Molecular Aspects of 2OG-Fe (II) - Dioxygenases and Response to Abiotic Stress in Rice

Thesis submitted in part fulfilment of the Degree of Doctor of Philosophy

School of Agriculture, Policy and Development

Raheema Wamaedeesa

January 2016

Abstract

In specific mammalian cells, the Ten-Eleven Translocation (TET) enzymes which are family members of 2-oxoglutarate (2-OG)-and Fe (II)-dependent dioxygenase (2ODD) are responsible for the modification of 5-mC to generate oxidised products such as 5-hmC, 5-caC and 5-fC, which are involved in oxidative demethylation. In rice, some reports have also revealed that 5-hmC was also present but the enzyme(s) responsible for oxidising 5-mC has not been identified. Here, possible *Tet* gene homologues in rice were identified by using various databases. There are 18 genes having a 2ODD domain most similar to TETs. Moreover, additional significant information such as the relatedness of protein sequence, protein interaction, nuclear localisation prediction, mutant analysis as well as gene annotation was obtained. Next, the spatial and temporal expression patterns of these genes were analysed within seedling root, mature leaf, young leaf and young panicle by analysing data from both microarray and qPCR. It was found that *Os01g0546900* had highest expression in young leaf and six genes showed highest expression in mature leaf, whereas the remaining genes showed low expression in various organs. Finally, selected contrasting cultivars of rice were used to examine the expression of three genes of 2ODD and *SLENDER RICE 1* (*SLR1*) under flooding and drought in order to evaluate the molecular aspects of *Submergence 1A* (*Sub1A*) regulation. The results showed that the *Sub1A* gene of tolerant rice was highly expressed under submergence but was not expressed under drought. *Os03g0803500*, *Os07g0194500* and *Os09g0489200,* and *SLR1* did not show any significant difference in expression in submerged and dehydrated plants of both IR64 and IR64.S1, nor in Japonica varieties which were used as susceptible check. To conclude, this finding indicates that *Os03g0803500*, *Os07g0194500*, *Os09g0489200* and *SLR1* may not have a correlation with *Sub1A* under flooding and drought stress.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

……..

Raheema Wamaedeesa

Acknowledgement

In the name of Allah, the Most Gracious, the Most Merciful. With these verses presented in the Al-Quran, *"*Read in the name of your Lord who created", "Nun*,* By the pen and what they inscribe", **"**Are the educated and uneducated equal?" and "Lord, increase my knowledge."**,** Those verses vigorously motivate me to continue my studying of Doctoral degree. During this PhD journey, I have learnt and obtained much more experiences of not only being a good scientific researcher but also how to live worth life, abroad. It is a great pleasure to thank the kind-hearted people who supported and helped me over the past few years.

First and foremost, I would like to express my hearties gratitude to my supervisor, Professor Jim Dunwell, for providing me with this great opportunity to do my PhD under his great supervision, helpful assistance, invaluable guidance and excellent encouragement from the beginning of my study to the final stages.

Special thanks to Professor Nick Battey, Dr Andy Wetten, Dr G. Gibbings, and Dr Ihsan Ullah for clarifying me on molecular biology techniques and information. I owe them my heartfelt gratitude for their friendly and kindly assistance given to me in the molecular biology laboratory.

The staff members of the school of Agriculture, Policy and Development particularly Matthew Richardson, Val Jasper, Christine Williams, Anna Mieleszko and Alp Volkan as well as the staff members of the School of Biological Sciences particularly Sue Rumsey, Fiona Brown, Lee, also deserve my sincerest thanks for their friendly and warm-hearted assistance.

I would like to thank my colleagues of plant science group particularly Dr Kulchana , Dr Anushaka, Dr Shridar, Dr Asaad Mahmood, Eddy and Dr Sujitra Techakote for their suggestions and the help. I would like to extend my thanks to Thai friends Dr Karnjana Wiangnon, Dr Benjawan Poonthananiwatkul, Dr Chaweewan Klongsiriwet and all friends

from Muslim Thai student Society in the UK who always gave me a spirit, cheer and great advice.

I would like also to express my thanks to the Former President, Asst. Prof. Dr. Jongrak Palasai, the Former Dean, Mongkol Wachiraampai, the Present Dean of faculty of Agricultural Sciences, Tawee Bunphirom, and all of my colleagues at Princess of Naradhiwas University, Thailand, for their encouragement and support.

Major recognition goes to my mom and dad for their love, encouragement, and the sacrifices that they have made for me throughout their life. Special gratitude is extended to my husband, my sisters and my children, who support me and make my life stable and warm.

I also gratefully acknowledge the funding support provided by the Ministry of Science and Technology of Thailand

Finally, all praise is due to Allah, who is most compassionate and most merciful. This work indeed could not have been achieved at all without his allowance. My life and my soul are for the sake of him.

List of contents

List of figures

List of tables

List of abbreviations

2ODDs	2OG-Fe (II)-dependent dioxygenases
5-caC	5-carboxycytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxy methylcytosine
5-mC	5-methylcytosine
ABA	Abscisic acid
ALKB	Alpha-ketoglutarate-dependent dioxygenase
ALKBH5	ALKB homolog 5
AP	APETALA
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
blastp	Protein blast (Search protein database using a protein query)
CD	Cysteine-rich domain
CDD	conserved domain database
cDNA	Complementary DNA
CD-search tool	Conserved Domain Search Service tool
CEL	Sixty-six cell intensity
CMML	Chronic myelomonocytic leukemia
CMT3	CHROMOMETHYLASE3
Cn3D tool	macromolecular structure viewer tool
Cnr	Colourless non-ripening
DNA	Deoxyribonucleic acid
dpi	days post infection

DREB1s	DEHYDRATION RESPONSIVE ELEMENT BINDING
DRM	DOMAINS REARRANGED METHYLTRANSFERASE
DSBH	Double-strand β-helix
dSpm	Defective-suppressor mutator
eFP	electronic Fluorescent Pictographic
ELM	Eukaryotic Linear Motif
EPO	erythropoietin
ERFs	Ethylene-responsive factors
ESC	Embryonic stem cell
EST database	Expressed Sequence Tags database
FAO	Food and Agriculture Organization (United Nations)
FR13A	Flood Resistant 13A
FTO	Fat mass and obesity-associated protein
GA	Gibberellic acid
gDNA	Genomic DNA
HIF	Hypoxia-inducible transcription factor
HIF1	Hypoxia-inducible factor 1
HPLC	High Performance Liquid Chromatography
IAA	Indole-3-acetic acid
Jmj	Jumonji
LC-MS/MS/MS	Liquid Chromatography-multi-stage mass spectrometry
LDH	lactate dehydrogenase
LEA	LATE EMBRYOGENESIS ABUNDANT
MAS	Microarray Suite software

MET1	DNA METHYLTRANSFERASE1
miRNA	Micro RNA
MTases	DNA methyltransferases
NCBI	The National Center for Biotechnology Information
NERP	N-end rule pathway
NLSs	Nuclear localisation sequences
nr protein	Non redundant protein
OGT	O-linked N-acetylglucosamine (O-GlcNAc) transferase
ORF	open reading frame
P4H	prolyl 4-hydroxylase
PCR	Polymerase chain reaction
PGK	phosphoglycerate kinase
piRNA	Piwi-interacting RNA
qPCR	Quantitative polymerase chain reaction
QTL	Quantitative trait locus
RdDM	RNA-directed DNA methylation
rDNA	Ribosomal DNA
RMA	Robust Multichip Average
RNA	Ribonucleic Acid
RT-PCR	reverse transcription PCR
SA	Salicylic acid
semi-qPCR	Semi-quantitative Real time PCR
SK1	SNORKEL1
SK2	SNORKEL2

SLR1	SLENDER RICE 1
SLRL1	SLR1 Like 1
SNP	single nucleotide polymorphism
snRNA	Small nuclear RNA
SRA	SET and RING finger associated domain
Sub 1	Submergence 1
Sub1A	Submergence 1A
Sub1B	Submergence 1B
Sub1C	Submergence 1C
SUVH4	Su(var)3-9 homologue protein 4
SVM	Support vector machine
Tblastn	Search translated nucleotide database using a protein query
T-DNA	Transfer DNA
TEs	Transposable elements
TET	Ten-eleven translocation
tRNA	Transfer RNA
TSSs	Transcription start sites
USDA	United States Department of Agriculture
VIM	VARIATION IN METHYLATION

Chapter 1 Introduction

1.1 The importance of rice

Considering its important position, the United Nations nominated the year 2004 as the International Year of Rice. It is the essential food of more than half the world's population and has fed a great number of people of Asia for a longer period of time than any other food crop. Furthermore, there is an acknowledged necessity to increase the awareness of the role of rice in relieving poverty and malnourishment (Changchui, 2004). Nowadays, rice consumption has been increasing gradually. It is estimated that 85% of total production is consumed by humans and more than 90% of this is consumed in Asia, where it is a principle food for a majority of the population, including the region's 560 million starved people (Figure 1-1).

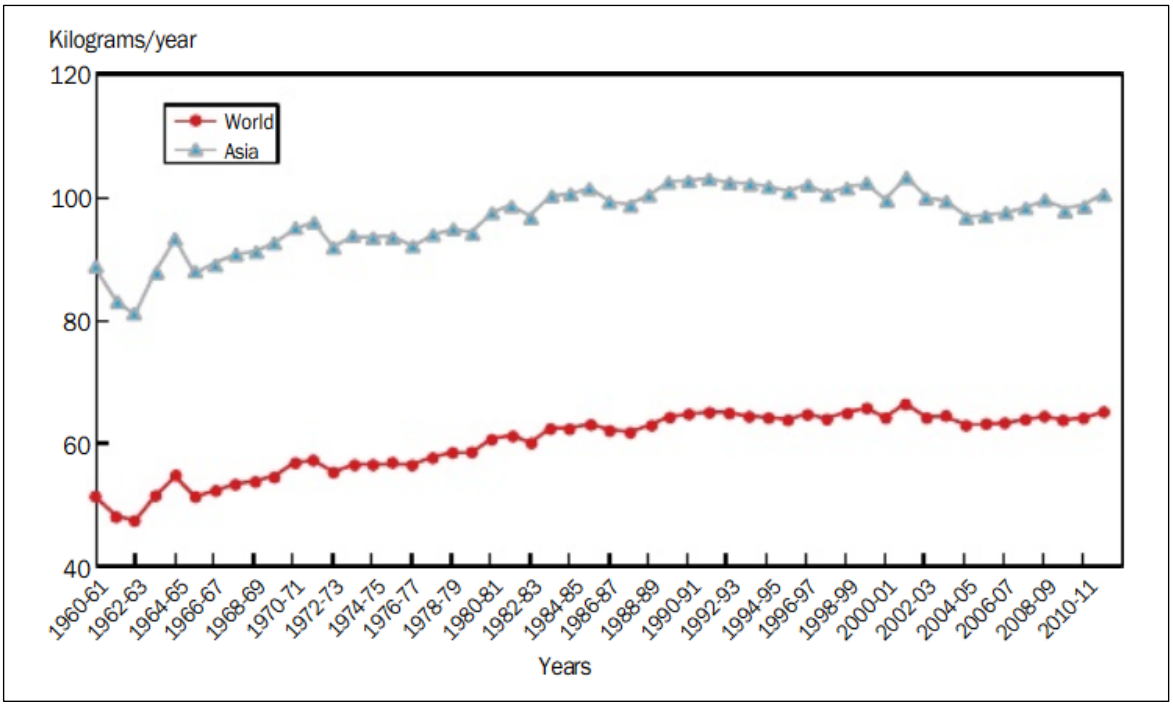


Figure 1-1 Global and Asian per capita rice consumption. (Mohanthy, 2013)

Specifically in Asia, more than 2 billion people gain 60 to 70 percent of their food energy from rice and its derivative products. Rice has a great influence on human nourishment and the fight against starvation all over the world. It provides 21% of global human per capita energy, 15% of per capita protein as well as minerals, vitamins, and fibre (Maclean, et al., 2002 ; Guimarães, 2009).

Moreover, 30% of the world cereal production today is rice production. This production has doubled in the last 30 years due to the development of new improved varieties but at the present time, this growth of production barely follows consumption. In 2025 there will be 4.6 billion people dependent on rice for their daily nourishment, compared with three billion today. Therefore, there is an urgent need an increase the production of rice. At present, rice is cultivated in 114 countries of which China is currently the world's leading rice-producer. Besides, its yields worldwide tend to be higher than 4 tonnes/ha on average (Table 1-1).

Table 1-1 Rice production estimates in the 10 leading-producing countries; five-year average 2006–2010 (Food and Agriculture Organization of the United Nation (FAO/UN), 2012)

Rank	Country	Production (Mt)	Area harvested (Mha)	Yield (t/ha)	World production (%)
1	China	191.5	29.5	6.5	28.6
2	India	137.3	42.0	3.3	20.5
3	Indonesia	60.5	12.4	4.9	9.0
4	Bangladesh	45.5	11.1	4.1	6.8
5	Vietnam	37.8	7.3	5.2	5.7
6	Myanmar	32.1	8.00	4.0	4.8
7	Thailand	31.4	10.7	2.9	4.7
8	Philippines	16.0	4.3	3.7	2.4
9	Brazil	11.7	2.8	4.2	1.7
10	Japan	10.7	1.6	6.7	1.6
Total		574.5	129.7	4.5[a]	85.8

a denotes average of rice yield among the 10 leading producing countries

According to the USDA, global rice consumed in 2014-15 is estimated to have increased by approximately 7 million tons, a 1.3% growth above the previous year. FAO also forecasts that an increase in consumption growth of 2.5% in 2014-15 (Mohanthy, 2014). However, on the supply side, over the medium and longer term, climate change seems to be the biggest threat to rice production in several countries. The consequence of this problem is also associated with the El Niño effect and is likely to result in severe drought and flooding. These phenomena lead to considerable uncertainty about the prospects of future rice production in many rice-growing countries. South and Southeast Asia are carefully optimistic on rainfall circulation that will regulate the fate of the biggest crop of the year. For example, In the case of India, the biggest rice exporter in the world, where the wet season crop accounts for more than 85% of the total crop, the southwest monsoon arrived in Kerala and quickly spread to the rest of the country in June, 2014. Apart from India, two Southeast Asian rice importers, Indonesia and the Philippines, are also bracing for weather disruptions from El Niño (Mohanthy, 2014). These examples of such weather events are likely to lead to reduce production of rice in those countries. Although, such disasters cannot be easily avoided, increasing rice yield is the greatest challenge for rice scientists to use their skills and knowledge in breeding methodologies and innovative techniques in order to guarantee that production rises sufficiently under changeable weather and thus permits sufficient consumption and allows people to avoid poverty and hunger.

1.2 Biology of rice

Oryza sativa L. is a diploid (2n = 24) and annual grass. In the tropics or suitable climatic conditions, it can survive as a perennial grass. It has hollow, round, jointed culms; flat, narrow, sessile leaf blades jointed to the leaf sheaths with collars; sickle-shaped ,well-defined, hairy auricles; small acute to acuminate or two cleft ligules and terminal panicles (Figure 1-2). Depending on the variety and the environment, the life cycle of rice cultivars ranges between 150 and 210 days from germination to maturity. The development of rice is related to temperature and day length and is divided into three phases (Figure 1-3) which are a vegetative, a reproductive and a ripening phase (Moldenhauer, et al., 2013).

