,351	88.7	14,622	11.3	129,973
4	92,280	92.6	7,406	7.4	99,686
5	82,754	92.5	6,712	7.5	89,466
6	114,578	96.2	4,571	3.8	119,149
7	113,844	96.5	4,082	3.5	117,926
8	92,834	98.3	1,598	1.7	94,432
9	87,933	98.1	1,678	1.9	89,611
10	83,513	85.6	14,010	14.4	97,523
11	123,151	97.9	2,689	2.1	125,840
12	102,312	98.3	1,781	1.7	104,093
Total or Ave.	1,234,336	91.2	119,822	8.8	1,354,158

Table 1-6. Determination of the *indica*/*japonica* genome origin of Tongil, based on a window size of 9.

Chromosome	Pseudomolecule	*indica* region (bp)	Ratio (%)	*japonica* region (bp)	Ratio (%)	Unknown region (bp)	Ratio (%)
1	45,038,604	36,563,905	81.2	7,596,808	16.9	877,891	2.0
2	36,792,247	27,235,850	74.0	9,544,379	25.9	12,018	0.0
3	37,312,367	33,336,733	89.3	3,748,667	10.1	226,967	0.6
4	36,060,865	33,898,364	94.0	2,150,911	6.0	11,590	0.0
5	30,073,438	28,436,341	94.6	1,637,097	5.4	-	-
6	32,124,789	31,619,689	98.4	499,676	1.6	5,424	0.0
7	30,357,780	29,667,148	97.7	690,632	2.3	-	-
8	28,530,027	28,487,631	99.9	333	-	42,063	0.2
9	23,895,721	23,592,877	98.7	302,844	1.3	-	-
10	23,703,430	20,504,662	86.5	3,198,768	13.5	-	-
11	31,219,694	30,345,040	97.2	846,802	2.7	27,852	0.1
12	27,679,166	27,679,166	100.0	-	-	-	-
Total	382,788,128	351,367,406	91.8	30,216,917	7.9	1,203,805	0.3

Gene distribution and gene ontology analysis in Tongil

We analyzed the gene content of Tongil to understand the relationship between the composition of the genome and genes (Open Reading Frames: ORFs), and also to elucidate the distribution of *indica*- and *japonica*-originated genes (alleles) in the Tongil genome. The gene distribution ratio according to *indica* or *japonica* genome composition was similar to the genome distribution ratio of Tongil (Table 1-6 and Table 1-7). The origin of genes from the *indica* and *japonica* parents were 88.3% and 11.4%, respectively, suggesting that the average gene composition was similar to the genome composition ratio of Tongil, although the distribution of parental origin varied across chromosomes. To identify a biological pattern in a list of the genes that belong to the *indica*, *japonica*, and unknown genomes, we performed gene ontology (GO) analysis of the Tongil genome according to three categories: cellular component, molecular function, and biological process (Figure 1-4 and Figure 1-5 and Figure 1-6). The results of GO analysis revealed that the average contribution of *indica* or *japonica* genome to each GO category was almost identical to the gene and genome distribution ratios. *Indica* and *japonica* contributed 86.8% and 12.7% of cellular components, 87.4% and 12.2% of molecular functions, and 87.3% and 12.2% of biological processes, respectively. However, in the 'molecular function' category, all 17 genes related to channel regulator activity belonged to *indica* regions, whereas in the biological process category, all adhesion-related genes in the 'biological process' category belonged to only *japonica* regions.

Table 1-7. Gene distribution of Tongil

Chromosome	*Indica*		*Japonica*		Unknown		Total
	No. of genes	%	No. of genes	%	No. of genes	%	
1	3,614	77.9	953	20.5	75	1.6	4,642
2	2,553	68.4	1,178	31.6	2	0.1	3,733
3	3,475	85.7	548	13.5	31	0.8	4,054
4	2,713	93.2	197	6.8	-	0.0	2,910
5	2,440	90.9	245	9.1	-	0.0	2,685
6	2,659	97.2	77	2.8	1	0.0	2,737
7	2,485	96.4	94	3.6	-	0.0	2,579
8	2,261	99.9	-	0.0	2	0.1	2,263
9	1,858	99.3	14	0.7	-	0.0	1,872
10	1,390	77.0	416	23.0	-	0.0	1,806
11	2,012	97.0	62	3.0	1	0.0	2,075
12	1,921	100.0	-	0.0	-	0.0	1,921
Total or Ave.	29,381	88.3	3,784	11.4	112	0.3	33,277

Figure 1-4. GO analysis by 'cellular component' category of Tongil genes, corresponding to *indica*/*japonica* sequence.

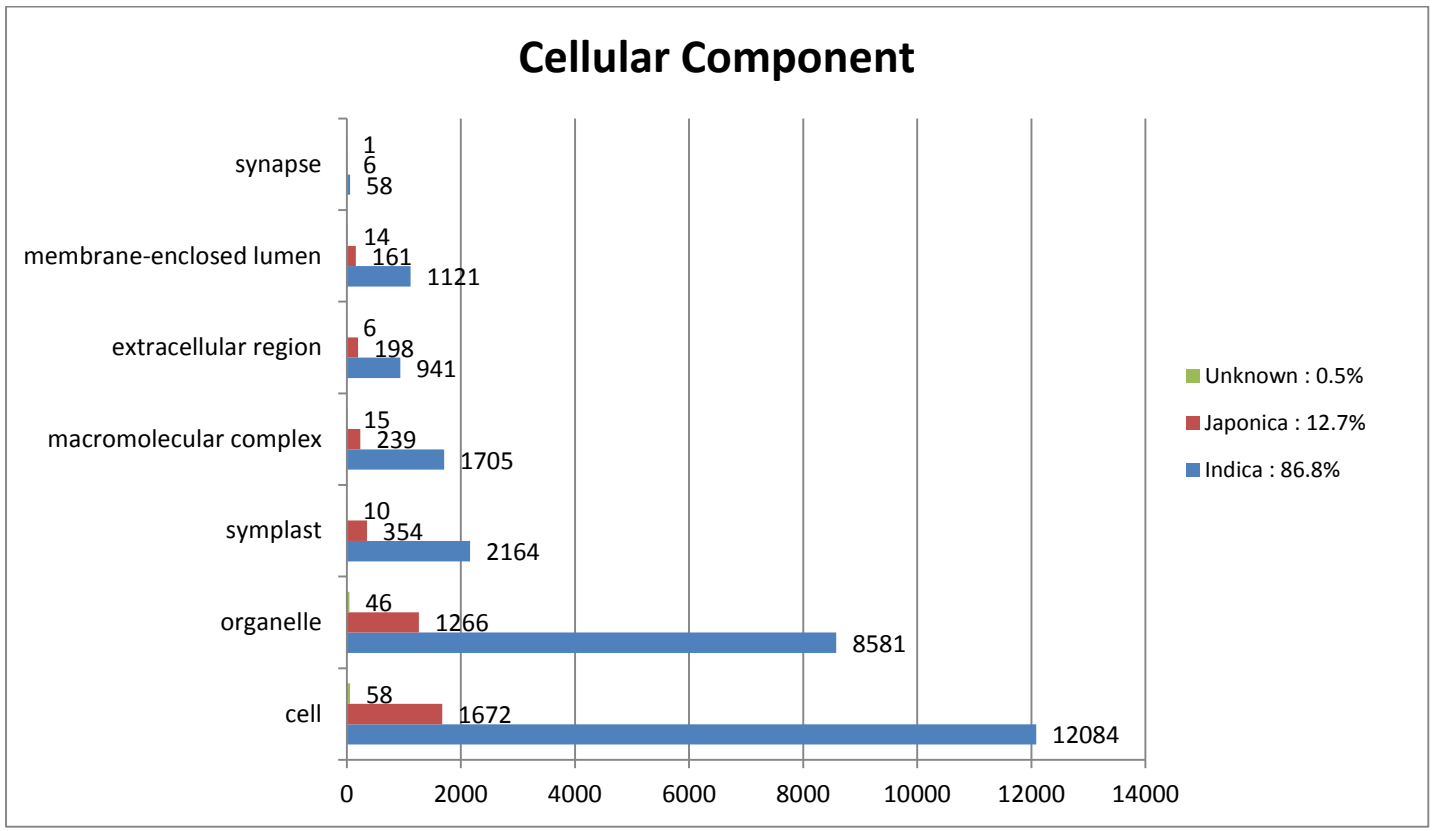


Figure 1-5. GO analysis by 'molecular function' category of Tongil genes, corresponding to *indica*/*japonica* sequence.

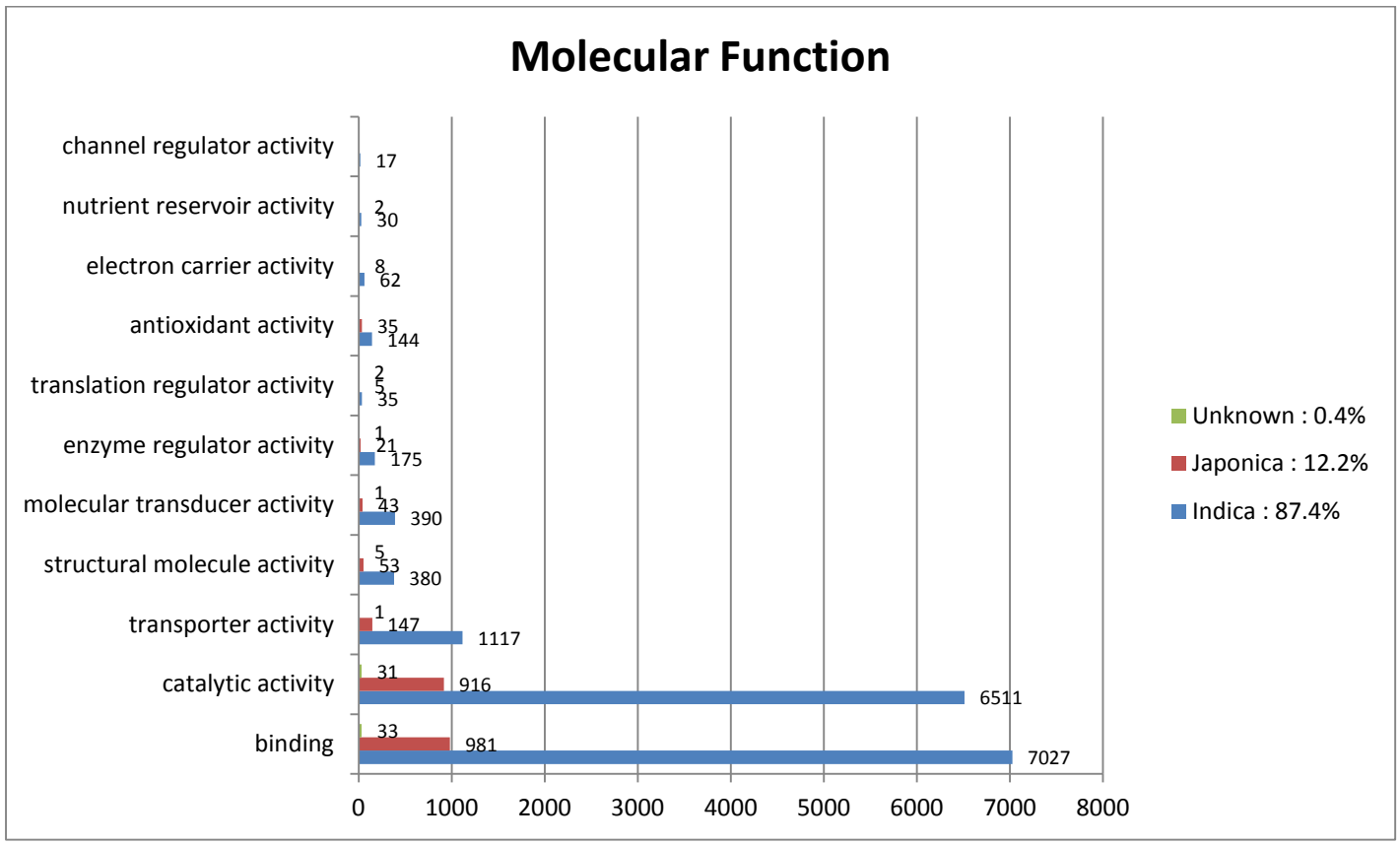
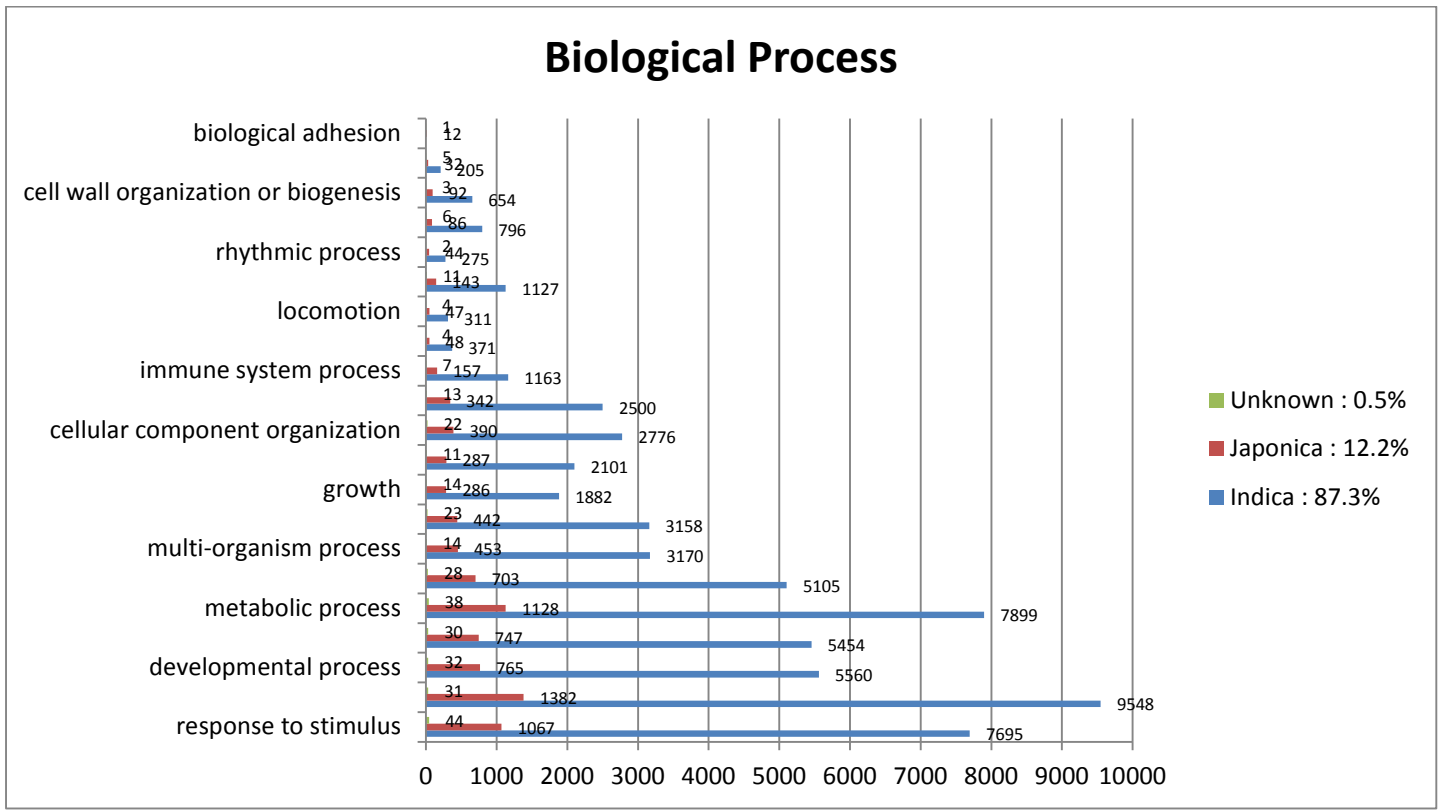


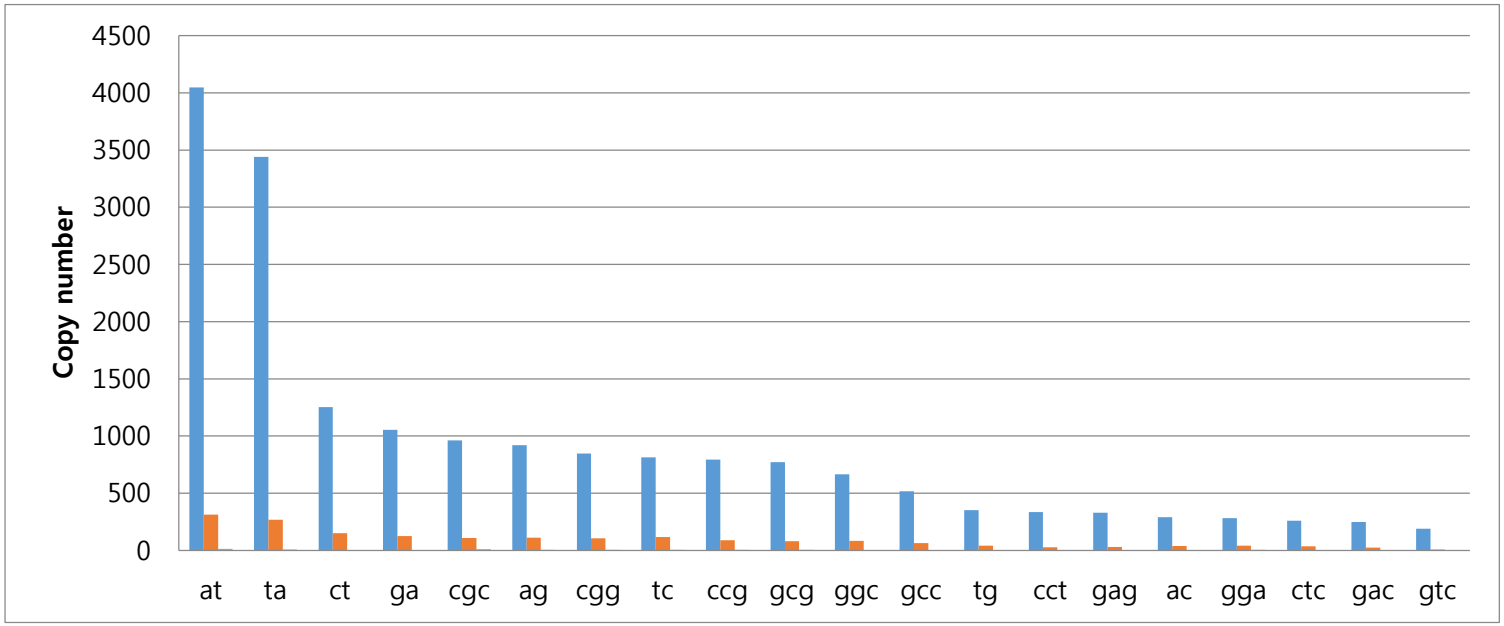
Figure 1-6. GO analysis by 'biological process' category of Tongil genes, corresponding to *indica*/*japonica* sequence



Simple sequence repeats (SSRs) in Tongil genome

A total of 177 distinctive motif families were annotated on the Tongil genome (Figure 1-7). Among the classified repeats, di-nucleotide repeats were predominant, and AT/TA repeats were the most abundant motif in both *indica* (29.09%) and *japonica* (21.8%) regions within the Tongil genome. The next most abundant motif relative to AT/TA was CT/GA, and CGC was the most abundant motif among tri-nucleotide repeats. The di-, tri-, and tetra-nucleotide repeat pattern was not similar to that of the Nipponbare genome (McCouch et al., 2002; Zhou et al., 2005), and it also differed from that of wheat (Weng et al., 2005). A total of 90.1% of SSR motifs in the Tongil genome were from *indica*, 9.6% were from *japonica*, and 0.3% were from an unknown genome (Figure 1-7).

Figure 1-7. Copy number of SSR motif families in Tongil



Indica (%)	4048 (92.6)	3440 (92.6)	1253 (89.2)	1053 (89.2)	963 (88.9)	919 (88.8)	848 (88.5)	815 (87.1)	794 (89.4)	773 (90.1)	666 (88.8)	517 (88.7)	352 (89.3)	335 (91.8)	329 (91.4)	290 (88.4)	282 (85.7)	261 (87.3)	249 (89.9)	191 (95.5)
Japonica (%)	312 (7.1)	267 (7.2)	151 (10.8)	125 (10.6)	110 (10.2)	111 (10.7)	105 (11.0)	117 (12.5)	90 (10.1)	80 (9.3)	83 (11.1)	65 (11.1)	42 (10.7)	29 (7.9)	30 (8.3)	38 (11.6)	42 (12.8)	35 (11.7)	26 (9.4)	9 (4.5)
Unknown (%)	13 (0.3)	8 (0.2)	0 (0.0)	2 (0.2)	10 (0.9)	5 (0.5)	5 (0.5)	4 (0.4)	4 (0.5)	5 (0.6)	1 (0.1)	1 (0.2)	0 (0.0)	1 (0.3)	1 (0.3)	0 (0.0)	5 (1.5)	3 (1.0)	2 (0.7)	0 (0.0)

Distribution of yield-related genes in the Tongil genome

One of the most important points of this study was to explore which regions of the *indica* and *japonica* parental genomes have been introgressed in the Tongil variety to provide its high-yield potential. Tongil is morphologically characterized by short plant height, lodging resistance, open plant architecture, medium-long and erect leaves, thick leaf sheaths and culms, relatively long panicles, and easily shattered grain (Chung and Heu, 1980) (Figure 1-1). Although these phenotypic characteristics must affect Tongil's high-yield potential, to date we have no molecular genetic evidence regarding the nature of these traits, with the exception of semi-dwarf gene 1 (*sd1*) (Chung and Heu, 1980). Therefore, we analyzed several well-characterized genes that are associated with high yield potential: *sd1* (Nagano et al., 2005), *Ghd7* (Liu et al., 2013; Xue et al., 2008), *Gn1a* (Ashikari et al., 2005), *qSW5* (Yan et al., 2011), *GS3* (Takano-Kai et al., 2009), and *GW2* (Li et al., 2010) in the Tongil genome.

sd1 : semi-dwarf stature

Semi-dwarf stature is one of the main genetic contributors to the green revolution. Introduction of semi-dwarf genes increased yield by conferring lodging resistance, which enabled more input of nitrogen fertilizer. Tongil was the first variety into which the *sd1* allele was introduced in South Korea. Analysis of the *sd1* gene, which encodes GA20ox-2 in Tongil and its parents, revealed that Tongil received its *sd1* from an *indica* parent, IR8 or TN1; this allele contains a 383 bp deletion resulting in a frame-shift to make a stop

codon (Nagano et al., 2005) (Figure 1-8A). We also confirmed other *sd1* alleles from the native semi-dwarf rice variety Jikkoku (G281T) and the γ-ray–induced varieties Reimei (C1045G) and Calose76 (C796T) (Monna et al., 2002; Sasaki et al., 2002). On the other hand, Yukara, the *japonica* parent of Tongil, did not have any *sd1* alleles.

Ghd7 : grain number, plant height, and heading date

A gene encoding a CCT domain protein, *Ghd7*, is an important regulator of potential yield, plant height, and heading date in rice. Plant height and panicle size are increased under long-day conditions by the delay in heading date resulting from increased *Ghd7* expression (Liu et al., 2013; Xue et al., 2008). *Ghd7* has five natural variant haplotypes. Tongil has the *Ghd7-1* allele (A-G-S-V-A) derived from *indica* parent IR8 or TN1 (Figure 1-8B), considered to be the original wild-type allele with full function; plants with this allele are relatively tall, late heading, have a large panicle, and are widely grown. By contrast, the *japonica* parent of Tongil, Yukara, has the *Ghd7-2* allele (A-E-G-D-P), which is weaker than *Ghd7-1* and is found in temperate *japonica* varieties.

Gn1a : Grain number on chromosome 1

Gn1a is one of the most effective QTLs for increasing grain number. It is predicted to encode a cytokinin oxidase/dehydrogenase (OsCKX2). Habataki, an *indica* rice variety, has a 16 bp deletion in the 5′ UTR, a 6 bp deletion in the first exon, and three amino acid substitutions in the first

and fourth exons of this gene. In addition, an 11 bp deletion in the third exon has been detected in the high-yielding rice variety 5150 (Ashikari et al., 2005). Comparison of DNA sequences between Tongil and parent varieties revealed that Tongil is identical to the TN1 allele, which has only a 6 bp deletion in the first exon and no 16 bp deletion in the 5′ UTR, as in Habataki. On the other hand, IR8 contains a 16 bp deletion in the 5′ UTR and a 12 bp deletion in the first exon, distinct from the pattern in the TN1 allele. We could not identify any variation in Yukara, which has the same allele as Nipponbare (Figure 1-8C).

qSW5 : QTL for seed width on chromosome 5

qSW5 is responsible for seed width; the product of this gene controls cell number in the outer glume of the rice flower. The gene product increases seed width and seed weight by enlarging sink size. The Nipponbare-type allele, which contains a 1,212 bp deletion, is loss-of-function relative to the Kasalath-type allele (Shomura et al., 2008). In addition, the *indica* Ⅱ-type allele has a 950 bp deletion relative to the Kasalath allele (Yan et al., 2011). Tongil, IR8, and TN1 have the *indica* Ⅱ-type allele, whereas Yukara contains the Nipponbare allele (Figure 1-8D).

GS3 : Grain length and weight (grain size 3)

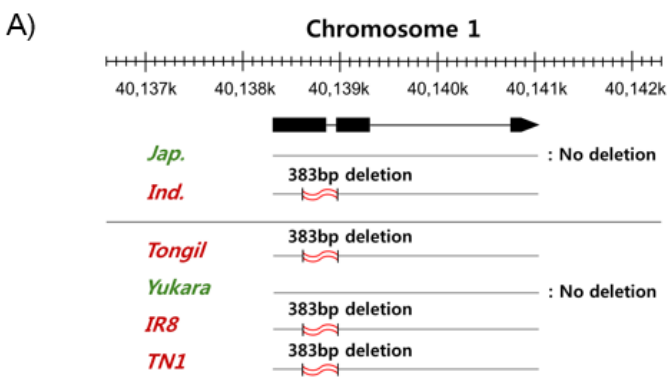
GS3, which encodes a PEPB-like domain protein, was cloned from a QTL for grain length and weight on chromosome 3 in rice. A C-to-A substitution in the second exon of the *GS3* gene is highly associated with grain length

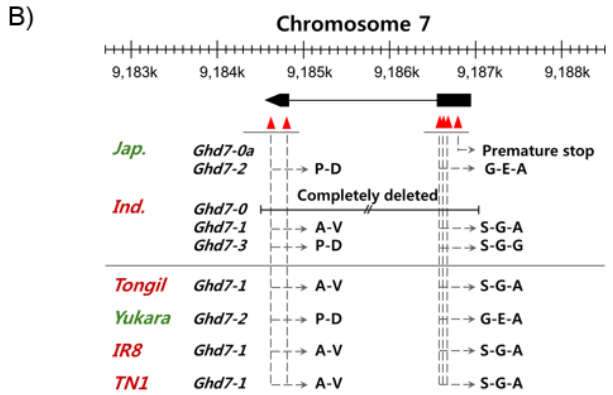
and width: the A-allele confers significantly longer and thinner grains than the C-allele. Tongil possesses an A-allele originating from an *indica* parent, IR8 or TN1, whereas the *japonica* parent, Yukara, has the C-allele (Figure 1-8E). In the case of another gene that controls grain width, *GW2*, there were no SNPs among any of the strains we sequenced or Nipponbare, indicating that *GW2* is a highly conserved gene in rice and even in *Zea mays* (Yan et al., 2011).

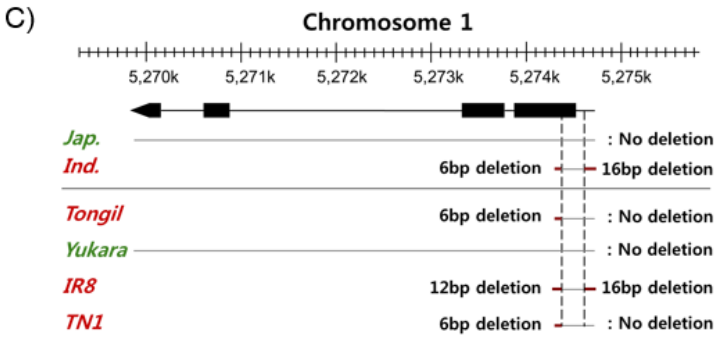
Table 1-8. Comparison of alleles of yield-related genes in Tongil and its parents

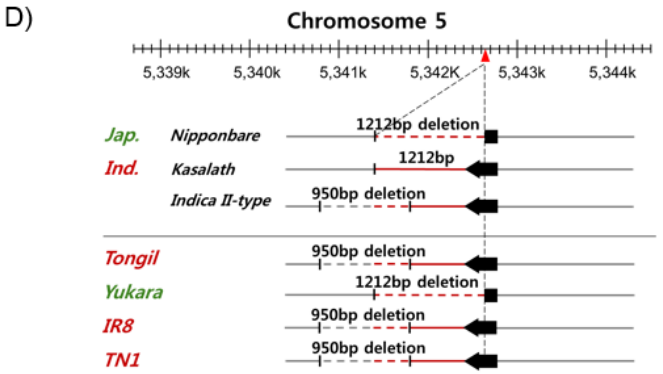
Gene	Gene name	Tongil	Yukara	TN1	IR8
qSW5	QTL for Seed width on chromosome 5	**I**[a]	**J**	**I**[a]	**I**[a]
Gn1a	Grain number in chromosome 1	**I**	**J**	**I**	**I**[b]
Ghd7	Grain number, plant height and heading date	**I**	**J**	**I**	**I**
sd1	Semi Dwarf 1	**I**	**J**	**I**	**I**
GS3	Grain size 3	**I**	**J**	**I**	**I**
GW2	Grain width and weight	**I**	**J**	**I**	**I**
S5	Hybrid Sterility 5	**I**	**J**	**I**	**I**

a: *indica* -II type
b: different *indica* type allele with Tongil and TN1


A)
Chromosome 1
40,137k
40,138k
40,139k
40,140k
40,141k
40,142k
Jap.
: No deletion
Ind.
383bp deletion
Tongil
383bp deletion
Yukara
: No deletion
IR8
383bp deletion
TN1
383bp deletion


B)
Chromosome 7
9,183k
9,184k
9,185k
9,186k
9,187k
9,188k
Jap.
Ghd7-0a
Premature stop
Ghd7-2
P-D
G-E-A
Ind.
Completely deleted
Ghd7-0
Ghd7-1
A-V
S-G-A
Ghd7-3
P-D
S-G-G
Tongil
Ghd7-1
A-V
S-G-A
Yukara
Ghd7-2
P-D
G-E-A
IR8
Ghd7-1
A-V
S-G-A
TN1
Ghd7-1
A-V
S-G-A


C)
Chromosome 1
5,270k
5,271k
5,272k
5,273k
5,274k
5,275k
Jap.
: No deletion
Ind.
6bp deletion
16bp deletion
Tongil
6bp deletion
: No deletion
Yukara
: No deletion
IR8
12bp deletion
16bp deletion
TN1
6bp deletion
: No deletion


D)
Chromosome 5
5,339k
5,340k
5,341k
5,342K
5,343k
5,344k
Jap.
Nipponbare
1212bp deletion
Ind.
Kasalath
1212bp
Indica II-type
950bp deletion
Tongil
950bp deletion
Yukara
1212bp deletion
IR8
950bp deletion
TN1
950bp deletion

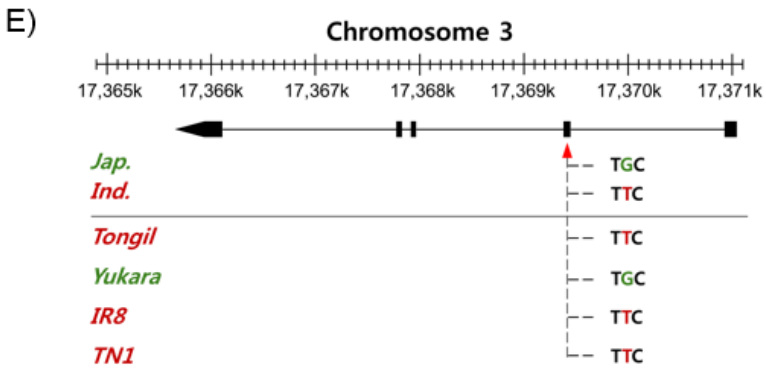


Figure 1-8. *Indica*/*japonica* region comparisons of high yield-related alleles or QTLs. A) *sd1*, B) *Ghd7*, C) *Gn1a*, D) *qSW5*, and E) *GS3*. Black arrows and box regions represent exons. Vertical, dashed lines refer to the same position in the genome or gene region.

DISCUSSION

indica/*japonica* genome composition of Tongil

In this study, we analyzed genome structure of Tongil using high-depth NGS tool to demonstrate that the Tongil genome is composed of 91.8% *indica*, 7.9% *japonica*, and 0.3% unknown genome (Figure 1-2 and Table 1-6). The amounts and types of genes and SSRs in the Tongil genome were very similar to its genomic composition with respect to *indica* or *japonica* origin. This deviation from the expectation that about one-fourth of the Tongil genome originated from the *japonica* parent is likely due to the results of selection during the breeding process and/or to segregation distortion in favor of the *indica* genome because *indica*-type alleles and plants are favored among hybrid progenies from *indica*/*japonica* crosses

These results suggest several possibilities, Linkage disequilibrium (LD) is a non-random association of alleles at two more different loci (Flint-Garcia et al., 2003). Independently segregating, in a large and random mated population without selection, mutation, migration, polymorphic loci will be in linkage equilibrium (Falconer et al., 1996). However, when rice varieties are developed, LD level is increased due to small population, artificial mating, and selection for desirable agronomic trait thus, biased polymorphic loci occurred in rice genome. In addition, plant genes have been considered to be randomly distributed through the genome. However, some resistant and metabolic genes have discovered to be clustered and inheritance together. For example, UV inducible gene, momilactone biosynthesis and chitin oligosaccharide elicitor formation related gene

cluster (Shimura et al., 2007), second multifunctional diterpenoid biosynthetic gene cluster (Swaminathan et al., 2009), and clustered genes containing *Pi-ta* for blast resistance inherited together from backcross progeny involving *indica* and *japonica* cross and large portion of chromosome fragment was maintained by artificial selection during breeding process (Jia, 2009). These gene clustering confers the selective advantage by combining the inheritance genes and facilitates regulation of gene expression by enabling localized changes in chromatin structure (Wegel et al., 2009). Some plant metabolic pathways might be clustered by evolutionary selection pressure (Field and Osbourn, 2008).

Linkage blocks (LBs), where a set of gene is inherited together is larger in *indica*/*japonica* cross than *indica*/*indica* cross due to the incompatibility between *indica* and *japonica* rice. The larger LBs, the higher recombination suppression occurred. Thus, crosses of *indica*/*japonica* are subject to large linkage block and have significant recombination suppressions (Jia et al., 2012). Similarly, linkage drag was studied by Hanson W. D. (Hanson, 1959). Linkage drag means that the rate of decrease of donor fragment is slower than for unlinked region due to the linked to the target gene from the donor plant in backcrossing (Collard and Mackill, 2008; Hospital, 2005) and linkage drag is influenced by location of chromosome, selection way, and recombination factors (Flint-Garcia et al., 2003; Naveira and Barbadilla, 1992), however the molecular mechanism of linkage drag is still unclear. The major factor of gene clustering, linkage blocks, and linkage drag is due to the selection for targeted gene or trait.

The aim of breeding of Tongil variety was development of high yielding and cold and blast resistant rice. High yield and stress tolerant related traits such as plant height, grain size, shape, panicle size, 1000 grain weight, sterility, and cold and blast disease tolerance were artificially selected by breeders, although some traits were eliminated during the shuttle breeding between Korea and IRRI. In our provided result, Tongil rice consists of 99.9% *indica* genome on chromosome 8 and no *japonica* genome sequences on chromosome 12 (Figure 1-2 and Table 1-6). These phenomenon will give an important information to study LD in the view of genome sequence and SNPs.

Segregation distortion is occurred in genetic mapping populations from *indica/japonica* cross and usually observed through genetic markers in rice thus, the segregation ratio is not fit to the Mendelian inheritance (Harushima et al., 1996). Segregation distortion is influenced by genetic and experimental factors such as population size, polymorphic markers, and genotyping error (Hackett and Broadfoot, 2003). It can arise either from gametophytic competition among gametes which gametophyte gene is expressed or from abortion of the gamete or zygote by hybrid sterility genes (Harushima et al., 1996). Three-way system at *S5* locus which responsible for hybrid sterility of inter-specific, namely *S5i*, *S5j* and *S5n* for *indica*, *japonica* and wide compatibility variety (WCV), respectively, was suggested by Ikehashi H. and Araki H. (Lin et al., 1992). WCV was used to overcome reproductive barriers and the cross combination containing *S5n* allele showed high fertile. (Chen et al., 2008b; Lin et al., 1992). The

interaction between *S5i* and *S5j* encoded aspartyl protease which has single amino acid difference was responsible for embryo sac abortion and *indica*/*japonica* hybrid sterility. Unlike *S5i* and *S5j*, *S5n* was not formed dimers due to a large deletion in the N terminus of AP protein thus, it might occurred wide compatibility (Ji et al., 2012). Killer-protector system at the *S5* locus that endoplasmic reticulum (ER) stress is triggered by ORF5+ (killer) and ORF4+(partner) while ORF3+(protector) prevents ER stress from the ORF5+ and ORF4+ to produce normal gamete was proposed (Yang et al., 2012). In contrast to female sterility genes, the *Sa*, *indica-japonica* hybrid male sterility related locus also cloned in rice. *Sa* locus comprises two adjacent genes, *Sa*M encodes a small ubiquitin-like modifier E3 ligase-like protein and *Sa*F encodes F-box, respectively. *Indica* (SaM+SaF+) and *japonica*(SaM-SaF-) hybrid male sterility is caused by selective abortion of pollen carrying SaM- therefore, interrupting its transmission into hybrid progeny and segregation distortion reveals in this allele (Long et al., 2008). These reproductive isolations through allele interactions might lead diverge into *indica* and *japonica* eventually, prevent gene exchange or gene flow between two subspecies. Segregation distortion ratio was higher in inter-specific population than intra-specific population in rice.

Our result indicates that Tongil genome was biased to *indica* unlike theoretical expectation (Table 1-6). We assumed that one of the main factor for *indica* biased genome in Tongil was the interaction *indica*/*japonica* hybrid sterility genes, although many other loci were involved in hybrid

incompatibility. The two *indica* parent varieties, IR8 and TN1 and a *japonica* parent variety, Yukara, of Tongil have a typical *indica* (*S5i*) and *japonica* (*S5j*) *S5* locus, respectively (data not shown). It might enables the embryo sac abortion and reproductive isolation between TN1 and Yukara at the first cross and then *indica*-compatible gamete was selected at the second cross between F_1 hybrid and IR8. In addition, chromosomal rearrangement such as a translocation, inversion, fusion, fission, and duplication might act as isolating barriers by suppressing recombination and thus reducing the gene flow. Most chromosomal rearrangements appear to have large effect on fertility of plants (Rieseberg, 2001; Widmer et al., 2009). For these reasons, Tongil rice might have the *indica* toward genome. Tongil rice is highly successful in terms of grain yield in South Korea, although Korean climatic environments are not favorable to the cultivation of typical *indica* varieties (Chung and Heu 1991). This may be attributable to its heightened adaptability compared to most *indica* varieties, perhaps due to the partial incorporation of the *japonica* parental genome.

Comparison of yield related alleles in Tongil

Rice yield is determined by four yield components. 1) Number of panicles per unit area 2) number of spikelet 3) number of filled grain and 4) 1000 grains weight. Seed size, one of the main factors of seed weight, is affected by following four parameters: seed length, width, thickness and grain filling degree. Tongil rice belongs to the heavy panicle type variety and has larger grain size than *japonica* variety. Seed weight is higher than that of other

japonica varieties and parental *japonica* variety. We manually sequenced and compared the *sd1*, *Gn1a*, *Ghd7*, *GS3*, *qSW5*, and *GW2* alleles in Tongil and its parents genome, which were reported and proved as yield related genes and QTLs to understand which alleles are from between *indica* and *japonica* parents, respectively. (Table 1-8). *GW2*, *GS3*, and *qSW5* alleles of Tongil, which are related in determining seed size all alleles, were from *indica* parent IR8 or TN1. Despite Tongil and its *indica* parent varieties commonly have three identified seed size controlling alleles, seed shape of Tongil is closer to TN1 than that of IR8 (Figure 1-1). This indicates that seed shape could be determined by another numerous genes or complex QTLs besides *GW2*, *GS3*, and *qSW5* alleles in rice. It might be due to an additive and epistatic effect between gene interactions and different gene expression by various genetic backgrounds and cross combination (Yan et al., 2011). *S5* allele also came from *indica* parents. For two times crossing, *indica* parent, TN1 or IR8 conferred their *S5* allele to Tongil and it enables Tongil could avoid embryo sac abortion and recover its fertility despite of *indica*/*japonica* hybrid. We assumed that this is one of the main reasons for increasing yield potential of Tongil as well as *sd1* allele introgressed from *indica* parents. In addition, *sd1*, *Gn1a*, and *Ghd7* alleles of Tongil showed *indica* or *indica*-type alleles. Gene pyramiding for increasing yield potential is difficult because the complicated gene interactions and undesirable effect on target trait. However, the information of combination and predominant of alleles could be used to develop new *indica*/*japonica* hybrid variety and gene pyramiding for high yielding rice breeding.