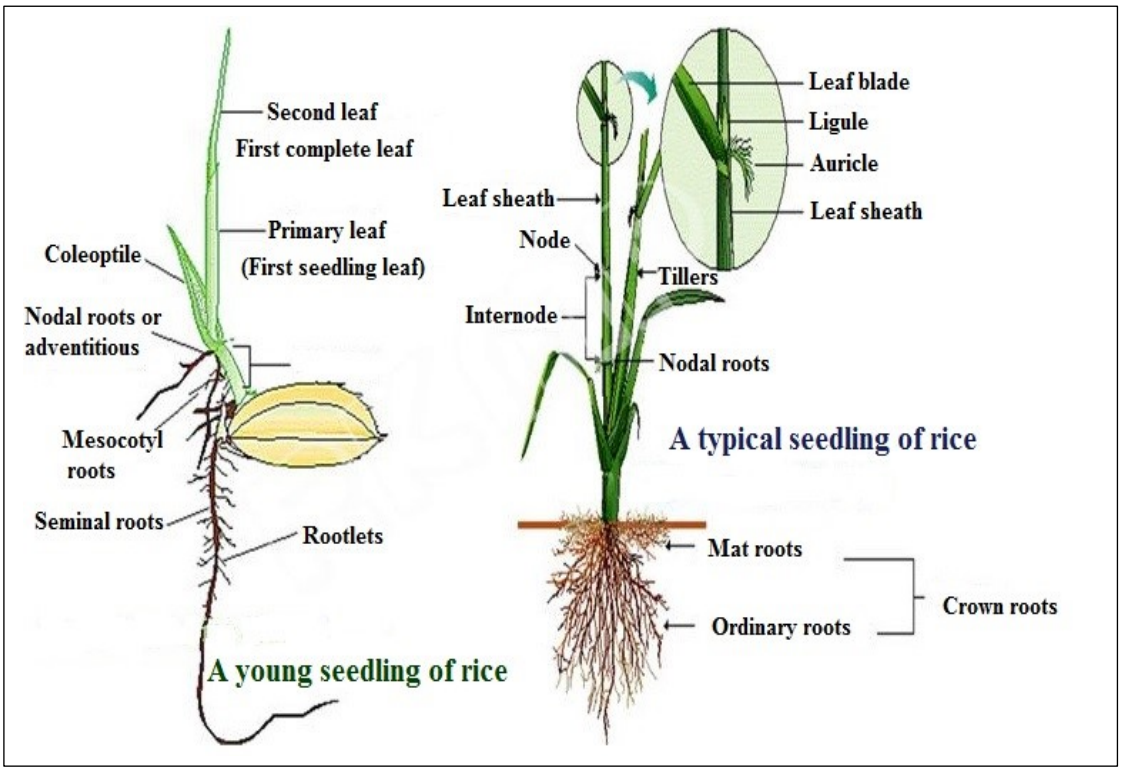


Figure 1-2 Morphology of a rice plant (derived from Rice Knowledge Management Portal (RKMP), 2011)

Vegetative phase:

This growth phase is categorised by active tillering, a regular increase in plant height and leaf emergence at regular intervals. The length of this phase primarily determines the growth duration of cultivars. The distinct steps occur during the vegetative stage comprise five stages which are:- 1) seed germination stage occurring when the seed coat has imbibed adequate water to become soft and elastic then allowing the coleorhiza (the sheath covering the radicle or embryonic primary root) to elongate slightly, emerging through the seed coat, and promoting the elongation of the coleoptile or primary leaf; 2) Seedling Emergence stage occurring when the first internode, called the mesocotyl, has elongated and pushed the tip of the rice coleoptile (epiblast) through the soil surface and followed by the emergence of the prophyll (first sheath in leaf) through the coleoptile; 3) Pre-Tillering which is the period from the development of the first- to fourth-leaf; 4) Tillering stage normally starts at the fifth-leaf stage after emergence of the first tiller from the axillary bud of the second leaf on the culm. Tillering continues when the sixth leaf emerges and the second tiller develops from the axillary bud of the third leaf; 5) Maximum tillering, which is the stage that tiller number increases in a sigmoidal-shaped curve until the maximum tiller number is reached; and 6) Vegetative Lag Phase is the period from the end of active tillering to the beginning of the reproductive phase.

Reproductive phase:

This phase is sometimes referred to as the internode elongation or jointing stage which can be characterized by five actual stages which are:- 1) Panicle Initiation, which occurs when the panicle primordia initiate the production of a panicle in the uppermost node of the culm; 2) Internode Elongation, which occurs when the panicle initiation continues until full plant height is reached and is followed by heading. The top five internodes are connected with the

final five leaves. This stage is also referred to as the jointing stage; 3) Panicle Differentiation, which is the stage of forming panicle branches. This stage is closely associated with the jointing or the internode elongation stage and is roughly equivalent to 1/2 to ¾ inch internode elongation. 4) Booting, which is the stage forming the collar on the flag leaf and also the stage during which meiosis occur. Full or late boot occurs with the appearance of the complete extension of the flag leaf; 5) Heading, which is the time when the panicle exsertion from the boot and the tip of panicle is above the collar of the flag leaf and 6) Anthesis or flowering, which generally begins upon panicle exertion or on the following day. This stage is also when one or more florets on the main stem panicle have reached anthesis.

Ripening phase:

This stage is the grain filling and ripening or maturation phase which follows ovary fertilization and is categorised by grain growth. During this period, the grain size increases in weight as the sugars and starch which have accumulated in the culms and leaf sheaths are transferred to the grain which then changes colour from green to gold or straw colour at maturity. The leaves of the plant also begin to senesce. This stage includes five steps which are:- 1) Milk stage, which is the stage of developing starch grains in the kernel. These grains are soft and the interior of the kernel is filled with a white liquid resembling milk; 2) Soft dough stage, which is the period when the starch in the grain is beginning to become firm but is still soft; 3) Hard dough stage, which is the phase including the end of grain filling and the grain-drying stage. During this stage, the whole grain is firm and almost ready for harvest and 4) Maturity stage, which is the point when the whole grain reaches approximately 20 to 22 percent moisture and is hard and ready for harvest.

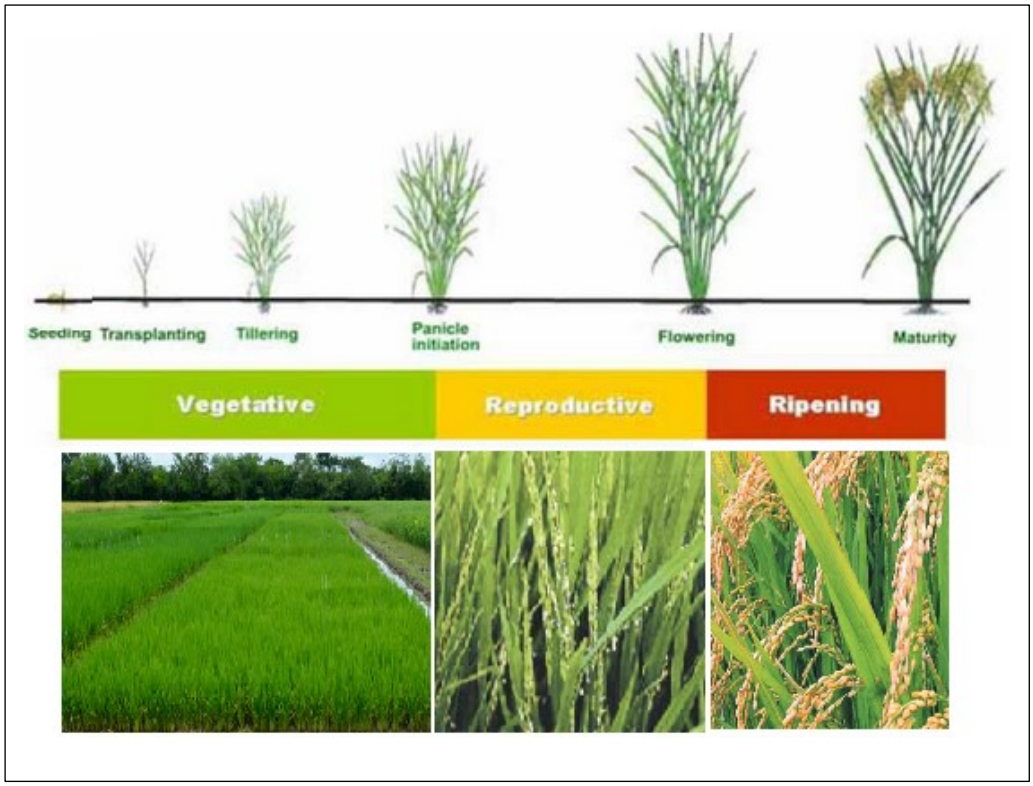


Figure 1-3 Rice plant development stages including vegetative phase, reproductive phase and ripening phase (Modified from Nathan Russell, (2014), Laborte, et al., (2012) and Tripathi, et al., (2011))

1.3 Cultivated rice specie: *Oryza sativa*

Oryza sativa is Asian rice and grown worldwide. It was first domesticated from the wild rice *Oryza rufipogon* in China between 8,200 and 13,500 years ago and then spread to South and Southeast Asia. It contains two major subspecies (GRiSP, 2013) which are indica having the non-sticky, long-grain and japonica possessing the sticky, short grain.

1.3.1 *Japonica*

Japonica rice is from northern and eastern China and grown widely in some areas of the world such as in dry fields, in temperate East Asia, upland areas of Southeast Asia and high elevations in South Asia. In addition, it is also found in the cooler regions of the subtropics and in the temperate zones. It is a relatively short plant which has narrow, dark green leaves

and medium-height tillers. Its grains are short and round, do not break easily and have low amylose content, allowing them to be sticky and moist when cooked. Its genome was the first cereal to be sequenced and assembled by whole-genome shotgun sequencing (Goff, et al., 2002) then was updated by revising and validating the minimal tiling path of clones with the optical map for rice using the Illumina Genome Analyzer II/IIx platform. The assembled sequence covers in 321 megabase genome and the prediction of genes in the assembled sequences showed its genome contains 97,751 gene transcripts with 35,679 coding genes and 55,401 non coding genes (Kawahara, et al., 2013).

1.3.2 *Indica*

Indica rice is the major type of rice grown in the tropics and subtropics, comprising the Philippines, India, Pakistan, Java, Sri Lanka, Indonesia, central and southern China, and in some African countries. Its plants are tall with broad to narrow, light green leaves and the grains are long to short, slender, flat, tend to break up more easily and have high amylose content, allowing them to be crumblier and drier than *japonica* varieties, when cooked. The assembly and annotation of the rice genome of *93-11*, a cultivar of *Oryza sativa ssp. Indica* was predicted that it contains 88,438 gene transcripts with 40,745 coding genes and 45,577 non coding genes (Zhao, et al., 2010).

1.4 Molecular aspects of genome variation in plants

1.4.1 The definition of genetic mutation and epigenetic modification

Previously, it was widely known that genetic mutation plays a significant role in regulating plant growth and development, and resulting in evolutionary change. However, there is mounting evidence that another process, namely epigenetic modification, occurs either transiently in plants in response to environmental change or can have more permanent effects

on plant morphogenesis and development. Genetic mutation and epigenetic modification are mainly distinguished by the specific changes in DNA (Figure 1-4).

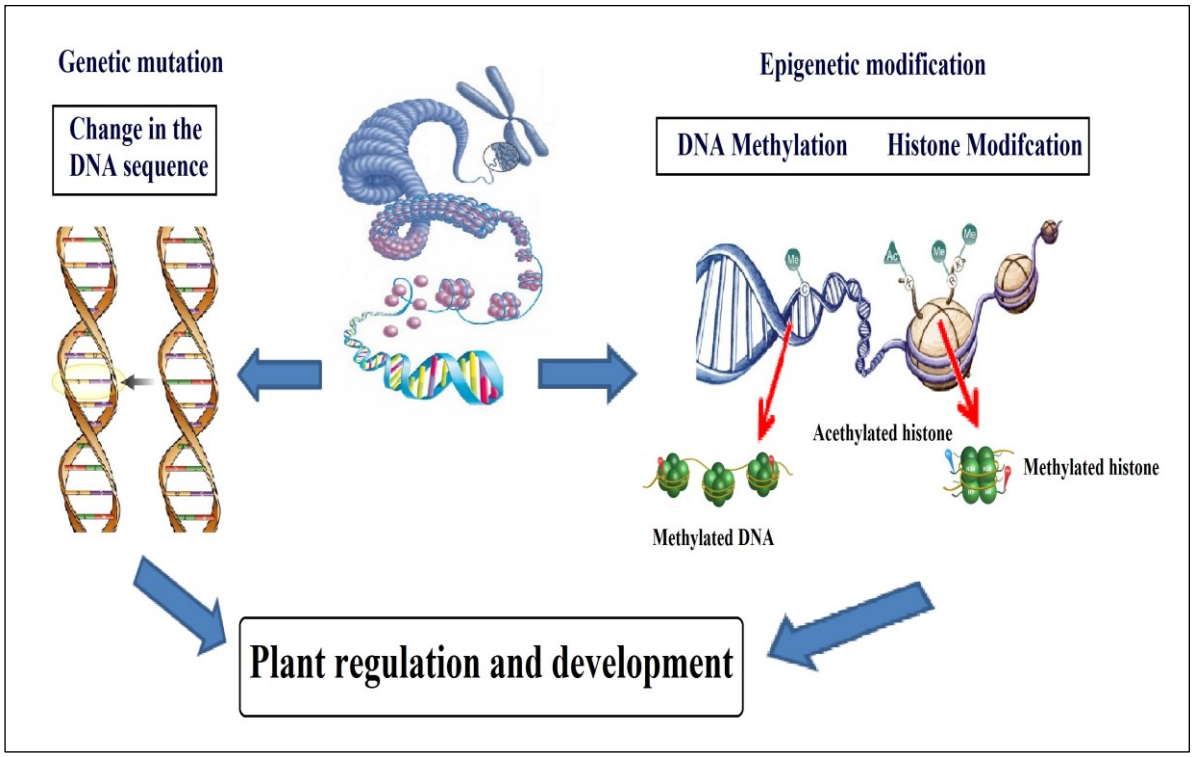


Figure 1-4 The difference between genetic mutation and epigenetic modification. Genetic mutation is the event occurring through the alteration in DNA sequences whereas epigenetic modification is the result from changing on Chromatin structure (Modified from Federuk, (2006); Gillam, (2015); Promega, (2016) and Epigenetik, (2016))

Genetic mutation can be defined as the mechanism that involves either specific base changes DNA sequences or more wide-scale changes in the structure of a gene. These alterations can occur spontaneously via deamination, depurination and by copying errors during DNA replication by slippage or shifting of the translational reading frame and can also be caused by exposure to either physical or chemical mutagens. Some mutations only affect the specific individual that carries the change, while others are inherited by subsequent generations (Loewe, 2008).

In addition to genetic mutation, Lauria and Rossi (2011) suggested that as the genome is compacted into chromatin, many mechanisms generate reversible changes in gene expression

by changing chromatin structure but not DNA sequence and these changes which may be inherited through mitotic or meiotic cell division, are called epigenetic modification.

In addition, Tammen, et al. (2012) described that although the somatic cells of eukaryotic organisms all have the same genome, the different cells within each tissue have specific characteristics. Interestingly, these differences are caused by the patterns of gene expression which are unique to each cell and are determined during cellular differentiation. These expression patterns can be affected by an organism's environment throughout its lifetime and sometimes lead to phenotypic changes. Not only cell-specific gene expression signatures but also environment mediated changes in expression patterns can be explained by a complicated system of modifications to the DNA, histone proteins and degree of DNA packaging called epigenetic marks.

1.4.2 Epigenetic modifications in plants

An increasing number of studies have shown that epigenetic modifications including DNA methylation, histone variants and modification, small RNA and positioning of nucleosomes represent important alterations in the regulation of plant growth and reproduction However in the genome of higher plants, one of the most important and widespread modifications is DNA methylation. Various aspects of plant development such as flowering time, stress response, gametogenesis, morphological change and light signalling are directly or indirectly modulated by epigenetic marks on their genomic DNA (Feng and Jacobsen, 2011). Such epigenetic modification can be generated by enzymatic activity, which itself is affected by environmental factors. As it has been mentioned above that epigenetic modification comprises many patterns, however here, we mainly focus on DNA modification.

1.4.2.1 Enzymatic involvement in epigenetic modification

Due to the complexity and diversity of methylation patterns in mammals and plants (Feng and Jacobsen, 2011), an understanding of the full range of DNA methylation requires a study of both DNA methylation and demethylation (Figure 1-5). Particularly, *Arabidopsis thaliana* has been used as a model organism to study DNA methylation in plants (Furne and Matzke, 2011).

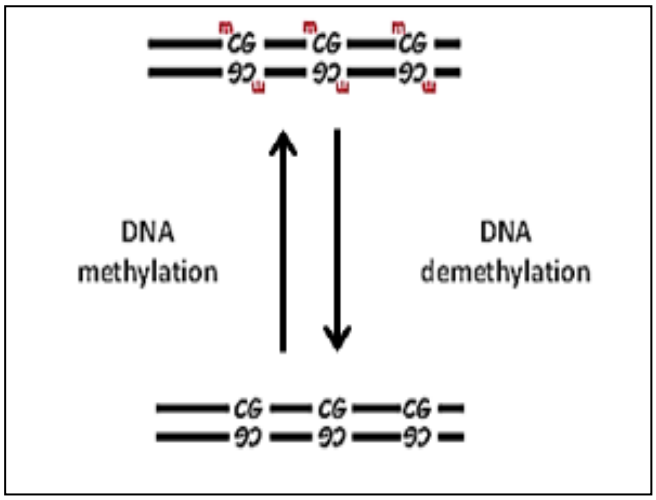


Figure 1-5 Reversible DNA methylation and demethylation (Modified from Easwaran, 2003)

1.4.2.1.1 DNA methylation

One of the most important epigenetic phenomena is DNA methylation which plays a major role in gene expression during plant development as well as in responses to environmental stimuli (Wang, et al., 2011). The pattern of DNA methylation is non-random, in that some sequence contexts or some genome regions are highly methylated and some of them are rarely methylated. Moreover, this pattern is also unstable and may change in response to environmental and developmental cues that might cause correlated changes in gene expression (Furne and Matzke, 2011). DNA methylation is not only specific to species, tissues, organelles and age but is also involved in DNA repair, gene transposition, embryonic

development and cell differentiation. It is associated with gene silencing and parental imprinting, and also controls expression of transgenes and foreign DNA in cells (Vanyushin, et al., 2011). In the plant genome, DNA cytosine methylation, which is the process of adding a methyl group to the cytosine of DNA (Figure 1-6), is widely found and it also is one of the most important epigenetic modifications which affect plant development. This view is supported by Zhang, et al., (2010) who suggested that in the nuclear genome of higher plants, cytosine bases are frequently extensively methylated and the silencing of both transposable elements (TEs) and endogenous genes is caused by such methylation.

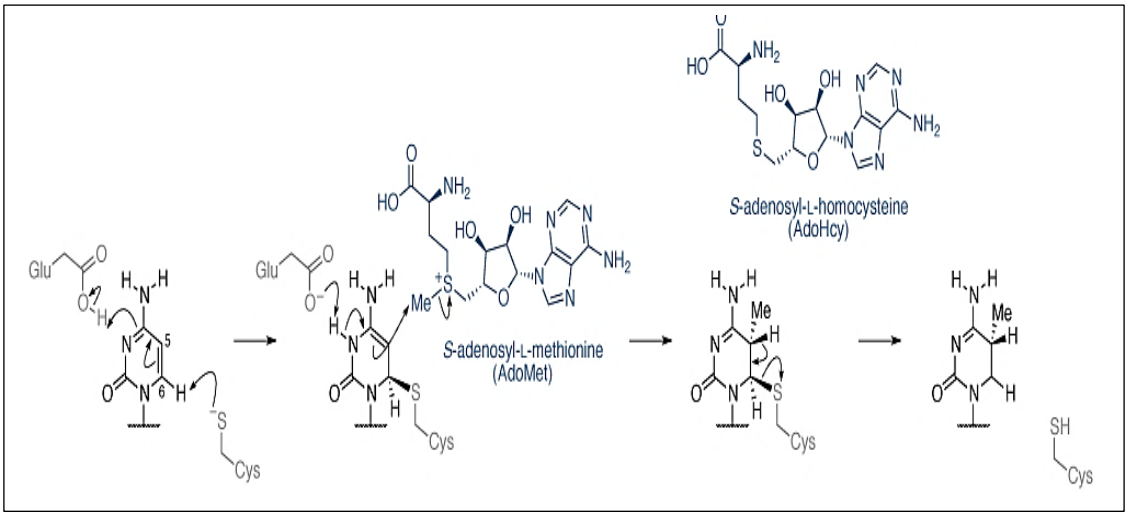


Figure 1-6 The mechanism of adding a methyl group to cytosine catalyzed by DNA methyltransferases. A strong nucleophilic cytosine is attacked by the cysteine thiolate at the C(6) atom allowing a covalent bond between the cysteine sulfur atom and the cytosine C(6) atom. The glutamate residue helps to stabilise the negative charge on cytosine. Then, the cofactor *S*-adenosyl-l-methionine (AdoMet; SAM) which contains the methyl group is taken place by nucleophilic attack converting to *S*-adenosyl-l-homocysteine (AdoHcy). Finally, β-elimination occurs through the C(5)-C(6) bond and the enzyme is releasesed. (Brown and Brown, 2016)

DNA cytosine methylation is regulated by two distinct complementary groups of enzymes, first “de novo” DNA methyltransferases which mediate the process by which unmethylated cytosine residues are methylated leading to the formation of newly methylated patterns and secondly “maintenance” DNA methyltransferases (MTases), which control the maintenance

of pre-existing cytosine methylation patterns, and are active after DNA replication (Figure 1-7) (Chen and Li, 2004; Zhang, et al., 2010).

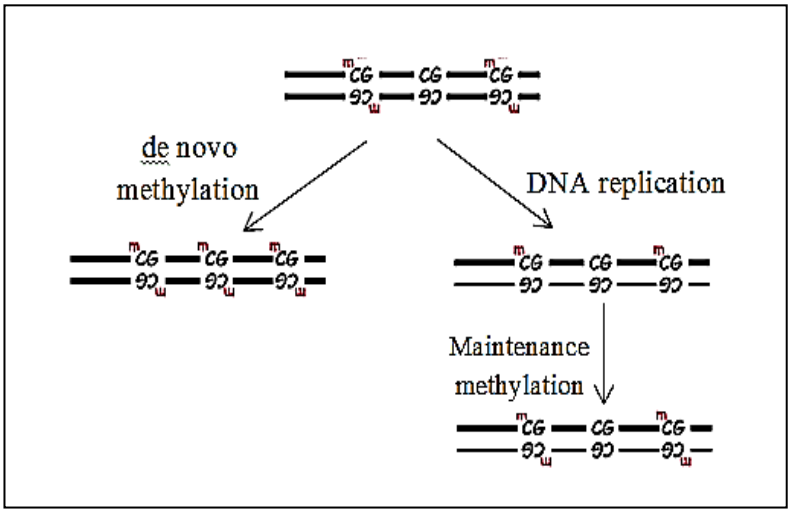


Figure 1-7 Two different types of DNA methylation (Modified from Easwaran, 2003)

Cytosines on transposable elements (TEs) and other repetitive DNA, which are mostly found in pericentromeric, heterochromic regions but also exist in small patches between genes in the euchromatic arms, are methylated frequently in the three sequence contexts CG, CHG and CHH (H can be A, G or T) (Feng and Jacobsen, 2011). Moreover, the percentages of methylation on CG, CHG and CHH sites are 24 %, 6.7 % and 1.7 %, respectively (Zhang, et al., 2010). This DNA modification is catalysed by at least three main families of DNA methyltransferases which have specificity to affect particular sequence contexts.

In *Arabidopsis*, the addition of methyl group in all sequence texts is controlled by the *de novo* methyltransferase, DOMAINS REARRANGED METHYLTRANSFERASE (DRM) activity which functions through the RNA-directed DNA methylation (RdDM) pathway (Cao and Jacobsen, 2002b; Zhang, et al., 2010). The maintenance of methylation at symmetric CG dinucleotide sites is catalysed by DNA METHYLTRANSFERASE1 (MET1). This enzyme functions with a conserved cofactor called VARIATION IN METHYLATION (VIM) which contains an SRA domain that binds methylated DNA. MET1 is a prerequisite for CG

methylation to occur, then S adenosyl homocysteine hydrolase, chromatin remodelling ATPase and binding to methylated DNA are required. Methylation of a symmetrical site of CHG is maintained by CHROMOMETHYLASES3 (CMT3) and a histone methyltransferase called KRYPTONITE (SUVH4). Methylation on the asymmetric CHH context results absolutely from the activities of the Domains Rearranged Methyltransferases (DRM1 and DRM2) and also must be affected by small RNAs to be created following DNA replication.

1.4.2.1.2 DNA demethylation

DNA demethylation patterns can be categorised into two groups (Ruiz, et al., 2006); (i) Passive DNA demethylation, which might take place due to the lack of maintenance methylation (the maintenance methytransferases are inactive) during several cycles of DNA replication resulting in a retention of the unmethylated state of the newly synthesized strand (Zhu, 2009) and (ii) active DNA demethylation, which might be present in the absence of replication (Figure 1-8). Some studies have revealed that the maternal genome of mammalian preimplantation embryos is demethylated by a passive process during cleavage stages but the parental genome is demethylated immediately after fertilization via an active mechanism. Moreover, site specific local demethylation also occurs throughout development and tissue differentiation (Ruiz, et al., 2006). Similarly, the explicit function of active DNA demethylation in plants, particular *Arabidopsis*, has been reported to play a major role in genome regulation and development such as counteraction of the activities of the RNA-directed DNA methylation pathway for prevention of the spreading of methylation from repetitive sequences to neighbouring genes, prevention of transcriptional silencing of transgenes and endogenous genes, regulation of imprinting and transposon activities, decondensation of 5S rDNA chromatin and response to environment stimuli such as biotic and abiotic stress.

Additionally, this modification contributes to many mechanisms in animals such as early development, reprogramming during gametogenesis and cloning, memory function and neurogenesis, immune response and tumorigenesis (Zhu, 2009). Besides, a well-known example of DNA demethylation is a loss of 5-methylcytosine (5-mC) in paternal DNA after fertilization of mammals whereas the erasure of the methylation mark in maternal DNA proceeds through passive DNA demethylation. Moreover, the increase of 5-hydroxymethylcytosine (5-hmC) level in the fertilised egg also correlates with the decrease of 5-mC, while the female pronucleus remains methylated and contains low levels of 5-hmC (Kriukiene, et al., 2012).

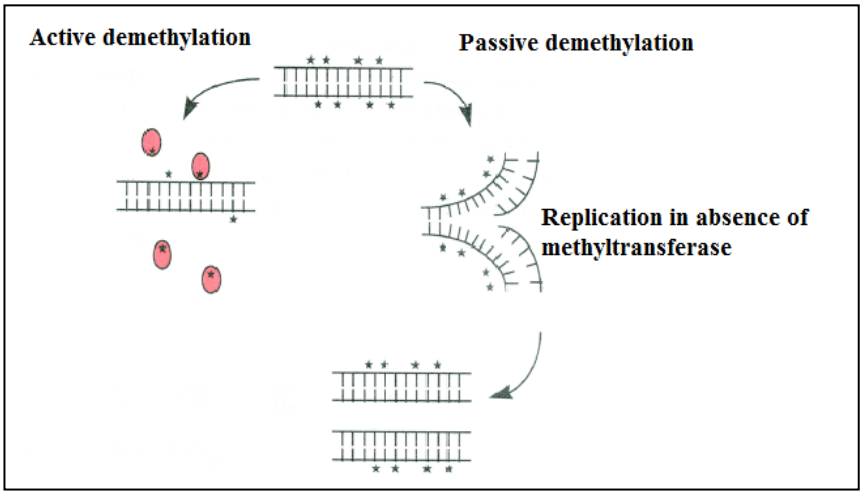


Figure 1-8 Categorisation of DNA demethylation patterns. Active demethylation occurs in the absence of replication and passive demethylation takes place due to the absence of methyltransferase during replication (Modified from Hajkova, 2003).

Demethylases have been classified as enzymes in the active demethylation pathway. Increasing evidence in *Arabidopsis* suggests that a subfamily of DNA glycosylases, which thus constitute one type of DNA demethylase, function to promote DNA demethylation through the base excision repair pathway. These specialized bifunctional DNA glycosylases (Figure 1-9) erase a 5-mC base and then cut off the DNA backbone at an abasic site; afterward a gap is induced and then an unmethylated cytosine nucleotide is added (Zhu,

2009). Furthermore, studies in *A. thaliana* provide strong evidence that Demeter (DME) and Repressor of silencing 1 (ROS1) are DNA demethylases because these proteins catalyse the removal of 5-mC through a glycosylase mechanism (Ruiz, et al., 2006). A few years ago, evidence also revealed that the mechanism of removing methyl groups from cytosine in animals also occurs through a base excision repair pathway which is initiated by two activities of 5-mC deaminase which converts 5-mC to T, and G/T mismatch DNA glycosylase which corrects the G/T mismatch (Zhu, 2009) (Figure 1-9).

To date, active DNA demethylation through base excision pathway has been well described in plants. In contrast, it has been reported that such active DNA demethylation in animals can occur through the oxidative demethylation pathway (Mohr, et al., 2011; Robertson, et al., 2012; Korlach and Turner, 2012)

In other words, recent evidence indicates that particularly in animals, the removal of such methyl groups critically depends on oxygenases. Hence, reversible epigenetic systems could only appear after accumulation of oxygen in the atmosphere. However, this oxidative modification reaction is still elusive in plants (Figure 1-9)

Despite the fact that 5-hmC, the first product of oxidised 5-mC that was discovered over 30 years ago, at that time it was not considered to be interesting and to be a rare and non-mutagenic DNA damage lesion. Since early 2009, this modification has been repeatedly identified and it has been shown that the enzyme which is responsible to catalyse the formation of 5-hmC from 5-mC by using dioxygen as a substrate is Ten-eleven translocation 1 (TET1) which is named for a common translocation in cancers as well as its translocation can occur between chromosomes 10 and 11 in acute myeloid leukemia (Lorsbach, et al., 2003). Stable levels of 5-hmC are present in specialized Purkinje neurones (Robertson, et al.,

2012), embryonic stem cells cellular development and carcinogenesis of various mammalian cells (Korlach and Turner, 2012)

Further research has also shown that TET family enzymes (TET1, TET2 and TET3) are able to catalyse the oxidation of 5-mC, leading to 5-hmC as well (Robertson, et al., 2012). A recent study by Korlach and Turner (2012) stated that the family of TET enzymes are also capable to catalyse the oxidation of 5-mC, creating two new DNA modifications namely 5-formylcytosine (5-fC) and 5-carboxycytosine (5-caC), in total bringing four chemical cytosine modifications (Figure 1-10) which not only might be intermediates in active DNA demethylation but could also be considered to be distinct epigenetic marks in their own right. More evidence about the TET family will be described in the introduction of Chapter 5.