The approach used in this study to determine the parental origins of specific genome segments is applicable to the genomic dissection of agricultural breeding lines or varieties of diverse parental origins

CHAPTER II.

Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

ABSTRACT

Yellow colored pericarp in rice (*Oryza sativa.* L) is rare in nature. We generated a yellow pericarp and embryo lethal mutant, *yel*, derived from chemical mutagenesis using N-methyl-N-nitrosourea (MNU) on a *japonica* rice cultivar, Hwacheong. In this study, we fine mapped and cloned the gene responsible for yellow pericap and embryo lethal phenotype using a map-based approach. Fine mapping revealed that the mutant gene (*yel*) was located on the long arm of chromosome 2. Sequencing of the candidate genes analysis between the parent and the mutant showed a 706-bp deletion in the first exon including 5'UTR and start codon of LOC_Os02g53140, *Os*COP1 (Constitutive Photomorphogenic 1) which encodes a protein of comprising RING-finger, coiled-coil and WD40 domains. To confirm the function of *yel* gene on yellow pigmentation and embryo lethality, we introduced YEL overexpression vector into the mutant heterozygous plant. Overexpression of YEL, containing normal RING-finger domain restored the pericarp color from yellow to white, while the embryo

color was still black and maintained lethality. This result demonstraiting that the *Os*COP1 gene is involved in yellow pericarp color. Gus staining signals were primarily detected in the embryo, pericarp, pollen and nodes. HPLC-MS analysis of the four candidate standard with extracts of embryo and endosperm of mutant grain revealed that orientin was the major pigment both in embryo and endosperm. In addition, extremely high level of orientin accumulation was observed in mutant embryo and endosperm. The antioxidant activities and total phenolic contents of embryo and endosperm of the mutant seed were significantly higher than those of wild-type. Fatty acid analysis showed the content of saturated fatty acid was increased in mutant embryo on the other hand, the content of unsaturated fatty acid was decreased than that of wild-type. These results indicated that *yellow embryo lethal* (*yel*) gene might be involved in the flavonoids biosynthetic pathway and other ingredient characteristics of rice seed.

KEYWORDS: Yellow pericarp, Embryo lethality, Orientin, Flavonoid, Antioxydant activity

INTRODUCTION

Flavonoids are one of the major class of compounds found in various crops as well as fruits, vegetables and beverages (tea, coffee, beer, wine and friut drink) and it has been important nutrient and beneficial sources for human health by providing essential vitamins and high pharmacological activities and antioxidant capacity. For these reasons, flavonoids have been studied extensively for their effects. Typically, flavonoids are divided into six major subclass including flavonols, flavonones, flavones, flavanols, anthocyanidins, and isoflavones according to the position of substitutes present on the parent molecules. Although flavonoids are distributed throughout the plant kingdom, some flavonoid classes are founded only a few plant species. The interest in increasing the flavonoid content of commonly consumed food crops for health benefits have grown (Butelli et al., 2008; Schijlen et al., 2006).

Flavonoid biosynthesis is well characterized in terms of its molecular genetics and the biochemical mechanisms of the enzymes involved in the flavonoid pathway (Dixon and Pasinetti, 2010). In plants, the biosynthesis of naringenin, a flavanone and central precursor of most flavonoids, involves the following five enzymes in the phenylpropanoid pathway: phenylalanine ammonia lyase (Matsumoto et al.), cinnamate 4-hydroxylase (C4H), 4-coumarate: coenzyme A ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI). The pathway diverges from naringenin into several side branches, each resulting in the production of a different class

of flavonoid. Flavonol synthesis is catalysed by flavanone 3-hydroxylase (F3H) and flavonol synthase (FLS), while isoflavone synthesis and flavone synthesis are catalysed by isoflavone synthase (IFS) and flavone synthase (FNS), respectively. Flavonoids also have many derivatives with hydroxyl groups and/or methoxy groups within a single class. Flavonoids are often glycosylated with one or more sugar units at various positions on the core aglycone structure. These glycosylated flavonoids are thought to be transported to and accumulated in the vacuole. The flavone *C*-glycosides are an important subclass of the flavonoids family, were found in some plants, such as flavone *C*-glycosides are present in food stuff and nutraceuticals and they include orientin, isoorientin, vitexin, and isovitexin. Orientin (*C*-glycosyl flavone, luteolin-8-*C*-glucoside) has not been yet studid in rice. orientin cardioprotection (Lu et al., 2011), antinociception (Da Silva et al., 2010)

Flavonoids play an important role in the grain color and flavor of rice. Rice bran contains a unique complex of naturally occurring antioxidant compounds (Mazza G, 2003). Oryzamutaic acids were detected in yellow endosperm mutant of the variety Hatsuyamabuki however, the biosynthetic pathway remains still unknown. This mutant showed the yellow endosperm which was caused by mutation of Oryzamutaic acid A and *OsALDH7* (Nakano et al., 2009; Nakano et al., 2011; Shen et al., 2012) Mutation of *OsALDH7* resulted in accumulation of yellow pigment in endosperm. Oryzamutaic acid A accumulates during late seed development and after a year-long storage in the endosperm. Pigmented rice also contains

anthocyanins in pericarp. Anthocyanins are responsible for cyanic color of pigmented rice and are regarded as important nutraceuticals mainly due to their antioxidant effect, which provide a potential to prevent various diseases associated with oxidative stress (Duthie et al., 2000; Kong et al., 2003)

Recently, the consumption of colored rice is getting increased to intake the health-promoting components such as vitamins, minerals and phytochemicals which include phenolic compound while eating rice (Dykes and Rooney, 2007). Research has shown that phenolic compounds have antioxidant properties and can protect against degenerative disease such as heart disease and cancer. These compunds are accumulated in the pericarp and aleurone layer in cereal crops. Many studies were conducted in fruits and vegetables but grain cereals are good material for phenolic compounds.

One feature of the constitutive photomorphogenesis phenotype of cop1 and spa mutants is the increased accumulation of anthocyanins (Deng et al., 1991; Hoecker et al., 1998; Laubinger et al., 2004). In Arabidopsis, anthocyanin accumulation occurs only in the light and this accumulation is further enhanced by a number of environmental stress factors, such as cold, drought, pathogen attack and nutrient depletion (Chalker-Scott, 1999). The biosynthesis of anthocyanins is regulated by transcription factors that induce the expression of structural genes that code for enzymes in the biosynthesis pathway (Broun, 2005; Li, 2014). These transcription factors include HY5 and a complex consisting of a WD-repeat protein, a bHLH

protein and a myeloblastosis (MYB) protein (Feller et al., 2011; Koes et al., 2005; Lee et al., 2007; Shin et al., 2007). In Arabidopsis, the WD-repeat protein TTG1 acts together with the bHLH proteins GL3, EGL3 or TT8 and the PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP) protein family of R2R3 MYB transcription factors (PAP1, PAP2, MYB113 and MYB114) (Gonzalez et al., 2008; Zhang et al., 2003; Zimmermann et al., 2004). The PAP protein family, in contrast, is required specifically for anthocyanin accumulation and, therefore, provides specificity to the WD-repeat/bHLH/MYB complex (Lillo et al., 2008; Shi and Xie, 2010; Tohge et al., 2005) The expression of PAP1, PAP2, TT8, GL3 and EGL3 is induced by light, while the expression of TTG1 is light-independent (Cominelli et al., 2008).

In this study, we report the yellow pericarp and embryo lethal mutant which has the yellow pigment, isoorientin which is a very rare compound in embryo and endosperm of rice seed. This mutant rice might be able to develop the yellow colored rice and use for functional food source.

MATERIALS AND METHODS

Plant materials

The yellow embryo lethal mutant was induced by N-methyl-N-nitrosourea (MNU) and it was derived from *japonica* cultivar, Hwacheongbyeo from Crop Molecular Breeding Lab. of Department of Plant Science, Seoul National University. The material was used from a M14 line, 75094, Hwacheong M-6-17N-1-3-2-1-1-1-2-1-1-2-2. This mutant was maintained as heterozygotes because of embryo lethality due to suppressed germination.

Sequence analysis of the candidate genes

The full-length genomic DNA sequence of candidate genes was sequenced by dividing them into several overlapping segments. Specific PCR primers for each segment were used to amplify genomic DNA from Hwacheongbyeo and yellow embryo lethal mutant. PCR products were purified using a PCR purification kit (iNtRON Biotechnology, Korea) for TA cloning. Purified PCR introduced into the pGEM-T Easy Vector (Promega, USA) and transformed into the *E. coli* strain DH5α. The obtained sequences were compared by CodonCode Aligner software (version 1.6.3; CodonCode Coporation). Sequence alignment was performed with the Blast network service in National Center for Biotechnology Information (NCBI), Gramene and The Rice Annotation Project Database (RAP DB).

Gus assay

Histochemical GUS-staining was performed according to the method of Jeon et al. (2000). Mature flowers of the primary transformant were sampled and stained, each tissue type was incubated at 37 for 2 days, then the staining solution was replaced with 95% and 70% (w/v) ethanol in series to remove the chlorophyll. Afterward, we examined the tissues under a dissecting microscope and analyzed their staining patterns.

Vector constructs and rice transformation

In order to generate overexpression vectors, PCR-amplified WT and *yel* full-length cDNAs were inserted into the pCAMBIA 1300-modified vector containing a 35S promoter and a nos terminator. The resulting WT cDNA overexpression construct was denoted 35S::COP1 and the mutant cDNA construct was denoted 35S::yel. In order to generate the RNAi::COP1 construct for COP1 gene suppression, a 342-bp fragment of *yel*-cDNA spanning nucleotides was first cloned into pDONR201 (Invitrogen) and then cloned in sense and antisense directions into the binary transformation vector pH7GWIWG (II) using the Gateway BP and LR clonase enzyme mixes (Invitrogen). pH7GWIWG (II) and derivatives contain the hygromycin resistance (Hyg) gene. For the promoter-GUS assay, the genomic sequence containing the putative promoter region of COP1 was amplified by PCR from the genomic DNA. The COP1 promoter fragment was cloned into the binary vector pHGWFS7. Transgenic plants carrying the

above constructs were generated using wild-type Dongjin (a *japonica* cultivar) seeds via agrobacterium-mediated co-culture methods (Hiei et al., 1997).

UHPLC-MS analysis for yellow pigment

Extracts of yellow mutant and wild-type seed were analyzed using an Agilent 1290 Infinity high performance liquid chromatography (UHPLC) coupled to an Agilent 6490 QQQ with Agilent Jet Stream ESI system. LC condition and MS condition are as follows LC condition; HPLC system : Agilent 1290 Infinity, Column : ACQUITY UPLC BEH Shield RP18 2.1 mm x 100 mm, 1.7 μm, Injection Volume : 2 μL, Temp : 35 ℃, Flow rate : 0.3 mL/min., 55 min. stop time, Mobile phase : A – 0.1 % formic acid in Water : B – 0.1% formic acid in ACN, Gradient condition : 0 min, 5% B; 2 min, 5% B; 40 min, 30% B; 41 min, 90% B; 47 min, 90% B; 47.1 min, 5% B; 55 min, 5% B; MS condition; Mass system: Agilent 6490 QQQ MS, Ion source: Agilent Jet Stream ESI, Nebulizer gas: Nitrogen, Polarity: Negative, Nebulizer pressure: 40 psi, Ion spray voltage: 4000 V, Drying gas temperature: 150 ^{o}C, Drying gas flow rate: 17 L/min, Sheath Gas temp: 300 ^{o}C, Sheath gas flow: 11 L/min, Fragmentor: 380 V, Nozzle voltage: 0 V, High Pressure RF: 150 V, Low Pressure RF: 60 V

Phenolic compounds analysis by UHPLC

The samples were analyzed by ultra-high performance liquid chromatography (UHPLC) analysis performed from Thermo ACCELA UHPLC system (ACCELA UHPLC system, Thermo Fisher Scientific Inc., USA) with a column (Thermo, C18 2.1 × 100 mm, 2.6 μm) used for analysis. The mobile phases were composed of 0.1 % glacial acetic acid in distilled water (solvent A) and 0.1 % glacial acetic acid in ACN (solvent B). The flow rate of the solvent was 500 μL / min and injection volume was 4 μL. The wavelength of the UV detector was 280 nm. The linear gradient of solvent was as follow; 0 min, 98% A: 2% B; 0.50 min, 95% A: 5% B; 2.20 min, 90 % A: 10% B; 5.00 min, 85 % A: 15% B; 7.50 min, 84.3% A: 15.7% B; 8.00 min, 83.4% A: 16.6% B; 9.00 min, 82.2% A: 17.8% B; 9.50 min, 76.1% A: A: 23.9 % B; 14.00 min, 55.0% A: 45.0% B; 15.00 min, 0% A: 100% B; 15.50 min, 0 % A: 100% B; 16.00 min, 98 % A: 2 % B; 25.00 min, 98 % A, 2 % B. The detected time was 15min and the rest time was for equilibrium.

Standards solution

Isoorientin, Orientin, Vitexin, Apigenin, Caffeic acid, Tricetin, Eriodictyol, Luteoin, Catechin, p-coumaric acid, Ferulic acid, m-coumaric acid, Naringin, Myricetin, Kaempferol, Hesperetin, and Biochenin A were purchased from Sigma Aldrich. Chemical Co. (St. Louis, MO, USA). All standards were made stock solutions of 100 ppm the phenolic standards except Biochenin A by dilution in methanol. Biochanin A were prepared by dilution in Dimethyl Sulfoxide (DMSO). Concentrations range for standard curve

was determined with comparing the value of each phenolic compound area in sample extract (3 - 7 point).

Total Phenol Content

Total polyphenol contents were determined as described by Singleton et al. with slight modifications. Briefly, 100 μL of 0.2 N Folin–Ciocalteu's phenol reagents were added to 20 μL of each sample placed in 96-well plates. After 5 min of reaction in the dark, 80 μL of a 7.5% sodium carbonate solution was added to the mixture, which was then incubated for 60 min at room temperature. The absorbance was measured at 750 nm using a microplate reader (Thermo Electron, Spectronic Genesys 6, Madison, WI, USA). The results are presented as gallic acid equivalent concentration (ppm). The standard solution concentration curve ranged from 31.25 ppm to 500 ppm.

Free radical scavenging by the use of the 2, 2 – diphenyl – 1 – picrylhydrazyl (DPPH)

The free radical scavenging activities of the yellow embryo lethal mutant and wild-type seed were measured using the DPPH assay. Reaction mixtures were prepared by mixing 2.5 mL of DPPH solution (0.35 mM DPPH dissolved in 50 % EtOH) and 0.2 mL of each sample. Reaction was incubated for 10 min at room temperature and the changes in DPPH

absorbance at 517 nm were measured with a spectrophotometer. The antioxidant activity was calculated in terms of the per cent inhibition caused by the hydrogen donor activity of each sample.

Free radical scavenging by the use of the ferric reducing antioxidant power (FRAP)

The reducing ability of each sample was determined using the FRAP assay (Hu and Xu, 2011; Kim et al., 2014). Briefly, the FRAP reagent prepared by mixing 300 mM sodium acetate buffer (pH 3.6), 10 mM 2,4,6-tripyridyls-triazine (TPTZ) solution in 40 mM HCl and 20 mM $FeCl_3$ solution at a ratio of 10 : 1 : 1 (v/v/v). Ferric-reducing antioxidant power reagent was prepared fresh daily and pre-warmed in a 37 ℃ water bath prior to use. Then, 0.1 mL of extract was mixed with 1.8 mL of FRAP reagent and 3.1 mL of ultrapure water (HPLC grade). The reaction mixture was incubated at 37 ℃ for 0.5 h and then, the absorption of the mixture was measured at 593 nm with a spectrophotometer. Ferric-reducing antioxidant power values were calculated on the basis of $FeSO_4$ standard curve (100-1000 μM) and reported as μmoles of Fe per gram of dry weight.

Free radical scavenging by the use of the trolox equivalent antioxidant capacity (TEAC)

The Trolox equivalent antioxidant capacity (TEAC) assay, which measures the reduction of the radical cation of ABTS by antioxidants, was carried out as previously described with slight modifications (Hu and Xu, 2011; Kim et al., 2014). Briefly, ABTS+ radical cation was generated by a reaction of 7 mmol/L ABTS and 2.45 mmol/L potassium persulfate. The reaction mixture was allowed to stand in the dark at room temperature for 16-24 h before use and used within 2 days. The ABTS+ solution was diluted with methanol to an absorbance of 0.700 (0.050 at 734 nm. All samples were diluted appropriately to provide 20-80% inhibition of the blank absorbance. Fifty microliters of the diluted extracts was mixed with 1.9 mL of diluted ABTS+ solution. The assay with the mixture was carried out in triplicate, the mixture was allowed to stand for 6 min at room temperature and the absorbance was immediately recorded at 734 nm. Trolox solution (100-1000 μM) was used as a reference standard. The results were expressed as micromoles Trolox equivalents (TE) per gram of DW.

The antioxidant capacity of total phenolic assay by Folin-Ciocalteu reagent

Embryo and endosperm were ground and 1 g of the powder extracted with 80% methyl alcohol for 1 day in a shaker-chamber at 25℃ with stirring. The extracts were filtered through a Whatman No. 42 filter paper and the quantitative analysis of total phenols was measured by the Folin-Dennis method (AOAC, 1980). Half ml of the extract sample, 5 ml of distilled water

and 5 ml of the Folin-Ciocalteau phenol reagent were mixed in a screw-top flask. After 3 min, 2 ml of 10% sodium carbonate (Na_2CO_3) was added and the mixed solution stirred in a lit Shaker-chamber at 30℃for 1 h. After the reaction, the absorbance of the solution was measured with a spectrophotometer (Hitachi Ltd., Tokyo, Japan) at 760 nm. The standard curve was measured with 1, 50 and 100 ppm of ferulic acid purchased from Sigma Aldrich (Mo., U.S.A.).

Amino acid analysis

About 0.3 g of each sample was weighed and mixed with 5 mL 6 N HCl. Hydrolysis was maintained for 24 hr at 110℃ in test tubes that were flushed with nitrogen gas. Afterwards, the samples were diluted to 100 mL with Milli-Q water and filtered through 0.45 μm syringe filters (Waters, Millipore, Milford, MA, USA). One mL of each hydrolysate was placed in an autosampler bottle and injected into an amino acid autoanalyzer (Hitachi L-8800; Hitachi Ltd., Tokyo, Japan). Amino acid concentrations of samples were calculated with reference to an amino acid standard (Ajnomoto-Takara Co., Kyoto, Japan).

Fatty acid analysis

Fatty acid analysis was conducted as described by Rafael and Mancha (1993)(Kim et al., 2013). The procedure was as follows: 0.5 g freeze-dried

GSG was heated with a reagent containing methanol:heptane:benzene:2,2-dimethoxypropane:H_2SO_4 (37:36:20:5:2, v/v). Simultaneous digestion and lipid transmethylation took place in a single phase at 80℃. The fatty acid analysis was performed on a gas chromatograph (6890 N series, Agilent Technologies Inc., Wilmington, DE, USA) equipped with a FID using a HP-Innowax capillary column (0.25 μm, 30 m, J&W Scientific, Agilent Technologies Inc., Wilmington, DE, USA). The initial oven temperature of 150℃ was ramped up by 4℃/min up to the final temperature of 280℃. During the analysis, the temperature of the injector and detector were maintained at 250 and 300℃, respectively, and the air, hydrogen and nitrogen (carrier gas) flow rates were set to 400, 30 and 2 ml/min, respectively. The standard methyl esters of the fatty acid mix (C14-C22) were obtained using a Supelco instrument (Bellefonte, USA). The data obtained from the gas chromatograph were converted to percentage of total fatty acids measured.

Analysis of crude protein and lipid

The protein content of rice was measured according to the Kjeldahl procedure using a protein auto analyzer (Tecator Kjeltec Auto Analyzer 2400; Foss Tecator, Sweden). The lipid content was measured using the Soxtherm Automatic System (Gerhardt, Germany).

RESULTS

Molecular cloning of the *yel* gene

The yellow embryo lethal mutant was obtained by chemical mutagenesis (MNU) and it showed a recessive phenotype of yellow pericarp and embryo lethality in grain (Figure 2-1), thus, it was designated as *yel*. In the previous study, we fine mapped the *yel* locus using a population derived from a cross between *yel* mutant and Milyang23 (*indica* type). The *yel* gene was located on the long arm of chromosome 2 between two flanking markers, S02k-1 and S02k, originating from the BAC clone of AP004058 and the physical distance was 55kb.

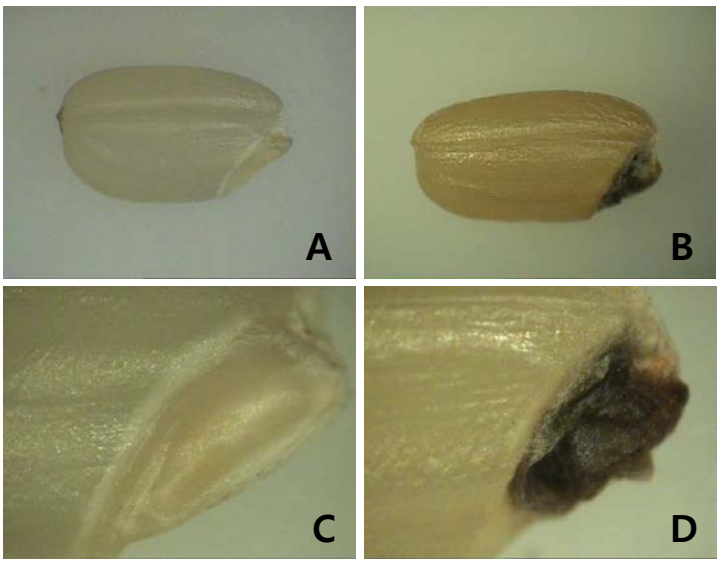


Figure 2-1. Comparison of the brown rice morphology between the wild-type and *yel* mutant. (A) brown rice of wild-type (B) brown rice of *yel* mutant (C) embryo of wild-type seed (D) embryo of *yel* mutant seed; *Yel* mutant seed showed yellow pericarp color and black colored embryo compared to wild-type

Total seven candidate genes were located in this region based on available rice genome annotation (http://tigr.org, http://gramene.org) (Table 2-1). To identify the *yel* gene, we sequenced seven of the candidate genes and compared the sequence variations between the *yel* mutant and its wild-type, Hwacheongbyeo. The sequence analysis indicated that 706-bp deletion was found from 5'UTR to first exon including start codon on the LOC_Os02g53140 (Figure 2-2).

Table 2- 1. Candidate genes analyzed for identification of *yel* gene

No	Locus Name	Gene Description
1	Loc_Os02g53130	nitrate reductase
2	Loc_Os02g53140	COP1 (constitutive photomorphogenesis 1)
3	Loc_Os02g53150	expressed protein
4	Loc_Os02g53160	tyrosine phosphatase family protein
5	Loc_Os02g53169	expressed protein
6	Loc_Os02g53180	1-aminocyclopropane-1-carboxylate oxidase protein
7	Loc_Os02g53200	glucan endo-1,3-beta-glucosidase precursor

In order to determine, whether the 706-bp deletion is associated with mutant phenotype, we conducted the co-segregation analysis with the mutant and wild-type individuals in the F_2 progenies using co-dominant marker (Figure 2-3). This result supported that the 706-bp deletion identified in the LOC_Os02g53140 gene was responsible for the yellow pericarp and embryo lethal phenotype. YEL encodes a protein of comprising RING-finger, coiled-coil, and WD40 domains. All RING-finger

domain as a result of 706-bp deletion was deleted in YEL gene (Figure 2-4).

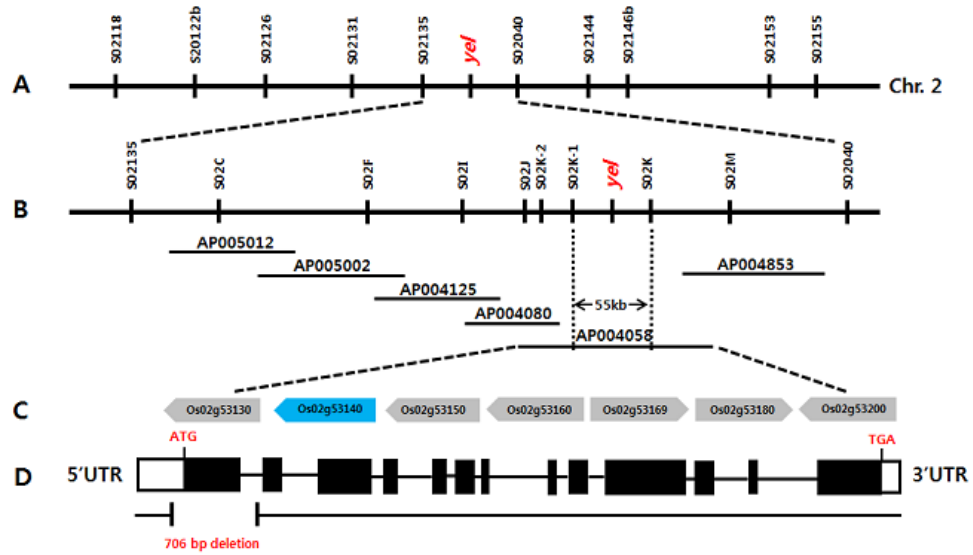


Figure 2-2. Map-based cloning of *yel* gene. (A) Genetic mapping of the *yel* locus with STS markers. (B) Fine mapping of the *yel* locus with additional STS markers. The *yel* locus is located to an approximately 55kb genomic DNA region between S02K-1 and S02K. (C) Seven candidate genes in the 55 kb genomic DNA region identified by fine mapping. (D) The *yel* gene structure at genomic sequence. Thirteen exons and eleven introns are indicated using black rectangle and black line, respectively; the 706-bp deletion is identified from the 5'UTR to first intron included entire first exon.

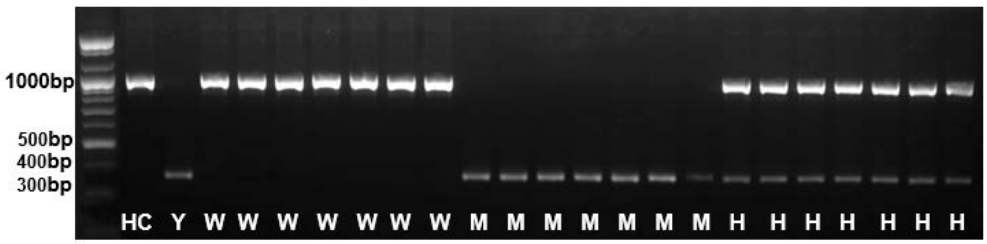


Figure 2-3. Co-segregation analysis in the F_2 progenies. A 706-bp deletion polymorphism were tested for co-segregation using a co-dominant marker. HC: Hwacheong (original parent), Y: *yel* mutant, W: wild-type phenotype, M: mutant phenotype, H: heterozygous phenotype in F_2 progenies.

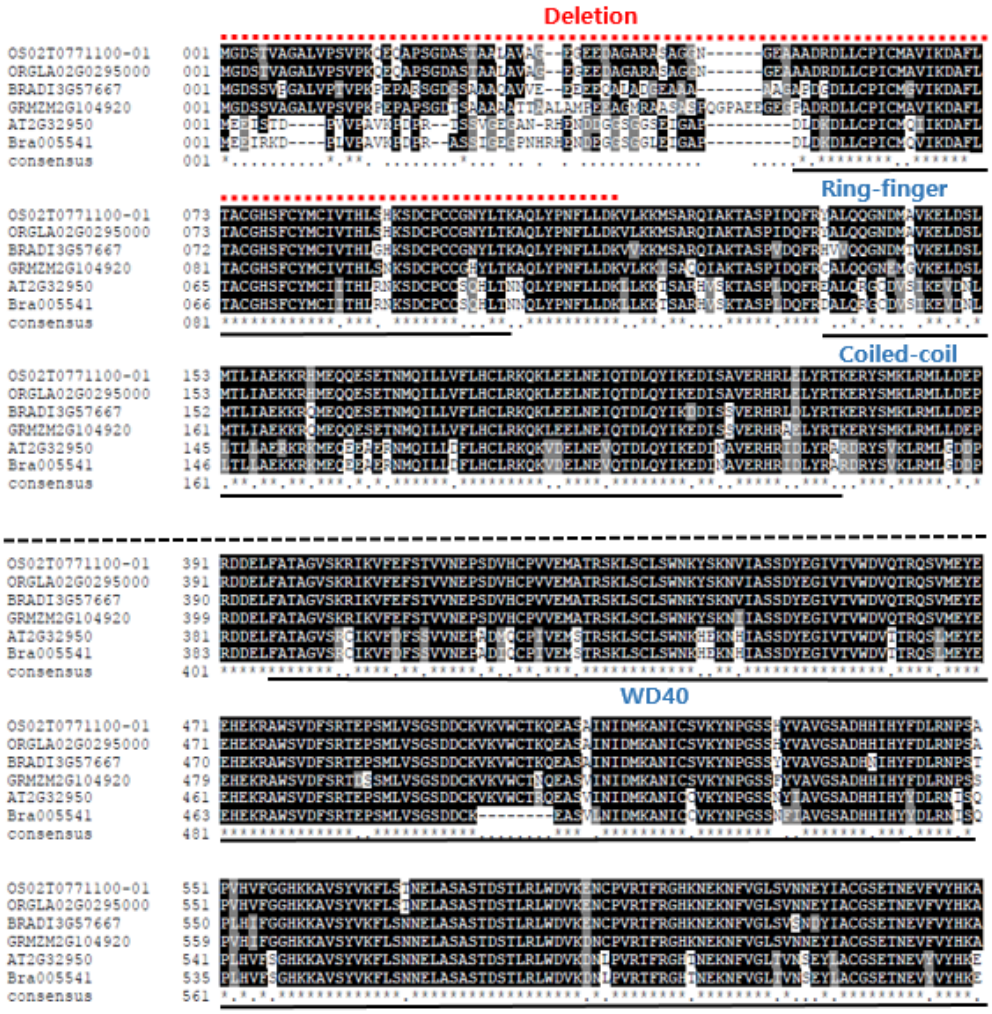


Figure 2-4. Multiple alignment of the amino acid sequences of *YEL* protein. Alignment of YEL with RING-finger, Coiled-coil, and WD40 domains from *Oryza sativa*. *japonica* (OS02T0771100-01), *Oryza glaberrima* (ORGLA02G0295000), *Brachypodium distachyon* (BRADI3G57667), *Zea mays* (GRMZM2G104920), *Arabidopsis thaliana* (AT2G32950), *Brassica rapa subsp. pekinensis* (Bra005541). The positions of domains are indicated by lines below the sequences. Residues

identical in all the sequences are shown by asterisks (*) below the sequences. Residues identical to those in YEL (*Oryza sativa*. *japonica*) with 100% similarity black background and at least 60% similarity are highlighted with a gray background.

Demonstration of the yellow pericarp phenotype conferred by *yel* in transgenic plants

In order to confirm that the 706-bp deletion in the *yel* gene was indeed responsible for the mutant phenotype, transgenic rice plants expressing the mutant *yel* allele under the 35S promoter (35S::*yel*) were generated in the wild-type (WT) Dongjin background (YEL) and their phenotypes were observed. However, all of the transgenic plants didn't show yellow pericarp phenotype or embryo lethal phenotype although, some RNAi transgenic plants showed dark green leaf and short height or sterile and semi-sterile phenotypes (Figure 2-5).

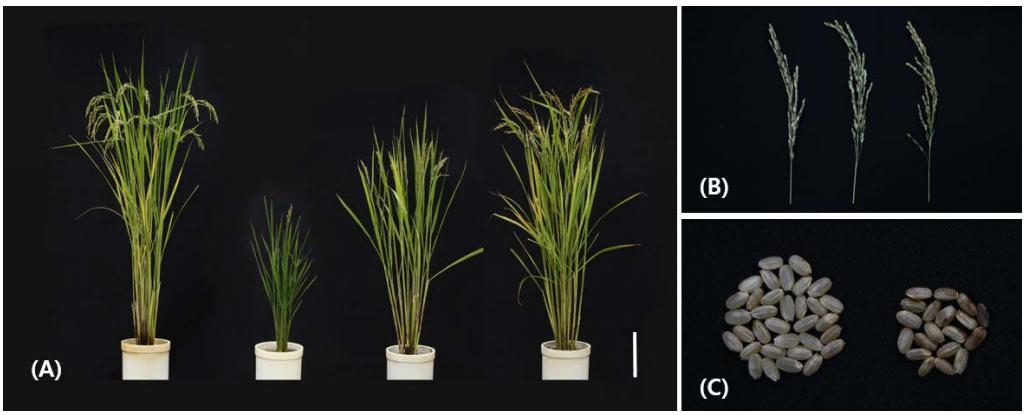


Figure 2-5. RNAi transgenic plants of *yel* mutant (A) Hwacheongbyeo (WT), short and dark green morphology line, sterile line, semi-sterile line (from left to right) (B) morphology of panicle of RNAi lines; steril panicle, semi-sterile panicle, fertile

panicle (from left to right) (C) Seeds from semi-sterile RNAi transgenic lines ; No yellow pericarp or embryo lethal seeds were found. Bar = 20 cm

In contrast, we constructed the transgenic Dongjin (YEL) plants overexpressing the wild-type YEL (35S::YEL) and then crossed with heterozygous plant which is maintained for segregating mutant due to the lethality of mutant type seed. We selected YEL-OE inserted plants among the F_1 plants that crossed between YEL-OE transgenic plant and heterozygous plant afterwards, we observed the segregation of F_2 seeds. Some white pericarp color but still has black embryo (lethal) seeds were segregated in YEL-OE inserted plant (YEL-OE/heterozygous) (Figure 5).

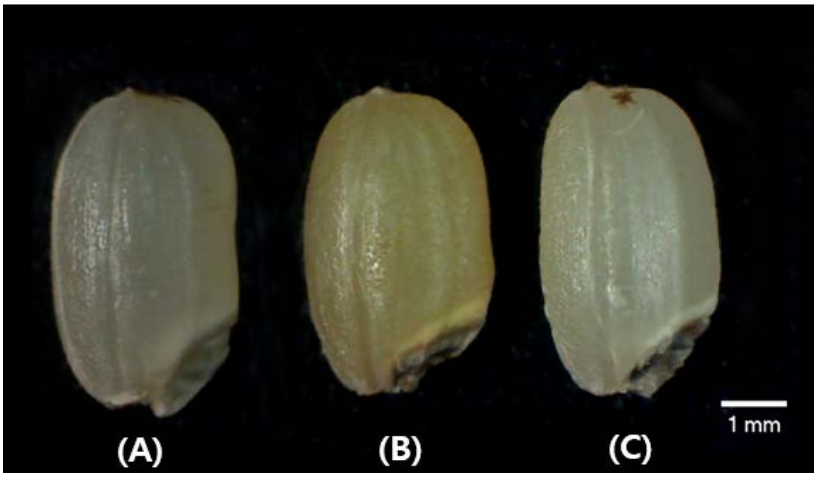


Figure 2-6. Segregated progenies from the cross between YEL-OE transgenic plant and *yel* heterozygous plant (A) Wild-type seed (B) *yel* homozygous seed (mutant type) (C) Recovered seed by YEL-OE

To confirm whether the all the white lethal seeds have YEL-OE and *yel* homozygous allele at the same time, we genotyped the F_2 seeds. The result showed that all the white lethal seeds had YEL-OE insert and *yel* allele. The

segregation ratio of yellow lethal, white lethal, and normal seeds was fitted to the expected ratio of 1 : 3 : 12 by chi-square test (Table 2). It is conceivable that the yellow pericarp phenotype may have been the result of loss of function in rice *yel* gene**.**

Table 2-2. The segregation ratio of F_2 seed in YEL-OE inserted plant

Panicle	Yellow lethal	White lethal	Normal	Total
1	8	22	60	90
2	5	16	59	80
3	8	15	55	78
4	4	3	44	51
5	4	13	44	61
6	4	6	33	43
7	3	7	35	45
8	3	13	64	80
Total	39	95	394	528

X^2-1.26, p-0.53

Expression pattern of *yel* in the different organs and tissues of rice plants

In order to better understand the role of the *yel* gene in rice development, the expression pattern of YEL in different organs and tissues was examined using a beta-glucuronidase reporter gene under the control of the YEL promoter. GUS staining was primarily detected in the pollen, embryo, pericarp, and node (Figure 2-7 C, E, G and H). Relatively strong

GUS staining signals were detected in the differentiation zone and meristematic region of the stem while no GUS expression was detected in the spikelet, leaf, stigma and endosperm (Figure 2-7 A, C and D).

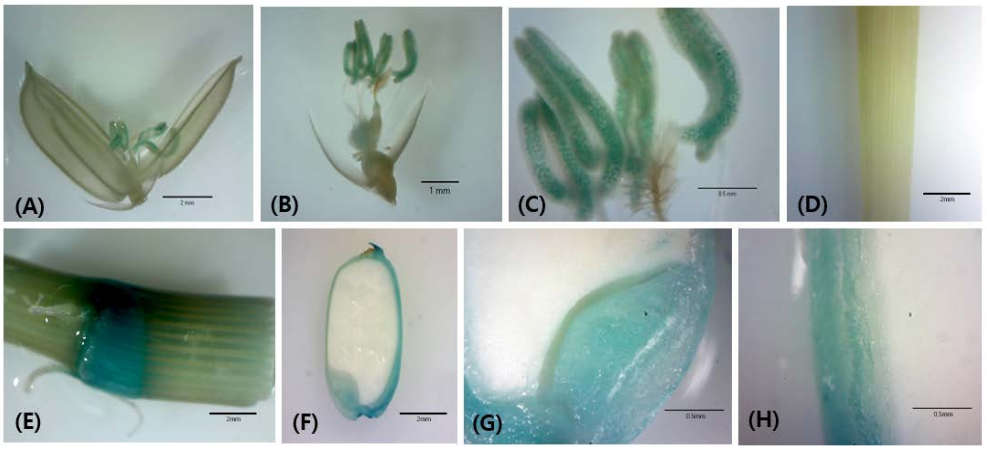
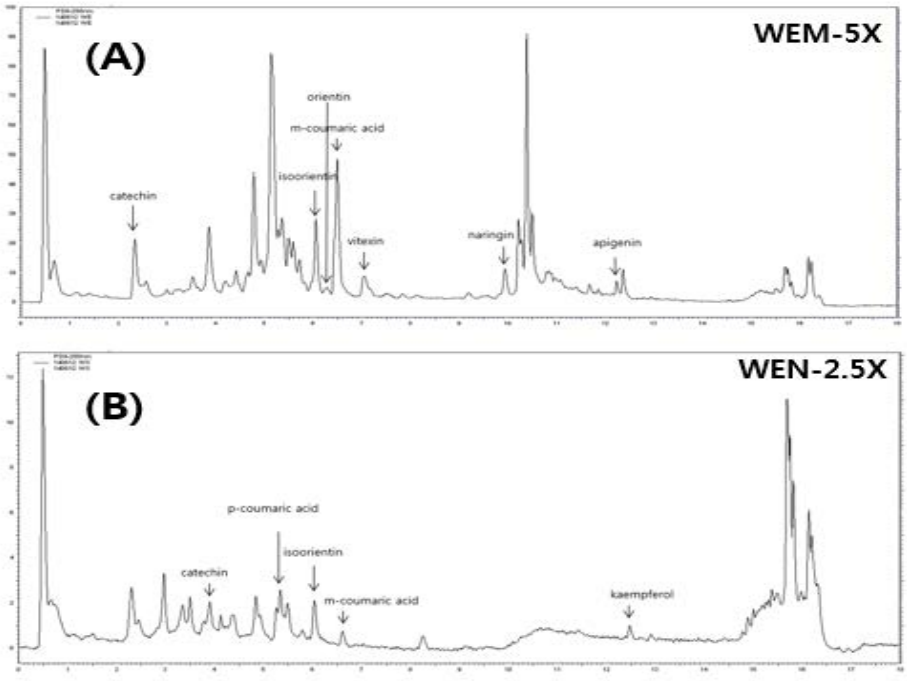


Figure 2-7. Expression patterns of the YEL gene. (A to H) GUS activity detected in the YEL promoter::GUS transgenic plants. (A) spikelet (B) floral organ (C) pollen (D) leaf (E) node (F) seed (G) embryo (H) pericarp

Analysis of phenolic compounds in embryo and endosperm

Yellow pericap color is very rare in rice grain. In order to screen and determine the phenolic compounds contents in the dried embryo (black color) and endosperm (yellow color), we compared UHPLC chromatograms of the embryo and endosperm extracts of wild-type and yellow mutant, respectively. This study was analyzed by ultra-high performance liquid chromatography (UHPLC) analysis performed using 16 phenolic standards such as catechin, p-coumaric acid, m-coumaric acid, naringin, apigenin, caffeic acid, myricetin, tricetin, isoorientin, orientin, vitexin, eriodictyol,

luteolin, kaempferol, hesperetin, biochanin A (Figure 2-8). To get the desirable calibration curve values, quantification to individual phenolic compound was conducted and all standards with each concentration were determined using in triplicate and average value. We assigned individual peak of compound using spike. The result showed that catechin, isoorientin/orientin, and vitexin/isovitexin were strongly detected in yellow embryo and endosperm than those of wild-type amomg the 16 phenolic compounds. Isoorientin/orientin content in yellow endosperm was higher than that of wild-type endosperm (Table 2-3). These results suggest that catechin, isoorientin/orientin, vitexin/isovitexin might play an important role in the yellow color of pericap and black color of embryo in mutant seed.



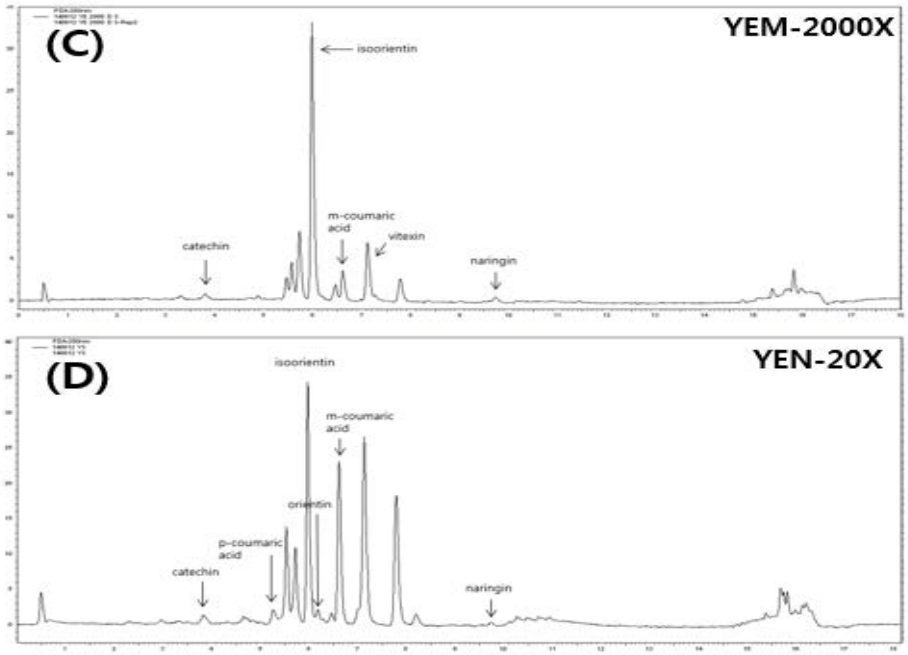


Figure 2-8. UHPLC chromatograms of 16 phenolic compounds in embryo and endosperm. (A) WEM (wild-type embryo), (B) WEN (wild-type endosperm), (C) YEM (yellow embryo), (D) YEN (yellow endosperm). X means dilution

Table 2-3. Contents of major phenolic compounds in embryo and endosperm of mutant and wild-type seed

	Catechin	p-coumaric acid	Isoorientin/Orientin	m-coumaric acid	Vitexin/Isovitexin	Naringin	Apigenin
WEM	143.6±22.7	nd	43.6±2.1	30.3±1.6	10.6±1.4	23.3±2.0	3±0.2
WEN	6.5±0.2	0.6±0	3.3±0.1	nd	nd	nd	nd
YEM	4614.5±152.8	nd	21681.8±396.5	nd	5006.4±164.6	666.7±125.7	nd
YEN	48±2.6	4.4±0.3	226.2±6.2	48.9±1.9	nd	7.11±0.5	nd

Unit: μg/g

Mean value ± SD (n = 3), nd: not detectable

Qualitative and Quantitative analysis of *yel* mutant seed

After screening the content of phenolic compound, we narrowed the candidate metabolites corresponding to yellow pericarp and black embryo color in *yel* mutant. Isoorientin and vitexin were the major compounds of the *yel* mutant grain thus, we analyzed quantitative and qualitative characteristics of isoorientin and vitexin include their isomers, orientin and isovitexin using UHPLC/MS system. To determin which pigment is correspond to yellow and black pigmentation in pericarp and embryo, respectively among the four candidate pigments and to confirm the accumulated amount of pigments at the different rice seed part, we analyzed the extracts of embryo and endosperm of *yel* mutant and wild-type seed, seperately using UHPLC/MS. The result showed that orientin extremely accumulated in the embryo (24965464.26±988976.36 μg g^{-1}) of mutant seed than that of wild-type. Although the total orientin content of endosperm in mutant seed was relatively lower than embryo, orientin was the major pigment in mutant endosperm as well as embryo. This results indicates that orientin plays an important role to have the black color of embryo and yellow color of endosperm in mutant seed (Figure 2-9). Isovitexin was the second highly accumulated pigment in embryo (4093988.68±216424.20 μg g^{-1}) and endosperm (139862.78±4978.00 μg g^{-1}) of mutant seed following orientin. The amount of isovitexin was relatively lower than orientin, however, the concentration of isovitexin also was extremely higher in mutant seed than wild-type. Isovitexin also one of the important pigment for yellowing mutant seed (Figure 2-9).

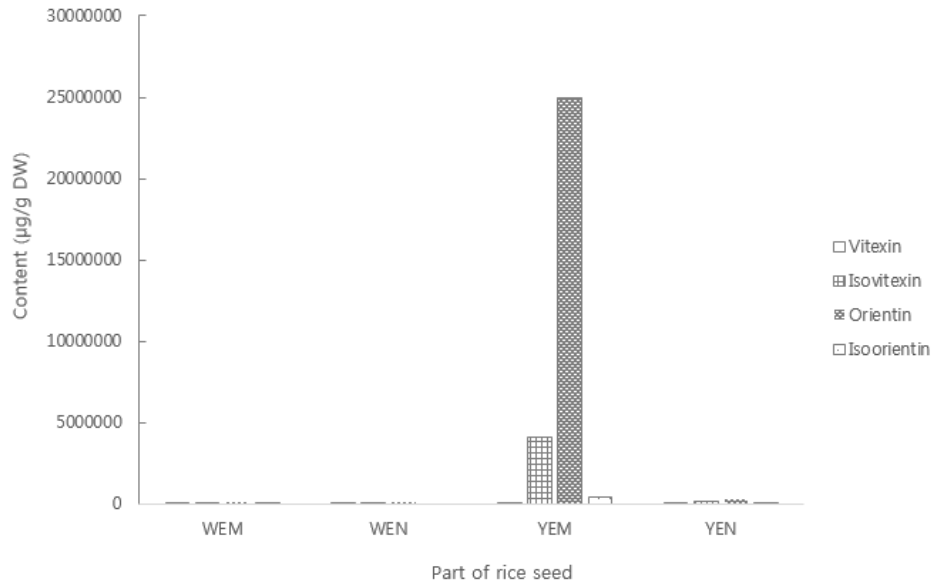


Figure 2-9. Accumulation of four candidate flavonoid in embryo and endosperm of *yel* mutant and wild-type seed. WEM (wild-type embryo), WEN (wild-type endosperm), YEM (yellow embryo), YEN (yellow endosperm).

Determination of antioxydant capacities by DPPH, FRAP, TAEC assay

The DPPH (1,1-Diphenyl-2-picrylhydrazyl) and FRAP (Ferric reducing antioxidant power) and TAEC (Trolox Equivalent Antioxidant Capacity) antioxidant activities were measured. The extracts from whole grain, embryo, and endosperm of wild-type and *yel* mutant were examined and

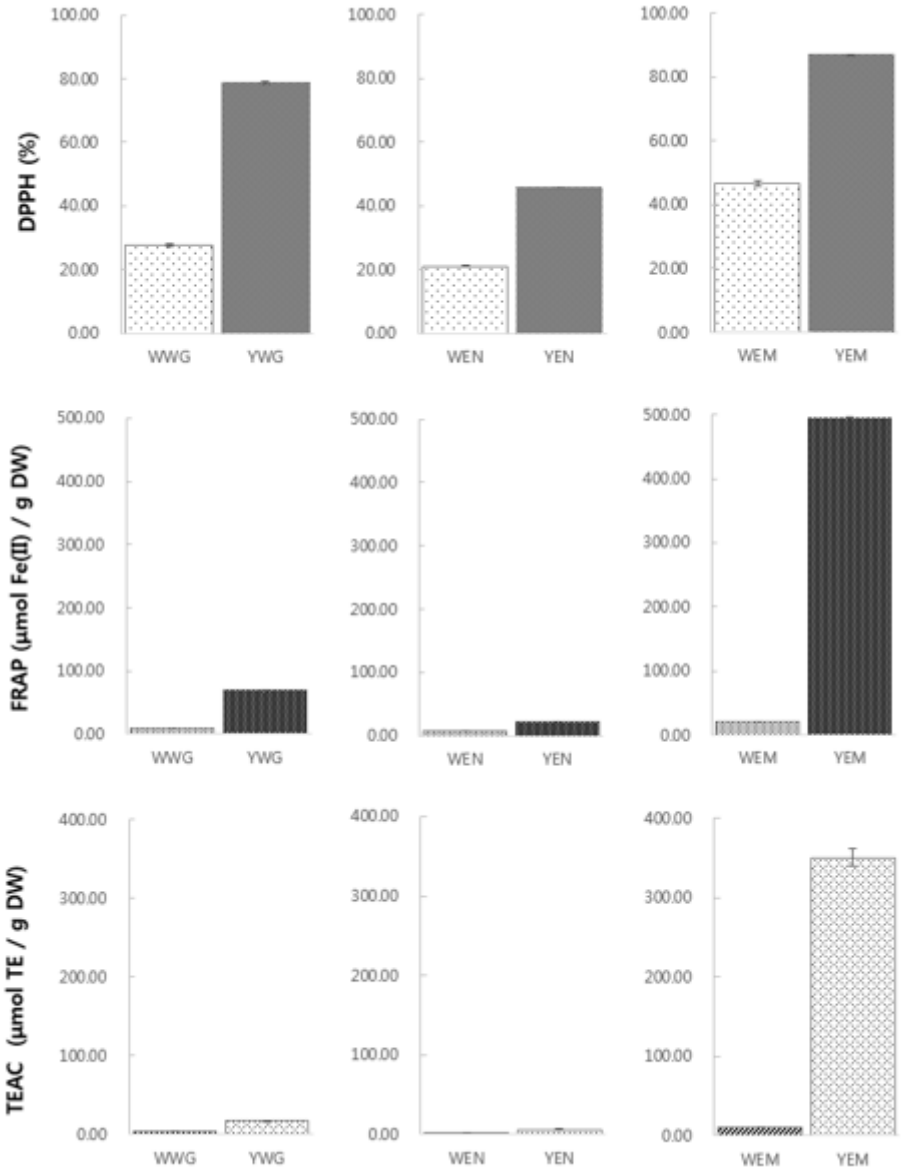


Figure 2-10. Antioxidant activity of *yel* mutant determined by DPPH, FRAP, and TEAC assay. WWG (wild-type whole grain), YWG(yellow whole grain) WEM (wild-type embryo), WEN (wild-type endosperm), YEM (yellow embryo), YEN (yellow endosperm).

compared to their free radical scavenging activities against DPPH• radical. As shown in Figure 2-10, the highest radical scavenging activity was found in the extract of *yel* mutant embryo (87.06±0.19 %). DPPH radicals scavenging activity was higher in all part such as whole grain, embryo, and endosperm of *yel* mutant. In addition, extract from mutant embryo had a greater reducing power (495.38±1.00 μmol Fe(II) / g) than extract from wild-type embryo (20.79±0.29 μmol Fe(II) / g) in the FRAP assay. Extremely high antioxidant activity was also observed in *yel* mutant embryo (350.03±11.18 μmol TE / g) in TEAC assay. These results showed that orientin and isovitexin which highly accumulated in *yel* mutant embryo, might be attribute to high level of antioxidant activity (Figure 2-10).

Total phenolic contents of *yel* mutant seed

Total phenolic compounds of *yel* mutant seed were also measured. Total contents of phenolic compounds were higher in embryo, endosperm, and whole grain of *yel* mutant seed than those of wild-type. In common with anitioxidant activity assay, extremely high level of total phenolic contents were detected in *yel* mutant embryo (46.28±0.87 mg/g, ferulic acid equivalent) (Table 2-4). This indicates that the phenolic compounds are the major constitutuents in mutant embryo and endosperm contributing to antioxidant activity such as DPPH, FRAP, and TEAC.

Table 2-4. Total phenolic contents of the *yel* mutant and wild-type seed

	Total phenol (mg/g, ferulic acid equivalent)	
	WT	*yel*
Whole grain	1.25±0.04	3.68±0.07
Endosperm	0.97±0.02	1.79±0.00
Embryo	2.62±0.05	46.28±0.87

Mean value ± SD (n = 3)

Content and composition of fatty acids in *yel* mutant seed

The fatty acid content of *yel* mutant seed is shown in Table 2-5. The content of saturated fatty acid was relatively high in *yel* mutant embryo, endosperm, and whole grain while the contents of unsaturated fatty acid was high in embryo, endosperm, and whole grain of wild-type. The contents of unsaturated fatty acid such as oleic acids (C18:1) and linoleic acid (C18:2) were dramatically decreased however, the content of saturated fatty acid, palmitic acid (C16:0) was increased more than two folds in the embryo of *yel* mutant than wild-type (Table 2-5).

Table 2-5. Contents of fatty acid of *yel* mutant and wild-type seed

	Contents of fatty acids (%)							
	Saturated fatty acid				**Unsaturated fatty acid**			
	Palmitic acid	Stearic acid	Arachidic acid	Total	Oleic acid	Linoleic acid	Linolenic acid	Total
	C16:0	C18:0	C20:0		C18:1	C18:2	C18:3	
WWG	22.7±0.6	1.4±0.4	1.3±0.4	25.4±0.7	36.7±1.1	36.7±0.4	1.2±0.0	74.6±0.7
YWG	30.7±0.4	2.2±0.9	2.4±0.9	35.2±2.1	33.8±1.5	29.8±0.6	1.2±0.0	64.8±2.1
WEN	25.9±0.6	1.3±0.5	1.1±0.6	28.2±0.9	34.6±0.8	36.2±0.7	1.0±0.0	71.8±0.9
YEN	29.4±2.1	1.4±0.1	1.4±0.0	32.2±2.2	31.3±2.6	35.0±0.3	1.6±0.1	67.8±2.2
WEM	22.0±0.7	1.4±0.3	0.6±0.1	24.0±0.4	33.8±2.6	40.2±3.1	2.0±0.2	76.0±0.4
YEM	48.8±1.0	5.5±0.3	4.4±0.3	58.7±0.6	17.5±0.6	20.3±0.3	3.6±0.1	41.3±0.6

Mean value ± SD (n = 3)

Crude fat and protein contents analysis of *yel* mutant seed

Crude fat and protein contents were measured in *yel* mutant seed. The content of crude fat in *yel* mutant embryo (3.12±0.10 %) was decreased than that of wild-type (19.24±0.14 %). Crude protein contents of the whole grain (7.04±0.07 %) and endosperm (6.47±0.04 %) of *yel* mutant were higher than that of wild-type while, the crude protein content in the embryo of wild-type (18.72±0.07 %) was higher than that of *yel* mutant embryo (18.36±0.08 %) (Table 2-6).

Table 2-6. Crude fat and protein content analysis of *yel* mutant and wild-type seed

	Crude fat content (%)		**Crude protein content** (%)	
	WT	*yel*	WT	*yel*
Whole grain	2.26±0.08	1.37±0.07	6.40±0.02	7.04±0.07
Endosperm	1.31±0.05	1.28±0.01	6.15±0.02	6.47±0.04
Embryo	19.24±0.14	3.12±0.10	18.72±0.07	18.36±0.08

Mean value ± SD (n = 3)

DISCUSSION

Comparison of *yel* mutant and other cop1 related phenotype

We isolated a *yellow embryo lethal* (*yel*) gene, which exhibits a yellow colored pericarp and lethal embryo. Map-based cloning revealed that the *YEL* is an ortholog of *Arabidopsis CONSTITUTIVE PHOTOMOTPHOGENIC 1* (*COP1*) gene encodes a transcription factor which plays an important role in light regulated growth, and has the Ring-finger, coiled-coil, and WD40 repeat domain. This COP1 protein is conserved in both higher plants and mammalians. In plant, COP1 acts as an E3 ubiquitin ligase to repress light signaling by targeting photoreceptors and downstream transcription factors for ubiquitination and degradation. COP1 is a known repressor of photomorphogenesis in darkness as well as in light. *cop1* mutants have a light-grown phenotype in darkness, show features of enhanced photomorphogenesis in light (Yi and Deng, 2005; Oravecz et al, 2006). At the molecular level, COP1 targets different photomorphogenesis-promoting transcription factors for degradation in darkness (Osterlund et al, 2000; Saijo et al, 2003).

Our study showed that unlike the *cop1* Arabidopsis mutant, the *yel* mutant didn't exhibit photomorphogenesis both under dark condition and light conditions. Instead of, *yel* mutant showed the yellow colored pericarp and lethal embryo in the seed. The yellow pigment in pericarp and embryo of *yel* mutant seed revealed that the yellow pigment is orientin, one of the flavonoid in plant. We couldn't observe COP1 related any function in our

mutant. Thus, YEL appears to have acquired novel functions associated with flavonoid biosynthesis during the evolution of rice.

Interestingly, one feature of the constitutive photomorphogenesis phenotype of *cop1* is the increased accumulation of anthocyanins (Deng et al., 1991; Hoecker et al., 1998; Laubinger et al., 2004). In Arabidopsis, anthocyanin accumulation occurs only in the light and this accumulation is further enhanced by a number of environmental stress factors, such as cold, drought, pathogen attack and nutrient depletion (Chalker-Scott, 1999). The biosynthesis of anthocyanins is regulated by transcription factors that induce the expression of structural genes that code for enzymes in the biosynthesis pathway (Broun, 2005; Li, 2014). These transcription factors include HY5 and a complex consisting of a WD-repeat protein, a bHLH protein, and a myeloblastosis (MYB) protein (Feller et al., 2011; Koes et al., 2005; Lee et al., 2007; Shin et al., 2007)

Another characteristics of *yel* mutant is the embryo lethality. The *yel* mutant gene was deleted 706-bp from 5′ UTR to first intron includes first exon and start codon. It means that the Ring-finger domain of N-terminal of YEL protein is deleted. To get the *yel* mutant plant, we conducted anther culture however, we were not able to get a homozygous mutant plant. All the generated plants were homozygous wild-type plants of YEL. Introduction of an N-terminal fragment of Arabidopsis COP1 (AtCOP1) containing the RING-finger and coiled-coil domains into a cop1 null allele rescued its lethal phenotype, indicating that the AtCOP1 N-terminal region alone is able to sustain a basal function during development (Stacey et al.,

2000). The AtCOP1 C-terminal WD40 domain, by contrast, led to repression of photomorphogenesis when expressed in a wild-type background but failed to complement a *cop1* loss-of-function allele (Stacey et al., 2000). We concluded that the absence of Ring-finger domain in the YEL protein caused embryo lethality or defect of cell division. The *pps*-2 mutant which consists of a single base change (G to A) in the eighth exon of the gene, generating a premature stop codon, also couldn't germinate and only ~5% of seeds could germinate on nutrient medium containing sucrose but died in a week (Tanaka et al., 2011).

Accumulation of Orientin, flavone-C-glycoside

Rice is a staple food throughout the world. However, rice seeds do not contain most of flavonoids, except for a small amount of tricin found in bran. The biosynthetic genes involved in flavonoid synthesis are hardly expressed in endosperm while in embryos, PAL, CHI, CHS, and FNS, which are thought to be involved in tricin biosynthesis, are expressed (RiceXpro) (Ogo et al., 2013).

Our UHPLC/MS analysis reveal that the yellow pigment in the pericarp and embryo of *yel* mutant seed was orientin (Luteolin-8-C-glucoside). Yellow colored pericarp rice was rarely known so far. Oryzamutaic acids were detected in yellow endosperm mutant of the variety Hatsuyamabuki however, the biosynthetic pathway remains still unknown. This mutant showed the yellow endosperm which was caused by mutation of

Oryzamutaic acid A and *OsALDH7* (Nakano et al., 2009; Nakano et al., 2011; Shen et al., 2012) Mutation of *OsALDH7* resulted in accumulation of yellow pigment in endosperm. Oryzamutaic acid A accumulates during late seed development and after a year-long storage in the endosperm. Moreover, high level of orientin such as the *yel* mutant was not reported in rice grain. Embryo of *yel* mutant colored black not yellow because extremely high level of yellow colored orientin accumulated in small area of embryo. Changes in biosynthetic pathways or related metabolism have been reported in various mutant plants and various novel compounds may accumulate in plants with different organs and concentrations. We still couldn't determine the reason of the lethality of *yel* mutant embryo. It could be result from high level of pigmentation in embryo cell or null fucntion of *yel* allele. Our data denomstrated that the single recessive gene controlling high concentration of orientin in embryo and endosperm. In order to determine the consequences of unscheduled accumulation of flavonoid, Misera et al. constructed double mutants of fusca with mutant alleles of several genes involved in flavonoid biosynthesis. Although the double mutants lacked flavonoids, the mutant seedlings were nonetheless lethal, indicating that the lethality of fusca seedlings is not caused by excessive amounts of flavonoids (Misera et al., 1994).

Antioxidant activity of *yel* mutant embryo

Several assays have been frequently used to estimate antioxidant

capacities in fresh fruits and vegetables and their products and foods for clinical studies including 2,2-azinobis (3-ethyl-benzothiazoline-6-sulfonic acid) (ABTS) (Leong and Shui, 2002; Miller and Rice-Evans, 1997), 2,2-diphenyl-1-picrylhydrazyl (DPPH) (Brand-Williams et al., 1995; Gil et al., 2002), ferric reducing antioxidant power (FRAP) (Benzie and Strain, 1999; Guo et al., 2003; Jimenez-Escrig et al., 2001), and the oxygen radical absorption capacity (ORAC) (Cao et al., 1993; Ou et al., 2001; Prior et al., 2003). Ou et al. (2002) reported no correlation of antioxidant activity between the FRAP and ORAC techniques among most of the 927 freeze-dried vegetable samples, whereas these methods revealed high correlation in blueberry fruit (Connor et al., 2002). Similarly, Awika et al. (2003) observed high correlation between ABTS, DPPH, and ORAC among sorghum and its products (Thaipong et al., 2006). In our study, there was a significant correlation between total phenolic content of the *yel* mutant extracts and their DPPH• scavening activities, FRAP, and TEAC. This result indicates that the phenolic compounds are the major constitutuents in mutant embryo and endosperm contributing to DPPH• sacvening activity, FRAP, and TEAC.

Reference

Abdel-Aal el, S.M., Young, J.C., and Rabalski, I. (2006). Anthocyanin composition in black, blue, pink, purple, and red cereal grains. Journal of agricultural and food chemistry *54*, 4696-4704.

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M.*, et al.* (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nature biotechnology *30*, 174-178.

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. New biotechnology *25*, 195-203.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H., and Matsuoka, M. (2005). Cytokinin oxidase regulates rice grain production. Science *309*, 741-745.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G., and Edwards, R. (2009). The C-glycosylation of flavonoids in cereals. The Journal of biological chemistry *284*, 17926-17934.

Broun, P. (2005). Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Current opinion in plant biology *8*, 272-279.

Brown, B.A., Cloix, C., Jiang, G.H., Kaiserli, E., Herzyk, P., Kliebenstein, D.J., and Jenkins, G.I. (2005). A UV-B-specific signaling component orchestrates plant UV protection. Proc Natl Acad Sci U S A *102*, 18225-18230.

Brown, M.E., and Funk, C.C. (2008). Climate - Food security under climate change. Science *319*, 580-581.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J.*, et al.* (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature biotechnology *26*, 1301-1308.

Catala, R., Medina, J., and Salinas, J. (2011). Integration of low temperature and light signaling during cold acclimation response in Arabidopsis. P Natl Acad Sci

USA *108*, 16475-16480.
Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol *70*, 1-9.
Chen, H., He, H., Zhou, F., Yu, H., and Deng, X.W. (2013). Development of genomics-based genotyping platforms and their applications in rice breeding. Current opinion in plant biology *16*, 247-254.
Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y.*, et al.* (2014). A high-density SNP genotyping array for rice biology and molecular breeding. Molecular plant *7*, 541-553.
Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J.*, et al.* (2008a). A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. Proc Natl Acad Sci U S A *105*, 11436-11441.
Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J.*, et al.* (2008b). A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice. Proceedings of the National Academy of Sciences *105*, 11436-11441.
Cheng, S.H., Zhuang, J.Y., Fan, Y.Y., Du, J.H., and Cao, L.Y. (2007). Progress in research and development on hybrid rice: a super-domesticate in China. Annals of botany *100*, 959-966.
Christie, J.M., Arvai, A.S., Baxter, K.J., Heilmann, M., Pratt, A.J., O'Hara, A., Kelly, S.M., Hothorn, M., Smith, B.O., Hitomi, K.*, et al.* (2012). Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. Science *335*, 1492-1496.
Chung, G.S., and Heu, M.H. (1980). Innovative approaches to rice breeding, pp. 139.
Chung, G.S., and Heu, M.H. (1991). Improvememt of Tongil-Type Rice Cultivars from *Indica*/*Japonica* Hybridization in Korea. In Biotechnology in Agriculture and Forestry 14, B. Y.P.S., ed. (Germany: Springer-Verlag Berlin Heidelberg), pp. 105-112.
Collard, B.C., and Mackill, D.J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical transactions of the Royal Society of London Series B, Biological sciences *363*, 557-572.

Cominelli, E., Gusmaroli, G., Allegra, D., Galbiati, M., Wade, H.K., Jenkins, G.I., and Tonelli, C. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in Arabidopsis thaliana. Journal of plant physiology *165*, 886-894.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674-3676.

Craig, D.W., Pearson, J.V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J.J., Pawlowski, T.L., Laub, T., Nunn, G., Stephan, D.A.*, et al.* (2008). Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods *5*, 887-893.

Crocco, C.D., Holm, M., Yanovsky, M.J., and Botto, J.F. (2010). AtBBX21 and COP1 genetically interact in the regulation of shade avoidance. The Plant journal : for cell and molecular biology *64*, 551-562.

Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res *36*.

Da Silva, R.Z., Yunes, R.A., de Souza, M.M., Delle Monache, F., and Cechinel-Filho, V. (2010). Antinociceptive properties of conocarpan and orientin obtained from Piper solmsianum C. DC. var. solmsianum (Piperaceae). Journal of natural medicines *64*, 402-408.

Dan, Z.W., Liu, P., Huang, W.C., Zhou, W., Yao, G.X., Hu, J., Zhu, R.S., Lu, B.R., and Zhu, Y.G. (2014). Balance between a Higher Degree of Heterosis and Increased Reproductive Isolation: A Strategic Design for Breeding Inter-Subspecific Hybrid Rice. Plos One *9*.

Deng, X.W., Caspar, T., and Quail, P.H. (1991). cop1: a regulatory locus involved in light-controlled development and gene expression in Arabidopsis. Genes Dev *5*, 1172-1182.

Deng, X.W., Matsui, M., Wei, N., Wagner, D., Chu, A.M., Feldmann, K.A., and Quail, P.H. (1992). COP1, an Arabidopsis regulatory gene, encodes a protein with both a zinc-binding motif and a G beta homologous domain. Cell *71*, 791-801.

Dietrych-Szostak, D., and Oleszek, W. (1999). Effect of processing on the flavonoid content in buckwheat (Fagopyrum esculentum Moench) grain. Journal of agricultural and food chemistry *47*, 4384-4387.

Dixon, R.A., and Pasinetti, G.M. (2010). Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. Plant Physiol *154*, 453-457.

Duthie, G.G., Duthie, S.J., and Kyle, J.A. (2000). Plant polyphenols in cancer and heart disease: implications as nutritional antioxidants. Nutrition research reviews *13*, 79-106.

Dyachok, J., Zhu, L., Liao, F., He, J., Huq, E., and Blancaflor, E.B. (2011). SCAR mediates light-induced root elongation in Arabidopsis through photoreceptors and proteasomes. Plant Cell *23*, 3610-3626.

Dykes, L., and Rooney, L.W. (2007). Phenolic compounds in cereal grains and their health benefits. Cereal Food World *52*, 105-111.

Falconer, D.S., Mackay, T.F., and Frankham, R. (1996). Introduction to Quantitative Genetics (4th edn). Trends in Genetics *12*, 280.

Famoso, A.N., Zhao, K., Clark, R.T., Tung, C.W., Wright, M.H., Bustamante, C., Kochian, L.V., and McCouch, S.R. (2011). Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS genetics *7*, e1002221.

Favory, J.J., Stec, A., Gruber, H., Rizzini, L., Oravecz, A., Funk, M., Albert, A., Cloix, C., Jenkins, G.I., Oakeley, E.J*., et al.* (2009). Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in Arabidopsis. The EMBO journal *28*, 591-601.

Feller, A., Machemer, K., Braun, E.L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. The Plant journal : for cell and molecular biology *66*, 94-116.

Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N., and Paterson, A.H. (2004). An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome research *14*, 1812-1819.

Field, B., and Osbourn, A.E. (2008). Metabolic diversification--independent assembly of operon-like gene clusters in different plants. Science *320*, 543-547.

Fitzgerald, M.A., McCouch, S.R., and Hall, R.D. (2009). Not just a grain of rice: the quest for quality. Trends Plant Sci *14*, 133-139.

Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S.t. (2003). Structure of linkage disequilibrium in plants. Annual review of plant biology *54*, 357-374.

Gao, Q., Yue, G.D., Li, W.Q., Wang, J.Y., Xu, J.H., and Yin, Y. (2012). Recent Progress Using High-throughput Sequencing Technologies in Plant Molecular Breeding. J Integr Plant Biol *54*, 215-227.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S., and McCouch, S. (2005). Genetic structure and diversity in Oryza sativa L. Genetics *169*, 1631-1638.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science *296*, 92-100.

Gonzalez, A., Zhao, M., Leavitt, J.M., and Lloyd, A.M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. The Plant journal : for cell and molecular biology *53*, 814-827.

Gross, B.L., and Zhao, Z.J. (2014). Archaeological and genetic insights into the origins of domesticated rice. P Natl Acad Sci USA *111*, 6190-6197.

Hackett, C.A., and Broadfoot, L.B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity *90*, 33-38.

Hanson, W.D. (1959). Early Generation Analysis of Lengths of Heterozygous Chromosome Segments around a Locus Held Heterozygous with Backcrossing or Selfing. Genetics *44*, 833-837.

Harborne, J.B. (1986). Nature, distribution and function of plant flavonoids. Progress in clinical and biological research *213*, 15-24.

Harushima, Y., Kurata, N., Yano, M., Nagamura, Y., Sasaki, T., Minobe, Y., and Nakagahra, M. (1996). Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. Theoret Appl Genetics *92*, 145-150.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent

advances in the transcriptional regulation of the flavonoid biosynthetic pathway. Journal of experimental botany *62*, 2465-2483.

Hiei, Y., Komari, T., and Kubo, T. (1997). Transformation of rice mediated by Agrobacterium tumefaciens. Plant Mol Biol *35*, 205-218.

Hoecker, U., and Quail, P.H. (2001). The phytochrome A-specific signaling intermediate SPA1 interacts directly with COP1, a constitutive repressor of light signaling in Arabidopsis. The Journal of biological chemistry *276*, 38173-38178.

Hoecker, U., Xu, Y., and Quail, P.H. (1998). SPA1: a new genetic locus involved in phytochrome A-specific signal transduction. Plant Cell *10*, 19-33.

Holm, M., Ma, L.G., Qu, L.J., and Deng, X.W. (2002). Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in Arabidopsis. Genes Dev *16*, 1247-1259.

Hospital, F. (2005). Selection in backcross programmes. Philosophical transactions of the Royal Society of London Series B, Biological sciences *360*, 1503-1511.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T.*, et al.* (2009). High-throughput genotyping by whole-genome resequencing. Genome research *19*, 1068-1076.

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

Huang, X., Ouyang, X., Yang, P., Lau, O.S., Li, G., Li, J., Chen, H., and Deng, X.W. (2012b). Arabidopsis FHY3 and HY5 positively mediate induction of COP1 transcription in response to photomorphogenic UV-B light. Plant Cell *24*, 4590-4606.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z.*, et al.* (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nature genetics *42*, 961-967.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C.*, et al.* (2012c). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature genetics *44*, 32-39.

Hultin, P.G. (2005). Bioactive C-glycosides from bacterial secondary metabolism.

Current topics in medicinal chemistry *5*, 1299-1331.
Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.
IRRI (2009). IRRI World Rice Statistics.
Jang, I.C., Yang, J.Y., Seo, H.S., and Chua, N.H. (2005). HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling. Gene Dev *19*, 593-602.
Jang, S., Marchal, V., Panigrahi, K.C., Wenkel, S., Soppe, W., Deng, X.W., Valverde, F., and Coupland, G. (2008). Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. The EMBO journal *27*, 1277-1288.
Ji, Q., Zhang, M., Lu, J., Wang, H., Lin, B., Liu, Q., Chao, Q., Zhang, Y., Liu, C., Gu, M., *et al.* (2012). Molecular basis underlying the S5-dependent reproductive isolation and compatibility of indica/japonica rice hybrids. Plant physiology *158*, 1319-1328.
Jia, Y. (2009). Artificial introgression of a large chromosome fragment around the rice blast resistance gene Pi-ta in backcross progeny and several elite rice cultivars. Heredity *103*, 333-339.
Jia, Y., Jia, M.H., Wang, X., and Liu, G. (2012). Indica and japonica crosses resulting in linkage block and recombination suppression on rice chromosome 12. PloS one *7*, e43066.
Jiang, Y., Cai, Z., Xie, W., Long, T., Yu, H., and Zhang, Q. (2012). Rice functional genomics research: progress and implications for crop genetic improvement. Biotechnology advances *30*, 1059-1070.
Kang, C.Y., Lian, H.L., Wang, F.F., Huang, J.R., and Yang, H.Q. (2009). Cryptochromes, phytochromes, and COP1 regulate light-controlled stomatal development in Arabidopsis. Plant Cell *21*, 2624-2641.
Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. Plant Mol Biol *35*, 25-34.
Khush, G.S. (1999). Green revolution: preparing for the 21st century. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada *42*, 646-655.

Khush, G.S. (2001). Green revolution: the way forward. Nat Rev Genet *2*, 815-822.
Kim, M.K., Kim, H.A., Koh, K., Kim, H.S., Lee, Y.S., and Kim, Y.H. (2008). Identification and quantification of anthocyanin pigments in colored rice. Nutrition research and practice *2*, 46-49.
Kim, S.L., Lee, J.E., Kwon, Y.U., Kim, W.H., Jung, G.H., Kim, D.W., Lee, C.K., Lee, Y.Y., Kim, M.J., Kim, Y.H.*, et al.* (2013). Introduction and nutritional evaluation of germinated soy germ. Food Chem *136*, 491-500.
Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. Cell *155*, 27-38.
Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci *10*, 236-242.
Kong, J.M., Chia, L.S., Goh, N.K., Chia, T.F., and Brouillard, R. (2003). Analysis and biological activities of anthocyanins. Phytochemistry *64*, 923-933.
Laubinger, S., Fittinghoff, K., and Hoecker, U. (2004). The SPA quartet: A family of WD-repeat proteins with a central role in suppression of photomorphogenesis in arabidopsis. Plant Cell *16*, 2293-2306.
Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., and Deng, X.W. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell *19*, 731-749.
Li, Q., Li, L., Yang, X., Warburton, M.L., Bai, G., Dai, J., Li, J., and Yan, J. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. BMC plant biology *10*, 143.
Li, S. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. Plant signaling & behavior *9*, e27522.
Li, Z., Pinson, S.R., Paterson, A.H., Park, W.D., and Stansel, J.W. (1997). Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (Oryza sativa L.) population. Genetics *145*, 1139-1148.
Lillo, C., Lea, U.S., and Ruoff, P. (2008). Nutrient depletion as a key factor for

manipulating gene expression and product formation in different branches of the flavonoid pathway. Plant Cell Environ *31*, 587-601.
Lin, S.Y., Ikehashi, H., Yanagihara, S., and Kawashima, A. (1992). Segregation distortion via male gametes in hybrids between Indica and Japonica or wide-compatibility varieties of rice (Oryza sativa L). Theoret Appl Genetics *84*, 812-818.
Liu, L.J., Zhang, Y.C., Li, Q.H., Sang, Y., Mao, J., Lian, H.L., Wang, L., and Yang, H.Q. (2008). COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. Plant Cell *20*, 292-306.
Liu, T., Liu, H., Zhang, H., and Xing, Y. (2013). Validation and characterization of Ghd7.1, a major QTL with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (Oryza sativa L.). Journal of integrative plant biology.
Long, Y., Zhao, L., Niu, B., Su, J., Wu, H., Chen, Y., Zhang, Q., Guo, J., Zhuang, C., Mei, M.*, et al.* (2008). Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proceedings of the National Academy of Sciences *105*, 18871-18876.
Lu, B.R., Zheng, K.L., Qian, H.R., and Zhuang, J.Y. (2002). Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics *106*, 101-106.
Lu, N., Sun, Y., and Zheng, X. (2011). Orientin-induced cardioprotection against reperfusion is associated with attenuation of mitochondrial permeability transition. Planta medica *77*, 984-991.
Luo, X.M., Lin, W.H., Zhu, S., Zhu, J.Y., Sun, Y., Fan, X.Y., Cheng, M., Hao, Y., Oh, E., Tian, M.*, et al.* (2010). Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis. Developmental cell *19*, 872-883.
Ma, J.X., and Bennetzen, J.L. (2004). Rapid recent growth and divergence of rice nuclear genomes. P Natl Acad Sci USA *101*, 12404-12410.
Matsumoto, T., Wu, J.Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B.A., Baba, T.*, et al.* (2005). The map-based sequence of the rice genome. Nature *436*, 793-800.
McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y.*, et al.* (2002). Development and mapping of 2240 new

SSR markers for rice (Oryza sativa L.). DNA research : an international journal for rapid publication of reports on genes and genomes *9*, 199-207.
McCouch, S.R., Zhao, K.Y., Wright, M., Tung, C.W., Ebana, K., Thomson, M., Reynolds, A., Wang, D., DeClerck, G., Ali, M.L.*, et al.* (2010). Development of genome-wide SNP assays for rice. Breeding Sci *60*, 524-535.
McNally, K.L., Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E.*, et al.* (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. P Natl Acad Sci USA *106*, 12273-12278.
Misera, S., Muller, A.J., Weiland-Heidecker, U., and Jurgens, G. (1994). The FUSCA genes of Arabidopsis: negative regulators of light responses. Molecular & general genetics : MGG *244*, 242-252.
Miura, K., Ashikari, M., and Matsuoka, M. (2011). The role of QTLs in the breeding of high-yielding rice. Trends Plant Sci *16*, 319-326.
Mizuta, Y., Harushima, Y., and Kurata, N. (2010). Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. Proc Natl Acad Sci U S A *107*, 20417-20422.
Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., Tanji, M., Sato, M., Nasu, S., and Minobe, Y. (2002). Positional cloning of rice semidwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA research : an international journal for rapid publication of reports on genes and genomes *9*, 11-17.
Morishima, H., and Oka, H.I. (1981). Phylogenetic Differentiation of Cultivated Rice .22. Numerical Evaluation of the Indica-Japonica Differentiation. Jpn J Breed *31*, 402-413.
Nagano, H., Onishi, K., Ogasawara, M., Horiuchi, Y., and Sano, Y. (2005). Genealogy of the "Green Revolution" gene in rice. Genes & genetic systems *80*, 351-356.
Nakano, H., Kosemura, S., Suzuki, T., Hirose, K., Kaji, R., and Sakai, M. (2009). Oryzamutaic acid A, a novel yellow pigment from an Oryza sativa mutant with yellow endosperm. Tetrahedron Lett *50*, 2003-2005.
Nakano, H., Tamura, K., Sakai, M., Ooka, N., Arakawa, M., Yagasaki, K., Minoda, T.,

Noda, S., and Arai, M. (2011). Isolation and identification of oryzamutaic acids A and D from an Oryza sativa spontaneous mutant with yellow endosperm. Breeding Sci *61*, 93-95.

Nam, S.H., Choi, S.P., Kang, M.Y., Koh, H.J., Kozukue, N., and Friedman, M. (2006). Antioxidative activities of bran extracts from twenty one pigmented rice cultivars. Food Chem *94*, 613-620.

Naveira, H., and Barbadilla, A. (1992). The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. Genetics *130*, 205-209.

Nelson, G.C., and International Food Policy Research Institute. (2009). Climate change : impact on agriculture and costs of adaptation (Washington, D.C.: International Food Policy Research Institute).

Ogo, Y., Ozawa, K., Ishimaru, T., Murayama, T., and Takaiwa, F. (2013). Transgenic rice seed synthesizing diverse flavonoids at high levels: a new platform for flavonoid production with associated health benefits. Plant biotechnology journal *11*, 734-746.

Olsen, K.M., Caicedo, A.L., Polato, N., McClung, A., McCouch, S., and Purugganan, M.D. (2006). Selection under domestication: evidence for a sweep in the rice waxy genomic region. Genetics *173*, 975-983.

Oravecz, A., Baumann, A., Mate, Z., Brzezinska, A., Molinier, J., Oakeley, E.J., Adam, E., Schafer, E., Nagy, F., and Ulm, R. (2006). CONSTITUTIVELY PHOTOMORPHOGENIC1 is required for the UV-B response in Arabidopsis. Plant Cell *18*, 1975-1990.