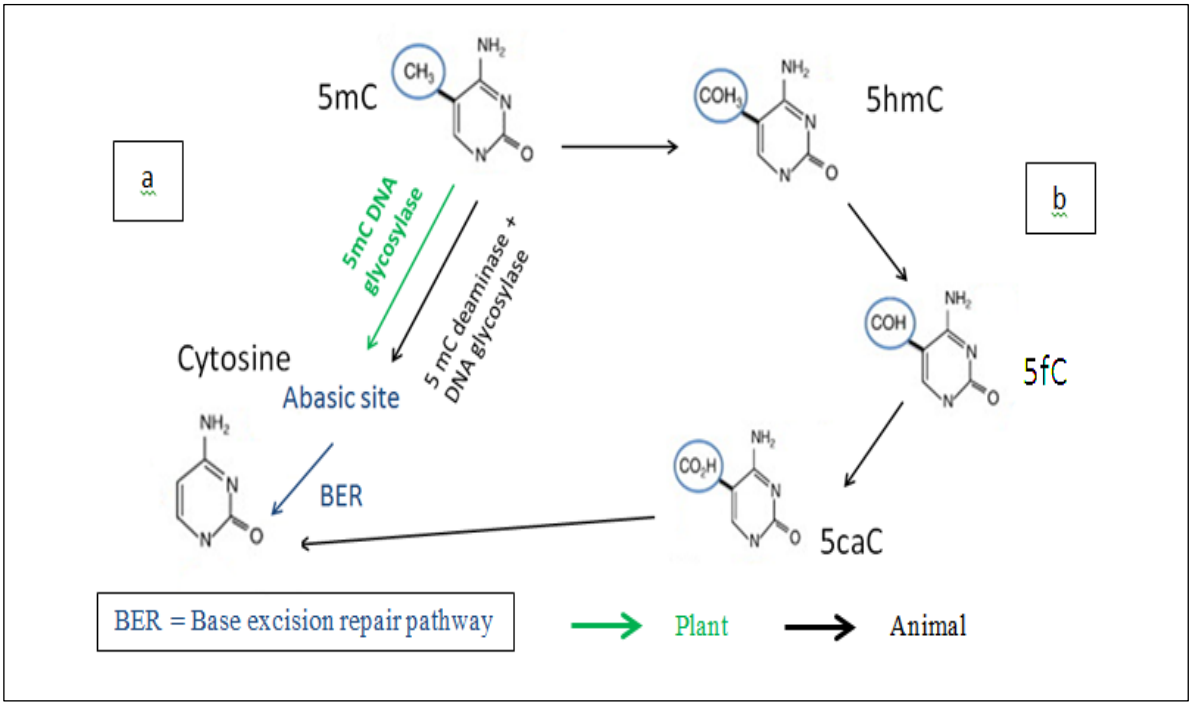


Figure 1-9 The removal of a methyl groups. (a) through base excision repair (b) through oxidation when 5-mC is hydroxylated to become 5-hmC which further can be oxidised to 5-fC and 5-caC (modified from Gong and Zhu, 2011)

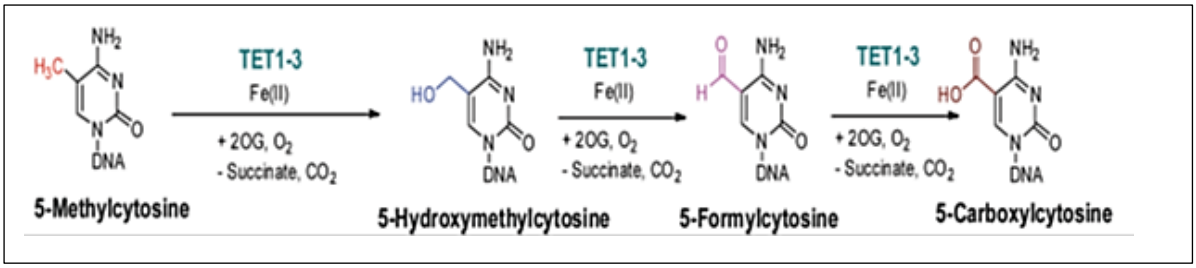


Figure 1-10 Reactions catalysed by TET family enzymes and involve four chemical cytosine modifications (Woon, et al., 2012)

1.4.2.2 Environmentally induced epigenetic modification

It is known that these epigenetic modifications in plants are mediated by both DNA methylation and stable chromatin modifications; the latter include histone tail modifications, exchange of histone variants, or nucleosome occupancy and larger chromatin configuration. Such modifications become biologically stabilised at specific stages of development and are maintained subsequently throughout the lifetime of the plant. In addition, there has been substantial interest in whether reversible DNA methylation is also triggered by environmental factors (extrinsic factors) which lead to epigenetic drift over time and consequently in the loss of DNA methylation (Figure 1-11) (Feil and Fraga, 2012)

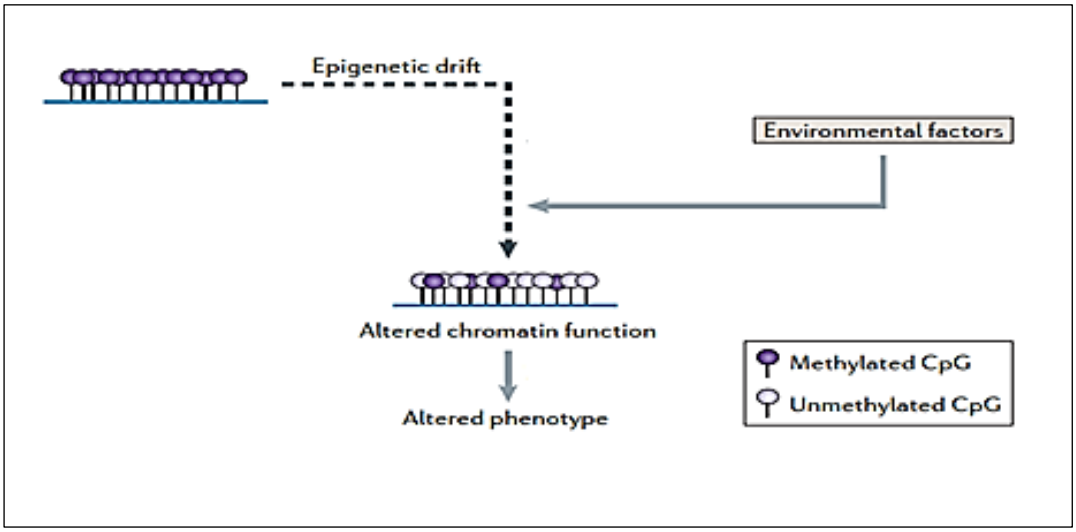


Figure 1-11 Epigenetic variations is modulated by environmental factors leading to the loss of DNA methylation (Feil and Fraga, 2012).

Recently, the effect of environment stress on epigenetic modification has been widely reported. Such stress can possibly bring about both permanent changes of gene expression and consequent adaptation that could have an evolutionary impact due to mitotic or meiotic inheritance of chromatin modifications. Stress signalling can thus cause the expression of stress-adapted genes by affecting the chromatin structure of responsive genes, either directly or indirectly (Figure 1-12**Error! Reference source not found.**) (Gutzat and Scheid, 2012).

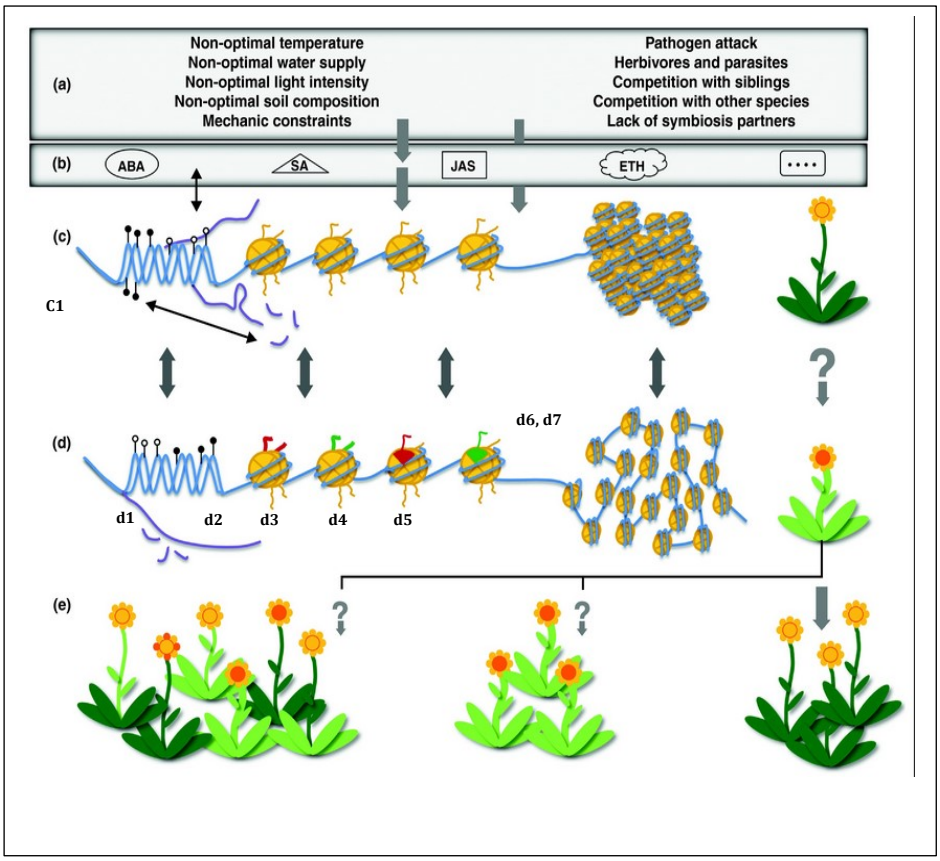


Figure 1-12 Environmental stress effects on chromatin structure. Biotic and abiotic stress conditions in plants (a). Those factors can alter gene expression with or without effects on related plant stress hormones (b). The alteration of chromatin structure through DNA methylation (c1, d1), histone tail modifications (d2, d3), histone variant replacements (d4, d5), or nucleosome loss and chromatin de-condensation (d6, d7) can be directly affected by transcription changes or stress factors (c, d). These alterations are generally reversible but under stress conditions, they can modify metabolism and/or plant morphology. Typically, the new phenotypes are not transferred to progeny. However, chromatin-associated changes are potentially inherited to progeny and might lead to permanent maintenance of new characteristics, and epigenetic diversity (e) (Gutzat and Scheid, 2012).

Seemingly, there are several studies revealing the mechanism of plant responses to abiotic stress as extreme temperatures, drought, flood, high salt and oxygen tension. However, these revelations only offer an explanation in term of physical adaptation for plant survival. Fascinatingly, many researchers have also found that such stress resistance of plants, plant adaptation and plant phenotypic variation are correlated with the altered molecular mechanisms which mostly involve changes in DNA methylation.

1.4.2.2.1 Association of abiotic stress and DNA methylation changes

In several studies, the environment induction initiates the changes in gene expression which are associated with the altered either DNA methylation pattern or histone modifications. Presently, environmental epigenetics are mainly focused on DNA methylation because this methylation is essential in plant development and is involved in many processes such as genomic imprinting and silencing transposable element.

The environmental signals can cause epigenetic transitions in individual plants subsequently leading to the stable alteration in DNA methylation. Besides, such environmental factors also have directly induced epigenetic phenotypes in plants. To date, several reports have revealed that there are many natural examples of stable alterations in DNA methylation that confer specific phenotypes. For instance, heritable hypermethylation at the Colourless non-ripening (Cnr) locus induces repression resulting in non-ripening of the fruit (Manning, et al., 2006). Similarly, levels of DNA methylation at the CYCLOIDEA locus, which determine adaxial (dorsal) flower identity in the bilaterally symmetric flowers of *Antirrhinum majus* (snapdragon) and comprise a retrotransposon, have heritable influences on flower symmetry in the toadflax *Linaria vulgaris* (Feil & Fraga, 2012). Moreover, flower morphology of the wild potato *Solanum ruiz-lealii* seems to be linked to differential DNA methylation and also in melon where the transition from male to female flowers is controlled by the DNA

methylation status of a transposon at the CmWIP1 transcription factor locus (Martin, et al., 2009). However in these examples, it is unclear whether DNA methylation states are influenced by the environment. Hence, the comparison of natural populations in different environments is the one way to address this important issue. For examples, cytosine methylation levels at multiple gene loci control the phenotype of selected wild populations of *Viola cazorlensis* (Herrera and Bazaga, 2010). In addition, differential DNA methylation between populations growing in different microenvironments was also reported in mangrove trees (Lira-Medeiros, et al., 2010) and in orchid species (Paun, et al., 2010).

In many wild plants, it has been challenging to verify that phenotypic variation is indeed initiated by differential epigenetic states rather than by being caused by genetic mutation. This epigenetic status allows plant to have an alternative approach for directly responding to any specific stresses. To understand the environmentally induced changes in DNA methylation, the associated changes in plant phenotype have been widely studied.

In one intriguing study on asexually reproduced dandelions (*Taraxacum officinale* Weber ex Wigg), plants exposed to two different abiotic treatments, namely high salt and low nutrients, demonstrated considerable changes in methylation throughout the genome. Consequently, this alteration was frequently transmitted to the progeny of the stressed plants (Verhoeven, et al., 2010). In addition, it was reported that cold-treatment leads to a reduction of 5-mC in the genome tobacco cells in culture (Burn, et al., 1993), Arabidopsis seedlings (Finnegan, et al., 1998), and maize root tissues (Steward, et al., 2002).

1.4.2.2.2 Association of gene expression and DNA methylation changes

Gene expression can be regulated by several causes. DNA methylation is one feature that is important for defining cellular identities and coordinating organism-wide developmental

programs in many organisms. In plants, variation of DNA methylation in response to environmental conditions signifies a potentially forceful mechanism to regulate gene expression networks. Biotic stress is also the factor that can trigger the alteration of DNA methylation leading to enhance or suppress gene expression. For instance, an investigation of DNA methylation in response to biotic stress by examining the DNA methylomes in leaf tissue from populations of plants which were exposed to a *Pseudomonas syringae* pv. *Tomato* DC3000 (*Pst*), a virulent bacteria or salicylic acid (SA) hormone at 5 days post infection (dpi) were conducted. The result showed that many of numerous stress-induced differentially methylated regions were closely associated with differentially expressed genes. Particularly, SA stress can regulate expression of adjacent genes by dynamic changes in DNA methylation within repetitive sequences or transposons (Dowena, et al., 2012). Furthermore, cytosine methylation in *A. thaliana* attacked by *Pst* DC3000 was monitored by isolating and characterising plant genomic fragments; reduced methylation was detected by high-performance liquid chromatography assays. The result showed that *Pst* induced hypomethylation occurred in the absence of DNA replication, suggesting that it involves an active demethylation mechanism (Pavet, et al., 2006). Lastly, Wada, et al. (2004) reported that tobacco mosaic virus infection of tobacco generates hypomethylation of genomic regions such as the pathogen-responsive *NtAlix1* gene, which becomes concomitantly activated.

1.4.2.3 Chemically induced epigenetic modification

Similarly, studies of chemical factors affecting plant DNA methylation have been reported. For example, it was shown that one line of the progeny of rice seedlings, which were treated with a chemical affecting DNA methylation (azacytidine, an inhibitor of DNA methyltransferase) and then used to initiate lines for further propagation over several generations, presented acquisition of resistance to a bacterial pathogen (*Xanthomonas oryzae*) that correlated with the hypomethylation of a resistance gene promoter (Akimoto, et al.,

2007). Another study on chemical inhibition of DNA methylation was conducted by applying zebularine, a more stable cytidine analogue with a similar mode of action that is successfully used as a methylation inhibitor in *Neurospora* and mammalian tumour cell lines, to Arabidopsis seed, which were grown in solid medium under tissue culture condition for 21 d and were transferred to drug-free growth medium after 14 or 21 d for recovery. Then a range of methodologies were used to assess the level of DNA methylation from both treated and untreated plants. The result showed that zebularine can significantly reduce DNA methylation in plants in a dose-dependent and transient manner independent of sequence context. This finding suggests that zebularine represents a promising new and versatile tool for further investigating the role of DNA methylation in plants with regard to transcriptional control, maintenance and formation of (hetero-) chromatin (Baubec, et al., 2009).

1.4.3 The crucial role of 5-mC and its oxidative products in plants

Regarding the vital role of cytosine modification in plants, 5-mC is well established in epigenetic regulation of gene expression. In particular, it has been shown to be involved in genome defence system, protecting the genome against both endogenous selfish DNA elements (predominantly transposable elements or TEs) and exogenous virus invasion (Yoder, et al., 1997; Zilberman, 2008; Zhang, et al., 2010). In addition, 5-mC is also implicated in controlling gene expression throughout plant development and biotic and abiotic stress encounters (Bird, 2002; Zhang and Bruice, 2006; Zilberman, et al., 2007; Zhang, et al., 2010). For instance, two biological processes in plants, vernalisation and genomic imprinting, are found to depend on epigenetic regulation at the transcriptional level because the mechanism involves 5-mC, which is able to attract methyl–binding proteins to form a complex which interrupts the binding of transcription factors (Fransz and de Jong, 2002; Zhang, et al., 2010).

In addition, some studies have shown that 5-hmC is involved in various DNA responses in animals, serves as an intermediate in DNA demethylation (Robertson, et al., 2012; Yu, et al., 2012), has a dual function in transcription (Robertson, et al., 2012) and could reduce binding of certain methyl CpG-binding proteins (Hashimoto, et al., 2012; Yu, et al., 2012). Similarly, both 5-mC and 5-hmC also play a crucial role in the maintenance of cellular identity (Kellinger, et al., 2012).

Although 5-fC and 5-caC are suggested to be part of the cytosine demethylation pathway catalysed by TET dioxygenases in animals, the precise role of 5-fC and 5-caC remains poorly understood. Interestingly, research has been conducted on the effect of these four cytosine modifications on the rate of Pol ll nucleotide incorporation and substrate specificity for DNA template having a site–specific C, 5-mC, 5-hmC, 5-fC and 5-caC. These studies have revealed that the pol ll polymerisation rate and specific constant for GTP incorporation of 5-fC and 5-caC are considerably decreased whereas these variables essentially do not change for 5-mC and 5-hmC templates indicating that 5-fC and 5-caC play a role in RNA polymerase ll transcription (Kellinger, et al., 2012).

While these forms of cytosine modifications and their functions have been well described in mammalian cells, there is no such previous evidence in plants. Fascinatingly, a recent study in *A. thaliana* has been revealed that 5-hmC is present in leaf and flower and it is recognised by a VIM-a methyl cytosine binding protein that plays a major role in maintaining CG methylation (Yao, et al., 2012).

1.4.4 The regulation of gene expression

The determination of a cell's structure and function, and the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism is specified by gene regulation which can be regulated at many of steps in the pathway leading from DNA to

RNA and to protein (Martinez and Walhout, 2009). Therefore, a cell is able to regulate the production of proteins by controlling many processes (1) the starting and duration of transcription from a specified gene (2) the splicing or processing of RNA transcript (3) the selection of mRNAs to export from the nucleus to the cytosol (4) the selective degradation of each mRNA molecule (5) the selection of mRNAs which are translated by ribosomes (Tomilin, 2008) and (6) the activity of protein by selecting activating or inactivating proteins after they have been made (Figure 1-13). However, in most genes, the significant site of control is transcription (step number 1 in Figure 1-13) because this is the step at which it can be confirmed that no unnecessary intermediates are synthesized, or in other words, it would mediate any downstream effect which might be caused by incorrect transcription resulting in an undesirable phenotype or death of the plant.

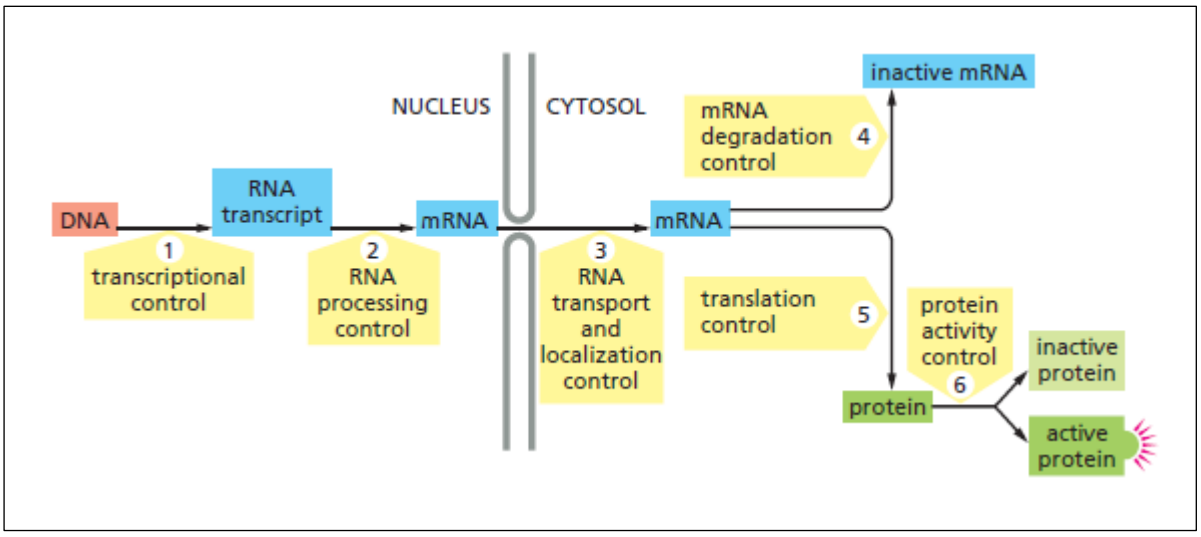


Figure 1-13 The regulation of gene expression at several steps. (Alberts, et al., 2008)

1.5 The Objectives

Several studies have revealed factors that affect the alteration of epigenetic modification. This study mainly focuses on two factors which are enzymatic factors and environmental stress factors that may cause the alteration of this modification.

Recently, there new epigenetic marks have been shown to be present in some plants such as 5-hmC which was detected from leaves and flower of *Arabidopsis thaliana* by the Yao group (2012) and 5-caC which was measured in leaves and panicles of rice and other species (Wang, et al., 2015; Ketsuwan, 2013). However, unlike the situation in animals the mechanism of this modification in plants is still unclear. This finding prompted an attempt to identify equivalent plant proteins which might have the ability to lose the imbalance of DNA methylation and demethylation by converting 5-mC to 5-hmC and its oxidative products. Additionally, another aspect of this study examine the effect of specific environmental stress on genes encoding 2ODD, possible *TET* gene homologues, which may also cause the epigenetic modification change or might relate to other biological mechanism.

This study identified *TET* gene homologues and their expression in order to generate a basic understanding and consequent investigations examined functional links to plant development and responses to stress condition.

The specific objectives of this research were as follow:

Specific objective 1: Studying the expression of possible *TET* gene homologues

- To identify the *TET* gene homologues which have conserved domains of 2ODD
- To study the gene expression during vegetative and reproductive development

Specific objective 2: Studying the expression of 2ODD gene (selected possible *TET* gene homologues) and genes relating with flooding tolerance (*SuB1A*, *SLR1*)

- To confirm the identity of tolerant and intolerant varieties

- To study the expression of selected 2ODD genes, together with *Sub1A* and *SLR1*, in tolerant and intolerant varieties under flooding and drought condition

Chapter 2
General materials and methods

2.1 Computational investigation

2.1.1 Conserved domain analysis

Conserved domains identification was performed by using both Conserved Domain Search Service tool (CD-search tool) in the NCBI's conserved domain database (CDD), and macromolecular structure viewer tool (Cn3D tool) which is a helper application to view 3-dimensional structures from NCBI's Entrez Structure database.

2.1.2 Identification of regions of similarity between biological sequences

The regions of local similarity between sequences were identified by using Basic Local Alignment Search Tool (BLAST) which is the program that can be operated by comparing or aligning nucleotide or protein query sequences against those proteins present in a selected target database and calculating the statistical significance of any matches.

2.1.2.1 Searching protein similarity by BlastP

Protein query sequences were used in a BlastP search against proteins both in the global Non-redundant (nr) protein sequence database and restricted to the organism of interest through http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Any specific sequence identified by this search was examined by focusing on particular conserved domains.

2.1.2.2 Examination of the activity of each gene by identifying transcripts through TblastN

The protein sequence identified from the BlastP algorithm was used as query search against the Expressed Sequence Tags (EST) database by using TblastN algorithm (the tool for searching translated nucleotide using a protein query) via http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=tblastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome. Those ESTs showing a perfect match with specific regions of the gene were selected.

2.1.2.3 Alignment of sequences

Protein or nucleotide sequences were aligned by using ClustalW and CLC Sequence Viewer 6.6.2 program for investigating a consistently conserved domain.

2.1.1 Gene expression analysis

Gene expression was examined using through Genevestigator software (Hruz, et al., 2008) which is an innovative online platform helping to interpret gene expression results, identify correlated genes, and explore the world's gene expression data available from the Rice electronic Fluorescent Pictographic (eFP) browser tool, which creates electronic fluorescent pictographic representation of gene expression patterns and can be accessed via http://bar.utoronto.ca/efprice/cgi-bin/efpWeb.cgi.

2.2 Cultivation of rice plants

2.2.1 General plant materials and conditions

Fertile wild type seeds of flooding-intolerant IR64 containing *Sub1A-2* and flooding-tolerant IR64.S1 containing *Sub1A-1* of *Oryza sativa* Indica were obtained from the International

Rice Research Institute (IRRI), Los Baños, the Philippines. Two varieties of *Oryza sativa* Japonica including Gleva and Nipponbare varieties were obtained from the Institute for Food and Agricultural Research and Technology (IRTA), Barcelona, Spain and IRRI, respectively. Both of them are flooding-intolerant varieties and utilised as susceptible controls. All plants were grown in controlled environment conditions in The University of Reading growth chamber (Fitotron plant growth chamber, Weiss Gallenkamp, UK). The photosynthetic photon flux density at the base of the cabinet was maintained at 350-560 µmol $m^{-2}s^{-1}$ using cool white fluorescent tubes. The photoperiod was 15 h, the humidity ranged from 75-80% and the temperature was maintained at 30°C at noon and 25°C at night. Plants were pre-germinated on wet paper in petri dishes for 7 d and transferred to a pot (15 cm. diam.). After transfer to soil, young leaves and young roots were collected when the seedlings were 14 d of age, 3-4 cm sections of panicles and the mature leaves were harvested from plants at 45 d of age. All samples were stored at -80°C until use.

2.3 Laboratory protocols

2.3.1 Nucleic acid extraction

2.3.1.1 Isolation of plant genomic DNA

Genomic DNA was extracted using the DNeasy® Plant Mini Kit (Qiagen, UK) following the manufacturer's instructions. Briefly, samples (approximately 100 mg) were ground to a fine powder under liquid nitrogen using a pre-cooled mortar and pestle and suspended in 400 µl of AP1 Buffer and 4 µl of RNase A stock solution (100 mg/ml) (provided with the kit). After 10 min of incubation at 65°C on a shaking incubation for mixing, 130 µl of AP2 buffer (supplied with the kit) was added to the lysate and incubated on ice for 5 min. After that, the lysate was centrifuged at 13,000 rpm for 5 min and the resultant lysate was transferred into a QIAshredder Mini spin column (provided with the kit) for centrifugation at 13,000 rpm for 2

min. The AP3/E buffer was directly added onto cleared lysate and transferred into a DNeasy mini spin column sitting in a 2 ml collection tube (supplied with the kit) and centrifuged for 1 min at 8000 rpm. After discarding the flow-through, 500 µl of AW1 buffer was added into a DNeasy mini spin column for washing and then centrifuged for 2 min at maximum speed to dry the membrane. Finally, 50-100 µl of AE buffer was directly added onto the DNeasy membrane and then incubated for 5 min at room temperature (20-25ºC) before centrifugation for 1 min at 8000 rpm to elute DNA which was subsequently used or stored at -20ºC.

2.3.1.2 Isolation of total RNA

Total RNA was extracted using a RNeasy® Plant Mini Kit (Qiagen, UK) according to manufacturer's protocol. In brief, samples ground in liquid nitrogen to fine powder and 450 µl of RLT buffer (provided with the kit) containing 4.5 µl β-mercaptoethanal were added to a 1.5 ml Eppendorf tube and homogenised immediately by vortexing vigorously. Consequently, the tissue sample was incubated at 56ºC for 3 min and the lysate was transferred to to a QIAshredder spin column placed in a 2 ml collection tube (provided with the kit) and centrifuged for 2 min at 23,000 rpm. The supernatant of the flow-through was carefully transferred to a new micro centrifuge tube without disturbing the cell-debris pellet in the collection tube and immediately mixed with 225 µl of 100% ethanol by pipetting. The sample was transferred to an RNeasy spin column (supplied with the kit) placed in a 2 ml collection tube and any flow-through was discarded. 350 µl of RW1 Buffer was added to the RNeasy spin column and then centrifuged for 15 s at 13,000 rpm to wash the spin column membrane and the flow-through was discarded. In order to first remove gDNA, 70 µl RDD Buffer (supplied with the RNase-Free DNase Set) mixed gently with 10 µl DNase I stock solution by inverting the tube was added directly to the RNeasy spin column membrane, and then placed on the bench top (20–30°C) for 15 min. Subsequently, 350 µl of RW1 buffer was added to repeat washing the RNeasy spin column and centrifuged for 15 s 13,000. The

RNeasy spin column was then washed twice by adding 500 µl of RPE buffer together with centrifugation at 13,000 rpm for 2 min. 30–50 µl RNase-free water was directly added to the spin column membrane and centrifuged for 1 min at 13,000 rpm to elute the RNA. In addition, gDNA, the DNase and divalent cations from the RNA preparation sample were subsequently removed by using TURBO DNA-free™ Kit. The 50 µl reaction mix including 43 µl of RNA sample, 2 µl of DNase and 5 µl of 10 x buffers was mixed gently and incubated at 37°C for 30 min. The 10 µl of resuspended DNase inactivation reagent was added and then incubated at room temperature (20–30°C) for 5 min along with occasional mixing. Finally, the RNA sample was transferred to a fresh tube after centrifugation at 13,000 rpm for 1.5 min and subsequently used or stored at -80ºC.

2.3.1.3 DNA and RNA quantification

The quality and quantity of extracted genomic DNA and RNA were determined using a Nanodrop^{TM}2000C spectrophotometer (Thermo Scientific, USA) which is able to measure DNA, RNA (A260) and concentrations and sample purity (260/280 ratio and 260/230 nm). Additionally, gel electrophoresis was used to further confirm the quality of eluted DNA and RNA.

2.3.2 Polymerase chain reaction (PCR) amplification

2.3.2.1 First-strand cDNA synthesis

- **By a SuperScript®III First Strand Synthesis SuperMix for qRT-PCR kit (Invitrogen, UK)**

First-strand cDNA was synthesised following the manufacture's protocol. In brief, reaction components given in Table 2-1 were combined in a 0.2 ml Eppendorf tube on ice.

Table 2-1 Reaction components for first stand cDNA synthesis

Components	**Volume (µl)**
2X RT Reaction Mix	10
RT Enzyme Mix	2
DEPC-treated water	to 20
RNA (up to 1 µg)*	x
Total per reaction	**20**

* Based on total RNA concentration, 300-500 µg of total RNA was used to construct cDNA

After mixing all reaction components, the tube was incubated at 25ºC for 10 min and at 50ºC for 30 min, respectively. Finally, the reaction was terminated by placing the tube at 85ºC for 5 min, and then chilled on ice. Afterwards, 1 µl (2 U) of *E. coli* RNase H was added into the reaction mix and incubated at 37ºC for another 20 min.

- **By High Capacity RNA-to-cDNA kit (Applied Biosystem, UK).**

Based on manufacture's protocol, all components from Table 2-2 were mixed to prepare both +RT reaction and –RT reaction mix.