Ouyang, Y.D., Liu, Y.G., and Zhang, Q.F. (2010). Hybrid sterility in plant: stories from rice. Current opinion in plant biology *13*, 186-192.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M.R., Zschocke, J., and Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. Briefings in bioinformatics *15*, 256-278.

Rafalski, J.A. (2010). Association genetics in crop improvement. Current opinion in plant biology *13*, 174-180.

Rieseberg, L.H. (2001). Chromosomal rearrangements and speciation. Trends in ecology & evolution *16*, 351-358.
Saijo, Y., Sullivan, J.A., Wang, H., Yang, J., Shen, Y., Rubio, V., Ma, L., Hoecker, U., and Deng, X.W. (2003). The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. Genes Dev *17*, 2642-2647.
Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., and Fernie, A.R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale *72*, 21-34.
Sano, R., and Morishima, H. (1992). Indica-Japonica differentiation of rice cultivars viewed from variations in key characters and isozymes, with special reference to landraces from the Himalayan hilly areas. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *84*, 266-274.
Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S.*, et al.* (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. Nature *416*, 701-702.
Schijlen, E., Ric de Vos, C.H., Jonker, H., van den Broeck, H., Molthoff, J., van Tunen, A., Martens, S., and Bovy, A. (2006). Pathway engineering for healthy phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant biotechnology journal *4*, 433-444.
Schwechheimer, C., and Deng, X.W. (2000). The COP/DET/FUS proteins-regulators of eukaryotic growth and development. Seminars in cell & developmental biology *11*, 495-503.
Seo, H.S., Yang, J.Y., Ishikawa, M., Bolle, C., Ballesteros, M.L., and Chua, N.H. (2003). LAF1 ubiquitination by COP1 controls photomorphogenesis and is stimulated by SPA1. Nature *423*, 995-999.
Shen, Y., Wan, Z., Coarfa, C., Drabek, R., Chen, L., Ostrowski, E.A., Liu, Y., Weinstock, G.M., Wheeler, D.A., Gibbs, R.A.*, et al.* (2010). A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome research *20*, 273-280.
Shen, Y., Zhang, Y., Yang, C., Lan, Y., Liu, L.L., Liu, S.J., Chen, Z.J., Ren, G.X., and Wan,

J.M. (2012). Mutation of OsALDH7 causes a yellow-colored endosperm associated with accumulation of oryzamutaic acid A in rice. Planta *235*, 433-441.
Shi, M.Z., and Xie, D.Y. (2010). Features of anthocyanin biosynthesis in pap1-D and wild-type Arabidopsis thaliana plants grown in different light intensity and culture media conditions. Planta *231*, 1385-1400.
Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N*., et al.* (2007). Identification of a biosynthetic gene cluster in rice for momilactones. The Journal of biological chemistry *282*, 34013-34018.
Shin, J., Park, E., and Choi, G. (2007). PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. The Plant journal : for cell and molecular biology *49*, 981-994.
Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., and Yano, M. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. Nature genetics *40*, 1023-1028.
Song, X., Qiu, S.Q., Xu, C.G., Li, X.H., and Zhang, Q.F. (2005). Genetic dissection of embryo sac fertility, pollen fertility, and their contributions to spikelet fertility of intersubspecific hybrids in rice. Theoretical and Applied Genetics *110*, 205-211.
Stacey, M.G., Kopp, O.R., Kim, T.H., and von Arnim, A.G. (2000). Modular domain structure of Arabidopsis COP1. Reconstitution of activity by fragment complementation and mutational analysis of a nuclear localization signal in planta. Plant Physiol *124*, 979-990.
Suzuki, G., Yanagawa, Y., Kwok, S.F., Matsui, M., and Deng, X.W. (2002). Arabidopsis COP10 is a ubiquitin-conjugating enzyme variant that acts together with COP1 and the COP9 signalosome in repressing photomorphogenesis. Gene Dev *16*, 554-559.
Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B., and Peters, R.J. (2009). CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. The Plant cell *21*, 3315-3325.

Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., Utsushi, H., Natsume, S., Kanzaki, H., Matsumura, H., *et al.* (2013). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. The New phytologist *200*, 276-283.

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., Padhukasahasram, B., Bustamante, C., Yoshimura, A., Doi, K., *et al.* (2009). Evolutionary history of GS3, a gene conferring grain length in rice. Genetics *182*, 1323-1334.

Takeda, S., and Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet *9*, 444-457.

Tanaka, N., Itoh, H., Sentoku, N., Kojima, M., Sakakibara, H., Izawa, T., Itoh, J., and Nagato, Y. (2011). The COP1 ortholog PPS regulates the juvenile-adult and vegetative-reproductive phase changes in rice. Plant Cell *23*, 2143-2154.

Tang, J.B., Xia, H.A., Cao, M.L., Zhang, X.Q., Zeng, W.Y., Hu, S.N., Tong, W., Wang, J., Wang, J., Yu, J., *et al.* (2004). A comparison of rice chloroplast genomes. Plant Physiol *135*, 412-420.

Thaipong, K., Boonprakob, U., Crosby, K., Cisneros-Zevallos, L., and Byrne, D.H. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. J Food Compos Anal *19*, 669-675.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., *et al.* (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. The Plant journal : for cell and molecular biology *42*, 218-235.

Tsuda, T., Horio, F., Uchida, K., Aoki, H., and Osawa, T. (2003). Dietary cyanidin 3-O-beta-D-glucoside-rich purple corn color prevents obesity and ameliorates hyperglycemia in mice. The Journal of nutrition *133*, 2125-2130.

Tung, C.W., Zhao, K.Y., Wright, M.H., Ali, M.L., Jung, J., Kimball, J., Tyagi, W., Thomson, M.J., McNally, K., Leung, H., *et al.* (2010). Development of a Research Platform for Dissecting Phenotype-Genotype Associations in Rice (Oryza spp.). Rice *3*, 205-217.

Vaughan, D.A., Lu, B.R., and Tomooka, N. (2008). The evolving story of rice evolution. Plant Sci *174*, 394-408.

Vitte, C., Ishii, T., Lamy, F., Brar, D., and Panaud, O. (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (Oryza sativa L.). Mol Genet Genomics *272*, 504-511.

Wang, L., Wang, A.H., Huang, X.H., Zhao, Q.A., Dong, G.J., Qian, Q.A., Sang, T., and Han, B. (2011). Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theoretical and Applied Genetics *122*, 327-340.

Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. The Plant cell *21*, 3926-3936.

Weng, D.X., Xu, S.C., Lin, R.M., Wan, A.M., Li, J.P., and Wu, L.R. (2005). [Microsatellite marker linked with stripe rust resistant gene Yr9 in wheat]. Yi chuan xue bao = Acta genetica Sinica *32*, 937-941.

Widmer, A., Lexer, C., and Cozzolino, S. (2009). Evolution of reproductive isolation in plants. Heredity *102*, 31-38.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. Current opinion in plant biology *5*, 218-223.

Wu, D., Hu, Q., Yan, Z., Chen, W., Yan, C., Huang, X., Zhang, J., Yang, P., Deng, H., Wang, J.*, et al.* (2012). Structural basis of ultraviolet-B perception by UVR8. Nature *484*, 214-219.

Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., Yu, S., Han, B., and Zhang, Q. (2010). Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci U S A *107*, 10578-10583.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L.*, et al.* (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature biotechnology *30*, 105-111.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li,

X.*, et al.* (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nature genetics *40*, 761-767.
Yan, S., Zou, G., Li, S., Wang, H., Liu, H., Zhai, G., Guo, P., Song, H., Yan, C., and Tao, Y. (2011). Seed size is determined by the combinations of the genes controlling different seed characteristics in rice. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *123*, 1173-1181.
Yang, J., Zhao, X., Cheng, K., Du, H., Ouyang, Y., Chen, J., Qiu, S., Huang, J., Jiang, Y., Jiang, L.*, et al.* (2012). A Killer-Protector System Regulates Both Hybrid Sterility and Segregation Distortion in Rice. Science *337*, 1336-1340.
Yi, C., and Deng, X.W. (2005). COP1 - from plant photomorphogenesis to mammalian tumorigenesis. Trends in cell biology *15*, 618-625.
Yu, H.H., Xie, W.B., Wang, J., Xing, Y.Z., Xu, C.G., Li, X.H., Xiao, J.H., and Zhang, Q.F. (2011). Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers. Plos One *6*.
Yu, J., Hu, S.N., Wang, J., Wong, G.K.S., Li, S.G., Liu, B., Deng, Y.J., Dai, L., Zhou, Y., Zhang, X.Q.*, et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science *296*, 79-92.
Yu, J.W., Rubio, V., Lee, N.Y., Bai, S., Lee, S.Y., Kim, S.S., Liu, L., Zhang, Y., Irigoyen, M.L., Sullivan, J.A.*, et al.* (2008). COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. Molecular cell *32*, 617-630.
Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development *130*, 4859-4869.
Zhao, C., Giusti, M.M., Malik, M., Moyer, M.P., and Magnuson, B.A. (2004). Effects of commercial anthocyanin-rich extracts on colonic cancer and nontumorigenic colonic cell growth. Journal of agricultural and food chemistry *52*, 6122-6128.
Zhao, J., and Dixon, R.A. (2010). The 'ins' and 'outs' of flavonoid transport. Trends Plant Sci *15*, 72-80.
Zhao, K., Tung, C.W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J.*, et al.* (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature

communications *2*, 467.
Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M.L., Tung, C.W., Reynolds, A.*, et al.* (2010a). Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PloS one *5*, e10780.
Zhao, Q.A., Huang, X.H., Lin, Z.X., and Han, B. (2010b). SEG-Map: A Novel Software for Genotype Calling and Genetic Map Construction from Next-generation Sequencing. Rice *3*, 98-102.
Zhou, L.J., Ao, G.H., Wu, X.J., and Li, S.G. (2005). [SSR markers linked with early stability in rice]. Yi chuan xue bao = Acta genetica Sinica *32*, 837-845.
Zimmermann, I.M., Heim, M.A., Weisshaar, B., and Uhrig, J.F. (2004). Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. The Plant journal : for cell and molecular biology *40*, 22-34.

초 록

통일벼의 유전체 분석과 벼에서 노란색 종피 및 배 치사 유전자 동정

통일벼는 인디카 벼 (IR8, TN1)와 자포니카 벼 (Yukara)의 삼원교배로 만들어진, 우리나라의 녹색혁명을 가져다 준 고수량 벼이다. 이러한 통일벼와 통일벼의 모본들을 각각 47X 와 30X 의 coverage 로 re-sequencing 하여 염기수준에서 유전체의 구조를 분석하였다. 통일벼 내에서 인디카-자포니카로 구성되는 유전체의 구조는 통일벼와 세 모본 사이에서 나타난 SNP 를 기반으로, sliding window 방법을 이용하여 결정하였다. 그 결과 통일벼 유전체의 91.8%는 인디카 모본에서 유래하였고, 7.9%는 자포니카 모본에서 유래 한 것을 알 수 있었다. SSR (Simple sequence repeat) motif, ORF gene 분포, GO (Gene ontology) 등의 유전체를 구성하는 요소들을 분석해 본 결과 통일벼를 이루고 있는 인디카-자포니카 유전체의 구성과 유사한 비율로 이들 요소들이 통일벼 유전체 내에 존재하고 있음을 알 수 있었다. 또한 벼의 수량과 밀접한 연관이 있는, 이미 밝혀진 유전자 중 7 개의 유전자에 대한 통일벼 내에서의 분포 양상을 살펴본 결과 7 개 유전자 모두 인디카 벼의 모본에서 유래하였음을 알 수 있었다.

이러한 결과들은 통일벼 유전체의 대부분이 인디카 벼에서 유래했다는 사실을 뒷받침 해주고, 유전체를 구성하는 요소들 또한 유전체의 구조가 변함에 따라 함께 변화한다는 사실은 알 수 있었다. 통일벼의 유전체 구조를 밝힘으로써 육종적 효용성뿐만 아니라 앞으로 고수량 또는

인디카-자포니카 교잡 품종을 육성함에 있어서 중요한 기초 자료가 되리라 생각된다.

MNU 의 처리를 통해 얻어진, 노란색 종피색을 가지며 동시에 배가 죽는 돌연변이의 유전분석과 고밀도 유전자지도를 작성한 결과 이러한 표현형과 관련된 유전자는 2 번 염색체에 존재하였고 애기장대의 COP1 (Constitutive Photomorphogenesis 1)과 이종상동성 유전자인 것으로 밝혀졌다. 모본과 돌연변이에 대한 해당 유전자 (LOC_Os02g53140)의 sequence 변이를 살펴본 결과 5'UTR 과 start codon 을 포함한 첫 번째 exon, intron 일부 등, 총 706 bp 의 염기가 deletion 된 것을 확인 하였다. 이로 인해 이 유전자가 가지는 3 가지 domain, RING-finger, Coiled-coil, WD40 domain 중에서 RING-finger domain 의 전체가 deletion 되었다. 정상 YEL 유전자를 과발현 시킨 결과 종피가 백색을 띠는 정상 종자로 회복되었지만 배에는 지속적인 물질 축적이 관찰되었고 배가 치사되는 형질은 회복되지 못했다. GUS assay 결과 화분과 배, 종피, 마디(절)에서 유전자가 발현됨을 관찰할 수 있었다. 또한 종피와 배에 축적되는 물질을 HPLC/MS 를 이용하여 분석한 결과 Orientin 이라는 물질로 확인하였고 특히 배에 고농도로 축적되어 있었다. 이 물질의 항산화 능력은 orientin 이 고농도로 축적된 배에서 높게 나타났고, *yel* 뮤턴트 종자의 total polyphenol 함량도 높게 나타났다. *yel* 뮤턴트의 배에서는 포화 지방산 함량이 증가하였고 반대로 불포화 지방산 함량은 감소하였다. *yel* 뮤턴트 종자의 배에서 조지방의 함량은 정상 배에 비해서 유의하게 감소 하는 것으로 나타났다.

위의 결과들을 통해 YEL 유전자는 벼에서 플라보노이드 (Orientin)f 의 생합성 대사 과정에 관여하고 배의 발달 또는 배를 구성하는 성분들의

생산에 영향을 미치고 있음을 밝혀, 벼에서 COP1 유전자의 새로운 기작을 구명하는데 중요한 기초 자료를 제공하는 데에도 그 의의가 있다.

주요어: 벼, 통일벼, 차세대유전체분석, 유전체 구조, 인디카-자포니카, 노란색 종피, 배 치사 돌연변이, Orientin, 항산화능력

학번: 2010-30291

A DISSERTATION FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

BY

BACKKI KIM

FEBRUARY, 2015

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

UNDER THE DIRECTION OF DR. HEE-JONG KOH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF SEOUL NATIONAL UNIVERSITY

BY
BACKKI KIM

MAJOR IN CROP SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE
NOVEMBER, 2014

APPROVED AS QUALIFIED DISSERTATION OF BACKKI KIM
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS
FEBRUARY, 2015

CHAIRMAN _____
Suk-Ha Lee, Ph.D.

VICE-CHAIRMAN _____
Hee-Jong Koh, Ph.D.

MEMBER _____
Byon-Woo Lee, Ph.D.

MEMBER _____
Tae-Jin Yang, Ph.D.

MEMBER _____
Joohyun Lee, Ph.D.

Genome Analysis of a Cultivar 'Tongil' and Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

BACKKI KIM

GENERAL ABSTRACT

Asian cultivated rice (*Oryza sativa* L.), divided into sub-species, *Oryza sativa. indica* and *Oryza sativa. japonica* has been cultivated in all over the world including Asia. Recent evidence suggests an ancient *indica* and *japonica* divided between 200,000 and 440,000 years ago based on nuclear genome sequence comparisons (Ma and Bennetzen, 2004; Tang et al., 2004; Vitte et al., 2004). During cultivation and domestication, these subspecies have developed the unique morphologies and characteristic agronomic traits. Many phenotypic differences are obvious between *O. sativa indica* and *japonica* and the beneficial traits from each subspecies are useful to develop a new varieties. However, various obstacles such as reproductive barriers, prevent gene exchange or gene flow between two subspecies. Although several studies have tried to explain the differences between *indica* and *japonica* at a certain developmental stage or molecular and

physiological level, data from these studies are quite limited to explain general differences between *indica* and *japonica*. Whole genome analysis to elucidate the differences between *indiica* and *japonica* genome will be useful to explain the genome structure that led to their distinct features.

In this study, we analyzed the nucleotide-level genome structure of Tongil rice which is a high-yielding rice variety derived from a three-way cross between *indica* and *japonica* and compared it to those of the parental varieties. Sequence data were obtained by whole-genome resequencing using the Illumina Hiseq. A total of 17.3 billion reads, 47× genome coverage, were generated for Tongil rice. Three parental varieties of Tongil rice, two *indica* types and one *japonica* type, were also sequenced at approximately 30x genome coverage. *Indica-japonica* genome composition was determined based on SNP data by comparing Tongil with three parental genome sequences using the sliding window approach. Analyses revealed that 91.8% of the Tongil genome originated from the *indica* parents and 7.9% from the *japonica* parent. Copy number of SSR motifs, ORF gene distribution throughout the whole genome, gene ontology (GO) annotation, and yield-related QTLs or genes variations were also comparatively analyzed between Tongil and parents using sequence-based tools. These results indicated that each genetic factor was transferred from parents into Tongil in proportion to the whole-genome composition.

The yellow colored pericarp and embryo lethal mutant was derived from chemical mutagenesis using N-methyl-N-nitrosourea (MNU) on a *japonica* rice cultivar, Hwacheong. In this study, we cloned the gene responsible for

yellow pericarp and lethal embryo phenotype using a map-based approach. Fine mapping revealed that the mutant gene, *yel* was located on the long arm of chromosome 2 and sequencing of candidate genes identified the gene responsible for the yellow pericarp and embryo lethal phenotype, *Os*COP1 (Constitutive Photomorphogenic 1), orthologs of *Arabidopsis* COP1 which encodes a protein of comprising RING-finger, coiled-coil, and WD40 domains. A 706-bp deletion in the first exon including 5′UTR and start codon of *yel* gene, LOC_Os02g53140 was found thus, it is conceivable that the yellow pericarp phenotype may have been the result of loss of function in rice *yel* gene. HPLC-MS analysis of the four candidate standards with extracts of embryo and endosperm of the mutant grain showed that extremely high level of orientin, luteolin-8-*C*-glucoside accumulated in mutant embryo and endosperm and was the major pigment both in embryo and endosperm. In addition, the antioxidant activities and total phenolic contents of embryo and endosperm of the mutant seed were significantly higher than those of wild-type. Fatty acid analysis showed that the content of saturated fatty acid was increased in mutant embryo on the other hand, the content of unsaturated fatty acid was decreased than that of wild-type. These results indicated that *yellow embryo lethal* (*yel*) gene might be involved in the flavonoids biosynthetic pathway and other ingredient characteristics of rice seed.

Keywords: Rice, Tongil Rice, NGS, Genome structure, Yellow pericarp, Embryo lethal Mutant, Orientin, Antioxidant capacity. *Indica-japonica*

Student Number: 2010-30291

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

COP1	Constitutive photomorphogenic 1
NGS	Next generation sequencing
GO	Gene ontology
ORF	Open reading frame
QTL	Quantitative trait loci
SNP	Single nucleotide polymorphism
LD	Linkage disequilibrium
WCV	Wide compatibility variety
MNU	N-methyl-N-nitrosourea
UHPLC	Ultra-high performance liquid chromatography
DPPH	2, 2 – diphenyl – 1 – picrylhydrazyl
FRAP	Ferric reducing antioxidant power
TEAC	Trolox equivalent antioxidant capacity
RNAi	RNA interference

GENERAL INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important staple food and provides over 21 percent of the caloric needs for more than half of the world's population, and also contributes about 30 percent of Asia's caloric supply (IRRI Statistics, 2009), specially up to 76 percent of the caloric intake of the population of southeast Asia (Fitzgerald et al., 2009; Khush, 1997). Asian cultivated rice divided into two sub-species of *O. sativa. indica* and *japonica*. Recent evidences suggest an ancient *indica*/*japonica* diverged between 200,000 and 440,000 years ago based on nuclear genome sequence comparisons and between 86,000 and 200,000 years ago based on chloroplast sequences, respectively prior to the domestication of rice (Ma and Bennetzen, 2004; Tang et al., 2004; Vitte et al., 2004). In terms of morphology and living environments, *indica* varieties exhibit greater plant heights, longer leaves, and heat tolerance but are sensitive to low temperatures thus, *indica* varieties are cultivated at low latitudes and in humid regions on the other hand, *japonica* varieties have lower plant heights and shorter leaves than that of *indica* but are tolerant to low temperatures therefore, *japonica* varieties are more suitable for high latitudes and for the lower latitudes of high altitude cultivation (Dan et al., 2014). Due to the deep genetic structure that evolved during its domestication and adaptation, current *oryza. sativa* cultivars and landraces can be sub-divided in more detail into five genetically differentiated groups: *indica*, aus, aromatic, temperate *japonica*, and tropical *japonica* (Garris et al., 2005).

Since the green revolution, many countries overcame food shortage and have developed high yielding rice varieties however, the productivity and yield of rice are still important factors in rice cultivation because demand from population growth and the needs of biofuels exceeds global food supply (Brown and Funk, 2008). In addition, presently consumptions of functional health promoting rice, such as black and red rice are rapidly growing due to their healthy functional food ingredients (Kim et al., 2008). Phytochemicals such as anthocyanins and flavonoids, which are found in a variety of crops including fruits and vegetables (Abdel-Aal el et al., 2006) and have been recognized as health-promoting functional food ingredients due to their antioxidant activity (Nam et al., 2006), anticancer (Zhao et al., 2004), hypoglycemic and anti-inflammatory effects (Tsuda et al., 2003). Rice also contains a various naturally occurring compounds in bran, thus rice grain is a valuable source to provide health promoting ingredients. To meet the demands deriving from rapid population growth, lack of planting area, worldwide climate change, and the industrial trend for human health or biofortification continuous efforts to increase rice production by using the genetic and genomic improvement technologies will be of great importance (Takeda and Matsuoka, 2008).

Rice, the first genome sequenced crop plant is an excellent model plant that has been comprehensively studied in functional genomics. In addition, rice has a relatively small genome size compared to other crops. Draft genome sequences of two rice subspecies, *O. sativa ssp. japonica* (cv. Nipponbare) and *O. sativa ssp. indica* (cv. 93-11) were reported (Goff et al.,

2002; Yu et al., 2002), and subsequently the International Rice Genome Sequencing Project completed the final sequence of the entire rice genome of Nipponbare (Matsumoto et al., 2005). These achievements have provided us with a vast amount of information on the rice genome and genome structure and also allowed us to perform detailed genetic analysis, functional studies, and association studies (Miura et al., 2011). Using this genome information, researchers have succeeded in isolating and identifying many important genes and QTLs, which related to useful agronomic traits and rice yield potential. In addition, next generation sequencing technologies have allowed genome-wide genetic variation genotyping and understanding the genetic basis of phenotype variation, and have facilitated reduced laborious works and saving time and increased resolution of genetic mapping. With the Nipponbare sequence as a reference, genome re-sequencing of a large number of rice accessions (The 3,000 rice genomes project, 2014) has led to the discovery of millions of SNPs and insertion/deletion sites (indels), enabling genome-wide association studies (GWAS) to identify agronomically important genes in rice (Huang et al., 2012c; Xu et al., 2012).

LITERATURE REVIEW

Development of "Tongil" and Tongil-type rice

Tongil rice is the first rice variety developed from a hybridization between *indica* and *japonica* in South Korea. It was developed from the progeny of IR667, which was derived from three-way cross (IR8//Yukara/Taichung Native 1) and released to the farmers in 1972. In addition, Tongil is the first Korean rice variety introduced semi-dwarf gene (*sd1*) which is well known as the "Green Revolution gene" and contributed to the dramatic increase in rice production in the 1960s and 1970s, specially, in Asia. Genetic basis of the short-stature of Tongil and Tongil-type rice that led the Green Revolution of South Korea was derived from the *sd1* gene of *indica* rice, IR8 and Taichung Native 1 (TN1). Tongil has short-statured in their plant architecture with erect leaves, high yield potential, and tolerance to nitrogen application and lodging.

The development of "Tongil" rice variety raised the milled rice productivity to 5.13 MT/ha, which was about 30 percent higher than that of the leading *japonica* rice variety at that time. Subsequently, 25 Tongil-type varieties were released by 1977, the area of cultivating Tongil-type rice including Tongil rice was rapidly increased. The area reached to 76.2 percent of the total rice farming areas in 1977, leading to the Korean Green Revolution and self-sufficiency in rice (Chung and Heu, 1980). During the

1970s and 1980s, total 40 varieties of Tongil-type rice were developed and cultivated. The productivity of Tongil-type varieties constantly increased to 5.76 MT/ha in 1976. The national average yield of milled rice in farmers' fields dramatically increased to 4.93 MT/ha, in 1977 as compared to *japonica* varieties that yielded 3.37~4.69 MT/ha (NARI, 2012). Total rice production reached 4.67 million MT in 1976, and 5.21 million MT in 1977 and 6.01 million MT in 1978, respectively (NARI, 2012). For the first time, rice production exceeded the demands and achieved self-sufficiency in rice production and the Green Revolution in South Korea. The success of Tongil and Tongil-type rice varieties was not only a new milestone for future rice breeding but also a practical opportunity to utilize *indica* germplasm in temperate region (NARI, 2012).

Next generation sequencing

The advent of the next generation sequencing technology changed the paradigm of genome research and have provided immeasurable opportunity for the genotyping and genomic study and also greatly enhanced rice functional genomics and molecular breeding studies (Gao et al., 2012; Xie et al., 2010). NGS techniques not only increase sequencing throughput but also allow sequencing a large number of samples using a multiplexed sequencing strategy, simultaneously (Craig et al., 2008; Cronn et al., 2008). These recent technical advances facilitated the development of a sequencing-based high-throughput genotyping that has advantages

of time and cost effectiveness, dense marker coverage, high mapping accuracy and resolution, and more comparable genome and genetic maps among mapping populations and organisms (Huang et al., 2009). Massive parallel sequencing technology has proven revolutionary, shifting the paradigm of genomics to address biological questions at a genome-wide scale (Koboldt et al., 2013). Three main NGS platforms, Illumina/Solexa, Roche/454 and ABI/SOLiD sequencing, which are known as high-throughput sequencing, can generate large amounts of sequence data in a single run (Ansorge, 2009).

NGS technology has been widely used in plant breeding area, such as whole genome sequencing of varieties and population re-sequencing, developing new and numerous molecular markers, SNP-based or InDel markers in a number of plant species, and constructing saturated high resolution genetic and physical maps for association study. The whole genome sequence of a crop species contains important information regarding of the origin, evolution, development, and physiology of that species. In addition, the systematic study of genome sequences can allow exploration of the gene sequences of the species and provide a method for identifying the genes that allowed the species to adapt to a specific environment. Whole genome re-sequencing is a useful method for studying genome to have a reference genome sequence such as Arabidopsis and rice. Besides re-sequencing, to obtain the genome sequence of a crop, the whole-genome de novo sequencing method, which involves the first genome sequence from the species which do not have

reference sequence can be used and subsequently can use bioinformatic tools to assemble the sequences and obtain the genomic map for that species (Gao et al., 2012). The new generation sequencing technologies have been provided and readily adapted to high-quality genome researches, and facilitated the sequencing of abundant diverse germplasm resources of crops.

SNP genotyping and application

Advanced genome sequencing technologies provide unprecedented opportunities to characterize individual genomic differences and identify variations. Specifically, whole genome sequencing using next generation sequencing technologies is widely used due to the moderate costs, manageable data amounts and straightforward interpretation of analysis results (Pabinger et al., 2014). For these reasons, next generation sequencing is now available even for single laboratory. Recently, several re-sequencing projects have conducted and provided a great deal of information about rice genomic structure and genetic diversity (Huang et al., 2010; Huang et al., 2012c; Xu et al., 2012). These sequencing projects found millions of polymorphisms along the entire rice genome, including InDels and several millions of SNPs. The recent achievements in rice genomics research is based on establishing a high-throughput genotyping system. Next generation sequencing and array-based SNP detection are two major high-throughput variation detection platforms. This high-

throughput genotyping system is important for gene identification and molecular breeding. Furthermore, various SNP based platforms have become attractive tools for genotyping and the high-throughput genotyping technologies have been developed rapidly. Recently, a high-resolution 44K Affymetrix custom array has been designed for rice SNP genotyping (McCouch et al., 2010; Zhao et al., 2011). These SNP genotyping assays were considered critically important for associating phenotype–genotype in rice (Tung et al., 2010). High-throughput genotyping platforms play important roles in genes and QTLs cloning, and GWAS analysis. By 2012, over 800 genes had been cloned that are responsible for useful agronomic traits such as yield, plant architecture, grain quality, stresses resistances, and nutrient-use efficiency (Chen et al., 2013; Jiang et al., 2012). The traditional method for gene identification is based on genetic mapping of naturally or artificially occurring mutations and the traits for natural variations. Newly developed genomics-based strategies such as re-sequencing have greatly improved the resolution and accuracy of genetic mapping (Wang et al., 2011; Yu et al., 2011). Recently, a method called MutMap was used to rapidly isolate genes by crossing the mutant to the wild-type line followed by whole genome re-sequencing of the bulked DNA from a segregating population (F_2 population) of plants showing the mutant phenotype (Abe et al., 2012). MutMap-Gap approach also reported for isolating genes by whole genome re-sequencing of bulked DNA of mutant F_2 progeny combined with de novo assembly of gap regions in rice (Takagi et al., 2013). With the development of high-

throughput genomics based platforms, GWAS has been more frequently used in rice (Huang et al., 2010; Huang et al., 2012c; Rafalski, 2010; Zhao et al., 2011). A diversity panel which were selected to dissect the genetic diversity of domesticated rice, consisting of 413 *O. sativa* accessions, was recently genotyped with 44K SNPs and used for the GWAS analysis of aluminum tolerance and other complex traits (Famoso et al., 2011). The new SNP-based and genomic based mapping approaches will show advantages in detecting power, resolution, and time compared to marker-assisted map-based cloning or linkage mapping of QTLs.

Indica-japonica hybridization in rice

Asian cultivated rice (*Oryza sativa* L.) was domesticated approximately 7000–8000 years before the present in several regions such as India, the Yangtze River area in China, the southern Himalayas, and coastal swamp habitats in Southeast Asia (Gross and Zhao, 2014). Since the beginning of domestication and followed by its lasting cultivation, Asian cultivated rice has experienced genetic differentiation, adapting to different ecological and environmental conditions under both natural and human selection. Abundant genetic diversity was generated by genetic differentiation in rice, such as *indica* and *japonica* ecotypes (Vaughan et al., 2008), lowland and upland ecotypes, and non-glutinous and glutinous grain quality types (Olsen et al., 2006). The most momentous genetic differentiation in rice was *indica–japonica* differentiation, as a result, Asian cultivated rice evolved

into two major ecotypes. The *indica* rice is mainly found in tropical and subtropical area, with either low latitudes or altitudes, whereas the *japonica* rice is mostly found in temperate regions with high latitudes. The *japonica* rice can be cultivated in high mountainous areas in some rice planting regions of low latitude. In consequence of adaptation to various ecological conditions, *indica* and *japonica* rice has been diverged according to plant morphology, agronomical characteristics, and physiological–biochemical features (Vaughan et al., 2008). Usually, *indica* and *japonica* rice varieties are distantly related in terms of genetic background, therefore, the inter-subspecific hybridization between *indica* and *japonica* rice will result in remarkable genetic recombination and variation. The research of the mechanism of genetic differentiation between *indica* and *japonica* rice provides critical information for better understanding of adaptive evolution in rice under the different environments.

Notably, the hybridization between *indica* and *japonica* rice varieties generate strong hybrid vigor (Khush, 2001). This characteristic enables to utilize heterosis from inter-subspecies hybridization between *indica* and *japonica* by breeders. The magnitude of hybrid vigor depends on the genetic diversity between the two parents of the hybrids. The greater the genetic difference between the parents, the higher the heterosis (Khush, 1996). The inter-subspecific hybridization between *indica* and *japonica* has strong heterosis compared to intra-subspecies hybridization between *japonica* and *japonica* or *indica* and *indica* (Jiang J, 2002). Thus, the *indica*–*japonica* hybrid rice may has great yield potential in rice. The utilization of

inter-subspecific hybridization between *indica* and *japonica* is the most feasible approach for high yielding rice. Hybrid rice between *indica* and *japonica* varieties have a yield increase of approximately 25 to 30%. However, the F_1 hybrids often have a low seed setting rate due to the reproductive isolation of inter-subspecies (Li et al., 1997; Ouyang et al., 2010). Over the past decades, several studies have conducted regarding the fertility or low seed setting rate in the F_2 progenies (Chen et al., 2008a; Mizuta et al., 2010; Song et al., 2005). The reasons for this phenomenon include pollen sterility, embryo sac abortion or incompatibility. Since the interaction of genes which involved in formation of zygote in *indica-japonica* hybridization is complex, reproductive isolation is still a major obstacle to breed inter-subspecific hybrid rice.

Flavonoids in plant

Flavonoids, secondary metabolites widely accumulate in plants and involved in several aspects of development and defense mechanism. They are classified in six major subgroups: chalcones, flavones, flavonols, flavadiols, anthocyasnins, and proanthocyanidins (Winkel-Shirley, 2002) and are low molecular weight compounds composed of a three-ring structure with various substitutions. Most of flavonoids accumulate in all organs and tissues, at different stages of development, and depending on internal plant condition or external environmental conditions. Flavonoids are synthesized in the cytosol and are transported to the vacuole for storage.

They can also be found in cell walls, the nucleus, chloroplasts, and even in the extracellular space (Zhao and Dixon, 2010). They represent the color and flavor of leaves, fruits, flowers, and seeds and also contribute to plant adaptation to environmental conditions such as cold or UV stresses, and pathogen attacks (Hichri et al., 2011). In addition, plant-derived flavonoids are important health promoting sources for human and provides various nutrition. Furthermore, these compounds possess pharmaceutical properties extremely attractive for human health. Over several decades, the genes and enzymes involved in flavonoid biosynthetic pathway have been characterized in several plant species such as maize (*Zea mays*), snapdragon (*Antirrhinum majus*), petunia (*Petunia hybrida*), and arabidopsis (*Arabidopsis thaliana*) (Saito et al., 2013).

Although flavonoids normally accumulate in plants as *O*-glycosylated derivatives, several species, including major cereal crops, predominantly synthesized flavone-*C*-glycosides, which are stable to hydrolysis and biologically active in planta and as dietary components (Brazier-Hicks et al., 2009). *C*-Glycosides are formed in microbes, plants, and insects, where they serve as antibiotics, antioxidants, attractants, and feeding deterrents (Hultin, 2005). Despite the common secondary metabolites in major cereal crops and medicinal plants, *C*-Glycosylation has received little attention. The most abundant *C*-glycosylated natural products in plants are flavonoids, a large group of polyphenolic compounds. Flavonoids normally accumulate as the *O*-linked glycosidic conjugates in the vacuoles of plant tissues however, flavonoids also accumulate as the *C*-glycosides in at least 20 families of

angiosperms (Harborne, 1986). These derivatives are major secondary metabolites in maize, wheat, and rice. In these cereals, *C*-glycosides of the simple flavones apigenin and/or luteolin predominate with conjugation occurring singly or doubly at the *C*-8 and/or *C*-6 position (Brazier-Hicks et al., 2009). Among these, four flavonoid *C*-glycosyl compounds (orientin, isoorientin, vitexin, and isovitexin) were first isolated compounds from bamboo leaves but are also found in numerous plants (Dietrych-Szostak and Oleszek, 1999). However, the effects of flavonoid *C*-glycosyl compounds; orientin (Luteolin-8-*C*-glucoside), isoorientin (Luteolin-6-*C*-glucoside), vitexin (Apigenin-8-*C*-glucoside) and isovitexin (Apigenin-6-*C*-glucoside), were not well-known.

COP1 (constitutive photomorphogenic 1)

The COP1 (constitutive photomorphogenic 1) acts as an E3 ubiquitin ligase at the central of light signaling to repress light signaling by targeting photoreceptors and downstream transcription factors for ubiquitination and degradation and is one of the best characterized locus among COP (Constitutive Photomorphogenic) / DET (De-Etiolated) / FUS (Fusca) locus subsequently cloned for the first time (Deng et al., 1992; Yi and Deng, 2005). COP1 and DET1 were found among the members of a group of genes termed the Constitutive Photomorphogenic / De-Etiolated (COP/DET) (Schwechheimer and Deng, 2000). At the beginning of study, these loci were identified through mutant screening in Arabidopsis (*Arabidopsis*

thaliana) at the seedling stage that display light-grown phenotypes under dark condition (COP/DET) or seeds that accumulate high levels of anthocyanin (FUS). COP1 mediates the degradation of various photomorphogenesis-promoting transcription factors by the ubiquitin-proteasome system, with the other COP/DET/FUS proteins (Holm et al., 2002; Saijo et al., 2003). In plants, the function of COP1 is closely related to the light signaling pathway. COP1 acts as a central repressor in light signal transduction, where it promotes the ubiquitination and degradation of the positive regulators and is itself regulated by various photoreceptors such as phyA, phyB, cry1, cry2, and UVR8 (Yi and Deng, 2005). Besides seedling photomorphogenesis related to light signaling, research in the past few years has expanded the role of COP1 in other pathway. These pathway include flowering (Jang et al., 2008; Liu et al., 2008; Yu et al., 2008), circadian rhythm (Yu et al., 2008), UV-B signaling (Oravecz et al., 2006; Wu et al., 2012), stomatal opening and development (Kang et al., 2009), shade avoidance response (Crocco et al., 2010), crosstalk between light and brassinosteroid signaling (Luo et al., 2010), cold acclimation response (Catala et al., 2011) and light-induced root elongation (Dyachok et al., 2011) in Arabidopsis, and juvenile–adult phase duration change in rice (Tanaka et al., 2011). These studies indicated that COP1 is an important regulator and integrator in light signal transduction.

The COP1 protein has three conserved domains: a RING finger, coiled-coil and seven WD40 domains, in both higher plants and vertebrates. These domains intermediate COP1 with other proteins and its self-dimerization

(Hoecker and Quail, 2001; Suzuki et al., 2002). The various signaling molecules have been reported to directly interact with COP1 and control its activity. In addition, the activity and location of COP1 in plant cells is altered to cytoplasmic and nuclear partitioning according to dark or light conditions. Under dark condition, COP1 targets these transcription factors, including HY5 (Elongated Hypocotyl 5), LAF1 (Long After Far-red Light 1) and HFR1 (Long Hypocotyl in Far-Red) for ubiquitination and degradation, leading to suppression of photomorphogenesis (Jang et al., 2005; Seo et al., 2003). In general, protein ubiquitination requires a specific E3 ubiquitin ligase, which can be a single protein or a protein complex. An E3 typically functions by recruiting ubiquitin-conjugating enzymes (E2s) through a RING-finger motif and the substrate.

In contrast to the typical roles in the photomorphogenesis, COP1 and HY5 are involved in long-wavelength UV-B induced photomorphogenesis. This response is initiated by absorbing UV-B through its internal chromophore tryptophan residues by the UV-B receptor, UVR8 (UV Resistance Locus 8) (Christie et al., 2012; Wu et al., 2012). UVR8 then monomerizes to interact with the UV-B inducible protein, COP1 for downstream signaling (Favory et al., 2009; Huang et al., 2012b). The physical changes of this process include anthocyanin accumulation, and tolerance against damaging UV-B. The loss of function mutant of either COP1 or HY5 showed the decreased activation of UV-B–responsive genes, impaired photomorphogenesis, and defective UV-B acclimation in previous studies (Brown et al., 2005; Oravecz et al., 2006).

CHAPTER I

Genome Analysis of a Cultivar 'Tongil' Rice

ABSTRACT

Tongil (IR667-98-1-2) rice, developed in 1972, is a high-yielding rice variety derived from a three-way cross between *indica* and *japonica*. Tongil contributed to the self-sufficiency of rice in Korea, during a period known as the 'Korean Green Revolution'. In this study, we analyzed the nucleotide-level genome structure of Tongil rice and compared it to those of the parental varieties. Sequence data were obtained by whole-genome resequencing using the Illumina Genome Analyzer. A total of 17.3 billion reads, 47X genome coverage, were generated for Tongil rice. Three parental accessions of Tongil rice, two *indica* types and one *japonica* type, were also sequenced at approximately 30x genome coverage. A total of 2,149,991 SNPs were detected between Tongil and Nipponbare; the average SNP frequency of Tongil was 5.77 per kb. Genome composition was determined based on SNP data by comparing Tongil with three parental genome sequences using the sliding window approach. Analyses revealed that 91.8% of the Tongil genome originated from the *indica* parents and 7.9% from the *japonica* parent. Copy number of SSR motifs, ORF gene distribution throughout the whole genome, gene ontology (GO) annotation, and some yield-related QTLs or genes variation were also comparatively analyzed

between Tongil and parents using sequence-based tools. Each genetic factor was transferred from the parents into Tongil rice in amounts that were in proportion to the whole genome composition and these analyses support the finding that the Tongil genome is primarily made up of the *indica* genome.

Keywords: Tongil rice, three-way cross, next-generation sequencing, SEG map*, indica*/*japonica* hybridization

INTRODUCTION

Rice (*Oryza sativa* L.) is a staple food for more than half of the world's population, providing about 19 percent of the world's and 29 percent of Asia's caloric supply (IRRI, 2009). Although demands on the nutritional and industrial functionality of rice are increasing, especially to improve human health and quality of life, improving the yield potential of rice is still a major challenge for rice breeders, who must address the rapid growth of the world population along with dramatic reductions in the amount of cultivated land (Khush, 1999), as well as environmental challenges(Nelson and International Food Policy Research Institute., 2009). Asian varieties of cultivated rice include two major subspecies, *O. sativa indica* and *O. s. japonica,* which are differentiated based on morphological and physiological characteristics and geographical distribution (Morishima and Oka, 1981; Sano and Morishima, 1992). *Oryza sativa. indica* cultivars have higher genetic diversity (Lu et al., 2002), a broader cultivation range, and stronger resistance to prominent diseases and insect pests compared to *Oryza sativa. japonica* cultivars (Chung and Heu, 1991). Inter-subspecific hybridization between *indica* and *japonica* rice cultivars may enrich allelic variation and facilitate hybrid vigor by creating new genetic recombination (Cheng et al., 2007). In spite of these advantages, the introduction of desirable *indica* traits into the *japonica* variety has not been successful due to reproductive barriers and incorporation of undesirable characteristics, such as low eating quality for people who prefer the taste of *japonica* rice (Chung and Heu, 1991).

Tongil rice (IR667-98-1-2) is the first semi-dwarf variety obtained by a three-way cross of *indica*/*japonica* varieties as part of a collaborative research project between the International Rice Research Institute (IRRI) and the government of South Korea. The development of Tongil rice resulted in a significant yield increase from 4 to 5 t ha^{-1}, corresponding to a 30% yield increase relative to the leading *japonica* varieties grown in Korea (Chung and Heu, 1980). After the introduction of Tongil rice in 1972, Korean rice production significantly increased and the South Korean government announced the achievement of rice self-sufficiency (the so-called 'Green Revolution') in 1977. However, the genome characterization and structure of Tongil rice have never been analyzed.

Rice is a useful model crop for studying genome structure due to its relatively small genome. Furthermore, its genetic and physical data have been extensively analyzed by the International Rice Genome Sequencing Project (IRGSP) (Matsumoto et al., 2005). The recent improvement of next-generation sequencing (NGS) technology has enabled high-throughput genotyping and elucidation of genome structures of various rice cultivars (Huang et al., 2009; Huang et al., 2012a). Most sequence-based rice genome analyses are based on DNA polymorphisms, single nucleotide polymorphisms (SNPs) and insertion-deletions (InDels). SNP detection is the first step for comparing DNA variation and is an effective tool to elucidate genome structure and composition (Chen et al., 2014; Feltus et al., 2004; McNally et al., 2009; Shen et al., 2010).

In this study, we sequenced the whole genomes of Tongil rice (*Oryza*

sativa L.) and its parental varieties to analyze the genome structure of Tongil in detail and to identify regions of the *indica* and *japonica* parental genomes that introgressed in the Tongil genome. In addition, we analyzed previously reported yield-related genes (*Gn1a*, *Ghd7*, *sd1*, *GS3* and *qSW5*), SSRs, GO annotation, and other genetic characteristics of the Tongil genome.

MATERIALS AND METHODS

Plant Materials

Plant lines subjected to whole-genome resequencing in the present study included Tongil (SNU accession no. 260697) and its parental lines: Yukara, an early maturing temperate *japonica* cultivar (RDA-Genebank Information Center accession no. IT004665); Taichung native 1 (TN1), the first semi-dwarf *indica* variety with high adaptability (RDA-Genebank Information Center accession no. IT004120); and IR8, an improved high-yielding semi-dwarf variety developed at the International Rice Research Institute (IRRI, IRTP 195). The Tongil variety was developed through a three-way cross, IR8//Yukara/TN1. With generation advancement after the cross, the most promising line, IR667-98-1-2, was selected and released to farmers in Korea under the name 'Tongil' (Chung and Heu, 1991).

Whole- genome DNA sequencing

Four rice varieties were sequenced: Tongil and its parental varieties, Yukara, IR8, and TN1. Whole-genome shotgun sequencing of the four rice genomes was performed using the Illumina/Solexa GAII system. DNA sequencing, including construction of shotgun DNA libraries, was performed according to the methods recommended by the manufacturer (Illumina, San Diego, CA, USA). Briefly, whole-genome DNA shotgun paired-end sequencing libraries were generated by fragmentation of DNA

into 500-bp segments using a Covaris DNA shearing machine (Covaris, CA, US), followed by ligation of paired-end adapters ligation of 53 and 68 bp for sequencing on the FlowCell, size selection of the adapter-ligated fragments within the desired size range (500–600 bp), and PCR enrichment using complete primer constructs required for binding and clustering on the FlowCell. Illumina GAII sequencing was performed by identifying the emission color of single-base extensions on the FlowCell.

DNA variation

Illumina whole-genome shotgun 100-bp paired-end DNA sequencing data were filtered to obtain high-quality sequence data and to map reads to the Nipponbare reference genome sequence, which as downloaded from NCBI. Briefly, high-quality sequence with at least QC20-justified phred quality score was mapped to the reference Nipponbare sequence using CLC NGS Cell software (http://www.clcbio.com). The DNA sequence variation DB was converted to text format, including DNA variation based on the reference position, for the analysis of genome structure.

SNP calling – probabilities

Genotype calling to identify regions originating from the *japonica* and *indica* genomes was performed using the sliding-window approach suggested by Huang et al. (Huang et al., 2009). In each window, the proportion of SNPs originating from each parent was examined for

genotype calling. Huang et al. determined optimum window size by calculating the probability of finding a specific number of *japonica* SNPs in a window based on SNP error rates. Recent improvements in sequencing technology, however, resulted in fewer errors in SNP identification. Thus, the method suggested by Huang et al. (Huang et al., 2009) was not directly applicable in this study. Even with a window size of 2, for example, calling accuracy could reach 99.99%. Instead of calculating this probability, the optimum window size was determined iteratively by comparing the portion of *japonica* SNPs (*O*) and the portion of the genome originating from *japonica* (*P*). Tongil was resequenced to obtain SNPs originating from its parents and to calculate the percentage of *japonica* SNPs in each chromosome. SEG-Map software (Huang et al., 2009) was also used for genotype calling on each chromosome. Because the optimum window size was unknown, a range of window sizes from 1 to 199 was used. Then, the Nash-Sutcliffe efficiency (*E*) between *O* and *P* was calculated as follows:

$$E = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - O_m)^2},$$

Here, an individual chromosome is denoted by *i*. The average percentage of *japonica* SNPs on each chromosome is denoted by O_m. The optimal window size was defined as that with a maximum value of *E*; values of *E* ranged from -29 to 0.963. This maximum value of E occurred with a window size of 9. The percentage of *indica* SNPs was at its second highest (0.966) with a window size of 9. At a window size of 10, the *E* value dropped

rapidly for *japonica* SNPs (0.037) and *indica* SNPs (-0.018). Thus, a window size of 9 was selected as the optimum for data analysis (Figure 1-3).

Parental genome composition of Tongil

We compared DNA variation between the parental and Tongil genomes. Genomic regions originating from the *japonica* (Yukara) and *indica* (TN1 or IR8) parents were identified by comparing the Tongil genome sequence to parental sequences. Estimated *indica* and *japonica* regions in the Tongil genome sequence were calculated based on the methods of Zhao et al (Zhao et al., 2010a).

Gene ontology and classification

Annotated Nipponbare reference genes were classified based on parental origin in the Tongil genome and assigned to the three main GO-term categories (cellular component, molecular function, and biological process) using BLAST2GO software (www.blast2go.com) (Conesa et al., 2005).

Simple sequence repeats (SSRs)

SSR loci were searched using SSR search software (Initiative, 2000) and classified with respect to their parental origin.

Accession codes

Raw sequence data obtained in our study have been submitted to the NCBI Short Read Archive with the following accession numbers: Tongil [SRA: SRR923809, SRA: SRR923810], IR8 [SRA: SRR921498], TN1 [SRA: SRR921505], and Yukara [SRA: SRR925387].

RESULTS

Genome structure of Tongil

The whole genomes of Tongil and its three parental varieties, Yukara, IR8, and TN1 (Taichung Native 1) (Figure 1-1), were sequenced on the Illumina-GAII platform. A large number of short reads were mapped onto the reference Nipponbare genome and then assembled into a consensus sequence. A total of 199,543,820 reads of the Tongil genome, corresponding to 17,339,883,560 bp (17.3 Gb), were generated, representing a 47-fold sequence depth and covering 88.8% of the Nipponbare pseudomolecules (Table 1-1 and Table 1-2). We detected a total of 2,149,991 SNPs between Tongil and Nipponbare sequences. The two *indica* parents of Tongil, IR8 and TN1, had 6.22 and 6.04 SNPs per kb, respectively, whereas the *japonica* parent of Tongil, Yukara, had only 0.49 SNP per kb (Table 1-3). Using the SNP data sets from Tongil and its parents, we defined the genomic origins of regions of the Tongil genome by SNP calling (Figure 1-3 and Table 1-4 and Table 1-5; see also the SNP calling section in the Materials and Methods), and then performed a SEG-Map analysis (Zhao et al., 2010b) of Tongil (Figure 1-2). The whole genome of Tongil consisted of an average contribution of 91.8% from *indica*, 7.9% from *japonica*, and 0.3% unknown (i.e., not defined as *indica* or *japonica* regions) (Figure 1-2 and Table 1-4). The contribution of *indica* to the Tongil genome varied across chromosomes, from 74% (Chr. 2) to 100% (Chr. 12). A relatively high proportion of the *japonica* genome was found on chromosomes 1, 2, and 3, whereas the *japonica* sequences were barely

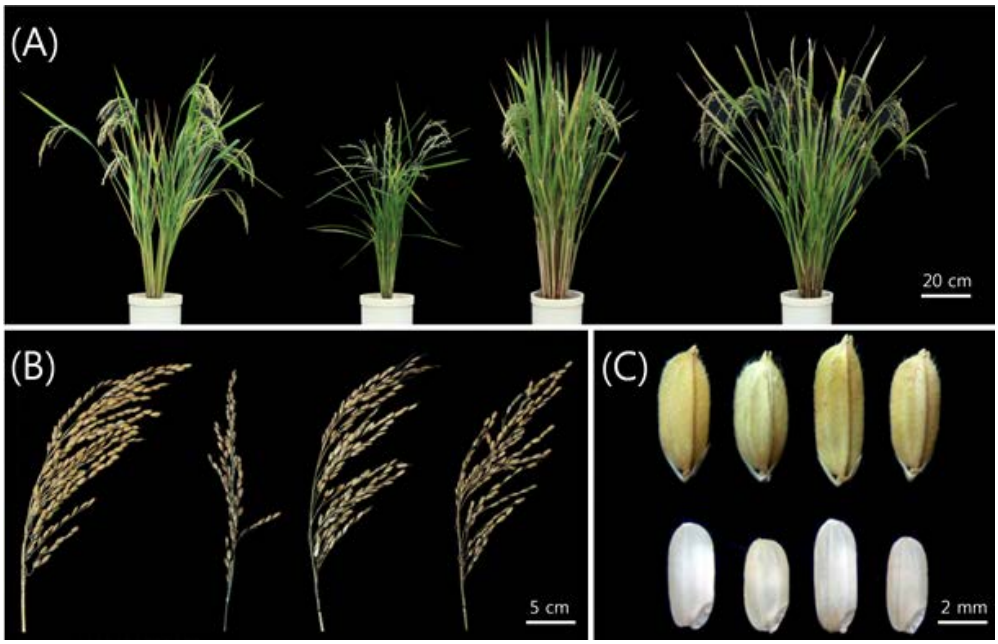


Figure 1-1. Morphological comparison of Tongil and its parental lines. From left to right: Tongil, Yukara, IR8, and TN1. (A) The plant architecture of Tongil, its *japonica* parent (Yukara), and its *indica* parent (IR8 and TN1). (B) The panicle phenotype of Tongil and its parents. (C) The brown rice shape and grain shape of Tongil and its parents. Scale bars are included in each panel.

Table 1-1. General sequencing statistics for Tongil and its parental genomes. Sequencing and mapping against Nipponbare reference genome.

Variety	Number of Reads	Total Read Length (bp)	Mapped Read Length (bp)	Sequencing Depth (×)	Coverage [a)] (%)	SNP Frequency (SNPs/kb)
Tongil	199,543,820	17,339,883,560	330,933,489	47	88.8	5.77
Yukara	114,615,268	12,429,060,750	345,058,384	34	92.6	0.49
IR8	109,304,614	11,790,909,253	327,065,806	32	87.7	6.22
TN1	105,708,026	11,299,286,038	326,132,058	30	87.5	6.04

[a)] Coverage to Nipponbare genome sequence

Table 1-2. Mapping coverage of Tongil rice and its three parents

Chromosome	Reference	Tongil		Yukara		IR8		TN1	
	Pseudomolecule	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)	Aligned Length (bp)	Coverage (%)
1	45,038,604	38,574,603	89.2	39,521,561	91.4	37,917,584	87.7	37,633,042	87.0
2	36,792,247	32,784,930	91.2	33,736,987	93.9	32,281,381	89.8	32,249,409	89.8
3	37,312,367	33,412,277	91.9	34,084,555	93.7	33,168,624	91.2	33,108,547	91.0
4	36,060,865	31,735,280	89.3	33,210,463	93.5	31,468,452	88.6	31,193,791	87.8
5	30,073,438	27,991,645	93.5	28,624,984	95.6	27,856,337	93.0	27,770,415	92.7
6	32,124,789	27,784,237	89.0	28,323,987	90.7	27,516,550	88.1	27,400,223	87.7
7	30,357,780	26,328,583	88.6	27,297,454	91.8	25,826,093	86.9	25,896,533	87.1
8	28,530,027	25,808,735	90.7	27,150,402	95.5	25,591,577	90.0	25,618,079	90.1
9	23,895,721	19,770,440	85.9	21,074,657	91.6	19,774,924	86.0	19,460,581	84.6
10	23,703,430	20,191,910	88.0	21,233,194	92.5	19,872,700	86.6	19,739,332	86.0
11	31,219,694	22,752,159	78.7	24,726,750	85.5	22,345,311	77.3	22,512,524	77.8
12	27,679,166	23,798,690	86.4	26,073,390	94.7	23,446,273	85.2	23,549,582	85.5
Total or Ave.	382,788,128	330,933,489	88.8	345,058,384	92.6	327,065,806	87.7	326,132,058	87.5

Table 1-3. SNPs and SNP frequency of Tongil and its three parents

	Reference	Tongil		Yukara		IR8		TN1	
Chromosome	Pseudomolecule	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)	Number of SNP	SNP Frequency (SNPs/Kbp)
1	45,038,604	222,438	5.14	19,609	0.45	267,113	6.18	262,349	6.07
2	36,792,247	159,177	4.43	7,129	0.20	225,762	6.28	209,537	5.83
3	37,312,367	183,617	5.05	6,785	0.19	205,566	5.65	192,339	5.29
4	36,060,865	177,722	5.00	24,186	0.68	190,976	5.38	182,173	5.13
5	30,073,438	132,916	4.44	3,276	0.11	141,789	4.73	145,396	4.85
6	32,124,789	184,344	5.90	23,105	0.74	187,503	6.00	184,005	5.89
7	30,357,780	198,051	6.66	21,460	0.72	205,384	6.91	191,745	6.45
8	28,530,027	162,946	5.73	15,459	0.54	163,856	5.76	162,534	5.71
9	23,895,721	154,565	6.72	11,273	0.49	143,014	6.22	150,864	6.56
10	23,703,430	146,848	6.40	6,749	0.29	163,379	7.12	165,561	7.22
11	31,219,694	226,457	7.83	28,203	0.98	227,571	7.87	216,814	7.50
12	27,679,166	200,910	7.30	17,131	0.62	198,404	7.21	190,243	6.91
Total or Ave.	382,788,128	2,149,991	5.77	184,365	0.49	2,320,317	6.22	2,253,560	6.04

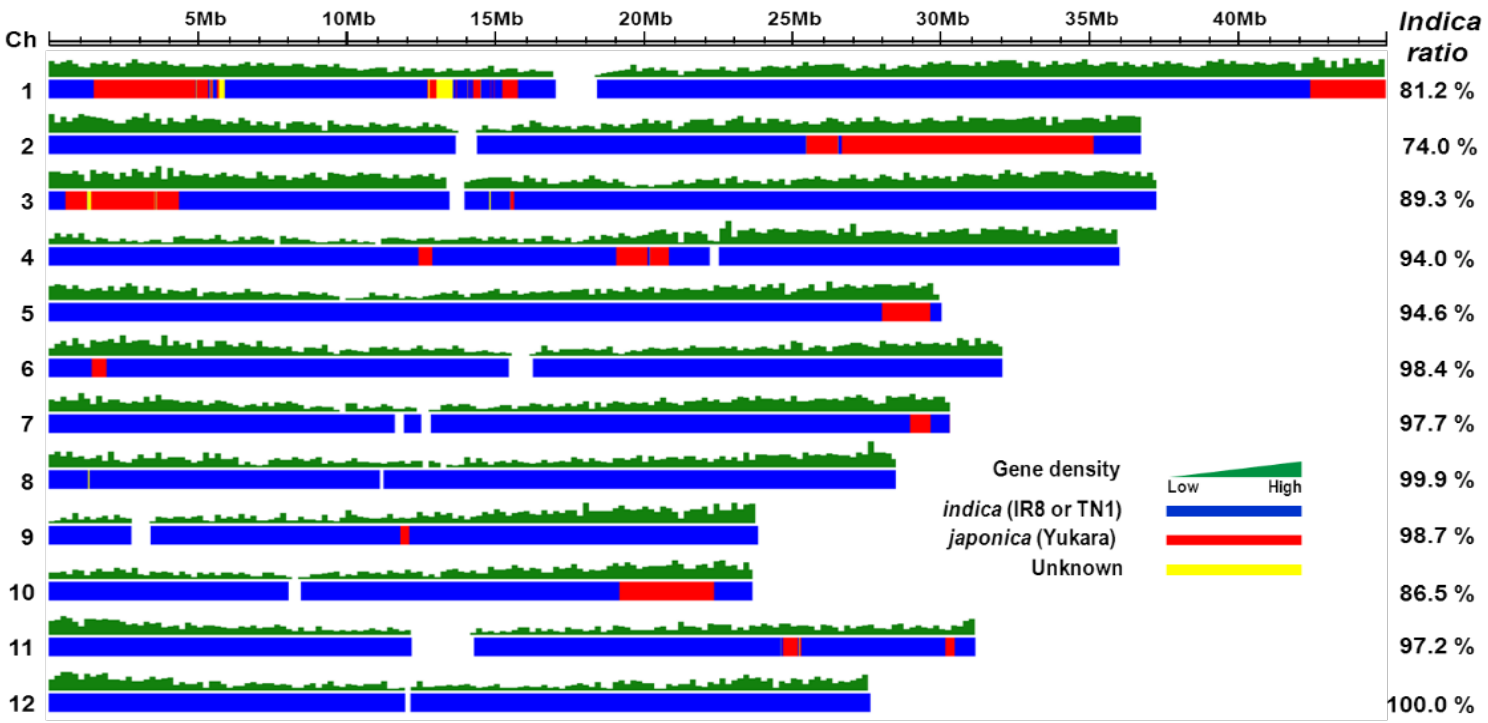


Figure 1-2. *Indica*/*japonica* genome organization on the 12 chromosomes of Tongil. Blue indicates the *indica* genome (TN1 and IR8); red indicates the *japonica* genome (Yukara); and yellow indicates a region from an unknown genome. The percentages indicate the proportion of *indica* contribution on each chromosome.

detectable on chromosomes 8 and 12. In addition, there were no differences in gene density between the *indica*- and *japonica*-derived genome regions of Tongil (Figure 1-2 and Table 1-6).

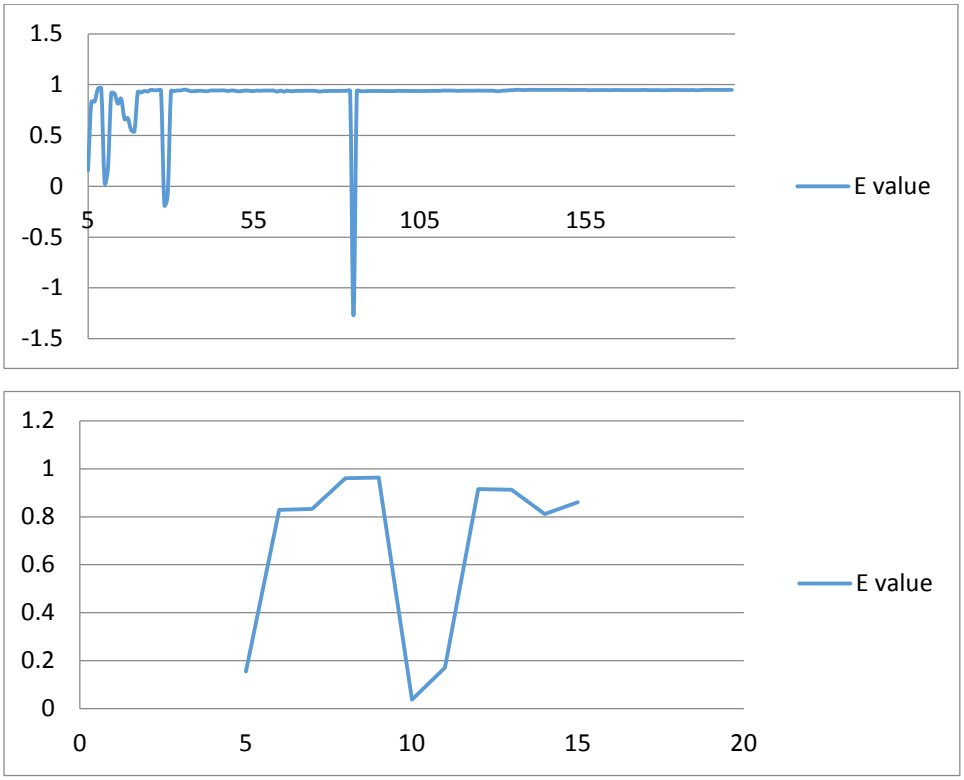


Figure 1-3. Determination of window size followed by *E*-value calculation

Table 1-4. Genome region definition by presence (O) or absence (X) of SNPs difference

Genome origin from / Define as	Presence of SNP or not		
	Yukara	IR8	TN1
IR8 / *indica*	O	X	O
TN1 / *indica*	O	O	X
IR8 and TN1 / *indica*	O	X	X
Yukara / *japonica*	X	O	O

Table 1-5. The ratio of *indica-japonica* specific SNPs determined by comparison of Tongil and its parental varieties.

Chromosome	*Indica* SNP	Ratio (%)	*Japonica* SNP	Ratio (%)	Total
1	129,021	82.1	28,045	17.9	157,066
2	96,765	74.8	32,628	25.2	129,393
3	115,351	88.7	14,622	11.3	129,973
4	92,280	92.6	7,406	7.4	99,686
5	82,754	92.5	6,712	7.5	89,466
6	114,578	96.2	4,571	3.8	119,149
7	113,844	96.5	4,082	3.5	117,926
8	92,834	98.3	1,598	1.7	94,432
9	87,933	98.1	1,678	1.9	89,611
10	83,513	85.6	14,010	14.4	97,523
11	123,151	97.9	2,689	2.1	125,840
12	102,312	98.3	1,781	1.7	104,093
Total or Ave.	1,234,336	91.2	119,822	8.8	1,354,158

Table 1-6. Determination of the *indica*/*japonica* genome origin of Tongil, based on a window size of 9.

Chromosome	Pseudomolecule	*indica* region (bp)	Ratio (%)	*japonica* region (bp)	Ratio (%)	Unknown region (bp)	Ratio (%)
1	45,038,604	36,563,905	81.2	7,596,808	16.9	877,891	2.0
2	36,792,247	27,235,850	74.0	9,544,379	25.9	12,018	0.0
3	37,312,367	33,336,733	89.3	3,748,667	10.1	226,967	0.6
4	36,060,865	33,898,364	94.0	2,150,911	6.0	11,590	0.0
5	30,073,438	28,436,341	94.6	1,637,097	5.4	-	-
6	32,124,789	31,619,689	98.4	499,676	1.6	5,424	0.0
7	30,357,780	29,667,148	97.7	690,632	2.3	-	-
8	28,530,027	28,487,631	99.9	333	-	42,063	0.2
9	23,895,721	23,592,877	98.7	302,844	1.3	-	-
10	23,703,430	20,504,662	86.5	3,198,768	13.5	-	-
11	31,219,694	30,345,040	97.2	846,802	2.7	27,852	0.1
12	27,679,166	27,679,166	100.0	-	-	-	-
Total	382,788,128	351,367,406	91.8	30,216,917	7.9	1,203,805	0.3

Gene distribution and gene ontology analysis in Tongil

We analyzed the gene content of Tongil to understand the relationship between the composition of the genome and genes (Open Reading Frames: ORFs), and also to elucidate the distribution of *indica*- and *japonica*-originated genes (alleles) in the Tongil genome. The gene distribution ratio according to *indica* or *japonica* genome composition was similar to the genome distribution ratio of Tongil (Table 1-6 and Table 1-7). The origin of genes from the *indica* and *japonica* parents were 88.3% and 11.4%, respectively, suggesting that the average gene composition was similar to the genome composition ratio of Tongil, although the distribution of parental origin varied across chromosomes. To identify a biological pattern in a list of the genes that belong to the *indica*, *japonica*, and unknown genomes, we performed gene ontology (GO) analysis of the Tongil genome according to three categories: cellular component, molecular function, and biological process (Figure 1-4 and Figure 1-5 and Figure 1-6). The results of GO analysis revealed that the average contribution of *indica* or *japonica* genome to each GO category was almost identical to the gene and genome distribution ratios. *Indica* and *japonica* contributed 86.8% and 12.7% of cellular components, 87.4% and 12.2% of molecular functions, and 87.3% and 12.2% of biological processes, respectively. However, in the 'molecular function' category, all 17 genes related to channel regulator activity belonged to *indica* regions, whereas in the biological process category, all adhesion-related genes in the 'biological process' category belonged to only *japonica* regions.

Table 1-7. Gene distribution of Tongil

Chromosome	*Indica*		*Japonica*		Unknown		Total
	No. of genes	%	No. of genes	%	No. of genes	%	
1	3,614	77.9	953	20.5	75	1.6	4,642
2	2,553	68.4	1,178	31.6	2	0.1	3,733
3	3,475	85.7	548	13.5	31	0.8	4,054
4	2,713	93.2	197	6.8	-	0.0	2,910
5	2,440	90.9	245	9.1	-	0.0	2,685
6	2,659	97.2	77	2.8	1	0.0	2,737
7	2,485	96.4	94	3.6	-	0.0	2,579
8	2,261	99.9	-	0.0	2	0.1	2,263
9	1,858	99.3	14	0.7	-	0.0	1,872
10	1,390	77.0	416	23.0	-	0.0	1,806
11	2,012	97.0	62	3.0	1	0.0	2,075
12	1,921	100.0	-	0.0	-	0.0	1,921
Total or Ave.	29,381	88.3	3,784	11.4	112	0.3	33,277

Figure 1-4. GO analysis by 'cellular component' category of Tongil genes, corresponding to *indica*/*japonica* sequence.

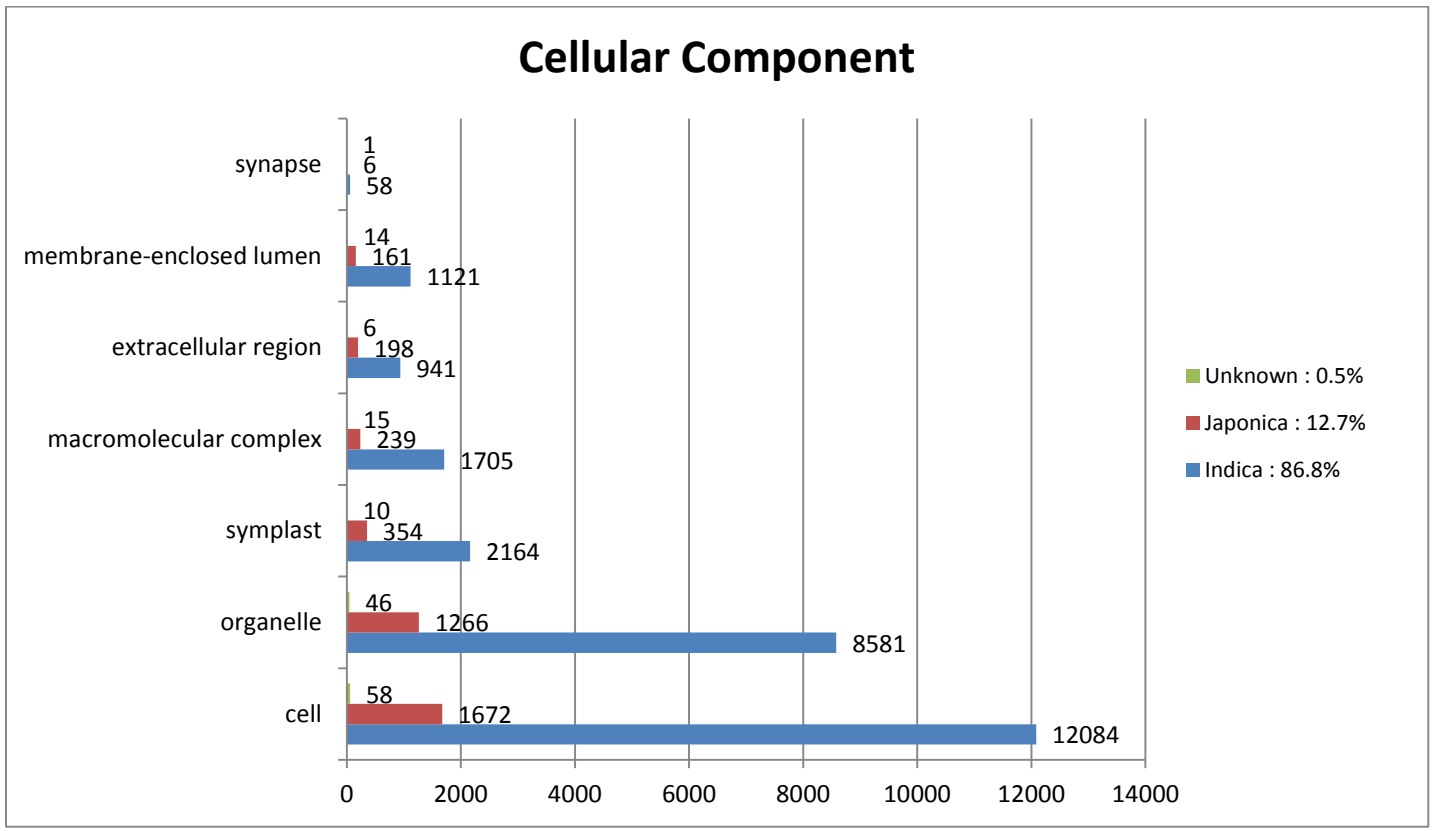


Figure 1-5. GO analysis by 'molecular function' category of Tongil genes, corresponding to *indica*/*japonica* sequence.

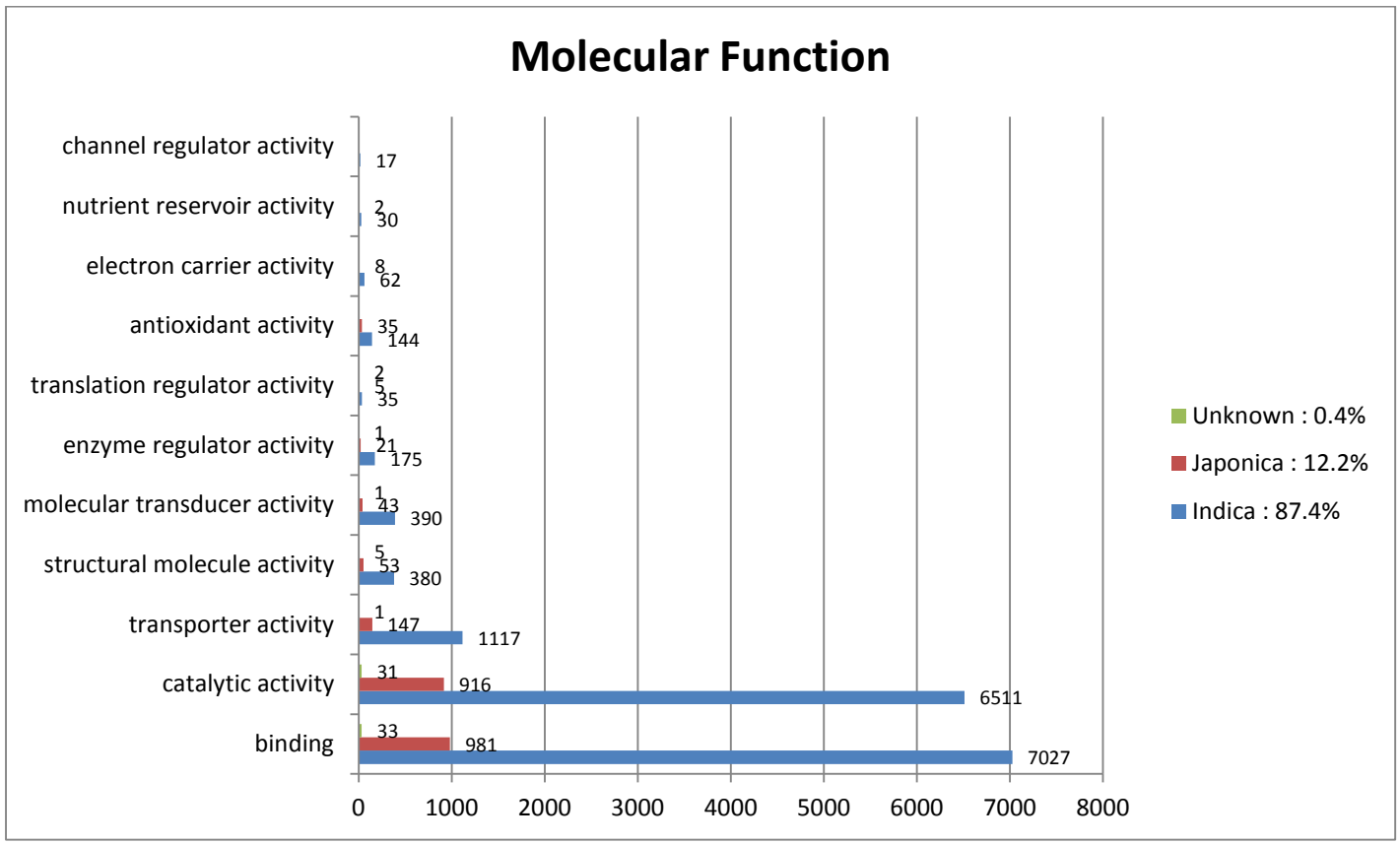
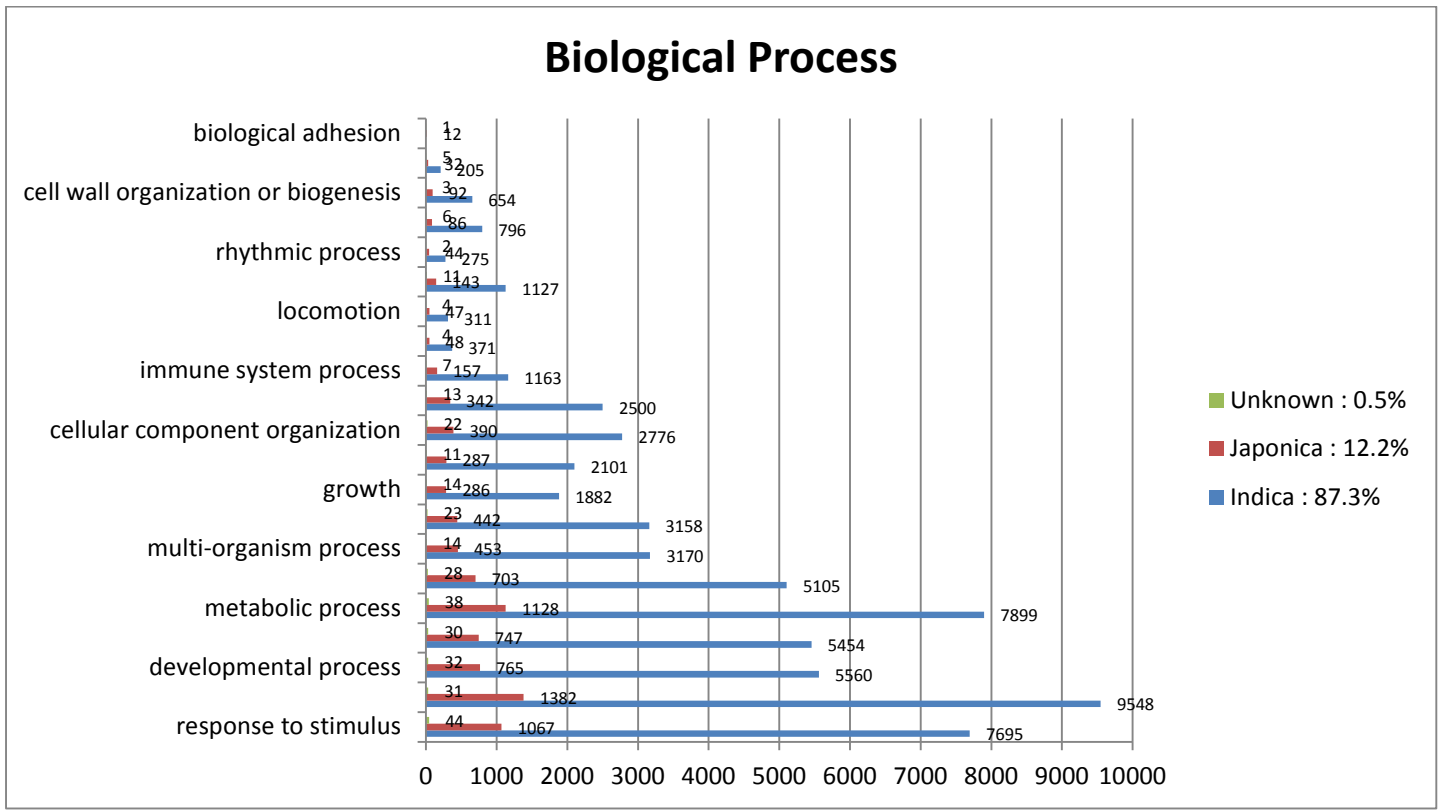


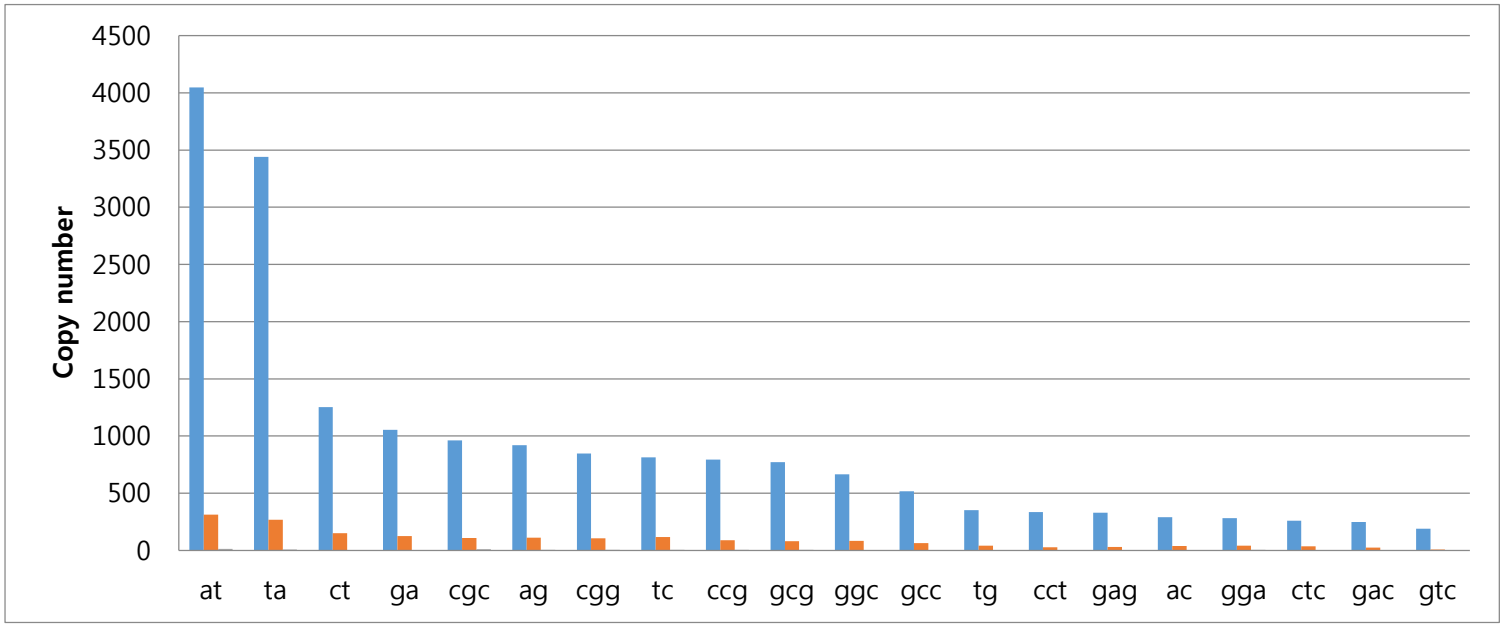
Figure 1-6. GO analysis by 'biological process' category of Tongil genes, corresponding to *indica*/*japonica* sequence



Simple sequence repeats (SSRs) in Tongil genome

A total of 177 distinctive motif families were annotated on the Tongil genome (Figure 1-7). Among the classified repeats, di-nucleotide repeats were predominant, and AT/TA repeats were the most abundant motif in both *indica* (29.09%) and *japonica* (21.8%) regions within the Tongil genome. The next most abundant motif relative to AT/TA was CT/GA, and CGC was the most abundant motif among tri-nucleotide repeats. The di-, tri-, and tetra-nucleotide repeat pattern was not similar to that of the Nipponbare genome (McCouch et al., 2002; Zhou et al., 2005), and it also differed from that of wheat (Weng et al., 2005). A total of 90.1% of SSR motifs in the Tongil genome were from *indica*, 9.6% were from *japonica*, and 0.3% were from an unknown genome (Figure 1-7).