Table 2-2 Reaction components for first stand cDNA synthesis

Components	Component Volume / Reaction (µl)	
	+RT reaction	**-RT reaction**
2X RT Buffer	10.0	10.0
20X Enzyme Mix	1.0	-
RNA Sample*	up to 9 µL	up to 9 µL
Nuclease-free H2O	Q.S.# to 20 µL	Q.S.# to 20 µL
Total per reaction	**20**	**20**

* Use up to 2 µg of total RNA per 20 µL reaction # Quantity sufficient

Reaction tubes were incubated at 37ºC for 60 min and the reaction was terminated by heating to 95ºC for 5 min and held at 4ºC. The cDNA was subsequently used or stored at -20ºC until required.

2.3.2.2 PCR of genomic DNA and reverse transcription PCR (RT-PCR) of cDNA

The PCR amplification were performed in a volume of 20 µL, comprising 10 µL of 2x Biomix PCR master mix (Bioline, UK) 0.5 µL of 0.25 µM forward and reverse primer (Invitrogen, UK), 1.0 µL of genomic DNA or cDNA and 8 µL of nuclease free water. The amplification was carried out in a GeneAmp PCR system (Applied Biosystem) using the following programme: 2 min at 94ºC; followed by 30 cycles of 30 s at 94ºC, 30 s at 55ºC, 30 s at 72ºC, 7 min at 72ºC for the final elongation and finally at 4ºC for keeping.

2.3.2.3 Semi-quantitative Real time PCR (semi-qPCR)

Real time quantitative PCR was performed using two different kits and machines as described below.

- **Using Rotor Gene™ 6000 (Corbett Life Science, UK) with SensiMix™ SYBR®No-ROX**

The reaction mix composition was prepared following the components as shown below

2x SensiMix™ SYBR®No-ROX	10 µl
Mixed forward and reverse primer (10 µM)	1.0 µl
cDNA template or 10 fold serial dilutions of cDNA*	1.0 µl
Nuclease free water	8.0 µl
Total per reaction	20.0 µl

* 10 dilutions were used to establish a relative standard curve

cDNA and 10 fold serial dilutions of cDNA were used to perform real time semi-quantitative PCR according these cycling condition:1 cycle of denaturation at 95ºC for 10 min, followed by 40 three-segment cycles of amplification (15 s at 95ºC, 15 s at 55ºC and 30 s at 72ºC) where the fluorescence was automatically measured during PCR and one three-segment cycle of product melting (ramp from 67-95 ºC by rising by 1 degree each step, at the first step, hold for 90 s and next step afterward, hold for 5 s). Output data were processed using the software provided.

- **Using Applied Biosystems StepOne™ (Life technology, UK) with 2x qPCRBIO SyGreen Mix Hi-ROX kit**

The reaction mix composition was prepared following the components as shown below

2x qPCRBIO SyGreen Mi	10.0 µl
Mixed forward and reverse primers (10 µM)	1.6 µl
cDNA template or five-fold serial dilutions of cDNA*	4.0 µl
Nuclease free water	4.4 µl
Total per reaction	20.0 µl

* Five dilutions were used to establish a relative standard curve

The mixed reaction components were used to conduct real time PCR amplification following three stages which are holding stage comprising one cycle of denaturation at 95ºC for 2 min, cycling stage including 40 three-segment cycles of amplification (15 s at 95ºC, 15 s at 59ºC) and melt curve stage which temperature would ramp from 67-95 ºC by rising for 1 degree each step; at the first step, hold for 90 s and the next step afterward, hold for 5 s. Output data were processed using the software provided with this machine.

2.3.2.4 Agarose gel electrophoresis

The amplicon size of each amplified PCR product was confirmed using agarose gel electrophoresis. 1% gel was prepared by adding 1 g agarose into100 mL 1x Tris-Acetate-EDTA (1x TAE) and microwaving until completely melted. It was allowed to cool and then 2 µl ethidium bromide was added for staining. The gel was poured into the tray and the comb was inserted to make wells. It was left on the bench till hard enough to be used. In order to check the quality and amplicon size of amplified products, the PCR product was used as a sample loading dye which is prepared by mixing 3 µl of it with 1 µl dye (40% (w/v) sucrose,

0.025 % (w/v) bromophenol blue) and then loading into the horizontal wells of the gel. Additionally, HyperLadder IV, a ready-to-use molecular weight marker (Bioline, UK) was also loaded. Both samples and HyperLadder IV were run at a voltage of 118 V for 45-50 min. Next, it was visualised on a UV illuminator (320 nm) (SynGene, UK) for observing the RNA bands which determine the purity of RNA. Lastly, the gel pictures were captured using the GeneSnap 6.00.19 system (SynGene, UK).

2.3.3 Sequencing of PCR products

After confirming the quality of un-purified PCR products by gel electrophoresis, PCR products were sent to be purified and sequenced at a commercial sequencing service (Source Bioscience, UK). Nucleotide sequences of all amplified product were determined by sequencing on both forward and reverse strands with our primer pairs.

It is necessary to proofread raw sequence data obtained from sequencing using sequence editing programs to obtain an accurate conclusions and higher quality of validation results, before commencing any DNA sequence data analysis. Basically, the quality read length ($Q \geq 20$) of each raw nucleotide sequence were generated by manual editing of sequences using freeware trace editor, FinchTV version 1.4.0 software (Geospiza, Inc., Seattle, WA) . Hence, the forward and reverse nucleotide sequence chromatograms provided by the Macrogen sequencing service for the submitted samples were visualised and edited manually using FinchTV software. Generally, quality read lengths begin around nucleotide base 20 with a clear distinct peak having little overlap and presenting a Q value greater than or equal to 20 ($Q \geq 20$). Therefore, unreliable bases at the beginning and the end of the each sequence were removed as shown in Figure 2-1

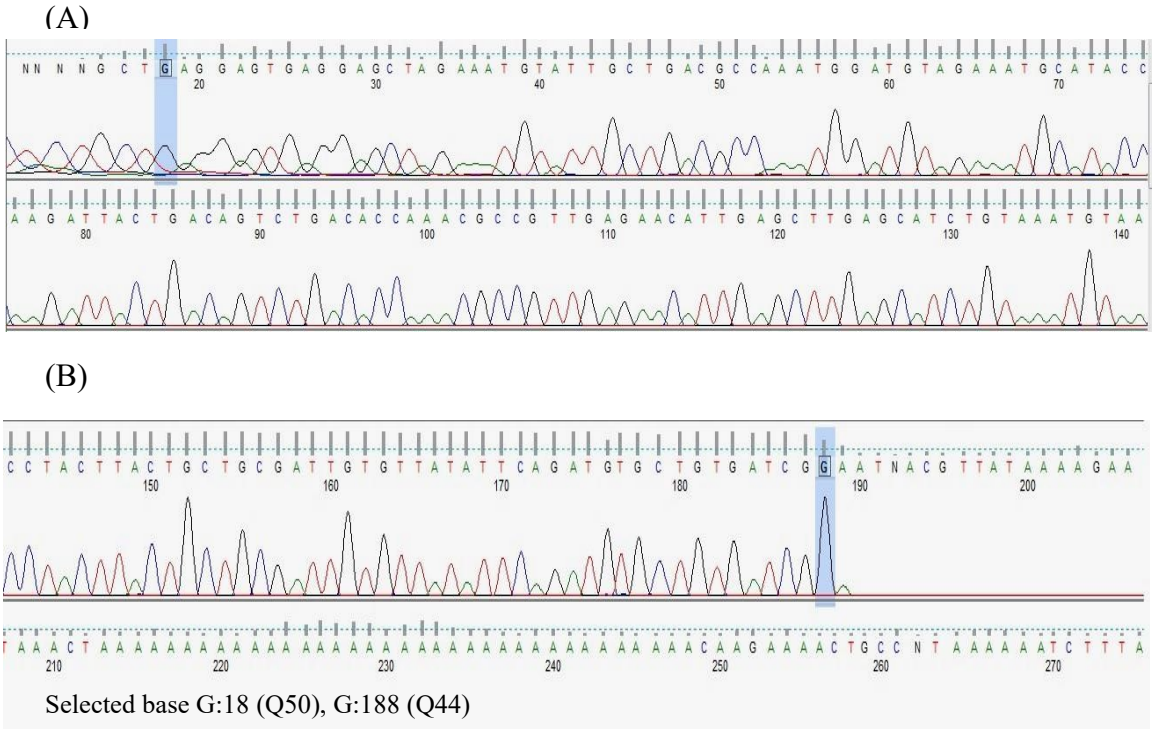


Figure 2-1 Raw sequence data in a chromatogram file. (A) The beginning of the sequence data; (B) The end of sequence data; black double headed arrows indicate the region trimmed from the beginning and the end of the raw sequence data using FinchTV 1.4.0 software

Thereafter, reverse sequences were generated to reverse complement sequence and aligned with forward sequences to produce a high quality full length nucleotide sequence for each amplified product using BioEdit version 7.0.9.0 software. Moreover, obtained nucleotide sequences were Blasted against the nonredundant (nr) database using the Blastn algorithm in order to ensure that these amplicons were amplified from the precise gene and also were generated from RNA.

2.3.4 Restriction analysis

The specific recognition site for a typical restriction endonuclease enzyme at the single nucleotide polymorphism (SNP) site of each nucleotide sequence was obtained by using DNASTAR software (DNASTAR, Inc., Wisconsin, USA). Thereafter, the selected endonuclease enzyme was used to cut the PCR product following manufacture's protocol. In

brief, before processing digestion step, BseNI was inactivated by incubation at 80°C for 20 min. Then, 20 µL of digestion reaction containing 10 µL of PCR reaction mixture, 18 µL of nuclease-free water, 1 µL of BseNI and 2 µL of 10X Buffer B (provided with BseNI enzyme) were mixed gently and incubated at 65ºC for 3 h. Finally, it was ready to use for gel electrophoresis in order to examine nucleotide sequence cleavage.

Chapter 3

Investigation of human *Tet* gene homologues encoding proteins with a 2ODD domain

3.1 Introduction

Enzymes containing 2-oxoglutarate (2-OG)-and Fe (II)-dependent dioxygenase (2ODD) domain are able to use O2 for oxidation of many substrates in various mechanisms which would be directly affected if O2 level changes. 2ODD domain is typically able to catalyse a wide range of reactions such as hydroxylation, oxidative ring closures and desaturations. The active site of such 2ODD comprise three specific amino acid residues, namely two histidines and an aspartate, a substrate binding site and a 2-oxoglutarate binding site (2OG) (Aravin and Koonin, 2001). These enzymes are able to oxidise an organic substrate by using a dioxygen molecule which mostly uses ferrous iron as the active site cofactor and 2OG as a co-substrate (Figure 3-1). In other words, the iron centre becomes poised to bind oxygen only when substrate complex (substrate+2OG) and cofactor (Fe2+) are present in the active site. Finally, such a dioxygen molecule and 2OG are decarboxylated to succinate and CO2 as final products.

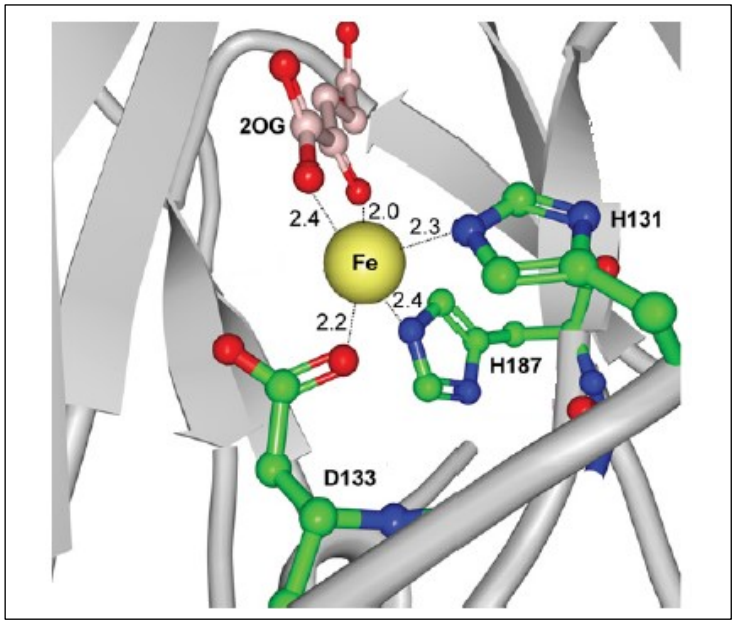


Figure 3-1 The catalytic centre of 2ODD conserved domain . Three amino acid residues which are indicated in black (D133=Aspartate, H187 and H131=Histidine) have a coordination with mononuclear non-haem Fe rendering this centre to be catalytically active site. This site contributes to the binding of 2OG (Bjørnstad, et al., 2011).

Recently, increasing evidence has revealed that certain 2ODD can act on histones and RNA as alternative substrates. For example, a dioxygenase-based JmjC-domain-containing histone demethylases is able to demethylate histone lysine H3K4. This family of demethylases utilise 2OG as a cosubstrate and Fe(II) as cofactors in the presence of dioxygen to hydroxylate the methylated histone. As a result, succinate, CO_2 and unstable carbinolamine are generated; subsequently, carbinolamine spontaneously releases formaldehyde and produces demethylated lysine. In addition, the fat mass and obesity-associated protein (FTO) and alkylation protein ALKB homolog 5 (ALKBH5), which are able to demethylase the adenosine N6 position (m6A) of RNA, are members of the ALKB subfamily of 2-ODDs (Dong, et al., 2014).

Similarly, there has been a major discovery of the diverse function of 2ODD family members in plants. For example, recently, two JmjC genes, *Jmj20* and *Jmj22*, encoding 2ODD were

reported that having a link to seed development in Arabidopsis. The function of *Jmj20* and *Jmj22* is to redundantly reverse repressive histone methylation, which allows transcription of the key gibberellin biosynthetic genes, gibberellin 3-oxidase and gibberellin 2-oxidase. The increase in level of active gibberellins subsequently stimulates seed germination (Cho, et al., 2012). Furthermore, a homolog of human *alkbh2* and *alkbh3* has been reported to be present in *A. thaliana*; the sequence and functional activity of the particular protein ALKBH2 show similarity with the human equivalent (Meza, et al., 2012). Additionally, auxin metabolism and reproductive development of rice also require the involvement of 2ODD (Zhao, et al., 2013). For example, the auxin oxidation (*DAO*) gene, encodes a putative 2ODD shown to catalyse the irreversible inactivation of IAA to its oxindole derivative (OxIAA); this compound is essential for anther dehiscence, pollen fertility, and seed initiation. Another interesting example of the functional diversity of 2ODD is the discovery of the hypoxia-inducible transcription factor (HIF) which is regulated by a proly 4-hydroxylase (P4H), which also has the usual characteristics of a 2ODD (Farrow and Facchini, 2014). In plants, the hypothetical functions of HIF-P4H homologs include the possibility that HIF-P4Hs modulate metabolic changes during iron starvation (Jin, et al,. 2007) and these changes are also implicated in the induction of 2ODD with roles in the reutilization of iron (Lan, et al., 2011).

In fact, plants have a characterised system for maintenance, deposition and removal of DNA methylation (Law and Jacobsen, 2010). The presence of 5-mC in promoters and transposable elements (TEs) is correlated with reduced transcription whereas removal of 5-mC increases transcription of silenced genes. There is some evidence reporting that plants are able to remove 5-mC through base excision repair. In particular, *A. thaliana* has four 5-mC DNA glycosylase enzymes involved in the removal of DNA methylation; these are ROS1, DML2,

DML3, and DME (Zhu, 2009). However, these DNA glycosylase gene homologues are unlikely to be present in animals.