Figure 1-7. Copy number of SSR motif families in Tongil



Indica (%)	4048 (92.6)	3440 (92.6)	1253 (89.2)	1053 (89.2)	963 (88.9)	919 (88.8)	848 (88.5)	815 (87.1)	794 (89.4)	773 (90.1)	666 (88.8)	517 (88.7)	352 (89.3)	335 (91.8)	329 (91.4)	290 (88.4)	282 (85.7)	261 (87.3)	249 (89.9)	191 (95.5)
Japonica (%)	312 (7.1)	267 (7.2)	151 (10.8)	125 (10.6)	110 (10.2)	111 (10.7)	105 (11.0)	117 (12.5)	90 (10.1)	80 (9.3)	83 (11.1)	65 (11.1)	42 (10.7)	29 (7.9)	30 (8.3)	38 (11.6)	42 (12.8)	35 (11.7)	26 (9.4)	9 (4.5)
Unknown (%)	13 (0.3)	8 (0.2)	0 (0.0)	2 (0.2)	10 (0.9)	5 (0.5)	5 (0.5)	4 (0.4)	4 (0.5)	5 (0.6)	1 (0.1)	1 (0.2)	0 (0.0)	1 (0.3)	1 (0.3)	0 (0.0)	5 (1.5)	3 (1.0)	2 (0.7)	0 (0.0)

Distribution of yield-related genes in the Tongil genome

One of the most important points of this study was to explore which regions of the *indica* and *japonica* parental genomes have been introgressed in the Tongil variety to provide its high-yield potential. Tongil is morphologically characterized by short plant height, lodging resistance, open plant architecture, medium-long and erect leaves, thick leaf sheaths and culms, relatively long panicles, and easily shattered grain (Chung and Heu, 1980) (Figure 1-1). Although these phenotypic characteristics must affect Tongil's high-yield potential, to date we have no molecular genetic evidence regarding the nature of these traits, with the exception of semi-dwarf gene 1 (*sd1*) (Chung and Heu, 1980). Therefore, we analyzed several well-characterized genes that are associated with high yield potential: *sd1* (Nagano et al., 2005), *Ghd7* (Liu et al., 2013; Xue et al., 2008), *Gn1a* (Ashikari et al., 2005), *qSW5* (Yan et al., 2011), *GS3* (Takano-Kai et al., 2009), and *GW2* (Li et al., 2010) in the Tongil genome.

sd1 : semi-dwarf stature

Semi-dwarf stature is one of the main genetic contributors to the green revolution. Introduction of semi-dwarf genes increased yield by conferring lodging resistance, which enabled more input of nitrogen fertilizer. Tongil was the first variety into which the *sd1* allele was introduced in South Korea. Analysis of the *sd1* gene, which encodes GA20ox-2 in Tongil and its parents, revealed that Tongil received its *sd1* from an *indica* parent, IR8 or TN1; this allele contains a 383 bp deletion resulting in a frame-shift to make a stop

codon (Nagano et al., 2005) (Figure 1-8A). We also confirmed other *sd1* alleles from the native semi-dwarf rice variety Jikkoku (G281T) and the γ-ray–induced varieties Reimei (C1045G) and Calose76 (C796T) (Monna et al., 2002; Sasaki et al., 2002). On the other hand, Yukara, the *japonica* parent of Tongil, did not have any *sd1* alleles.

Ghd7 : grain number, plant height, and heading date

A gene encoding a CCT domain protein, *Ghd7*, is an important regulator of potential yield, plant height, and heading date in rice. Plant height and panicle size are increased under long-day conditions by the delay in heading date resulting from increased *Ghd7* expression (Liu et al., 2013; Xue et al., 2008). *Ghd7* has five natural variant haplotypes. Tongil has the *Ghd7-1* allele (A-G-S-V-A) derived from *indica* parent IR8 or TN1 (Figure 1-8B), considered to be the original wild-type allele with full function; plants with this allele are relatively tall, late heading, have a large panicle, and are widely grown. By contrast, the *japonica* parent of Tongil, Yukara, has the *Ghd7-2* allele (A-E-G-D-P), which is weaker than *Ghd7-1* and is found in temperate *japonica* varieties.

Gn1a : Grain number on chromosome 1

Gn1a is one of the most effective QTLs for increasing grain number. It is predicted to encode a cytokinin oxidase/dehydrogenase (OsCKX2). Habataki, an *indica* rice variety, has a 16 bp deletion in the 5′ UTR, a 6 bp deletion in the first exon, and three amino acid substitutions in the first

and fourth exons of this gene. In addition, an 11 bp deletion in the third exon has been detected in the high-yielding rice variety 5150 (Ashikari et al., 2005). Comparison of DNA sequences between Tongil and parent varieties revealed that Tongil is identical to the TN1 allele, which has only a 6 bp deletion in the first exon and no 16 bp deletion in the 5′ UTR, as in Habataki. On the other hand, IR8 contains a 16 bp deletion in the 5′ UTR and a 12 bp deletion in the first exon, distinct from the pattern in the TN1 allele. We could not identify any variation in Yukara, which has the same allele as Nipponbare (Figure 1-8C).

qSW5 : QTL for seed width on chromosome 5

qSW5 is responsible for seed width; the product of this gene controls cell number in the outer glume of the rice flower. The gene product increases seed width and seed weight by enlarging sink size. The Nipponbare-type allele, which contains a 1,212 bp deletion, is loss-of-function relative to the Kasalath-type allele (Shomura et al., 2008). In addition, the *indica* Ⅱ-type allele has a 950 bp deletion relative to the Kasalath allele (Yan et al., 2011). Tongil, IR8, and TN1 have the *indica* Ⅱ-type allele, whereas Yukara contains the Nipponbare allele (Figure 1-8D).

GS3 : Grain length and weight (grain size 3)

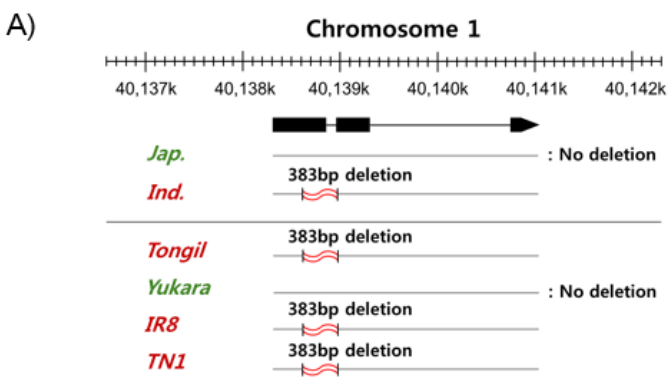
GS3, which encodes a PEPB-like domain protein, was cloned from a QTL for grain length and weight on chromosome 3 in rice. A C-to-A substitution in the second exon of the *GS3* gene is highly associated with grain length

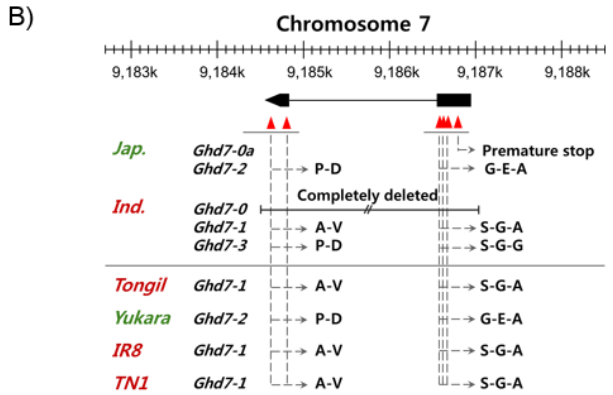
and width: the A-allele confers significantly longer and thinner grains than the C-allele. Tongil possesses an A-allele originating from an *indica* parent, IR8 or TN1, whereas the *japonica* parent, Yukara, has the C-allele (Figure 1-8E). In the case of another gene that controls grain width, *GW2*, there were no SNPs among any of the strains we sequenced or Nipponbare, indicating that *GW2* is a highly conserved gene in rice and even in *Zea mays* (Yan et al., 2011).

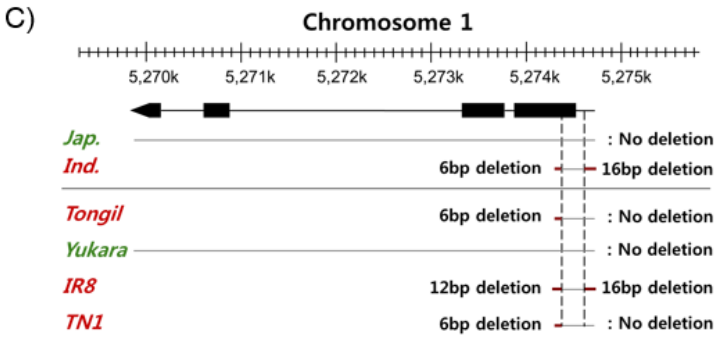
Table 1-8. Comparison of alleles of yield-related genes in Tongil and its parents

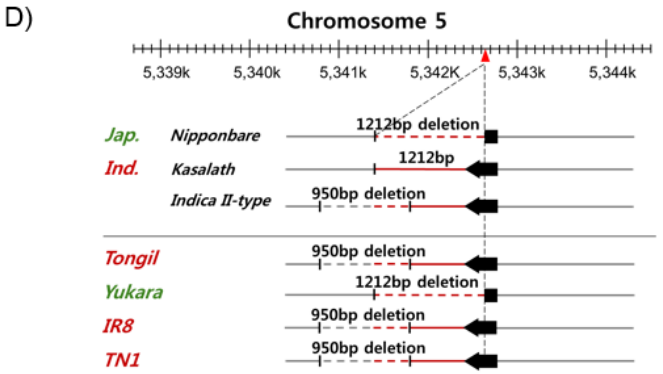
Gene	Gene name	Tongil	Yukara	TN1	IR8
qSW5	QTL for Seed width on chromosome 5	**I**[a]	**J**	**I**[a]	**I**[a]
Gn1a	Grain number in chromosome 1	**I**	**J**	**I**	**I**[b]
Ghd7	Grain number, plant height and heading date	**I**	**J**	**I**	**I**
sd1	Semi Dwarf 1	**I**	**J**	**I**	**I**
GS3	Grain size 3	**I**	**J**	**I**	**I**
GW2	Grain width and weight	**I**	**J**	**I**	**I**
S5	Hybrid Sterility 5	**I**	**J**	**I**	**I**

a: *indica* -II type
b: different *indica* type allele with Tongil and TN1


A)
Chromosome 1
40,137k
40,138k
40,139k
40,140k
40,141k
40,142k
Jap.
: No deletion
Ind.
383bp deletion
Tongil
383bp deletion
Yukara
: No deletion
IR8
383bp deletion
TN1
383bp deletion


B)
Chromosome 7
9,183k
9,184k
9,185k
9,186k
9,187k
9,188k
Jap.
Ghd7-0a
Premature stop
Ghd7-2
P-D
G-E-A
Ind.
Completely deleted
Ghd7-0
Ghd7-1
A-V
S-G-A
Ghd7-3
P-D
S-G-G
Tongil
Ghd7-1
A-V
S-G-A
Yukara
Ghd7-2
P-D
G-E-A
IR8
Ghd7-1
A-V
S-G-A
TN1
Ghd7-1
A-V
S-G-A


C)
Chromosome 1
5,270k
5,271k
5,272k
5,273k
5,274k
5,275k
Jap.
: No deletion
Ind.
6bp deletion
16bp deletion
Tongil
6bp deletion
: No deletion
Yukara
: No deletion
IR8
12bp deletion
16bp deletion
TN1
6bp deletion
: No deletion


D)
Chromosome 5
5,339k
5,340k
5,341k
5,342K
5,343k
5,344k
Jap.
Nipponbare
1212bp deletion
Ind.
Kasalath
1212bp
Indica II-type
950bp deletion
Tongil
950bp deletion
Yukara
1212bp deletion
IR8
950bp deletion
TN1
950bp deletion

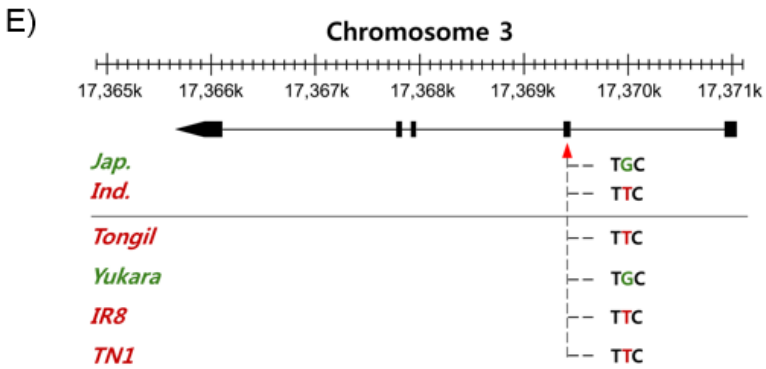


Figure 1-8. *Indica*/*japonica* region comparisons of high yield-related alleles or QTLs. A) *sd1*, B) *Ghd7*, C) *Gn1a*, D) *qSW5*, and E) *GS3*. Black arrows and box regions represent exons. Vertical, dashed lines refer to the same position in the genome or gene region.

DISCUSSION

indica/*japonica* genome composition of Tongil

In this study, we analyzed genome structure of Tongil using high-depth NGS tool to demonstrate that the Tongil genome is composed of 91.8% *indica*, 7.9% *japonica*, and 0.3% unknown genome (Figure 1-2 and Table 1-6). The amounts and types of genes and SSRs in the Tongil genome were very similar to its genomic composition with respect to *indica* or *japonica* origin. This deviation from the expectation that about one-fourth of the Tongil genome originated from the *japonica* parent is likely due to the results of selection during the breeding process and/or to segregation distortion in favor of the *indica* genome because *indica*-type alleles and plants are favored among hybrid progenies from *indica*/*japonica* crosses

These results suggest several possibilities, Linkage disequilibrium (LD) is a non-random association of alleles at two more different loci (Flint-Garcia et al., 2003). Independently segregating, in a large and random mated population without selection, mutation, migration, polymorphic loci will be in linkage equilibrium (Falconer et al., 1996). However, when rice varieties are developed, LD level is increased due to small population, artificial mating, and selection for desirable agronomic trait thus, biased polymorphic loci occurred in rice genome. In addition, plant genes have been considered to be randomly distributed through the genome. However, some resistant and metabolic genes have discovered to be clustered and inheritance together. For example, UV inducible gene, momilactone biosynthesis and chitin oligosaccharide elicitor formation related gene

cluster (Shimura et al., 2007), second multifunctional diterpenoid biosynthetic gene cluster (Swaminathan et al., 2009), and clustered genes containing *Pi-ta* for blast resistance inherited together from backcross progeny involving *indica* and *japonica* cross and large portion of chromosome fragment was maintained by artificial selection during breeding process (Jia, 2009). These gene clustering confers the selective advantage by combining the inheritance genes and facilitates regulation of gene expression by enabling localized changes in chromatin structure (Wegel et al., 2009). Some plant metabolic pathways might be clustered by evolutionary selection pressure (Field and Osbourn, 2008).

Linkage blocks (LBs), where a set of gene is inherited together is larger in *indica*/*japonica* cross than *indica*/*indica* cross due to the incompatibility between *indica* and *japonica* rice. The larger LBs, the higher recombination suppression occurred. Thus, crosses of *indica*/*japonica* are subject to large linkage block and have significant recombination suppressions (Jia et al., 2012). Similarly, linkage drag was studied by Hanson W. D. (Hanson, 1959). Linkage drag means that the rate of decrease of donor fragment is slower than for unlinked region due to the linked to the target gene from the donor plant in backcrossing (Collard and Mackill, 2008; Hospital, 2005) and linkage drag is influenced by location of chromosome, selection way, and recombination factors (Flint-Garcia et al., 2003; Naveira and Barbadilla, 1992), however the molecular mechanism of linkage drag is still unclear. The major factor of gene clustering, linkage blocks, and linkage drag is due to the selection for targeted gene or trait.

The aim of breeding of Tongil variety was development of high yielding and cold and blast resistant rice. High yield and stress tolerant related traits such as plant height, grain size, shape, panicle size, 1000 grain weight, sterility, and cold and blast disease tolerance were artificially selected by breeders, although some traits were eliminated during the shuttle breeding between Korea and IRRI. In our provided result, Tongil rice consists of 99.9% *indica* genome on chromosome 8 and no *japonica* genome sequences on chromosome 12 (Figure 1-2 and Table 1-6). These phenomenon will give an important information to study LD in the view of genome sequence and SNPs.

Segregation distortion is occurred in genetic mapping populations from *indica/japonica* cross and usually observed through genetic markers in rice thus, the segregation ratio is not fit to the Mendelian inheritance (Harushima et al., 1996). Segregation distortion is influenced by genetic and experimental factors such as population size, polymorphic markers, and genotyping error (Hackett and Broadfoot, 2003). It can arise either from gametophytic competition among gametes which gametophyte gene is expressed or from abortion of the gamete or zygote by hybrid sterility genes (Harushima et al., 1996). Three-way system at *S5* locus which responsible for hybrid sterility of inter-specific, namely *S5i*, *S5j* and *S5n* for *indica*, *japonica* and wide compatibility variety (WCV), respectively, was suggested by Ikehashi H. and Araki H. (Lin et al., 1992). WCV was used to overcome reproductive barriers and the cross combination containing *S5n* allele showed high fertile. (Chen et al., 2008b; Lin et al., 1992). The

interaction between *S5i* and *S5j* encoded aspartyl protease which has single amino acid difference was responsible for embryo sac abortion and *indica*/*japonica* hybrid sterility. Unlike *S5i* and *S5j*, *S5n* was not formed dimers due to a large deletion in the N terminus of AP protein thus, it might occurred wide compatibility (Ji et al., 2012). Killer-protector system at the *S5* locus that endoplasmic reticulum (ER) stress is triggered by ORF5+ (killer) and ORF4+(partner) while ORF3+(protector) prevents ER stress from the ORF5+ and ORF4+ to produce normal gamete was proposed (Yang et al., 2012). In contrast to female sterility genes, the *Sa*, *indica-japonica* hybrid male sterility related locus also cloned in rice. *Sa* locus comprises two adjacent genes, *Sa*M encodes a small ubiquitin-like modifier E3 ligase-like protein and *Sa*F encodes F-box, respectively. *Indica* (SaM+SaF+) and *japonica*(SaM-SaF-) hybrid male sterility is caused by selective abortion of pollen carrying SaM- therefore, interrupting its transmission into hybrid progeny and segregation distortion reveals in this allele (Long et al., 2008). These reproductive isolations through allele interactions might lead diverge into *indica* and *japonica* eventually, prevent gene exchange or gene flow between two subspecies. Segregation distortion ratio was higher in inter-specific population than intra-specific population in rice.

Our result indicates that Tongil genome was biased to *indica* unlike theoretical expectation (Table 1-6). We assumed that one of the main factor for *indica* biased genome in Tongil was the interaction *indica*/*japonica* hybrid sterility genes, although many other loci were involved in hybrid

incompatibility. The two *indica* parent varieties, IR8 and TN1 and a *japonica* parent variety, Yukara, of Tongil have a typical *indica* (*S5i*) and *japonica* (*S5j*) *S5* locus, respectively (data not shown). It might enables the embryo sac abortion and reproductive isolation between TN1 and Yukara at the first cross and then *indica*-compatible gamete was selected at the second cross between F_1 hybrid and IR8. In addition, chromosomal rearrangement such as a translocation, inversion, fusion, fission, and duplication might act as isolating barriers by suppressing recombination and thus reducing the gene flow. Most chromosomal rearrangements appear to have large effect on fertility of plants (Rieseberg, 2001; Widmer et al., 2009). For these reasons, Tongil rice might have the *indica* toward genome. Tongil rice is highly successful in terms of grain yield in South Korea, although Korean climatic environments are not favorable to the cultivation of typical *indica* varieties (Chung and Heu 1991). This may be attributable to its heightened adaptability compared to most *indica* varieties, perhaps due to the partial incorporation of the *japonica* parental genome.

Comparison of yield related alleles in Tongil

Rice yield is determined by four yield components. 1) Number of panicles per unit area 2) number of spikelet 3) number of filled grain and 4) 1000 grains weight. Seed size, one of the main factors of seed weight, is affected by following four parameters: seed length, width, thickness and grain filling degree. Tongil rice belongs to the heavy panicle type variety and has larger grain size than *japonica* variety. Seed weight is higher than that of other

japonica varieties and parental *japonica* variety. We manually sequenced and compared the *sd1*, *Gn1a*, *Ghd7*, *GS3*, *qSW5*, and *GW2* alleles in Tongil and its parents genome, which were reported and proved as yield related genes and QTLs to understand which alleles are from between *indica* and *japonica* parents, respectively. (Table 1-8). *GW2*, *GS3*, and *qSW5* alleles of Tongil, which are related in determining seed size all alleles, were from *indica* parent IR8 or TN1. Despite Tongil and its *indica* parent varieties commonly have three identified seed size controlling alleles, seed shape of Tongil is closer to TN1 than that of IR8 (Figure 1-1). This indicates that seed shape could be determined by another numerous genes or complex QTLs besides *GW2*, *GS3*, and *qSW5* alleles in rice. It might be due to an additive and epistatic effect between gene interactions and different gene expression by various genetic backgrounds and cross combination (Yan et al., 2011). *S5* allele also came from *indica* parents. For two times crossing, *indica* parent, TN1 or IR8 conferred their *S5* allele to Tongil and it enables Tongil could avoid embryo sac abortion and recover its fertility despite of *indica*/*japonica* hybrid. We assumed that this is one of the main reasons for increasing yield potential of Tongil as well as *sd1* allele introgressed from *indica* parents. In addition, *sd1*, *Gn1a*, and *Ghd7* alleles of Tongil showed *indica* or *indica*-type alleles. Gene pyramiding for increasing yield potential is difficult because the complicated gene interactions and undesirable effect on target trait. However, the information of combination and predominant of alleles could be used to develop new *indica*/*japonica* hybrid variety and gene pyramiding for high yielding rice breeding.

The approach used in this study to determine the parental origins of specific genome segments is applicable to the genomic dissection of agricultural breeding lines or varieties of diverse parental origins

CHAPTER II.

Isolation of the *yellow embryo lethal* (*yel*) Gene in Rice

ABSTRACT

Yellow colored pericarp in rice (*Oryza sativa.* L) is rare in nature. We generated a yellow pericarp and embryo lethal mutant, *yel*, derived from chemical mutagenesis using N-methyl-N-nitrosourea (MNU) on a *japonica* rice cultivar, Hwacheong. In this study, we fine mapped and cloned the gene responsible for yellow pericap and embryo lethal phenotype using a map-based approach. Fine mapping revealed that the mutant gene (*yel*) was located on the long arm of chromosome 2. Sequencing of the candidate genes analysis between the parent and the mutant showed a 706-bp deletion in the first exon including 5'UTR and start codon of LOC_Os02g53140, *Os*COP1 (Constitutive Photomorphogenic 1) which encodes a protein of comprising RING-finger, coiled-coil and WD40 domains. To confirm the function of *yel* gene on yellow pigmentation and embryo lethality, we introduced YEL overexpression vector into the mutant heterozygous plant. Overexpression of YEL, containing normal RING-finger domain restored the pericarp color from yellow to white, while the embryo

color was still black and maintained lethality. This result demonstraiting that the *Os*COP1 gene is involved in yellow pericarp color. Gus staining signals were primarily detected in the embryo, pericarp, pollen and nodes. HPLC-MS analysis of the four candidate standard with extracts of embryo and endosperm of mutant grain revealed that orientin was the major pigment both in embryo and endosperm. In addition, extremely high level of orientin accumulation was observed in mutant embryo and endosperm. The antioxidant activities and total phenolic contents of embryo and endosperm of the mutant seed were significantly higher than those of wild-type. Fatty acid analysis showed the content of saturated fatty acid was increased in mutant embryo on the other hand, the content of unsaturated fatty acid was decreased than that of wild-type. These results indicated that *yellow embryo lethal* (*yel*) gene might be involved in the flavonoids biosynthetic pathway and other ingredient characteristics of rice seed.

KEYWORDS: Yellow pericarp, Embryo lethality, Orientin, Flavonoid, Antioxydant activity

INTRODUCTION

Flavonoids are one of the major class of compounds found in various crops as well as fruits, vegetables and beverages (tea, coffee, beer, wine and friut drink) and it has been important nutrient and beneficial sources for human health by providing essential vitamins and high pharmacological activities and antioxidant capacity. For these reasons, flavonoids have been studied extensively for their effects. Typically, flavonoids are divided into six major subclass including flavonols, flavonones, flavones, flavanols, anthocyanidins, and isoflavones according to the position of substitutes present on the parent molecules. Although flavonoids are distributed throughout the plant kingdom, some flavonoid classes are founded only a few plant species. The interest in increasing the flavonoid content of commonly consumed food crops for health benefits have grown (Butelli et al., 2008; Schijlen et al., 2006).

Flavonoid biosynthesis is well characterized in terms of its molecular genetics and the biochemical mechanisms of the enzymes involved in the flavonoid pathway (Dixon and Pasinetti, 2010). In plants, the biosynthesis of naringenin, a flavanone and central precursor of most flavonoids, involves the following five enzymes in the phenylpropanoid pathway: phenylalanine ammonia lyase (Matsumoto et al.), cinnamate 4-hydroxylase (C4H), 4-coumarate: coenzyme A ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI). The pathway diverges from naringenin into several side branches, each resulting in the production of a different class

of flavonoid. Flavonol synthesis is catalysed by flavanone 3-hydroxylase (F3H) and flavonol synthase (FLS), while isoflavone synthesis and flavone synthesis are catalysed by isoflavone synthase (IFS) and flavone synthase (FNS), respectively. Flavonoids also have many derivatives with hydroxyl groups and/or methoxy groups within a single class. Flavonoids are often glycosylated with one or more sugar units at various positions on the core aglycone structure. These glycosylated flavonoids are thought to be transported to and accumulated in the vacuole. The flavone *C*-glycosides are an important subclass of the flavonoids family, were found in some plants, such as flavone *C*-glycosides are present in food stuff and nutraceuticals and they include orientin, isoorientin, vitexin, and isovitexin. Orientin (*C*-glycosyl flavone, luteolin-8-*C*-glucoside) has not been yet studid in rice. orientin cardioprotection (Lu et al., 2011), antinociception (Da Silva et al., 2010)

Flavonoids play an important role in the grain color and flavor of rice. Rice bran contains a unique complex of naturally occurring antioxidant compounds (Mazza G, 2003). Oryzamutaic acids were detected in yellow endosperm mutant of the variety Hatsuyamabuki however, the biosynthetic pathway remains still unknown. This mutant showed the yellow endosperm which was caused by mutation of Oryzamutaic acid A and *OsALDH7* (Nakano et al., 2009; Nakano et al., 2011; Shen et al., 2012) Mutation of *OsALDH7* resulted in accumulation of yellow pigment in endosperm. Oryzamutaic acid A accumulates during late seed development and after a year-long storage in the endosperm. Pigmented rice also contains

anthocyanins in pericarp. Anthocyanins are responsible for cyanic color of pigmented rice and are regarded as important nutraceuticals mainly due to their antioxidant effect, which provide a potential to prevent various diseases associated with oxidative stress (Duthie et al., 2000; Kong et al., 2003)

Recently, the consumption of colored rice is getting increased to intake the health-promoting components such as vitamins, minerals and phytochemicals which include phenolic compound while eating rice (Dykes and Rooney, 2007). Research has shown that phenolic compounds have antioxidant properties and can protect against degenerative disease such as heart disease and cancer. These compunds are accumulated in the pericarp and aleurone layer in cereal crops. Many studies were conducted in fruits and vegetables but grain cereals are good material for phenolic compounds.

One feature of the constitutive photomorphogenesis phenotype of cop1 and spa mutants is the increased accumulation of anthocyanins (Deng et al., 1991; Hoecker et al., 1998; Laubinger et al., 2004). In Arabidopsis, anthocyanin accumulation occurs only in the light and this accumulation is further enhanced by a number of environmental stress factors, such as cold, drought, pathogen attack and nutrient depletion (Chalker-Scott, 1999). The biosynthesis of anthocyanins is regulated by transcription factors that induce the expression of structural genes that code for enzymes in the biosynthesis pathway (Broun, 2005; Li, 2014). These transcription factors include HY5 and a complex consisting of a WD-repeat protein, a bHLH

protein and a myeloblastosis (MYB) protein (Feller et al., 2011; Koes et al., 2005; Lee et al., 2007; Shin et al., 2007). In Arabidopsis, the WD-repeat protein TTG1 acts together with the bHLH proteins GL3, EGL3 or TT8 and the PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP) protein family of R2R3 MYB transcription factors (PAP1, PAP2, MYB113 and MYB114) (Gonzalez et al., 2008; Zhang et al., 2003; Zimmermann et al., 2004). The PAP protein family, in contrast, is required specifically for anthocyanin accumulation and, therefore, provides specificity to the WD-repeat/bHLH/MYB complex (Lillo et al., 2008; Shi and Xie, 2010; Tohge et al., 2005) The expression of PAP1, PAP2, TT8, GL3 and EGL3 is induced by light, while the expression of TTG1 is light-independent (Cominelli et al., 2008).

In this study, we report the yellow pericarp and embryo lethal mutant which has the yellow pigment, isoorientin which is a very rare compound in embryo and endosperm of rice seed. This mutant rice might be able to develop the yellow colored rice and use for functional food source.

MATERIALS AND METHODS

Plant materials

The yellow embryo lethal mutant was induced by N-methyl-N-nitrosourea (MNU) and it was derived from *japonica* cultivar, Hwacheongbyeo from Crop Molecular Breeding Lab. of Department of Plant Science, Seoul National University. The material was used from a M14 line, 75094, Hwacheong M-6-17N-1-3-2-1-1-1-2-1-1-2-2. This mutant was maintained as heterozygotes because of embryo lethality due to suppressed germination.

Sequence analysis of the candidate genes

The full-length genomic DNA sequence of candidate genes was sequenced by dividing them into several overlapping segments. Specific PCR primers for each segment were used to amplify genomic DNA from Hwacheongbyeo and yellow embryo lethal mutant. PCR products were purified using a PCR purification kit (iNtRON Biotechnology, Korea) for TA cloning. Purified PCR introduced into the pGEM-T Easy Vector (Promega, USA) and transformed into the *E. coli* strain DH5α. The obtained sequences were compared by CodonCode Aligner software (version 1.6.3; CodonCode Coporation). Sequence alignment was performed with the Blast network service in National Center for Biotechnology Information (NCBI), Gramene and The Rice Annotation Project Database (RAP DB).

Gus assay

Histochemical GUS-staining was performed according to the method of Jeon et al. (2000). Mature flowers of the primary transformant were sampled and stained, each tissue type was incubated at 37 for 2 days, then the staining solution was replaced with 95% and 70% (w/v) ethanol in series to remove the chlorophyll. Afterward, we examined the tissues under a dissecting microscope and analyzed their staining patterns.

Vector constructs and rice transformation

In order to generate overexpression vectors, PCR-amplified WT and *yel* full-length cDNAs were inserted into the pCAMBIA 1300-modified vector containing a 35S promoter and a nos terminator. The resulting WT cDNA overexpression construct was denoted 35S::COP1 and the mutant cDNA construct was denoted 35S::yel. In order to generate the RNAi::COP1 construct for COP1 gene suppression, a 342-bp fragment of *yel*-cDNA spanning nucleotides was first cloned into pDONR201 (Invitrogen) and then cloned in sense and antisense directions into the binary transformation vector pH7GWIWG (II) using the Gateway BP and LR clonase enzyme mixes (Invitrogen). pH7GWIWG (II) and derivatives contain the hygromycin resistance (Hyg) gene. For the promoter-GUS assay, the genomic sequence containing the putative promoter region of COP1 was amplified by PCR from the genomic DNA. The COP1 promoter fragment was cloned into the binary vector pHGWFS7. Transgenic plants carrying the

above constructs were generated using wild-type Dongjin (a *japonica* cultivar) seeds via agrobacterium-mediated co-culture methods (Hiei et al., 1997).

UHPLC-MS analysis for yellow pigment

Extracts of yellow mutant and wild-type seed were analyzed using an Agilent 1290 Infinity high performance liquid chromatography (UHPLC) coupled to an Agilent 6490 QQQ with Agilent Jet Stream ESI system. LC condition and MS condition are as follows LC condition; HPLC system : Agilent 1290 Infinity, Column : ACQUITY UPLC BEH Shield RP18 2.1 mm x 100 mm, 1.7 μm, Injection Volume : 2 μL, Temp : 35 ℃, Flow rate : 0.3 mL/min., 55 min. stop time, Mobile phase : A – 0.1 % formic acid in Water : B – 0.1% formic acid in ACN, Gradient condition : 0 min, 5% B; 2 min, 5% B; 40 min, 30% B; 41 min, 90% B; 47 min, 90% B; 47.1 min, 5% B; 55 min, 5% B; MS condition; Mass system: Agilent 6490 QQQ MS, Ion source: Agilent Jet Stream ESI, Nebulizer gas: Nitrogen, Polarity: Negative, Nebulizer pressure: 40 psi, Ion spray voltage: 4000 V, Drying gas temperature: 150 ^{o}C, Drying gas flow rate: 17 L/min, Sheath Gas temp: 300 ^{o}C, Sheath gas flow: 11 L/min, Fragmentor: 380 V, Nozzle voltage: 0 V, High Pressure RF: 150 V, Low Pressure RF: 60 V

Phenolic compounds analysis by UHPLC

The samples were analyzed by ultra-high performance liquid chromatography (UHPLC) analysis performed from Thermo ACCELA UHPLC system (ACCELA UHPLC system, Thermo Fisher Scientific Inc., USA) with a column (Thermo, C18 2.1 × 100 mm, 2.6 μm) used for analysis. The mobile phases were composed of 0.1 % glacial acetic acid in distilled water (solvent A) and 0.1 % glacial acetic acid in ACN (solvent B). The flow rate of the solvent was 500 μL / min and injection volume was 4 μL. The wavelength of the UV detector was 280 nm. The linear gradient of solvent was as follow; 0 min, 98% A: 2% B; 0.50 min, 95% A: 5% B; 2.20 min, 90 % A: 10% B; 5.00 min, 85 % A: 15% B; 7.50 min, 84.3% A: 15.7% B; 8.00 min, 83.4% A: 16.6% B; 9.00 min, 82.2% A: 17.8% B; 9.50 min, 76.1% A: A: 23.9 % B; 14.00 min, 55.0% A: 45.0% B; 15.00 min, 0% A: 100% B; 15.50 min, 0 % A: 100% B; 16.00 min, 98 % A: 2 % B; 25.00 min, 98 % A, 2 % B. The detected time was 15min and the rest time was for equilibrium.

Standards solution

Isoorientin, Orientin, Vitexin, Apigenin, Caffeic acid, Tricetin, Eriodictyol, Luteoin, Catechin, p-coumaric acid, Ferulic acid, m-coumaric acid, Naringin, Myricetin, Kaempferol, Hesperetin, and Biochenin A were purchased from Sigma Aldrich. Chemical Co. (St. Louis, MO, USA). All standards were made stock solutions of 100 ppm the phenolic standards except Biochenin A by dilution in methanol. Biochanin A were prepared by dilution in Dimethyl Sulfoxide (DMSO). Concentrations range for standard curve

was determined with comparing the value of each phenolic compound area in sample extract (3 - 7 point).

Total Phenol Content

Total polyphenol contents were determined as described by Singleton et al. with slight modifications. Briefly, 100 μL of 0.2 N Folin–Ciocalteu's phenol reagents were added to 20 μL of each sample placed in 96-well plates. After 5 min of reaction in the dark, 80 μL of a 7.5% sodium carbonate solution was added to the mixture, which was then incubated for 60 min at room temperature. The absorbance was measured at 750 nm using a microplate reader (Thermo Electron, Spectronic Genesys 6, Madison, WI, USA). The results are presented as gallic acid equivalent concentration (ppm). The standard solution concentration curve ranged from 31.25 ppm to 500 ppm.

Free radical scavenging by the use of the 2, 2 – diphenyl – 1 – picrylhydrazyl (DPPH)

The free radical scavenging activities of the yellow embryo lethal mutant and wild-type seed were measured using the DPPH assay. Reaction mixtures were prepared by mixing 2.5 mL of DPPH solution (0.35 mM DPPH dissolved in 50 % EtOH) and 0.2 mL of each sample. Reaction was incubated for 10 min at room temperature and the changes in DPPH

absorbance at 517 nm were measured with a spectrophotometer. The antioxidant activity was calculated in terms of the per cent inhibition caused by the hydrogen donor activity of each sample.

Free radical scavenging by the use of the ferric reducing antioxidant power (FRAP)

The reducing ability of each sample was determined using the FRAP assay (Hu and Xu, 2011; Kim et al., 2014). Briefly, the FRAP reagent prepared by mixing 300 mM sodium acetate buffer (pH 3.6), 10 mM 2,4,6-tripyridyls-triazine (TPTZ) solution in 40 mM HCl and 20 mM $FeCl_3$ solution at a ratio of 10 : 1 : 1 (v/v/v). Ferric-reducing antioxidant power reagent was prepared fresh daily and pre-warmed in a 37 ℃ water bath prior to use. Then, 0.1 mL of extract was mixed with 1.8 mL of FRAP reagent and 3.1 mL of ultrapure water (HPLC grade). The reaction mixture was incubated at 37 ℃ for 0.5 h and then, the absorption of the mixture was measured at 593 nm with a spectrophotometer. Ferric-reducing antioxidant power values were calculated on the basis of $FeSO_4$ standard curve (100-1000 μM) and reported as μmoles of Fe per gram of dry weight.

Free radical scavenging by the use of the trolox equivalent antioxidant capacity (TEAC)

The Trolox equivalent antioxidant capacity (TEAC) assay, which measures the reduction of the radical cation of ABTS by antioxidants, was carried out as previously described with slight modifications (Hu and Xu, 2011; Kim et al., 2014). Briefly, ABTS+ radical cation was generated by a reaction of 7 mmol/L ABTS and 2.45 mmol/L potassium persulfate. The reaction mixture was allowed to stand in the dark at room temperature for 16-24 h before use and used within 2 days. The ABTS+ solution was diluted with methanol to an absorbance of 0.700 (0.050 at 734 nm. All samples were diluted appropriately to provide 20-80% inhibition of the blank absorbance. Fifty microliters of the diluted extracts was mixed with 1.9 mL of diluted ABTS+ solution. The assay with the mixture was carried out in triplicate, the mixture was allowed to stand for 6 min at room temperature and the absorbance was immediately recorded at 734 nm. Trolox solution (100-1000 μM) was used as a reference standard. The results were expressed as micromoles Trolox equivalents (TE) per gram of DW.

The antioxidant capacity of total phenolic assay by Folin-Ciocalteu reagent

Embryo and endosperm were ground and 1 g of the powder extracted with 80% methyl alcohol for 1 day in a shaker-chamber at 25℃ with stirring. The extracts were filtered through a Whatman No. 42 filter paper and the quantitative analysis of total phenols was measured by the Folin-Dennis method (AOAC, 1980). Half ml of the extract sample, 5 ml of distilled water

and 5 ml of the Folin-Ciocalteau phenol reagent were mixed in a screw-top flask. After 3 min, 2 ml of 10% sodium carbonate (Na_2CO_3) was added and the mixed solution stirred in a lit Shaker-chamber at 30℃for 1 h. After the reaction, the absorbance of the solution was measured with a spectrophotometer (Hitachi Ltd., Tokyo, Japan) at 760 nm. The standard curve was measured with 1, 50 and 100 ppm of ferulic acid purchased from Sigma Aldrich (Mo., U.S.A.).

Amino acid analysis

About 0.3 g of each sample was weighed and mixed with 5 mL 6 N HCl. Hydrolysis was maintained for 24 hr at 110℃ in test tubes that were flushed with nitrogen gas. Afterwards, the samples were diluted to 100 mL with Milli-Q water and filtered through 0.45 μm syringe filters (Waters, Millipore, Milford, MA, USA). One mL of each hydrolysate was placed in an autosampler bottle and injected into an amino acid autoanalyzer (Hitachi L-8800; Hitachi Ltd., Tokyo, Japan). Amino acid concentrations of samples were calculated with reference to an amino acid standard (Ajnomoto-Takara Co., Kyoto, Japan).

Fatty acid analysis

Fatty acid analysis was conducted as described by Rafael and Mancha (1993)(Kim et al., 2013). The procedure was as follows: 0.5 g freeze-dried

GSG was heated with a reagent containing methanol:heptane:benzene:2,2-dimethoxypropane:H_2SO_4 (37:36:20:5:2, v/v). Simultaneous digestion and lipid transmethylation took place in a single phase at 80℃. The fatty acid analysis was performed on a gas chromatograph (6890 N series, Agilent Technologies Inc., Wilmington, DE, USA) equipped with a FID using a HP-Innowax capillary column (0.25 μm, 30 m, J&W Scientific, Agilent Technologies Inc., Wilmington, DE, USA). The initial oven temperature of 150℃ was ramped up by 4℃/min up to the final temperature of 280℃. During the analysis, the temperature of the injector and detector were maintained at 250 and 300℃, respectively, and the air, hydrogen and nitrogen (carrier gas) flow rates were set to 400, 30 and 2 ml/min, respectively. The standard methyl esters of the fatty acid mix (C14-C22) were obtained using a Supelco instrument (Bellefonte, USA). The data obtained from the gas chromatograph were converted to percentage of total fatty acids measured.

Analysis of crude protein and lipid

The protein content of rice was measured according to the Kjeldahl procedure using a protein auto analyzer (Tecator Kjeltec Auto Analyzer 2400; Foss Tecator, Sweden). The lipid content was measured using the Soxtherm Automatic System (Gerhardt, Germany).

RESULTS

Molecular cloning of the *yel* gene

The yellow embryo lethal mutant was obtained by chemical mutagenesis (MNU) and it showed a recessive phenotype of yellow pericarp and embryo lethality in grain (Figure 2-1), thus, it was designated as *yel*. In the previous study, we fine mapped the *yel* locus using a population derived from a cross between *yel* mutant and Milyang23 (*indica* type). The *yel* gene was located on the long arm of chromosome 2 between two flanking markers, S02k-1 and S02k, originating from the BAC clone of AP004058 and the physical distance was 55kb.

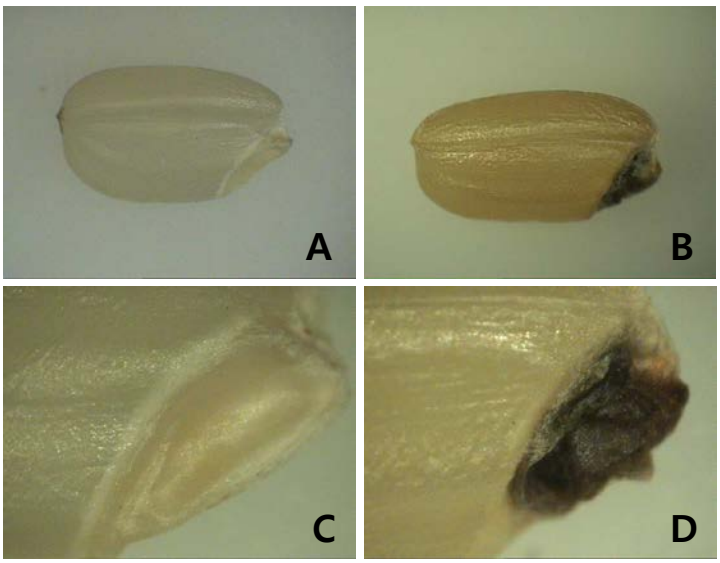


Figure 2-1. Comparison of the brown rice morphology between the wild-type and *yel* mutant. (A) brown rice of wild-type (B) brown rice of *yel* mutant (C) embryo of wild-type seed (D) embryo of *yel* mutant seed; *Yel* mutant seed showed yellow pericarp color and black colored embryo compared to wild-type

Total seven candidate genes were located in this region based on available rice genome annotation (http://tigr.org, http://gramene.org) (Table 2-1). To identify the *yel* gene, we sequenced seven of the candidate genes and compared the sequence variations between the *yel* mutant and its wild-type, Hwacheongbyeo. The sequence analysis indicated that 706-bp deletion was found from 5'UTR to first exon including start codon on the LOC_Os02g53140 (Figure 2-2).

Table 2- 1. Candidate genes analyzed for identification of *yel* gene

No	Locus Name	Gene Description
1	Loc_Os02g53130	nitrate reductase
2	Loc_Os02g53140	COP1 (constitutive photomorphogenesis 1)
3	Loc_Os02g53150	expressed protein
4	Loc_Os02g53160	tyrosine phosphatase family protein
5	Loc_Os02g53169	expressed protein
6	Loc_Os02g53180	1-aminocyclopropane-1-carboxylate oxidase protein
7	Loc_Os02g53200	glucan endo-1,3-beta-glucosidase precursor

In order to determine, whether the 706-bp deletion is associated with mutant phenotype, we conducted the co-segregation analysis with the mutant and wild-type individuals in the F_2 progenies using co-dominant marker (Figure 2-3). This result supported that the 706-bp deletion identified in the LOC_Os02g53140 gene was responsible for the yellow pericarp and embryo lethal phenotype. YEL encodes a protein of comprising RING-finger, coiled-coil, and WD40 domains. All RING-finger

domain as a result of 706-bp deletion was deleted in YEL gene (Figure 2-4).

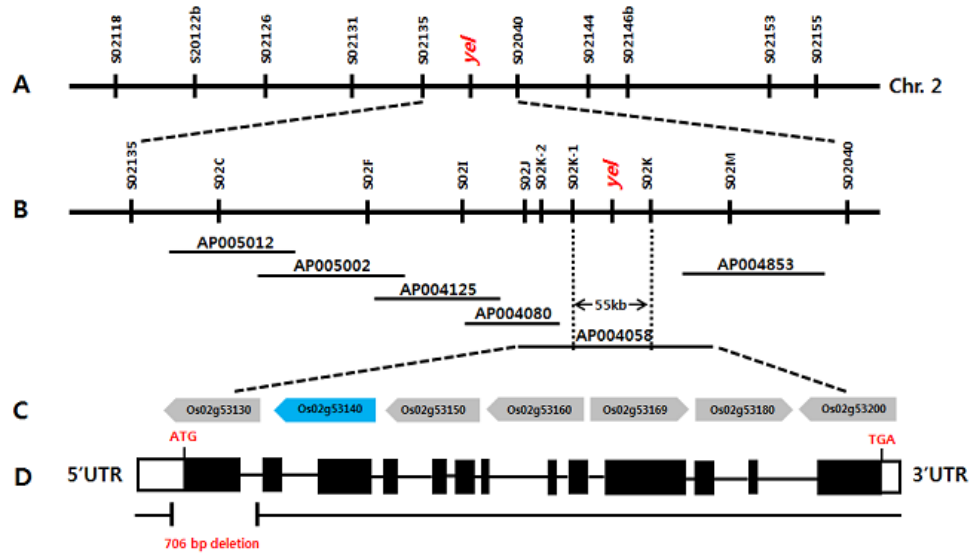


Figure 2-2. Map-based cloning of *yel* gene. (A) Genetic mapping of the *yel* locus with STS markers. (B) Fine mapping of the *yel* locus with additional STS markers. The *yel* locus is located to an approximately 55kb genomic DNA region between S02K-1 and S02K. (C) Seven candidate genes in the 55 kb genomic DNA region identified by fine mapping. (D) The *yel* gene structure at genomic sequence. Thirteen exons and eleven introns are indicated using black rectangle and black line, respectively; the 706-bp deletion is identified from the 5'UTR to first intron included entire first exon.

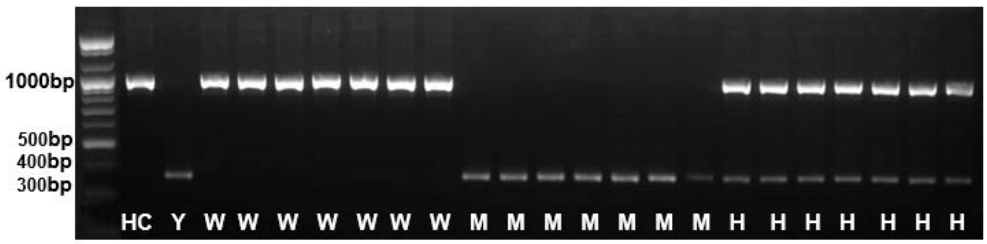


Figure 2-3. Co-segregation analysis in the F_2 progenies. A 706-bp deletion polymorphism were tested for co-segregation using a co-dominant marker. HC: Hwacheong (original parent), Y: *yel* mutant, W: wild-type phenotype, M: mutant phenotype, H: heterozygous phenotype in F_2 progenies.

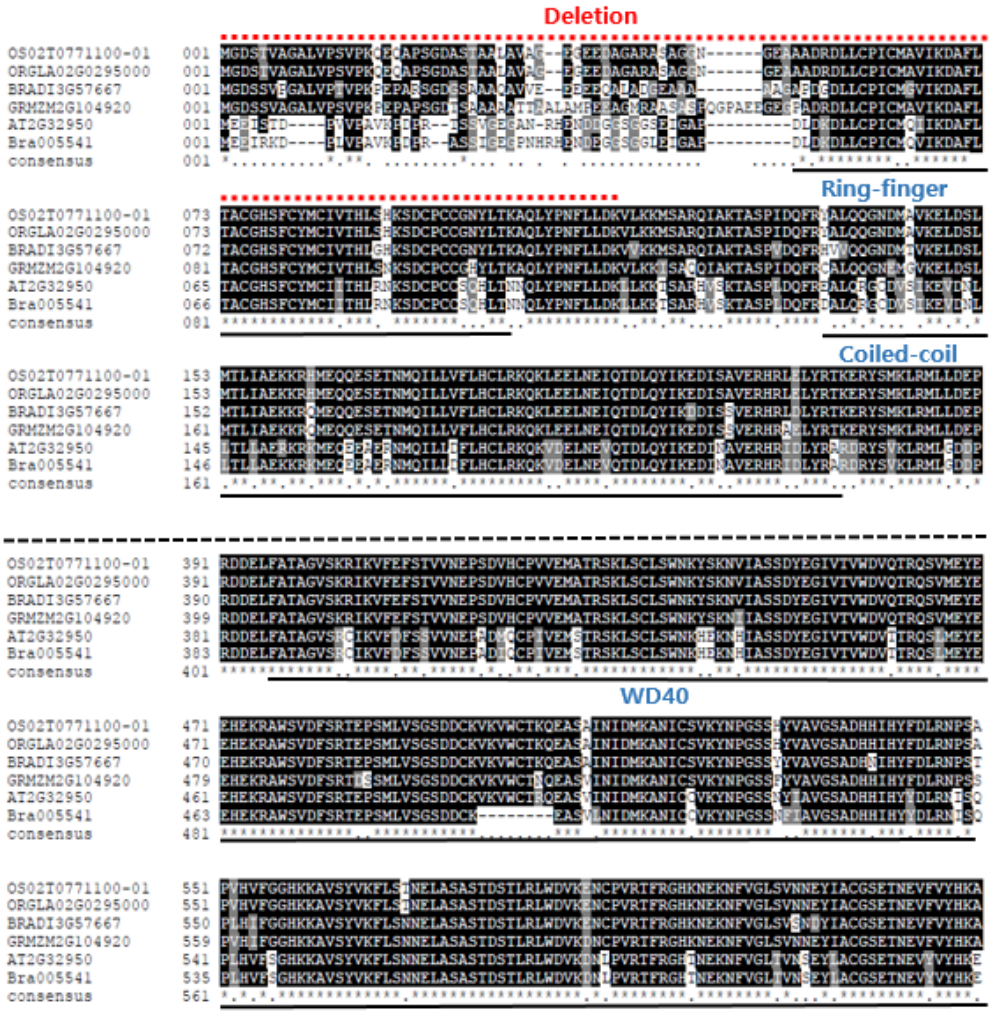


Figure 2-4. Multiple alignment of the amino acid sequences of *YEL* protein. Alignment of YEL with RING-finger, Coiled-coil, and WD40 domains from *Oryza sativa*. *japonica* (OS02T0771100-01), *Oryza glaberrima* (ORGLA02G0295000), *Brachypodium distachyon* (BRADI3G57667), *Zea mays* (GRMZM2G104920), *Arabidopsis thaliana* (AT2G32950), *Brassica rapa subsp. pekinensis* (Bra005541). The positions of domains are indicated by lines below the sequences. Residues

identical in all the sequences are shown by asterisks (*) below the sequences. Residues identical to those in YEL (*Oryza sativa*. *japonica*) with 100% similarity black background and at least 60% similarity are highlighted with a gray background.

Demonstration of the yellow pericarp phenotype conferred by *yel* in transgenic plants

In order to confirm that the 706-bp deletion in the *yel* gene was indeed responsible for the mutant phenotype, transgenic rice plants expressing the mutant *yel* allele under the 35S promoter (35S::*yel*) were generated in the wild-type (WT) Dongjin background (YEL) and their phenotypes were observed. However, all of the transgenic plants didn't show yellow pericarp phenotype or embryo lethal phenotype although, some RNAi transgenic plants showed dark green leaf and short height or sterile and semi-sterile phenotypes (Figure 2-5).

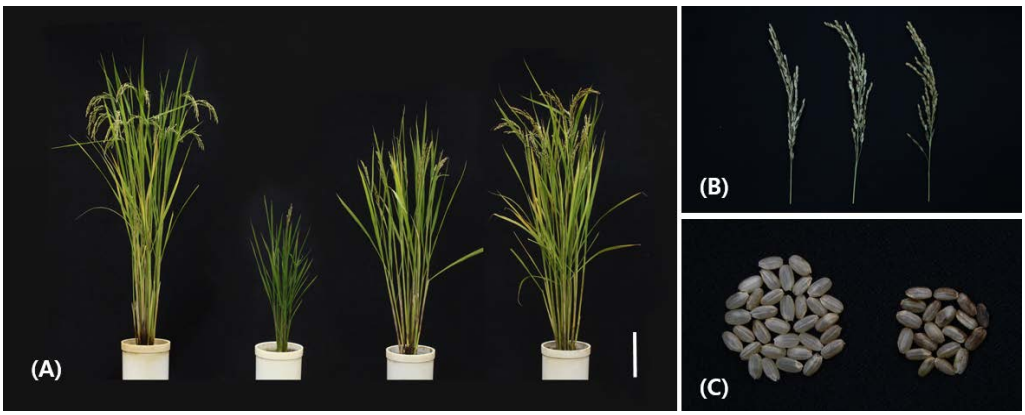


Figure 2-5. RNAi transgenic plants of *yel* mutant (A) Hwacheongbyeo (WT), short and dark green morphology line, sterile line, semi-sterile line (from left to right) (B) morphology of panicle of RNAi lines; steril panicle, semi-sterile panicle, fertile

panicle (from left to right) (C) Seeds from semi-sterile RNAi transgenic lines ; No yellow pericarp or embryo lethal seeds were found. Bar = 20 cm

In contrast, we constructed the transgenic Dongjin (YEL) plants overexpressing the wild-type YEL (35S::YEL) and then crossed with heterozygous plant which is maintained for segregating mutant due to the lethality of mutant type seed. We selected YEL-OE inserted plants among the F_1 plants that crossed between YEL-OE transgenic plant and heterozygous plant afterwards, we observed the segregation of F_2 seeds. Some white pericarp color but still has black embryo (lethal) seeds were segregated in YEL-OE inserted plant (YEL-OE/heterozygous) (Figure 5).

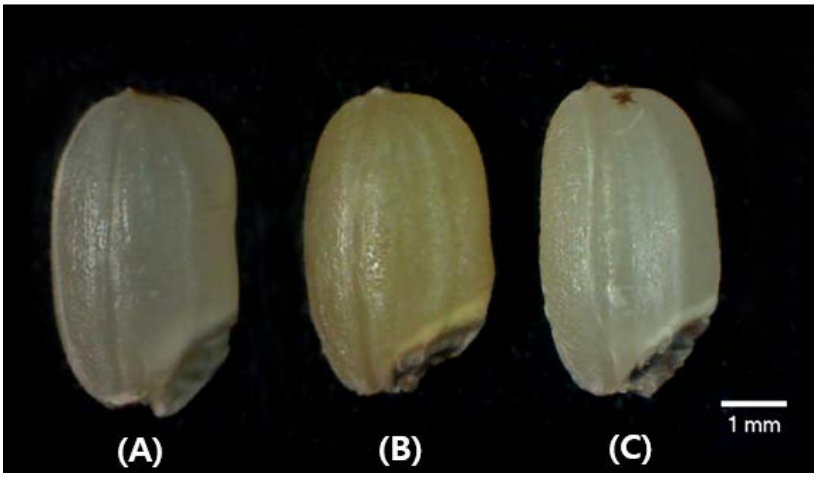


Figure 2-6. Segregated progenies from the cross between YEL-OE transgenic plant and *yel* heterozygous plant (A) Wild-type seed (B) *yel* homozygous seed (mutant type) (C) Recovered seed by YEL-OE

To confirm whether the all the white lethal seeds have YEL-OE and *yel* homozygous allele at the same time, we genotyped the F_2 seeds. The result showed that all the white lethal seeds had YEL-OE insert and *yel* allele. The

segregation ratio of yellow lethal, white lethal, and normal seeds was fitted to the expected ratio of 1 : 3 : 12 by chi-square test (Table 2). It is conceivable that the yellow pericarp phenotype may have been the result of loss of function in rice *yel* gene**.**

Table 2-2. The segregation ratio of F_2 seed in YEL-OE inserted plant

Panicle	Yellow lethal	White lethal	Normal	Total
1	8	22	60	90
2	5	16	59	80
3	8	15	55	78
4	4	3	44	51
5	4	13	44	61
6	4	6	33	43
7	3	7	35	45
8	3	13	64	80
Total	39	95	394	528

X^2-1.26, p-0.53

Expression pattern of *yel* in the different organs and tissues of rice plants

In order to better understand the role of the *yel* gene in rice development, the expression pattern of YEL in different organs and tissues was examined using a beta-glucuronidase reporter gene under the control of the YEL promoter. GUS staining was primarily detected in the pollen, embryo, pericarp, and node (Figure 2-7 C, E, G and H). Relatively strong

GUS staining signals were detected in the differentiation zone and meristematic region of the stem while no GUS expression was detected in the spikelet, leaf, stigma and endosperm (Figure 2-7 A, C and D).

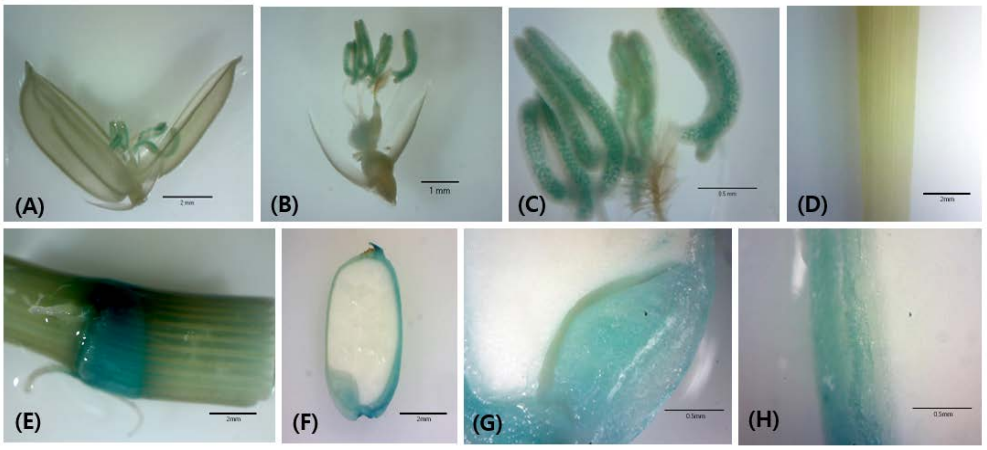
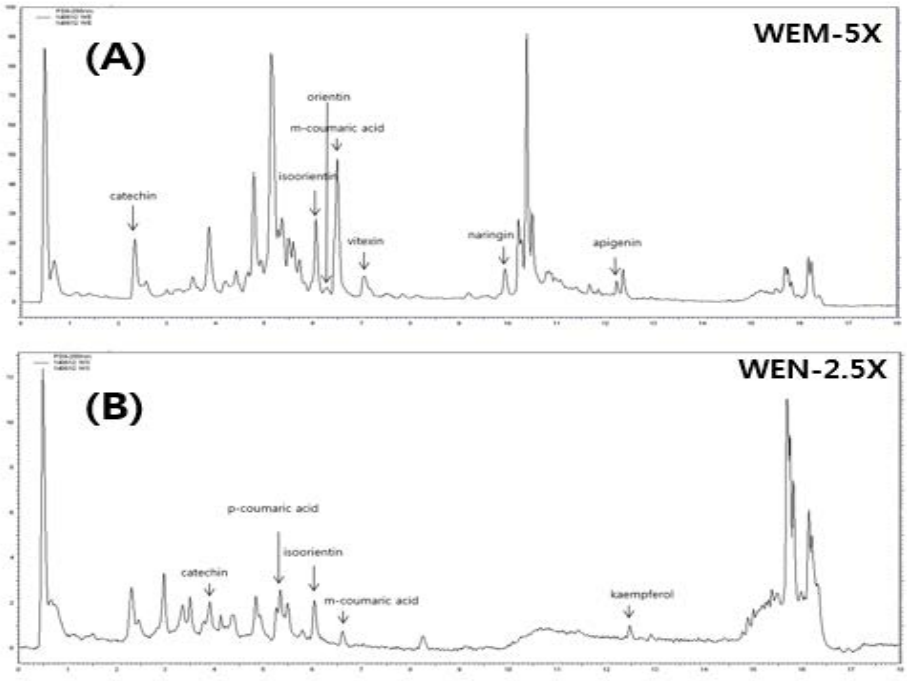


Figure 2-7. Expression patterns of the YEL gene. (A to H) GUS activity detected in the YEL promoter::GUS transgenic plants. (A) spikelet (B) floral organ (C) pollen (D) leaf (E) node (F) seed (G) embryo (H) pericarp

Analysis of phenolic compounds in embryo and endosperm

Yellow pericap color is very rare in rice grain. In order to screen and determine the phenolic compounds contents in the dried embryo (black color) and endosperm (yellow color), we compared UHPLC chromatograms of the embryo and endosperm extracts of wild-type and yellow mutant, respectively. This study was analyzed by ultra-high performance liquid chromatography (UHPLC) analysis performed using 16 phenolic standards such as catechin, p-coumaric acid, m-coumaric acid, naringin, apigenin, caffeic acid, myricetin, tricetin, isoorientin, orientin, vitexin, eriodictyol,

luteolin, kaempferol, hesperetin, biochanin A (Figure 2-8). To get the desirable calibration curve values, quantification to individual phenolic compound was conducted and all standards with each concentration were determined using in triplicate and average value. We assigned individual peak of compound using spike. The result showed that catechin, isoorientin/orientin, and vitexin/isovitexin were strongly detected in yellow embryo and endosperm than those of wild-type amomg the 16 phenolic compounds. Isoorientin/orientin content in yellow endosperm was higher than that of wild-type endosperm (Table 2-3). These results suggest that catechin, isoorientin/orientin, vitexin/isovitexin might play an important role in the yellow color of pericap and black color of embryo in mutant seed.



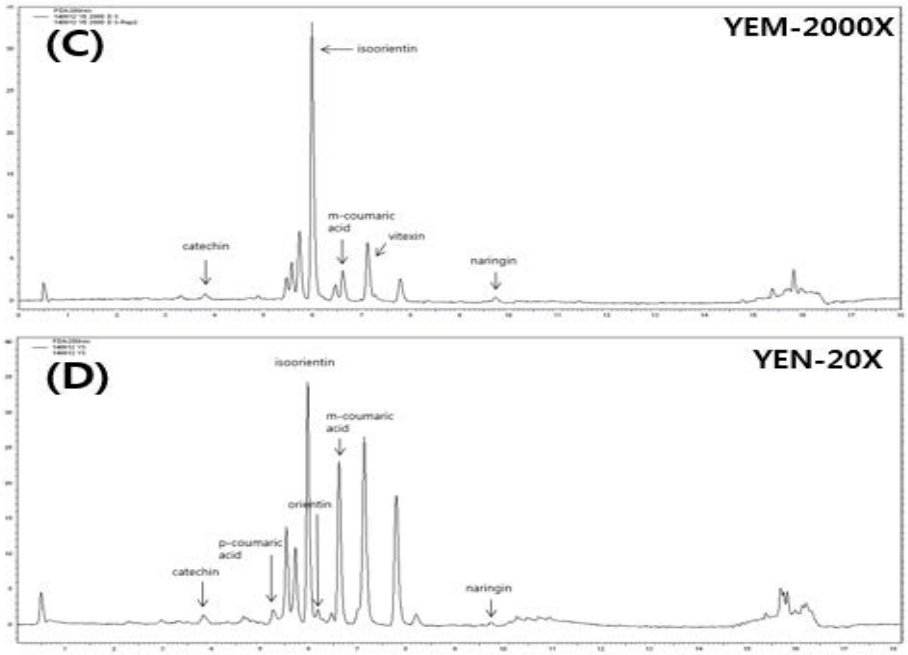


Figure 2-8. UHPLC chromatograms of 16 phenolic compounds in embryo and endosperm. (A) WEM (wild-type embryo), (B) WEN (wild-type endosperm), (C) YEM (yellow embryo), (D) YEN (yellow endosperm). X means dilution

Table 2-3. Contents of major phenolic compounds in embryo and endosperm of mutant and wild-type seed

	Catechin	p-coumaric acid	Isoorientin/Orientin	m-coumaric acid	Vitexin/Isovitexin	Naringin	Apigenin
WEM	143.6±22.7	nd	43.6±2.1	30.3±1.6	10.6±1.4	23.3±2.0	3±0.2
WEN	6.5±0.2	0.6±0	3.3±0.1	nd	nd	nd	nd
YEM	4614.5±152.8	nd	21681.8±396.5	nd	5006.4±164.6	666.7±125.7	nd
YEN	48±2.6	4.4±0.3	226.2±6.2	48.9±1.9	nd	7.11±0.5	nd

Unit: μg/g

Mean value ± SD (n = 3), nd: not detectable

Qualitative and Quantitative analysis of *yel* mutant seed

After screening the content of phenolic compound, we narrowed the candidate metabolites corresponding to yellow pericarp and black embryo color in *yel* mutant. Isoorientin and vitexin were the major compounds of the *yel* mutant grain thus, we analyzed quantitative and qualitative characteristics of isoorientin and vitexin include their isomers, orientin and isovitexin using UHPLC/MS system. To determin which pigment is correspond to yellow and black pigmentation in pericarp and embryo, respectively among the four candidate pigments and to confirm the accumulated amount of pigments at the different rice seed part, we analyzed the extracts of embryo and endosperm of *yel* mutant and wild-type seed, seperately using UHPLC/MS. The result showed that orientin extremely accumulated in the embryo (24965464.26±988976.36 μg g^{-1}) of mutant seed than that of wild-type. Although the total orientin content of endosperm in mutant seed was relatively lower than embryo, orientin was the major pigment in mutant endosperm as well as embryo. This results indicates that orientin plays an important role to have the black color of embryo and yellow color of endosperm in mutant seed (Figure 2-9). Isovitexin was the second highly accumulated pigment in embryo (4093988.68±216424.20 μg g^{-1}) and endosperm (139862.78±4978.00 μg g^{-1}) of mutant seed following orientin. The amount of isovitexin was relatively lower than orientin, however, the concentration of isovitexin also was extremely higher in mutant seed than wild-type. Isovitexin also one of the important pigment for yellowing mutant seed (Figure 2-9).

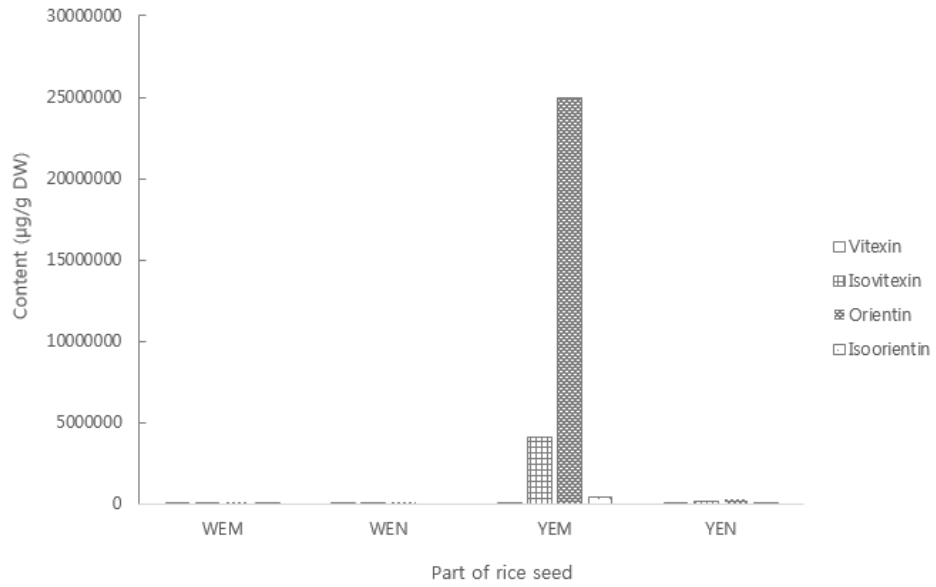


Figure 2-9. Accumulation of four candidate flavonoid in embryo and endosperm of *yel* mutant and wild-type seed. WEM (wild-type embryo), WEN (wild-type endosperm), YEM (yellow embryo), YEN (yellow endosperm).

Determination of antioxydant capacities by DPPH, FRAP, TAEC assay

The DPPH (1,1-Diphenyl-2-picrylhydrazyl) and FRAP (Ferric reducing antioxidant power) and TAEC (Trolox Equivalent Antioxidant Capacity) antioxidant activities were measured. The extracts from whole grain, embryo, and endosperm of wild-type and *yel* mutant were examined and

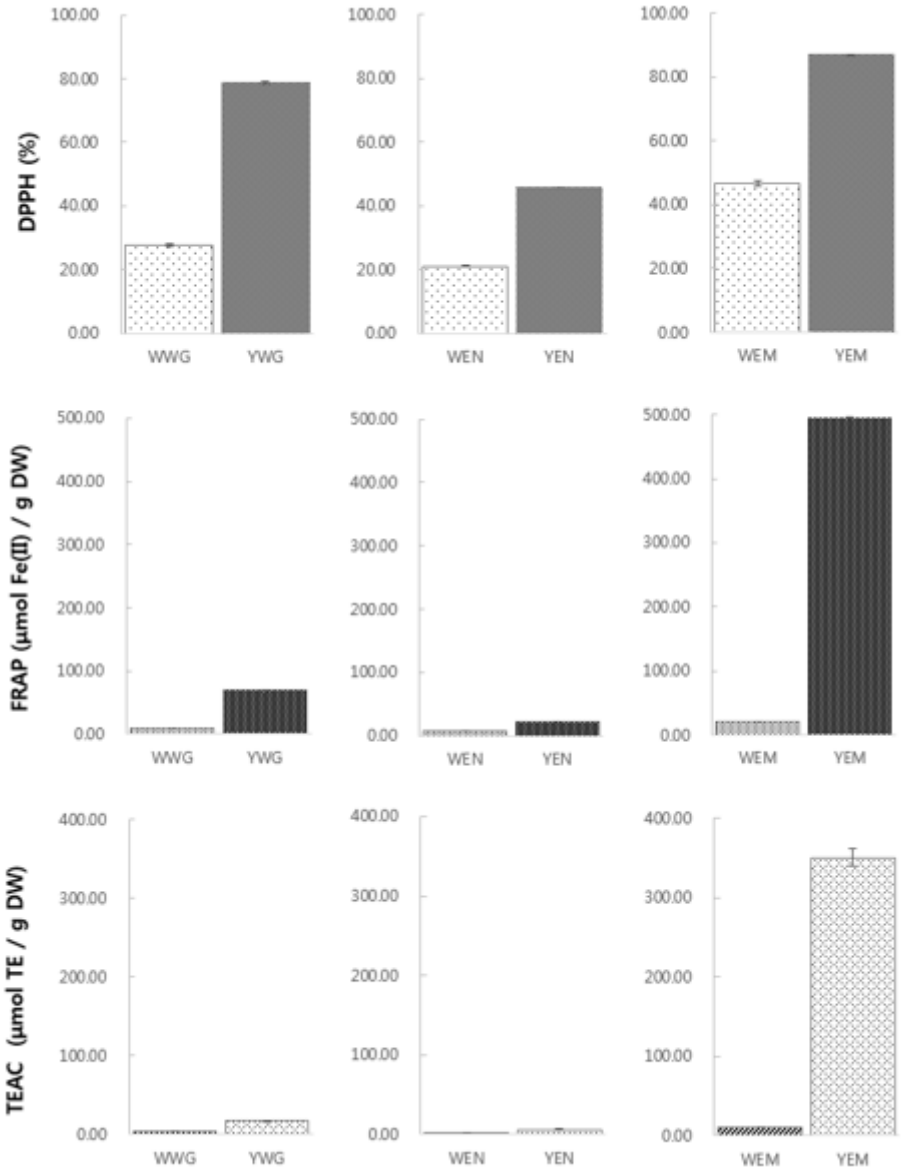


Figure 2-10. Antioxidant activity of *yel* mutant determined by DPPH, FRAP, and TEAC assay. WWG (wild-type whole grain), YWG(yellow whole grain) WEM (wild-type embryo), WEN (wild-type endosperm), YEM (yellow embryo), YEN (yellow endosperm).

compared to their free radical scavenging activities against DPPH• radical. As shown in Figure 2-10, the highest radical scavenging activity was found in the extract of *yel* mutant embryo (87.06±0.19 %). DPPH radicals scavenging activity was higher in all part such as whole grain, embryo, and endosperm of *yel* mutant. In addition, extract from mutant embryo had a greater reducing power (495.38±1.00 μmol Fe(II) / g) than extract from wild-type embryo (20.79±0.29 μmol Fe(II) / g) in the FRAP assay. Extremely high antioxidant activity was also observed in *yel* mutant embryo (350.03±11.18 μmol TE / g) in TEAC assay. These results showed that orientin and isovitexin which highly accumulated in *yel* mutant embryo, might be attribute to high level of antioxidant activity (Figure 2-10).

Total phenolic contents of *yel* mutant seed

Total phenolic compounds of *yel* mutant seed were also measured. Total contents of phenolic compounds were higher in embryo, endosperm, and whole grain of *yel* mutant seed than those of wild-type. In common with anitioxidant activity assay, extremely high level of total phenolic contents were detected in *yel* mutant embryo (46.28±0.87 mg/g, ferulic acid equivalent) (Table 2-4). This indicates that the phenolic compounds are the major constitutuents in mutant embryo and endosperm contributing to antioxidant activity such as DPPH, FRAP, and TEAC.

Table 2-4. Total phenolic contents of the *yel* mutant and wild-type seed

	Total phenol (mg/g, ferulic acid equivalent)	
	WT	*yel*
Whole grain	1.25±0.04	3.68±0.07
Endosperm	0.97±0.02	1.79±0.00
Embryo	2.62±0.05	46.28±0.87

Mean value ± SD (n = 3)

Content and composition of fatty acids in *yel* mutant seed

The fatty acid content of *yel* mutant seed is shown in Table 2-5. The content of saturated fatty acid was relatively high in *yel* mutant embryo, endosperm, and whole grain while the contents of unsaturated fatty acid was high in embryo, endosperm, and whole grain of wild-type. The contents of unsaturated fatty acid such as oleic acids (C18:1) and linoleic acid (C18:2) were dramatically decreased however, the content of saturated fatty acid, palmitic acid (C16:0) was increased more than two folds in the embryo of *yel* mutant than wild-type (Table 2-5).

Table 2-5. Contents of fatty acid of *yel* mutant and wild-type seed

	Contents of fatty acids (%)							
	Saturated fatty acid				**Unsaturated fatty acid**			
	Palmitic acid	Stearic acid	Arachidic acid	Total	Oleic acid	Linoleic acid	Linolenic acid	Total
	C16:0	C18:0	C20:0		C18:1	C18:2	C18:3	
WWG	22.7±0.6	1.4±0.4	1.3±0.4	25.4±0.7	36.7±1.1	36.7±0.4	1.2±0.0	74.6±0.7
YWG	30.7±0.4	2.2±0.9	2.4±0.9	35.2±2.1	33.8±1.5	29.8±0.6	1.2±0.0	64.8±2.1
WEN	25.9±0.6	1.3±0.5	1.1±0.6	28.2±0.9	34.6±0.8	36.2±0.7	1.0±0.0	71.8±0.9
YEN	29.4±2.1	1.4±0.1	1.4±0.0	32.2±2.2	31.3±2.6	35.0±0.3	1.6±0.1	67.8±2.2
WEM	22.0±0.7	1.4±0.3	0.6±0.1	24.0±0.4	33.8±2.6	40.2±3.1	2.0±0.2	76.0±0.4
YEM	48.8±1.0	5.5±0.3	4.4±0.3	58.7±0.6	17.5±0.6	20.3±0.3	3.6±0.1	41.3±0.6

Mean value ± SD (n = 3)

Crude fat and protein contents analysis of *yel* mutant seed

Crude fat and protein contents were measured in *yel* mutant seed. The content of crude fat in *yel* mutant embryo (3.12±0.10 %) was decreased than that of wild-type (19.24±0.14 %). Crude protein contents of the whole grain (7.04±0.07 %) and endosperm (6.47±0.04 %) of *yel* mutant were higher than that of wild-type while, the crude protein content in the embryo of wild-type (18.72±0.07 %) was higher than that of *yel* mutant embryo (18.36±0.08 %) (Table 2-6).

Table 2-6. Crude fat and protein content analysis of *yel* mutant and wild-type seed

	Crude fat content (%)		**Crude protein content** (%)	
	WT	*yel*	WT	*yel*
Whole grain	2.26±0.08	1.37±0.07	6.40±0.02	7.04±0.07
Endosperm	1.31±0.05	1.28±0.01	6.15±0.02	6.47±0.04
Embryo	19.24±0.14	3.12±0.10	18.72±0.07	18.36±0.08

Mean value ± SD (n = 3)

DISCUSSION

Comparison of *yel* mutant and other cop1 related phenotype

We isolated a *yellow embryo lethal* (*yel*) gene, which exhibits a yellow colored pericarp and lethal embryo. Map-based cloning revealed that the *YEL* is an ortholog of *Arabidopsis CONSTITUTIVE PHOTOMOTPHOGENIC 1* (*COP1*) gene encodes a transcription factor which plays an important role in light regulated growth, and has the Ring-finger, coiled-coil, and WD40 repeat domain. This COP1 protein is conserved in both higher plants and mammalians. In plant, COP1 acts as an E3 ubiquitin ligase to repress light signaling by targeting photoreceptors and downstream transcription factors for ubiquitination and degradation. COP1 is a known repressor of photomorphogenesis in darkness as well as in light. *cop1* mutants have a light-grown phenotype in darkness, show features of enhanced photomorphogenesis in light (Yi and Deng, 2005; Oravecz et al, 2006). At the molecular level, COP1 targets different photomorphogenesis-promoting transcription factors for degradation in darkness (Osterlund et al, 2000; Saijo et al, 2003).

Our study showed that unlike the *cop1* Arabidopsis mutant, the *yel* mutant didn't exhibit photomorphogenesis both under dark condition and light conditions. Instead of, *yel* mutant showed the yellow colored pericarp and lethal embryo in the seed. The yellow pigment in pericarp and embryo of *yel* mutant seed revealed that the yellow pigment is orientin, one of the flavonoid in plant. We couldn't observe COP1 related any function in our

mutant. Thus, YEL appears to have acquired novel functions associated with flavonoid biosynthesis during the evolution of rice.

Interestingly, one feature of the constitutive photomorphogenesis phenotype of *cop1* is the increased accumulation of anthocyanins (Deng et al., 1991; Hoecker et al., 1998; Laubinger et al., 2004). In Arabidopsis, anthocyanin accumulation occurs only in the light and this accumulation is further enhanced by a number of environmental stress factors, such as cold, drought, pathogen attack and nutrient depletion (Chalker-Scott, 1999). The biosynthesis of anthocyanins is regulated by transcription factors that induce the expression of structural genes that code for enzymes in the biosynthesis pathway (Broun, 2005; Li, 2014). These transcription factors include HY5 and a complex consisting of a WD-repeat protein, a bHLH protein, and a myeloblastosis (MYB) protein (Feller et al., 2011; Koes et al., 2005; Lee et al., 2007; Shin et al., 2007)

Another characteristics of *yel* mutant is the embryo lethality. The *yel* mutant gene was deleted 706-bp from 5′ UTR to first intron includes first exon and start codon. It means that the Ring-finger domain of N-terminal of YEL protein is deleted. To get the *yel* mutant plant, we conducted anther culture however, we were not able to get a homozygous mutant plant. All the generated plants were homozygous wild-type plants of YEL. Introduction of an N-terminal fragment of Arabidopsis COP1 (AtCOP1) containing the RING-finger and coiled-coil domains into a cop1 null allele rescued its lethal phenotype, indicating that the AtCOP1 N-terminal region alone is able to sustain a basal function during development (Stacey et al.,

2000). The AtCOP1 C-terminal WD40 domain, by contrast, led to repression of photomorphogenesis when expressed in a wild-type background but failed to complement a *cop1* loss-of-function allele (Stacey et al., 2000). We concluded that the absence of Ring-finger domain in the YEL protein caused embryo lethality or defect of cell division. The *pps*-2 mutant which consists of a single base change (G to A) in the eighth exon of the gene, generating a premature stop codon, also couldn't germinate and only ~5% of seeds could germinate on nutrient medium containing sucrose but died in a week (Tanaka et al., 2011).