It has been found that in specific animal cells, the removal of methyl groups from cytosine is due to the oxidative demethylation caused by TET enzymes, which have been shown to be 2ODD. Apparently, the TET enzymes (TET1, TET2 and TET3) play a major role in animals but have not yet been functionally characterised in plants. In specific animal cells TETs oxidise 5-mC in DNA by using dioxygen and Fe(II) as coenzyme and 2OG as a cosubstrate, and generate new epigenetic marks such as 5-hmC, 5-caC and 5-fC, The presence of these marks is implicated in various aspects of animal cell development as well as disease (Tan and Shi, 2012). However, whether such TET gene homologues exist in plants has been a controversial topic.

On the one hand, some reports have argued that 5-hmC is present in Arabidopsis. One study of 5-hmC using a reversed-phase HPLC coupled with tandem mass spectrometry method has concluded that 5-hmC is not present in biologically relevant amounts (Liu, et al., 2013). Similarly, another study on Arabidopsis found that DME and ROS1 have a significant 5-hmC excision activity in vitro, whereas 5-hmC was not detected; this suggests that it is very unlikely that plants employ 5-hmC as a DNA demethylation intermediate (Jang, et al., 2014).

On the other hand, there is also some evidence to confirm that oxidative products of 5-mC is present in particular rice lines. One study that employed a dot-blot assay, and further quantified 5-hmC levels in DNA from different rice tissues using liquid chromatography-multi-stage mass spectrometry (LC-MS/MS/MS) has found large inter-tissue variation in the levels of 5-hmC. Moreover, this study also obtained a genome-wide profile of 5-hmC modification in three different rice cultivars using a sensitive chemical labelling followed by

next-generation sequencing method. The result showed thousands of 5-hmC peaks were detected (Wang, et al., 2015).

Due to the conflicting evidence concerning the presence of 5-hmC in plants and the fact that there is still no evidence for an enzyme that might precisely remove a methyl group from the cytosine of DNA by oxidation, a search was conducted to identify possible *Tet* gene homologues. This study on rice is an initial step to identify such homologues which encode the 2ODD domain. Various bioinformatics software programmes were used to produce a detailed analysis. *Arabidopsis* protein sequences containing the characteristic 2ODD motif (HXD…H) (McDonough, et al., 2010), retrieved by conducting a blast search with the conserved domains of human 2ODD, were used as a query in a BlastP search restricted to rice. Selected genes were subsequently used in a tblastn search in order to examine the activity of each gene by identifying transcripts from libraries prepared from different tissues. Moreover, other important gene databases were utilised in order to produce further detail and possible functional annotation.

3.2 Objectives

In light of the previous comments that 5-hmC and 5-caC have been discovered in rice, the first hypothesis was that such compounds might be generated in specific rice lines/tissues and the associated oxidative demethylation probably also occur in plants and results in epigenetic changes equivalent to those occurring in animal cells. This aims of preliminary study were to identify possible *Tet* gene homologues as well as to assemble details and annotation of selected genes in order to subject these genes to subsequent investigation of their expression.

3.3 Materials and methods

3.3.1 Identification of rice genes which have conserved domain of 2ODD

The conserved domain of 2ODD was identified by using Conserved Domain Search Service tool (CD-search tool) in the conserved domain database and macromolecular structure viewer tool (Cn3D tool) in the Molecular Modelling Database.

- The sequence of 2ODD in *Arabidopsis thaliana* retrieved by conducting a blast search with the conserved domains of 2ODD in human was used in the query shown below to perform a BlastP search restricted to *Oryza sativa* (Japonica cultivar group) in the Non redundant (nr) protein sequence database

>gi|30686940 GENE ID: 828665 AT4G25600

MACLSRIFLILMITMSSSSPPFCSGGSRKELRDKEITSKSDDTQASYVLGSKFVDPTRV
LQLSWLPRVFLYRGFLSEEECDHLISLRKETTEVYSVDADGKTQLDPVVAGIEEKVS
AWTFLPGENGGSIKVRSYTSEKSGKKLDYFGEEPSSVLHESLLATVVLYLSNTTQGG
ELLFPNSEMKPKNSCLEGGNILRPVKGNAILFFTRLLNASLDGKSTHLRCPVVKGELL
VATKLIYAKKQARIEESGECSDEDENCGRWAKLGECKKNPVYMIGSPDYYGTCRKS
CNAC

- Before alignment, the sequence identified by the blast search was selected by focusing on conserved domains which have “HFD” “HYD” or “KLD” subsequence of amino acids which are conserved residues and represent the catalytic or active site identified in previous studies.
- Examination of the activity of each gene by identifying transcripts from libraries prepared from different tissues through Tblastn using the method as follows:-

A. The protein sequence identified above is used as query.
B. These are searched against the Expressed Sequence Tags (EST) database.
C. Those ESTs showing a perfect match with regions of the gene are selected.
D. Blast result has been aligned by ClustalW and CLC Sequence Viewer 6.6.2 program.
E. A consistently conserved domain shown after alignment was investigated.
F. The same sequences with different names are removed in order to produce a non- redundant list

3.3.2 Protein interaction analyses

Protein sequences were used for analysing the evolutionary relationship by performing CLC Sequence Viewer 6.6.2 program and an interaction of every protein were investigated by using STRING 9.1 (Search Tool for the Retrieval of Interacting Genes/Proteins version 9.1) database.

3.3.3 Nuclear localisation sequences (NLSs) prediction

The protein sequences of selected genes are used to predict the NLSs by using several bioinformatics tool as follow:

- The eukaryotic Linear Motif (ELM) resource at http://www.elm.eu.org/
- A SVM based method for subcellular localisation prediction of rice proteins (RSLpred) which can be accessed via http://www.imtech.res.in/raghava/rslpred/ (Kaundal and Raghava, 2009)
- The Arabidopsis Subcellular Localization Prediction Server (AtSubP) searched through http://bioinfo3.noble.org/AtSubP/?dowhat=AtSubP.

3.3.4 Mutant analysis

Mutant genes which encode proteins having the consistently conserved domain can be identified using OrygenesDB tool which comprises various mutant database such as Postech T-DNA Insertion Mutant, UCD Rice Transposon Flanking Sequence Tag Database with Ds KO lines, Oryza Tag Line (OTL) Database with Tos17 and T-DNA KO lines, Rice Mutant Database (RMD) with T-DNA KO lines and Taiwan Rice Insertional Mutants Database (TRIM) with T-DNA KO lines.

3.3.5 Gene annotation profiling

For detailing each single gene, 5 rice annotation websites as shown below were observed in order to profile these genes.

- http://www.ncbi.nlm.nih.gov/
- http://plants.ensembl.org/Oryza_sativa/Info/Index
- http://www.uniprot.org/
- http://www.genome.jp/dbget-bin/www_bfind?T02163
- http://rapdb.dna.affrc.go.jp/

3.4 Result and discussion

3.4.1 Identification of genes which encode proteins with conserved domains of 2ODD

The investigation of the 2ODD conserved domain by performing Cn3D tool which is capable to visual three-dimensional structures and simultaneously displays sequence, structure, and alignment by considering on interaction of examined sequence-structure relationships and superposition of geometrically similar structures was conducted. The result showed that this domain structure comprises of three amino acids conserved motif as HXD and H (H is Histidine, D is Aspartate and X is any of amino acid) binding site which coordinate with Fe metal and also contain 2OG binding site (Figure 3-2). Moreover, the alignment sequence of 2ODD superfamily likewise displayed that its family members such as crystal structure of catalytic complex of the oxidative DNA/RNA repair enzyme alkb from *Escherichia Coli* K12 (2FDG_A), RNA-dependent RNA polymerase from cherry green ring mottle virus (gi 81924228) as well as replicase from Grapevine rupestris stem pitting-associated virus 1 (gi 81924262) certainly contain HXD…X motif (Figure 3-3).

Several evidence has been revealed that 2ODD are widely found in aerobic and facultative anaerobic organisms (Aik, et al., 2012). 2ODD is able to use 2OG and molecular oxygen as co-substrates, and Fe(II) as a co-factor to catalyze a wide range of chemical reactions including demethylations, hydroxylations epimerizations, cyclizations, desaturations, rearrangement, halogenation (Kershaw, et al., 2005). In plant , many 2OG oxygenases have been identified, which have diverse cellular functions including in DNA repair, histone demethylation, post-translational modification, iron sensing, ethylene biosynthesis and auxin catabolism (Farrow and Facchini, 2014).

The functional character of 2OG oxygenases has been confirmed by Clifton, et al., (2006) that this enzyme family members have a core double-stranded β-helix-fold (DSBH), which was observed to allow two histidinyl residues and an aspartyl-residue binding with iron which characteristically possess a part of HXD…H motif. Moreover, Aravin and Koonin (2001) also revealed clear evidence for supporting that catalytic core of some family members of 2ODD such as clavaminic acid synthase (CAS) consists of DSBH containing a HXD dyad and a conserved carboxy-terminal histidine which together bind with Fe(II) atom.

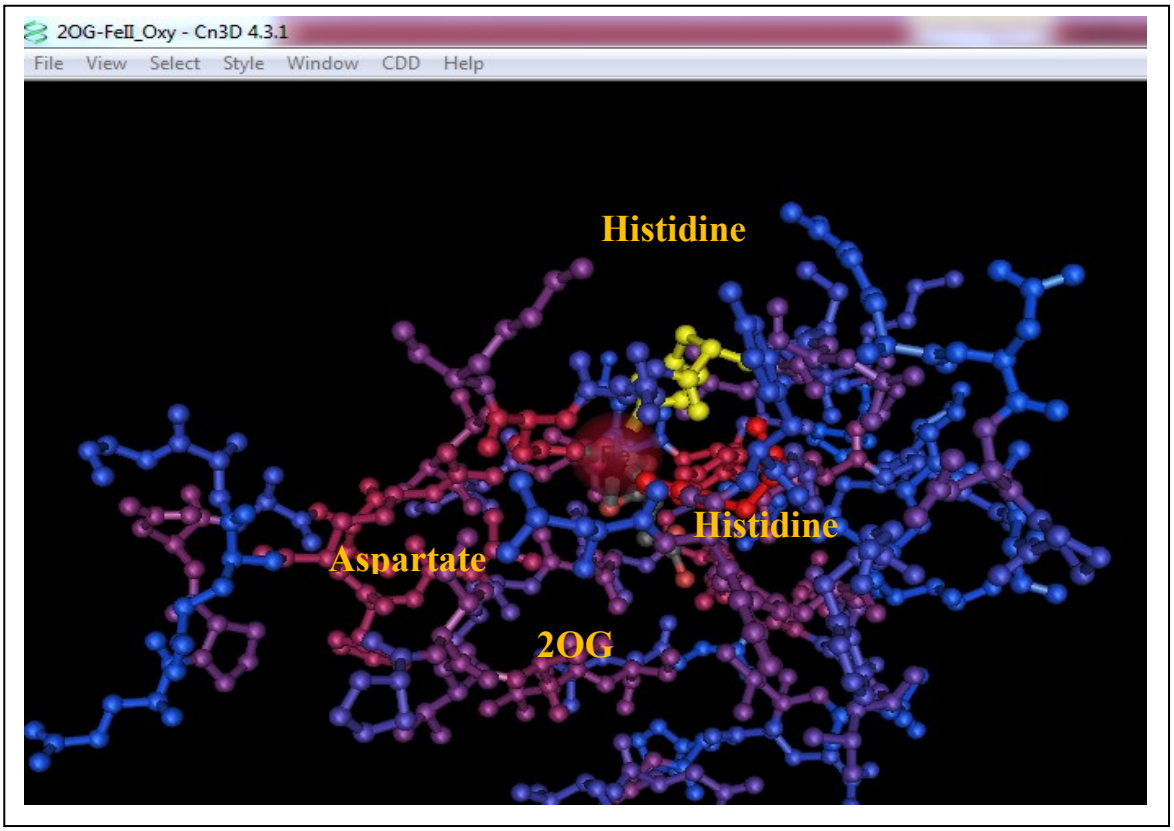


Figure 3-2 3D structure of 2ODD conserved domain retrieved by using Cn3D 4.3.1 tool. This catalytic domain comprises three amino acid (two histidines and an aspartate) binding site and a 2OG binding site which have coordination with Fe metal.

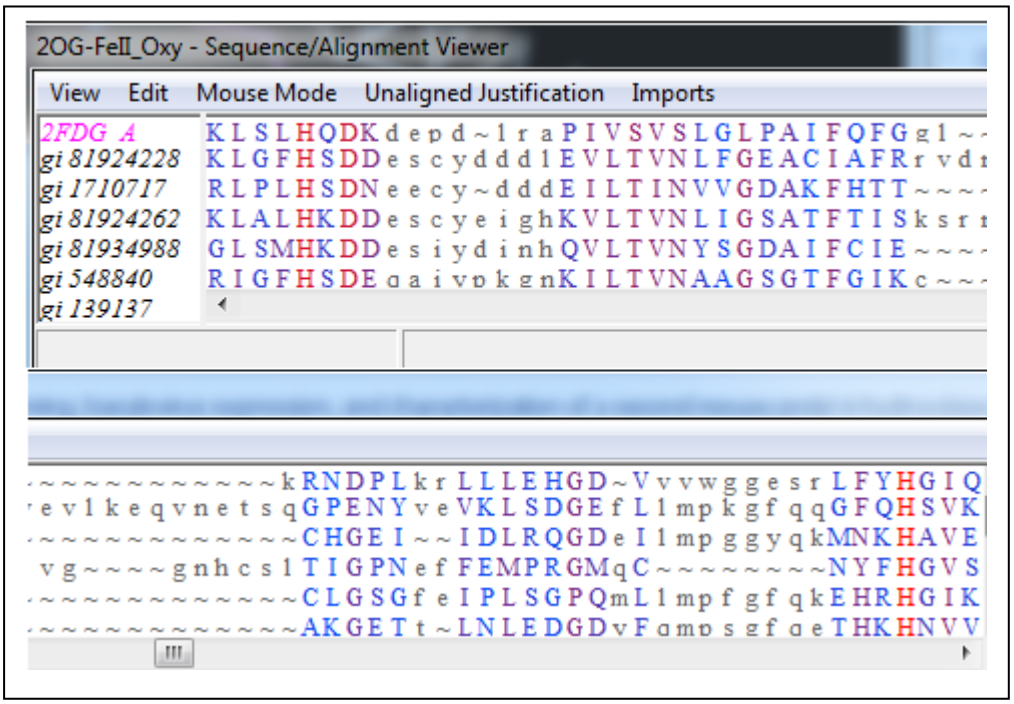


Figure 3-3 The alignment viewer of 2ODD conserved domain containing HXD…H motif from various organisms. 2FDGA is crystal structure of catalytic complex of the oxidative DNA/RNA repair enzyme alkb from *Escherichia Coli* K12, gi 81924228 is RNA-dependent RNA polymerase from cherry green ring mottle virus, gi 1710717 is RNA-directed RNA polymerase from apple chlorotic leaf spot virus, gi 81924262 is replicase from grapevine rupestris stem pitting-associated virus 1, gi 81934988 is RNA replication polyprotein from apple stem pitting virus PA66 and gi 548840 is RNA replication protein from shallot virus X

To investigate possible *Tet* gene homologues, the non-redundant protein sequence database from NCBI was searched using blastP tool together with searching restricted to *Oryza sativa* (Japonica cultivar group). The sequences of 2ODD were selected by considering to conserved domains which have "HXD…H". Additionally, the activity of each gene was examined through Tblastn in order to identify the transcripts of genes in different tissues by searching against the Expressed Sequence Tags (EST), a short sub-sequence of a cDNA sequence, database of single selected sequences. Those ESTs showing a perfect match or displaying highest percent identity with regions of the gene were selected. The result shows that there are 18 genes encoding 2ODD which have HXD…H conserved domain. (Table 3-1).

Table 3-1 18 genes encoding protein having HXD...H motifs were selected after using BlastP tool search against proteins both in the global Non-redundant (nr) protein sequence database and restricted to the *Oryza sativa* (Japonica cultivar group).