Accumulation of Orientin, flavone-C-glycoside

Rice is a staple food throughout the world. However, rice seeds do not contain most of flavonoids, except for a small amount of tricin found in bran. The biosynthetic genes involved in flavonoid synthesis are hardly expressed in endosperm while in embryos, PAL, CHI, CHS, and FNS, which are thought to be involved in tricin biosynthesis, are expressed (RiceXpro) (Ogo et al., 2013).

Our UHPLC/MS analysis reveal that the yellow pigment in the pericarp and embryo of *yel* mutant seed was orientin (Luteolin-8-C-glucoside). Yellow colored pericarp rice was rarely known so far. Oryzamutaic acids were detected in yellow endosperm mutant of the variety Hatsuyamabuki however, the biosynthetic pathway remains still unknown. This mutant showed the yellow endosperm which was caused by mutation of

Oryzamutaic acid A and *OsALDH7* (Nakano et al., 2009; Nakano et al., 2011; Shen et al., 2012) Mutation of *OsALDH7* resulted in accumulation of yellow pigment in endosperm. Oryzamutaic acid A accumulates during late seed development and after a year-long storage in the endosperm. Moreover, high level of orientin such as the *yel* mutant was not reported in rice grain. Embryo of *yel* mutant colored black not yellow because extremely high level of yellow colored orientin accumulated in small area of embryo. Changes in biosynthetic pathways or related metabolism have been reported in various mutant plants and various novel compounds may accumulate in plants with different organs and concentrations. We still couldn't determine the reason of the lethality of *yel* mutant embryo. It could be result from high level of pigmentation in embryo cell or null fucntion of *yel* allele. Our data denomstrated that the single recessive gene controlling high concentration of orientin in embryo and endosperm. In order to determine the consequences of unscheduled accumulation of flavonoid, Misera et al. constructed double mutants of fusca with mutant alleles of several genes involved in flavonoid biosynthesis. Although the double mutants lacked flavonoids, the mutant seedlings were nonetheless lethal, indicating that the lethality of fusca seedlings is not caused by excessive amounts of flavonoids (Misera et al., 1994).

Antioxidant activity of *yel* mutant embryo

Several assays have been frequently used to estimate antioxidant

capacities in fresh fruits and vegetables and their products and foods for clinical studies including 2,2-azinobis (3-ethyl-benzothiazoline-6-sulfonic acid) (ABTS) (Leong and Shui, 2002; Miller and Rice-Evans, 1997), 2,2-diphenyl-1-picrylhydrazyl (DPPH) (Brand-Williams et al., 1995; Gil et al., 2002), ferric reducing antioxidant power (FRAP) (Benzie and Strain, 1999; Guo et al., 2003; Jimenez-Escrig et al., 2001), and the oxygen radical absorption capacity (ORAC) (Cao et al., 1993; Ou et al., 2001; Prior et al., 2003). Ou et al. (2002) reported no correlation of antioxidant activity between the FRAP and ORAC techniques among most of the 927 freeze-dried vegetable samples, whereas these methods revealed high correlation in blueberry fruit (Connor et al., 2002). Similarly, Awika et al. (2003) observed high correlation between ABTS, DPPH, and ORAC among sorghum and its products (Thaipong et al., 2006). In our study, there was a significant correlation between total phenolic content of the *yel* mutant extracts and their DPPH• scavening activities, FRAP, and TEAC. This result indicates that the phenolic compounds are the major constitutuents in mutant embryo and endosperm contributing to DPPH• sacvening activity, FRAP, and TEAC.

Reference

Abdel-Aal el, S.M., Young, J.C., and Rabalski, I. (2006). Anthocyanin composition in black, blue, pink, purple, and red cereal grains. Journal of agricultural and food chemistry *54*, 4696-4704.

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M.*, et al.* (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nature biotechnology *30*, 174-178.

Ansorge, W.J. (2009). Next-generation DNA sequencing techniques. New biotechnology *25*, 195-203.

Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., Angeles, E.R., Qian, Q., Kitano, H., and Matsuoka, M. (2005). Cytokinin oxidase regulates rice grain production. Science *309*, 741-745.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G., and Edwards, R. (2009). The C-glycosylation of flavonoids in cereals. The Journal of biological chemistry *284*, 17926-17934.

Broun, P. (2005). Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in Arabidopsis. Current opinion in plant biology *8*, 272-279.

Brown, B.A., Cloix, C., Jiang, G.H., Kaiserli, E., Herzyk, P., Kliebenstein, D.J., and Jenkins, G.I. (2005). A UV-B-specific signaling component orchestrates plant UV protection. Proc Natl Acad Sci U S A *102*, 18225-18230.

Brown, M.E., and Funk, C.C. (2008). Climate - Food security under climate change. Science *319*, 580-581.

Butelli, E., Titta, L., Giorgio, M., Mock, H.P., Matros, A., Peterek, S., Schijlen, E.G., Hall, R.D., Bovy, A.G., Luo, J.*, et al.* (2008). Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. Nature biotechnology *26*, 1301-1308.

Catala, R., Medina, J., and Salinas, J. (2011). Integration of low temperature and light signaling during cold acclimation response in Arabidopsis. P Natl Acad Sci

USA *108*, 16475-16480.
Chalker-Scott, L. (1999). Environmental significance of anthocyanins in plant stress responses. Photochem Photobiol *70*, 1-9.
Chen, H., He, H., Zhou, F., Yu, H., and Deng, X.W. (2013). Development of genomics-based genotyping platforms and their applications in rice breeding. Current opinion in plant biology *16*, 247-254.
Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y.*, et al.* (2014). A high-density SNP genotyping array for rice biology and molecular breeding. Molecular plant *7*, 541-553.
Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J.*, et al.* (2008a). A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. Proc Natl Acad Sci U S A *105*, 11436-11441.
Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J.*, et al.* (2008b). A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica–japonica hybrids in rice. Proceedings of the National Academy of Sciences *105*, 11436-11441.
Cheng, S.H., Zhuang, J.Y., Fan, Y.Y., Du, J.H., and Cao, L.Y. (2007). Progress in research and development on hybrid rice: a super-domesticate in China. Annals of botany *100*, 959-966.
Christie, J.M., Arvai, A.S., Baxter, K.J., Heilmann, M., Pratt, A.J., O'Hara, A., Kelly, S.M., Hothorn, M., Smith, B.O., Hitomi, K.*, et al.* (2012). Plant UVR8 photoreceptor senses UV-B by tryptophan-mediated disruption of cross-dimer salt bridges. Science *335*, 1492-1496.
Chung, G.S., and Heu, M.H. (1980). Innovative approaches to rice breeding, pp. 139.
Chung, G.S., and Heu, M.H. (1991). Improvememt of Tongil-Type Rice Cultivars from *Indica*/*Japonica* Hybridization in Korea. In Biotechnology in Agriculture and Forestry 14, B. Y.P.S., ed. (Germany: Springer-Verlag Berlin Heidelberg), pp. 105-112.
Collard, B.C., and Mackill, D.J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical transactions of the Royal Society of London Series B, Biological sciences *363*, 557-572.

Cominelli, E., Gusmaroli, G., Allegra, D., Galbiati, M., Wade, H.K., Jenkins, G.I., and Tonelli, C. (2008). Expression analysis of anthocyanin regulatory genes in response to different light qualities in Arabidopsis thaliana. Journal of plant physiology *165*, 886-894.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674-3676.

Craig, D.W., Pearson, J.V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J.J., Pawlowski, T.L., Laub, T., Nunn, G., Stephan, D.A.*, et al.* (2008). Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods *5*, 887-893.

Crocco, C.D., Holm, M., Yanovsky, M.J., and Botto, J.F. (2010). AtBBX21 and COP1 genetically interact in the regulation of shade avoidance. The Plant journal : for cell and molecular biology *64*, 551-562.

Cronn, R., Liston, A., Parks, M., Gernandt, D.S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res *36*.

Da Silva, R.Z., Yunes, R.A., de Souza, M.M., Delle Monache, F., and Cechinel-Filho, V. (2010). Antinociceptive properties of conocarpan and orientin obtained from Piper solmsianum C. DC. var. solmsianum (Piperaceae). Journal of natural medicines *64*, 402-408.

Dan, Z.W., Liu, P., Huang, W.C., Zhou, W., Yao, G.X., Hu, J., Zhu, R.S., Lu, B.R., and Zhu, Y.G. (2014). Balance between a Higher Degree of Heterosis and Increased Reproductive Isolation: A Strategic Design for Breeding Inter-Subspecific Hybrid Rice. Plos One *9*.

Deng, X.W., Caspar, T., and Quail, P.H. (1991). cop1: a regulatory locus involved in light-controlled development and gene expression in Arabidopsis. Genes Dev *5*, 1172-1182.

Deng, X.W., Matsui, M., Wei, N., Wagner, D., Chu, A.M., Feldmann, K.A., and Quail, P.H. (1992). COP1, an Arabidopsis regulatory gene, encodes a protein with both a zinc-binding motif and a G beta homologous domain. Cell *71*, 791-801.

Dietrych-Szostak, D., and Oleszek, W. (1999). Effect of processing on the flavonoid content in buckwheat (Fagopyrum esculentum Moench) grain. Journal of agricultural and food chemistry *47*, 4384-4387.

Dixon, R.A., and Pasinetti, G.M. (2010). Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. Plant Physiol *154*, 453-457.

Duthie, G.G., Duthie, S.J., and Kyle, J.A. (2000). Plant polyphenols in cancer and heart disease: implications as nutritional antioxidants. Nutrition research reviews *13*, 79-106.

Dyachok, J., Zhu, L., Liao, F., He, J., Huq, E., and Blancaflor, E.B. (2011). SCAR mediates light-induced root elongation in Arabidopsis through photoreceptors and proteasomes. Plant Cell *23*, 3610-3626.

Dykes, L., and Rooney, L.W. (2007). Phenolic compounds in cereal grains and their health benefits. Cereal Food World *52*, 105-111.

Falconer, D.S., Mackay, T.F., and Frankham, R. (1996). Introduction to Quantitative Genetics (4th edn). Trends in Genetics *12*, 280.

Famoso, A.N., Zhao, K., Clark, R.T., Tung, C.W., Wright, M.H., Bustamante, C., Kochian, L.V., and McCouch, S.R. (2011). Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. PLoS genetics *7*, e1002221.

Favory, J.J., Stec, A., Gruber, H., Rizzini, L., Oravecz, A., Funk, M., Albert, A., Cloix, C., Jenkins, G.I., Oakeley, E.J*., et al.* (2009). Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in Arabidopsis. The EMBO journal *28*, 591-601.

Feller, A., Machemer, K., Braun, E.L., and Grotewold, E. (2011). Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. The Plant journal : for cell and molecular biology *66*, 94-116.

Feltus, F.A., Wan, J., Schulze, S.R., Estill, J.C., Jiang, N., and Paterson, A.H. (2004). An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. Genome research *14*, 1812-1819.

Field, B., and Osbourn, A.E. (2008). Metabolic diversification--independent assembly of operon-like gene clusters in different plants. Science *320*, 543-547.

Fitzgerald, M.A., McCouch, S.R., and Hall, R.D. (2009). Not just a grain of rice: the quest for quality. Trends Plant Sci *14*, 133-139.

Flint-Garcia, S.A., Thornsberry, J.M., and Buckler, E.S.t. (2003). Structure of linkage disequilibrium in plants. Annual review of plant biology *54*, 357-374.

Gao, Q., Yue, G.D., Li, W.Q., Wang, J.Y., Xu, J.H., and Yin, Y. (2012). Recent Progress Using High-throughput Sequencing Technologies in Plant Molecular Breeding. J Integr Plant Biol *54*, 215-227.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S., and McCouch, S. (2005). Genetic structure and diversity in Oryza sativa L. Genetics *169*, 1631-1638.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science *296*, 92-100.

Gonzalez, A., Zhao, M., Leavitt, J.M., and Lloyd, A.M. (2008). Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. The Plant journal : for cell and molecular biology *53*, 814-827.

Gross, B.L., and Zhao, Z.J. (2014). Archaeological and genetic insights into the origins of domesticated rice. P Natl Acad Sci USA *111*, 6190-6197.

Hackett, C.A., and Broadfoot, L.B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity *90*, 33-38.

Hanson, W.D. (1959). Early Generation Analysis of Lengths of Heterozygous Chromosome Segments around a Locus Held Heterozygous with Backcrossing or Selfing. Genetics *44*, 833-837.

Harborne, J.B. (1986). Nature, distribution and function of plant flavonoids. Progress in clinical and biological research *213*, 15-24.

Harushima, Y., Kurata, N., Yano, M., Nagamura, Y., Sasaki, T., Minobe, Y., and Nakagahra, M. (1996). Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. Theoret Appl Genetics *92*, 145-150.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent

advances in the transcriptional regulation of the flavonoid biosynthetic pathway. Journal of experimental botany *62*, 2465-2483.

Hiei, Y., Komari, T., and Kubo, T. (1997). Transformation of rice mediated by Agrobacterium tumefaciens. Plant Mol Biol *35*, 205-218.

Hoecker, U., and Quail, P.H. (2001). The phytochrome A-specific signaling intermediate SPA1 interacts directly with COP1, a constitutive repressor of light signaling in Arabidopsis. The Journal of biological chemistry *276*, 38173-38178.

Hoecker, U., Xu, Y., and Quail, P.H. (1998). SPA1: a new genetic locus involved in phytochrome A-specific signal transduction. Plant Cell *10*, 19-33.

Holm, M., Ma, L.G., Qu, L.J., and Deng, X.W. (2002). Two interacting bZIP proteins are direct targets of COP1-mediated control of light-dependent gene expression in Arabidopsis. Genes Dev *16*, 1247-1259.

Hospital, F. (2005). Selection in backcross programmes. Philosophical transactions of the Royal Society of London Series B, Biological sciences *360*, 1503-1511.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T.*, et al.* (2009). High-throughput genotyping by whole-genome resequencing. Genome research *19*, 1068-1076.

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

Huang, X., Ouyang, X., Yang, P., Lau, O.S., Li, G., Li, J., Chen, H., and Deng, X.W. (2012b). Arabidopsis FHY3 and HY5 positively mediate induction of COP1 transcription in response to photomorphogenic UV-B light. Plant Cell *24*, 4590-4606.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z.*, et al.* (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nature genetics *42*, 961-967.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C.*, et al.* (2012c). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature genetics *44*, 32-39.

Hultin, P.G. (2005). Bioactive C-glycosides from bacterial secondary metabolism.

Current topics in medicinal chemistry *5*, 1299-1331.
Initiative, T.A.G. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.
IRRI (2009). IRRI World Rice Statistics.
Jang, I.C., Yang, J.Y., Seo, H.S., and Chua, N.H. (2005). HFR1 is targeted by COP1 E3 ligase for post-translational proteolysis during phytochrome A signaling. Gene Dev *19*, 593-602.
Jang, S., Marchal, V., Panigrahi, K.C., Wenkel, S., Soppe, W., Deng, X.W., Valverde, F., and Coupland, G. (2008). Arabidopsis COP1 shapes the temporal pattern of CO accumulation conferring a photoperiodic flowering response. The EMBO journal *27*, 1277-1288.
Ji, Q., Zhang, M., Lu, J., Wang, H., Lin, B., Liu, Q., Chao, Q., Zhang, Y., Liu, C., Gu, M., *et al.* (2012). Molecular basis underlying the S5-dependent reproductive isolation and compatibility of indica/japonica rice hybrids. Plant physiology *158*, 1319-1328.
Jia, Y. (2009). Artificial introgression of a large chromosome fragment around the rice blast resistance gene Pi-ta in backcross progeny and several elite rice cultivars. Heredity *103*, 333-339.
Jia, Y., Jia, M.H., Wang, X., and Liu, G. (2012). Indica and japonica crosses resulting in linkage block and recombination suppression on rice chromosome 12. PloS one *7*, e43066.
Jiang, Y., Cai, Z., Xie, W., Long, T., Yu, H., and Zhang, Q. (2012). Rice functional genomics research: progress and implications for crop genetic improvement. Biotechnology advances *30*, 1059-1070.
Kang, C.Y., Lian, H.L., Wang, F.F., Huang, J.R., and Yang, H.Q. (2009). Cryptochromes, phytochromes, and COP1 regulate light-controlled stomatal development in Arabidopsis. Plant Cell *21*, 2624-2641.
Khush, G.S. (1997). Origin, dispersal, cultivation and variation of rice. Plant Mol Biol *35*, 25-34.
Khush, G.S. (1999). Green revolution: preparing for the 21st century. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada *42*, 646-655.

Khush, G.S. (2001). Green revolution: the way forward. Nat Rev Genet *2*, 815-822.
Kim, M.K., Kim, H.A., Koh, K., Kim, H.S., Lee, Y.S., and Kim, Y.H. (2008). Identification and quantification of anthocyanin pigments in colored rice. Nutrition research and practice *2*, 46-49.
Kim, S.L., Lee, J.E., Kwon, Y.U., Kim, W.H., Jung, G.H., Kim, D.W., Lee, C.K., Lee, Y.Y., Kim, M.J., Kim, Y.H.*, et al.* (2013). Introduction and nutritional evaluation of germinated soy germ. Food Chem *136*, 491-500.
Koboldt, D.C., Steinberg, K.M., Larson, D.E., Wilson, R.K., and Mardis, E.R. (2013). The next-generation sequencing revolution and its impact on genomics. Cell *155*, 27-38.
Koes, R., Verweij, W., and Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci *10*, 236-242.
Kong, J.M., Chia, L.S., Goh, N.K., Chia, T.F., and Brouillard, R. (2003). Analysis and biological activities of anthocyanins. Phytochemistry *64*, 923-933.
Laubinger, S., Fittinghoff, K., and Hoecker, U. (2004). The SPA quartet: A family of WD-repeat proteins with a central role in suppression of photomorphogenesis in arabidopsis. Plant Cell *16*, 2293-2306.
Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., Tongprasit, W., Zhao, H., Lee, I., and Deng, X.W. (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell *19*, 731-749.
Li, Q., Li, L., Yang, X., Warburton, M.L., Bai, G., Dai, J., Li, J., and Yan, J. (2010). Relationship, evolutionary fate and function of two maize co-orthologs of rice GW2 associated with kernel size and weight. BMC plant biology *10*, 143.
Li, S. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. Plant signaling & behavior *9*, e27522.
Li, Z., Pinson, S.R., Paterson, A.H., Park, W.D., and Stansel, J.W. (1997). Genetics of hybrid sterility and hybrid breakdown in an intersubspecific rice (Oryza sativa L.) population. Genetics *145*, 1139-1148.
Lillo, C., Lea, U.S., and Ruoff, P. (2008). Nutrient depletion as a key factor for

manipulating gene expression and product formation in different branches of the flavonoid pathway. Plant Cell Environ *31*, 587-601.
Lin, S.Y., Ikehashi, H., Yanagihara, S., and Kawashima, A. (1992). Segregation distortion via male gametes in hybrids between Indica and Japonica or wide-compatibility varieties of rice (Oryza sativa L). Theoret Appl Genetics *84*, 812-818.
Liu, L.J., Zhang, Y.C., Li, Q.H., Sang, Y., Mao, J., Lian, H.L., Wang, L., and Yang, H.Q. (2008). COP1-mediated ubiquitination of CONSTANS is implicated in cryptochrome regulation of flowering in Arabidopsis. Plant Cell *20*, 292-306.
Liu, T., Liu, H., Zhang, H., and Xing, Y. (2013). Validation and characterization of Ghd7.1, a major QTL with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (Oryza sativa L.). Journal of integrative plant biology.
Long, Y., Zhao, L., Niu, B., Su, J., Wu, H., Chen, Y., Zhang, Q., Guo, J., Zhuang, C., Mei, M.*, et al.* (2008). Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proceedings of the National Academy of Sciences *105*, 18871-18876.
Lu, B.R., Zheng, K.L., Qian, H.R., and Zhuang, J.Y. (2002). Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theoretical and Applied Genetics *106*, 101-106.
Lu, N., Sun, Y., and Zheng, X. (2011). Orientin-induced cardioprotection against reperfusion is associated with attenuation of mitochondrial permeability transition. Planta medica *77*, 984-991.
Luo, X.M., Lin, W.H., Zhu, S., Zhu, J.Y., Sun, Y., Fan, X.Y., Cheng, M., Hao, Y., Oh, E., Tian, M.*, et al.* (2010). Integration of light- and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis. Developmental cell *19*, 872-883.
Ma, J.X., and Bennetzen, J.L. (2004). Rapid recent growth and divergence of rice nuclear genomes. P Natl Acad Sci USA *101*, 12404-12410.
Matsumoto, T., Wu, J.Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., Mizuno, H., Yamamoto, K., Antonio, B.A., Baba, T.*, et al.* (2005). The map-based sequence of the rice genome. Nature *436*, 793-800.
McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y.*, et al.* (2002). Development and mapping of 2240 new

SSR markers for rice (Oryza sativa L.). DNA research : an international journal for rapid publication of reports on genes and genomes *9*, 199-207.
McCouch, S.R., Zhao, K.Y., Wright, M., Tung, C.W., Ebana, K., Thomson, M., Reynolds, A., Wang, D., DeClerck, G., Ali, M.L.*, et al.* (2010). Development of genome-wide SNP assays for rice. Breeding Sci *60*, 524-535.
McNally, K.L., Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E.*, et al.* (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. P Natl Acad Sci USA *106*, 12273-12278.
Misera, S., Muller, A.J., Weiland-Heidecker, U., and Jurgens, G. (1994). The FUSCA genes of Arabidopsis: negative regulators of light responses. Molecular & general genetics : MGG *244*, 242-252.
Miura, K., Ashikari, M., and Matsuoka, M. (2011). The role of QTLs in the breeding of high-yielding rice. Trends Plant Sci *16*, 319-326.
Mizuta, Y., Harushima, Y., and Kurata, N. (2010). Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. Proc Natl Acad Sci U S A *107*, 20417-20422.
Monna, L., Kitazawa, N., Yoshino, R., Suzuki, J., Masuda, H., Maehara, Y., Tanji, M., Sato, M., Nasu, S., and Minobe, Y. (2002). Positional cloning of rice semidwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA research : an international journal for rapid publication of reports on genes and genomes *9*, 11-17.
Morishima, H., and Oka, H.I. (1981). Phylogenetic Differentiation of Cultivated Rice .22. Numerical Evaluation of the Indica-Japonica Differentiation. Jpn J Breed *31*, 402-413.
Nagano, H., Onishi, K., Ogasawara, M., Horiuchi, Y., and Sano, Y. (2005). Genealogy of the "Green Revolution" gene in rice. Genes & genetic systems *80*, 351-356.
Nakano, H., Kosemura, S., Suzuki, T., Hirose, K., Kaji, R., and Sakai, M. (2009). Oryzamutaic acid A, a novel yellow pigment from an Oryza sativa mutant with yellow endosperm. Tetrahedron Lett *50*, 2003-2005.
Nakano, H., Tamura, K., Sakai, M., Ooka, N., Arakawa, M., Yagasaki, K., Minoda, T.,

Noda, S., and Arai, M. (2011). Isolation and identification of oryzamutaic acids A and D from an Oryza sativa spontaneous mutant with yellow endosperm. Breeding Sci *61*, 93-95.

Nam, S.H., Choi, S.P., Kang, M.Y., Koh, H.J., Kozukue, N., and Friedman, M. (2006). Antioxidative activities of bran extracts from twenty one pigmented rice cultivars. Food Chem *94*, 613-620.

Naveira, H., and Barbadilla, A. (1992). The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species. Genetics *130*, 205-209.

Nelson, G.C., and International Food Policy Research Institute. (2009). Climate change : impact on agriculture and costs of adaptation (Washington, D.C.: International Food Policy Research Institute).

Ogo, Y., Ozawa, K., Ishimaru, T., Murayama, T., and Takaiwa, F. (2013). Transgenic rice seed synthesizing diverse flavonoids at high levels: a new platform for flavonoid production with associated health benefits. Plant biotechnology journal *11*, 734-746.

Olsen, K.M., Caicedo, A.L., Polato, N., McClung, A., McCouch, S., and Purugganan, M.D. (2006). Selection under domestication: evidence for a sweep in the rice waxy genomic region. Genetics *173*, 975-983.

Oravecz, A., Baumann, A., Mate, Z., Brzezinska, A., Molinier, J., Oakeley, E.J., Adam, E., Schafer, E., Nagy, F., and Ulm, R. (2006). CONSTITUTIVELY PHOTOMORPHOGENIC1 is required for the UV-B response in Arabidopsis. Plant Cell *18*, 1975-1990.

Ouyang, Y.D., Liu, Y.G., and Zhang, Q.F. (2010). Hybrid sterility in plant: stories from rice. Current opinion in plant biology *13*, 186-192.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M.R., Zschocke, J., and Trajanoski, Z. (2014). A survey of tools for variant analysis of next-generation genome sequencing data. Briefings in bioinformatics *15*, 256-278.

Rafalski, J.A. (2010). Association genetics in crop improvement. Current opinion in plant biology *13*, 174-180.

Rieseberg, L.H. (2001). Chromosomal rearrangements and speciation. Trends in ecology & evolution *16*, 351-358.
Saijo, Y., Sullivan, J.A., Wang, H., Yang, J., Shen, Y., Rubio, V., Ma, L., Hoecker, U., and Deng, X.W. (2003). The COP1-SPA1 interaction defines a critical step in phytochrome A-mediated regulation of HY5 activity. Genes Dev *17*, 2642-2647.
Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., and Fernie, A.R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale *72*, 21-34.
Sano, R., and Morishima, H. (1992). Indica-Japonica differentiation of rice cultivars viewed from variations in key characters and isozymes, with special reference to landraces from the Himalayan hilly areas. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *84*, 266-274.
Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Swapan, D., Ishiyama, K., Saito, T., Kobayashi, M., Khush, G.S.*, et al.* (2002). Green revolution: a mutant gibberellin-synthesis gene in rice. Nature *416*, 701-702.
Schijlen, E., Ric de Vos, C.H., Jonker, H., van den Broeck, H., Molthoff, J., van Tunen, A., Martens, S., and Bovy, A. (2006). Pathway engineering for healthy phytochemicals leading to the production of novel flavonoids in tomato fruit. Plant biotechnology journal *4*, 433-444.
Schwechheimer, C., and Deng, X.W. (2000). The COP/DET/FUS proteins-regulators of eukaryotic growth and development. Seminars in cell & developmental biology *11*, 495-503.
Seo, H.S., Yang, J.Y., Ishikawa, M., Bolle, C., Ballesteros, M.L., and Chua, N.H. (2003). LAF1 ubiquitination by COP1 controls photomorphogenesis and is stimulated by SPA1. Nature *423*, 995-999.
Shen, Y., Wan, Z., Coarfa, C., Drabek, R., Chen, L., Ostrowski, E.A., Liu, Y., Weinstock, G.M., Wheeler, D.A., Gibbs, R.A.*, et al.* (2010). A SNP discovery method to assess variant allele probability from next-generation resequencing data. Genome research *20*, 273-280.
Shen, Y., Zhang, Y., Yang, C., Lan, Y., Liu, L.L., Liu, S.J., Chen, Z.J., Ren, G.X., and Wan,

J.M. (2012). Mutation of OsALDH7 causes a yellow-colored endosperm associated with accumulation of oryzamutaic acid A in rice. Planta *235*, 433-441.
Shi, M.Z., and Xie, D.Y. (2010). Features of anthocyanin biosynthesis in pap1-D and wild-type Arabidopsis thaliana plants grown in different light intensity and culture media conditions. Planta *231*, 1385-1400.
Shimura, K., Okada, A., Okada, K., Jikumaru, Y., Ko, K.W., Toyomasu, T., Sassa, T., Hasegawa, M., Kodama, O., Shibuya, N*., et al.* (2007). Identification of a biosynthetic gene cluster in rice for momilactones. The Journal of biological chemistry *282*, 34013-34018.
Shin, J., Park, E., and Choi, G. (2007). PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in Arabidopsis. The Plant journal : for cell and molecular biology *49*, 981-994.
Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., and Yano, M. (2008). Deletion in a gene associated with grain size increased yields during rice domestication. Nature genetics *40*, 1023-1028.
Song, X., Qiu, S.Q., Xu, C.G., Li, X.H., and Zhang, Q.F. (2005). Genetic dissection of embryo sac fertility, pollen fertility, and their contributions to spikelet fertility of intersubspecific hybrids in rice. Theoretical and Applied Genetics *110*, 205-211.
Stacey, M.G., Kopp, O.R., Kim, T.H., and von Arnim, A.G. (2000). Modular domain structure of Arabidopsis COP1. Reconstitution of activity by fragment complementation and mutational analysis of a nuclear localization signal in planta. Plant Physiol *124*, 979-990.
Suzuki, G., Yanagawa, Y., Kwok, S.F., Matsui, M., and Deng, X.W. (2002). Arabidopsis COP10 is a ubiquitin-conjugating enzyme variant that acts together with COP1 and the COP9 signalosome in repressing photomorphogenesis. Gene Dev *16*, 554-559.
Swaminathan, S., Morrone, D., Wang, Q., Fulton, D.B., and Peters, R.J. (2009). CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. The Plant cell *21*, 3315-3325.

Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., Utsushi, H., Natsume, S., Kanzaki, H., Matsumura, H., *et al.* (2013). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. The New phytologist *200*, 276-283.

Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., Padhukasahasram, B., Bustamante, C., Yoshimura, A., Doi, K., *et al.* (2009). Evolutionary history of GS3, a gene conferring grain length in rice. Genetics *182*, 1323-1334.

Takeda, S., and Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. Nat Rev Genet *9*, 444-457.

Tanaka, N., Itoh, H., Sentoku, N., Kojima, M., Sakakibara, H., Izawa, T., Itoh, J., and Nagato, Y. (2011). The COP1 ortholog PPS regulates the juvenile-adult and vegetative-reproductive phase changes in rice. Plant Cell *23*, 2143-2154.

Tang, J.B., Xia, H.A., Cao, M.L., Zhang, X.Q., Zeng, W.Y., Hu, S.N., Tong, W., Wang, J., Wang, J., Yu, J., *et al.* (2004). A comparison of rice chloroplast genomes. Plant Physiol *135*, 412-420.

Thaipong, K., Boonprakob, U., Crosby, K., Cisneros-Zevallos, L., and Byrne, D.H. (2006). Comparison of ABTS, DPPH, FRAP, and ORAC assays for estimating antioxidant activity from guava fruit extracts. J Food Compos Anal *19*, 669-675.

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., *et al.* (2005). Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. The Plant journal : for cell and molecular biology *42*, 218-235.

Tsuda, T., Horio, F., Uchida, K., Aoki, H., and Osawa, T. (2003). Dietary cyanidin 3-O-beta-D-glucoside-rich purple corn color prevents obesity and ameliorates hyperglycemia in mice. The Journal of nutrition *133*, 2125-2130.

Tung, C.W., Zhao, K.Y., Wright, M.H., Ali, M.L., Jung, J., Kimball, J., Tyagi, W., Thomson, M.J., McNally, K., Leung, H., *et al.* (2010). Development of a Research Platform for Dissecting Phenotype-Genotype Associations in Rice (Oryza spp.). Rice *3*, 205-217.

Vaughan, D.A., Lu, B.R., and Tomooka, N. (2008). The evolving story of rice evolution. Plant Sci *174*, 394-408.

Vitte, C., Ishii, T., Lamy, F., Brar, D., and Panaud, O. (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (Oryza sativa L.). Mol Genet Genomics *272*, 504-511.

Wang, L., Wang, A.H., Huang, X.H., Zhao, Q.A., Dong, G.J., Qian, Q.A., Sang, T., and Han, B. (2011). Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theoretical and Applied Genetics *122*, 327-340.

Wegel, E., Koumproglou, R., Shaw, P., and Osbourn, A. (2009). Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. The Plant cell *21*, 3926-3936.

Weng, D.X., Xu, S.C., Lin, R.M., Wan, A.M., Li, J.P., and Wu, L.R. (2005). [Microsatellite marker linked with stripe rust resistant gene Yr9 in wheat]. Yi chuan xue bao = Acta genetica Sinica *32*, 937-941.

Widmer, A., Lexer, C., and Cozzolino, S. (2009). Evolution of reproductive isolation in plants. Heredity *102*, 31-38.

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. Current opinion in plant biology *5*, 218-223.

Wu, D., Hu, Q., Yan, Z., Chen, W., Yan, C., Huang, X., Zhang, J., Yang, P., Deng, H., Wang, J.*, et al.* (2012). Structural basis of ultraviolet-B perception by UVR8. Nature *484*, 214-219.

Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., Yu, S., Han, B., and Zhang, Q. (2010). Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci U S A *107*, 10578-10583.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L.*, et al.* (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nature biotechnology *30*, 105-111.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li,

X.*, et al.* (2008). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nature genetics *40*, 761-767.
Yan, S., Zou, G., Li, S., Wang, H., Liu, H., Zhai, G., Guo, P., Song, H., Yan, C., and Tao, Y. (2011). Seed size is determined by the combinations of the genes controlling different seed characteristics in rice. TAG Theoretical and applied genetics Theoretische und angewandte Genetik *123*, 1173-1181.
Yang, J., Zhao, X., Cheng, K., Du, H., Ouyang, Y., Chen, J., Qiu, S., Huang, J., Jiang, Y., Jiang, L.*, et al.* (2012). A Killer-Protector System Regulates Both Hybrid Sterility and Segregation Distortion in Rice. Science *337*, 1336-1340.
Yi, C., and Deng, X.W. (2005). COP1 - from plant photomorphogenesis to mammalian tumorigenesis. Trends in cell biology *15*, 618-625.
Yu, H.H., Xie, W.B., Wang, J., Xing, Y.Z., Xu, C.G., Li, X.H., Xiao, J.H., and Zhang, Q.F. (2011). Gains in QTL Detection Using an Ultra-High Density SNP Map Based on Population Sequencing Relative to Traditional RFLP/SSR Markers. Plos One *6*.
Yu, J., Hu, S.N., Wang, J., Wong, G.K.S., Li, S.G., Liu, B., Deng, Y.J., Dai, L., Zhou, Y., Zhang, X.Q.*, et al.* (2002). A draft sequence of the rice genome (Oryza sativa L. ssp indica). Science *296*, 79-92.
Yu, J.W., Rubio, V., Lee, N.Y., Bai, S., Lee, S.Y., Kim, S.S., Liu, L., Zhang, Y., Irigoyen, M.L., Sullivan, J.A.*, et al.* (2008). COP1 and ELF3 control circadian function and photoperiodic flowering by regulating GI stability. Molecular cell *32*, 617-630.
Zhang, F., Gonzalez, A., Zhao, M., Payne, C.T., and Lloyd, A. (2003). A network of redundant bHLH proteins functions in all TTG1-dependent pathways of Arabidopsis. Development *130*, 4859-4869.
Zhao, C., Giusti, M.M., Malik, M., Moyer, M.P., and Magnuson, B.A. (2004). Effects of commercial anthocyanin-rich extracts on colonic cancer and nontumorigenic colonic cell growth. Journal of agricultural and food chemistry *52*, 6122-6128.
Zhao, J., and Dixon, R.A. (2010). The 'ins' and 'outs' of flavonoid transport. Trends Plant Sci *15*, 72-80.
Zhao, K., Tung, C.W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J.*, et al.* (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature

communications *2*, 467.
Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M.L., Tung, C.W., Reynolds, A.*, et al.* (2010a). Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. PloS one *5*, e10780.
Zhao, Q.A., Huang, X.H., Lin, Z.X., and Han, B. (2010b). SEG-Map: A Novel Software for Genotype Calling and Genetic Map Construction from Next-generation Sequencing. Rice *3*, 98-102.
Zhou, L.J., Ao, G.H., Wu, X.J., and Li, S.G. (2005). [SSR markers linked with early stability in rice]. Yi chuan xue bao = Acta genetica Sinica *32*, 837-845.
Zimmermann, I.M., Heim, M.A., Weisshaar, B., and Uhrig, J.F. (2004). Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. The Plant journal : for cell and molecular biology *40*, 22-34.

초 록

통일벼의 유전체 분석과 벼에서 노란색 종피 및 배 치사 유전자 동정

통일벼는 인디카 벼 (IR8, TN1)와 자포니카 벼 (Yukara)의 삼원교배로 만들어진, 우리나라의 녹색혁명을 가져다 준 고수량 벼이다. 이러한 통일벼와 통일벼의 모본들을 각각 47X 와 30X 의 coverage 로 re-sequencing 하여 염기수준에서 유전체의 구조를 분석하였다. 통일벼 내에서 인디카-자포니카로 구성되는 유전체의 구조는 통일벼와 세 모본 사이에서 나타난 SNP 를 기반으로, sliding window 방법을 이용하여 결정하였다. 그 결과 통일벼 유전체의 91.8%는 인디카 모본에서 유래하였고, 7.9%는 자포니카 모본에서 유래 한 것을 알 수 있었다. SSR (Simple sequence repeat) motif, ORF gene 분포, GO (Gene ontology) 등의 유전체를 구성하는 요소들을 분석해 본 결과 통일벼를 이루고 있는 인디카-자포니카 유전체의 구성과 유사한 비율로 이들 요소들이 통일벼 유전체 내에 존재하고 있음을 알 수 있었다. 또한 벼의 수량과 밀접한 연관이 있는, 이미 밝혀진 유전자 중 7 개의 유전자에 대한 통일벼 내에서의 분포 양상을 살펴본 결과 7 개 유전자 모두 인디카 벼의 모본에서 유래하였음을 알 수 있었다.

이러한 결과들은 통일벼 유전체의 대부분이 인디카 벼에서 유래했다는 사실을 뒷받침 해주고, 유전체를 구성하는 요소들 또한 유전체의 구조가 변함에 따라 함께 변화한다는 사실은 알 수 있었다. 통일벼의 유전체 구조를 밝힘으로써 육종적 효용성뿐만 아니라 앞으로 고수량 또는

인디카-자포니카 교잡 품종을 육성함에 있어서 중요한 기초 자료가 되리라 생각된다.

MNU 의 처리를 통해 얻어진, 노란색 종피색을 가지며 동시에 배가 죽는 돌연변이의 유전분석과 고밀도 유전자지도를 작성한 결과 이러한 표현형과 관련된 유전자는 2 번 염색체에 존재하였고 애기장대의 COP1 (Constitutive Photomorphogenesis 1)과 이종상동성 유전자인 것으로 밝혀졌다. 모본과 돌연변이에 대한 해당 유전자 (LOC_Os02g53140)의 sequence 변이를 살펴본 결과 5'UTR 과 start codon 을 포함한 첫 번째 exon, intron 일부 등, 총 706 bp 의 염기가 deletion 된 것을 확인 하였다. 이로 인해 이 유전자가 가지는 3 가지 domain, RING-finger, Coiled-coil, WD40 domain 중에서 RING-finger domain 의 전체가 deletion 되었다. 정상 YEL 유전자를 과발현 시킨 결과 종피가 백색을 띠는 정상 종자로 회복되었지만 배에는 지속적인 물질 축적이 관찰되었고 배가 치사되는 형질은 회복되지 못했다. GUS assay 결과 화분과 배, 종피, 마디(절)에서 유전자가 발현됨을 관찰할 수 있었다. 또한 종피와 배에 축적되는 물질을 HPLC/MS 를 이용하여 분석한 결과 Orientin 이라는 물질로 확인하였고 특히 배에 고농도로 축적되어 있었다. 이 물질의 항산화 능력은 orientin 이 고농도로 축적된 배에서 높게 나타났고, *yel* 뮤턴트 종자의 total polyphenol 함량도 높게 나타났다. *yel* 뮤턴트의 배에서는 포화 지방산 함량이 증가하였고 반대로 불포화 지방산 함량은 감소하였다. *yel* 뮤턴트 종자의 배에서 조지방의 함량은 정상 배에 비해서 유의하게 감소 하는 것으로 나타났다.

위의 결과들을 통해 YEL 유전자는 벼에서 플라보노이드 (Orientin)f 의 생합성 대사 과정에 관여하고 배의 발달 또는 배를 구성하는 성분들의

생산에 영향을 미치고 있음을 밝혀, 벼에서 COP1 유전자의 새로운 기작을 구명하는데 중요한 기초 자료를 제공하는 데에도 그 의의가 있다.

주요어: 벼, 통일벼, 차세대유전체분석, 유전체 구조, 인디카-자포니카, 노란색 종피, 배 치사 돌연변이, Orientin, 항산화능력

학번: 2010-30291