No.	The Selected Sequences with labelled HXD...H
1	1.>gi|297719687|ref|NP_001172205.1|Os01g0180900 MPLPRRAATVRIRRGQIHRVTAMPPPPQQAPEPAGLQLLLQPEVVAPNHPPPAPPAPAAVPAPPQPQGEA RYRRPLVRLQAVPNEDHVPDNYGDGPDELGITPAVYQALERHLPPDLAGAPAEVKRYFMRSVLRNYVPSP SQRIRTQNQREYRERILSAYQPLHPELYTNDPSTFILPAFLQAINGNTEESITSIMMEPAPGVFAFPMLK PSFCQMLMSEVNNFLRWAQSANQRIMRPTSLDRHGRGAALSDFGLQEMLDNLMKDFISPMSTVLFPEVGG NTLDSHHTFVLEYGEADGARGFHVDDSEVTLNICLGKHFTGADMYFRGIRCGNHVNSGTHDEEYFVHPNV PGQVLLHHGSHRHGVFSVTSGRRVNMVMWCKSSVFREMKKFMTDFSGFCCECQFQRTARQVQHLQELTAR ISISGQESEDETP
2	2.>gi|115436784|ref|NP_001043135.1|Os01g0502400 MALDGPAERRDEAQAAMGNGNGVAPPPLRPAGRPSGAPACADRRLRLNPNMEHKPQDYSDVRGEYAPAVY SALERHLPPSVLDANREIKLQLMREVLGHYWPHGERNKVQRHREYRQRILNHYKPLHKELYKMRPSSFFL PTFLEAIRTNTEESFRSIMTEPVPGVYSFAMLQPNFCEMLLQEVENFEKWVHTMKFKIMRPNTMNKYGAV LDDFGLEVMLNQFMEQFIAPMSTVLYPEVGGGTLDTHHAFVVEYGKDRDVELGFHVDDSEVTLNVCLGKQ FSGGELYFRGIRCENHVNSETQHEEMFDYSHVPGRAVLHRGRHRHGARPTSSGLRINLLLWCRSSVFREM KKYQKDFSGWCGECKREKKERQIHAVKATKLAFLRGAGGATI
3	3.>gi|297596978|ref|NP_001043285.2|Os01g0546900 MSMREFVDHWAASSSNGDSDGSLLYLKDWHFVKEYPGYVAYTTPTFFADDWLNMYLDSHPIHRDSDIANH TNEINCADYRFVYMGPKGTWTPLHADVFRSYSWSANVCGRKLWLFLPPSQSHFVFDRNLRSSVYNINDDV SEKQFPEFNNTKWLECTQEQNEIIFVPSGWYHQVHNLEDTISINHNWFNGYNLHWVWNLLHEDYKVAKDY IEDIRDICDDFEGLCQRNLAANTGMNFYDFFVFITRFALANIVELYHIQNPKDTDFISAETANHFVYNLM SIRDVASKMVSTEAFNTENICNISEQNRSAFSDIIKILEEESFRRLLVALSKAYNYIDRGQKDCLKMKDS SQKGCLSVTCLKPDCNVVGDIISFMREIHGPMDLVTLIDSALSDR
4	4.>gi|115455509|ref|NP_001051355.1|Os03g0761900 MKGGVIRSGGGVGGIGVGGGGGGGGAGGGLMRTRLRLPVVLLSCSLFFLAGFFGSILFTQDPQGEEELDT PMRRERLMEAAWPGMAYGESGEPEPSLIPYQILSWQPRALYFPQFATSQQCENIVKTAKQRLMPSTLALR KGETEESTKGIRTSSGTFLSSDEDPTGTLAEVEKKIAKATMIPRHHGEPFNILRYEIGQRYASHYDAFDP AQYGPQKSQRVASFLLYLTDVEEGGETMFPYENGENMDIGYDYEKCIGLKVKPRKGDGLLFYSLMVNGTI DPTSLHGSCPVIKGEKWVATKWIRDKSKAV
5	5.>gi|115456019|ref|NP_001051610.1|Os03g0803500 MAPLRTLVIFLLLLLALVPALSRPDGGGGGFYDPARVTQLSWRPRAFLYSGFLSHDECDHLVNLAKGRME KSMVADNDSGKSIMSQVRTSSGTFLSKHEDDIVSGIEKRVAAWTFLPEENAESIQILHYELGQKYDAHFD YFHDKNNLKRGGHRVATVLMYLTDVKKGGETVFPNAAGRHLQLKDETWSDCARSGLAVKPKKGDALLFFS LHVNATTDPASLHGSCPVIEGEKWSATKWIHVRSFDNPPDVSLDLPCSDENERCTRWAAVGECYRNPKYM VGTKDSLGFCRKSCGVCDA
6	6.>gi|115456185|ref|NP_001051693.1|Os03g0816500 MDSSKQQLAVWPVSAAKKYAGFDGKAAARERSGLANASFRVYYSLRAGAVPFLWESSPGTPKAAAVSSPA RESSSAARGALPPISPPPSYQSVEMKKGGRCRPRSSWPAAAAAGDIAPLKGESVKPFPALVLASALTLAA SSMVGGGGGGGHGVRTPANFPPRQKARLPRGPVHEKSLEQQKKGPSSSSPSVSSNKSPLQLAAAIVQPQK

No.	The Selected Sequences with labelled HXD...H
	PLESPQHMVTPVRLQESPGPRTIPCSSGSVGSGSGAAPFDICIKRDDKCSIKLSRSLLEINREKRREREQ LSKEAAPLQYLRPGMVLLKKFLKHDDQVDIIRRCQKLGIGSGGFYTPGYRDGGKLSLQMMCLGKNWDPNS RSYGDTRPFDGAQPPSIPEVFSKIVKDAIQASNEFLRQKARPANDVEELPPLSPDICLVNFYTSSGKLGL HQDKDETKPSLHKGLPVVSFSLGDTAEFLYGDVNDVDKASKVDLESGDVLIFGGKSRLIFHGVSRIKPKT APNWLTDEAKLRPGRLNLTFRQH
7	7.>gi|115457822|ref|NP_001052511.1|Os04g0346000 MGMARARARRLLPLLTFVTLGMILGSLLQLAFFRRIDDHSNVTHLENDQEAAFLRLGLVKPEVISWSPRI IVFHNFLSSEECDYLRSIARPRLQISTVVDVATGKGVKSNVRTSSGMFVSSEERKLPVIQSIEKRISVYS QIPEENGELIQVLRYEPSQYYRPHHDYFSDTFNIKRGGQRVATMLMYLTDGVEGGETHFPQAGDGECSCG GKMVKGLCVKPNKGDAVLFWSMGLDGETDSNSIHGGCPVLEGEKWSATKWMRQKEFV
8	8.>gi|297603304|ref|NP_001053775.2|Os04g0602700 MAGSGYTRPPPPPPLGEGAAPAPSAALYVANCGPAVGLTHDDIRAAFAAFGEVAGVHGADGSGVRVIVRF REPAAAEAAMSALHGRPCAGLAGRVLHIRYSVPAKPKAPVGGSLPVATSASELGVPGIYLVPDFVTAAEE QELLAAVDNRPWKSLAKRRVQHYGFEFLYETRNVDSKQFLGELPPFVSKIIDKIMSFPGANKCTSKLVDQ LTVNEYPCGVGLSPHIDTHSAFEEMIFSLSLAGPCIMEFRKYPKGSWRAPSMVSGTDKDSIEEPQCIRKA VFLPPRSMLLMSGEGRYAWHHYIPHHKIDDVGGQVIKRNTRRVSFTFRKVRMGLCDCEYGQFCDSQSK
9	9.>gi|115464581|ref|NP_001055890.1|Os05g0489100 MAPVRLGAPTSGALLLLVLLLICGGGGGGGVAAGGGGGGGGKGSSVYPAPVVYPHHSRQISWKPRVFLY QHFLSDDEANHLVSLARTELKRSAVADNLSGKSELSDARTSSGTFIRKSQDPIVAGIEEKIAAWTFLPKE NGEDIQVLRYKHGEKYERHYDYFSDNVNTLRGGHRIATVLMYLTDVAEGGETVFPLAEEFTESGTNNEDS TLSECAKKGVAVKPRKGDALLFFNLSPDASKDSLSLHAGCPVIKGEKWSATKWIRVASFDKVYHTQGNCT DDNESCEKWAALGECIKNPEYMIGTAALPGYCRKSCNIC
10	10.>gi|115466278|ref|NP_001056738.1|Os06g0138200 MDASSSASAGGEADAMALVQGYNADELAIAGEFLTTWLPFLSAGLCASCADSLRSRVSSLLPPQAEESPS SPPPRIDQIEPSGWESDPATAHPQHLPFEPSGWDSDPPQLPPEQEQQKQKPQPAEKPRKSWADMAQEDEL AAAAEEDAAAAAADDGEEGSEAGRPGVQLTREQRELRRFRNVVRRKDFICFERVNGRLVNILAGLELHCG VFSAAEQKRIVDYVYDLQEMGKHGELGDRTYTEPQRWMRGKGRVTIQFGCCYNYATDKNGNPPGIIRTIA SDPMPSLFKIMIKRLVRWHVLPKTCIPDSCIVNIYDPGDCIPPHIDSHDFVRPFCTVSFLSECNILFGST LKIAGPGEFTGSLPIPLPVGSVLILNGNGADVAKHCVPAVPTKRISITFRKMDPAKRPFNFRDDPELLNI IPLETAVQETGRSSDEGKGKQPDIQIRNPSKAHRNKKSKVRTSPGKGGRGGILGDGPPQYAQAQVTGISS QQNFHGQPTISGSSAERERRPVGPLRESRYQQDAPGMQSNMDGIRERANWLAQERMHGNSMNSIDDGTES QERRQRMEHRQILMINRTINDDMDSLSIGSHESDQTRVSVRTLYNKPRRTRVNLDE
11	11.>gi|297724717|ref|NP_001174722.1|Os06g0286310 KILAQVSHSLTVLQPRDTCYVADEGLTDLRYSGHQPHAHSWDEFPVLKDILKAVHEALPGSHFNSLLLNR YKTGSDYVSWHADDEPLYGPTPEIASVTLGCEREFLLRKKPTKSQASLGSGEVAPKRLKVSAPQQHSFLL KHGSLLVMRGYTQRDWQHSVPKRAKASSPRINLTFRRVL
12	12.>gi|115471029|ref|NP_001059113.1|Os07g0194500 MVAVVMMAMRKRMRGALLALALLLTATAVVPLLLLGEAGDDGVGAVAAAPPFNASRVRAVSWRPRVFVYK GFLSDDECDHLVKLGKRKMQRSMVADNKSGKSVMSEVRTSSGMFLDKRQDPVVSRIEKRIAAWTFLPEEN AENIQILRYEHGQKYEPHFDYFHDKVNQALGGHRYATVLMYLSTVEKGGETVFPNAEGWENQPKDDTFSE CAQKGLAVKPVKGDTVLFFSLHIDGVPDPLSLHGSCPVIEGEKWSAPKWIRIRSYEHPPVSKVTEGCSDN SARCAKWAEAGECEKNPVYMVGAEGLPGNCRKSCGVCDS

No.	The Selected Sequences with labelled HXD...H
13	13.>gi|297609695|ref|NP_001063533.2|Os09g0489200 MAGGEPPAPATEAEEGRRAALLREITEEGGFAFVASAEKAACGDLRAAEAAREMAWEQLHSGPWSEVGAA WRDAYALACLHVARLRRLGAAAADRRAALRALDMGLIMGGNLLRADLEAAIARIVADPGGGGDAEAVDEE TRRWREGLERNRDVADALNILPAKSLSCKKVERRSCISLEEFICDYFLRESPVIISGSIDHWPARTKWKD IQYLKKIAGDRTVPVEVGKNYVCSEWKQELITFSQFLERMWSAGCPSNLTYLAQHPLFEQIKELHEDIMV PDYCYAGGGELQSLNAWFGPHGTVTPLHHDPHHNILAQVLGRKYIRLYPASISEDLYPHTETMLSNTSQV DLDNVDLKEFPRVENLDFLDCILEEGDLLYIPPKWWHYSYVNKLRENYASLKSDACKTIEGFT
14	14.>gi|115481998|ref|NP_001064592.1|Os10g0413500 MAALSSRRRHGSLRLLAVALLVLAGVASSAAAAGSGRGAFDPSRVVQLSWRPRAFLHKGFLTDAECEHLI SLAKDKLEKSMVADNESGKSVMSEVRTSSGMFLEKKQDEVVARIEERIAAWTFLPPDNGESIQILHYQNG EKYEPHYDYFHDKNNQALGGHRIATVLMYLSDVGKGGETIFPEAEGKLLQPKDDTWSDCAKNGYAVKPVK GDALLFFSLHPDATTDSDSLHGSCPVIEGQKWSATKWIHVRSFDISVKQGASTDGCEDENVLCPQWAAVG ECAKNPNYMVGTNEAPGFCRKSCNVCAQ
15	15.>gi|115482056|ref|NP_001064621.1|Os10g0420000 MEDDKTTQESSAAAAPPLRSLADYAVGPIPTLLYVPGFISDAEQSQLLHHIYQAPAPKWKSLKNRRLQNW GGVVHEKGLLPQALPSWLTKITDRICQWTGLFPSAINHVLINEYHPNQGIMPHQDGPAYFPVVAIISLAS PVVIDFTPHQRLKGEDFTDPQNAHSGESQATTTESNGSHNLEGANETDPASSSLLLMPCSLLIFKDQAYT GYRRTARRVSLTCRLVLKVHKKLFKM
16	16.>gi|115482738|ref|NP_001064962.1|Os10g0497800 MAGSRAAGRGGRPLLGGGGGKRGGGGGGGGGGGGKSSSYTTTVILAALLLASVALLLLVALGALSLPSGG GVGVGHAGIGLPRPRPRFRRSAAFESGLEMRGGEKGEPWTEVLSWEPRAFLYHNFLSKEECEYLISLAKP HMKKSTVVDASTGGSKDSRVRTSSGMFLGRGQDKIIRTIEKRISDYTFIPVENGEGLQVLHYEVGQKYEP HFDYFHDEFNTKNGGQRIATLLMYLSDVEEGGETIFPSSKANSSSSPFYNELSECAKKGLAVKPKMGDAL LFWSMRPDGSLDATSLHGGCPVIKGNKWSSTKWMRVHEYKI
17	17.>gi|115485553|ref|NP_001067920.1|Os11g0488500 YGDTEPAAAAAAERTAFRRAEKQYKLYKPLKPKGGRARSKPGGGGGGGGLEDLSAVVDFHALLAAAGGGG GELPAGIGRRDVAGFDLFCFLDRPGFYFIPGALSIEEQCYWIRESLKTFPQPPNRTNLTALYGPIFDLLT AAKSGKILVEVGNPNDQERSEQNSDGRKSNSFKFLEDTEIHKGEVCRSTAATTLVRKLRWSTLGLQFDWS KRNYDVSLPHNKIPDLLSALAKKMAIPAMPSGEEFKPEAAIVNYYGPSDMLGGHVDDMEADWSKPIVSIS LGCKCIFLLGGKTRDEVPTAMFLRSGDIVLMAGEARECFHGNAFY
18	18.>gi|115486503|ref|NP_001068395.1|Os11g0657200 MAAAADDDELRRRAAALREVFGDSSDSEADDLPVGGAGREQWRWEAVEGVRGLWLCAAFLSADEQSRLLT AIRREGWFSDARNQAMRFGDLPSWAVELSALIHEAICFGDVRVGCGLELKNEDEDACPLPSDLLWRKPLF DQMIANRYEPGEGICAHVDLMRFDDGIAIVSLESPCVMHFSRAEQEVPICETLESVHAEPTKIPVYLNPG SLVLMSGDARYLWKHEINRKPGAQQWGGRELEQQIRTSITLRKLLPSPN

Consequently, blasted result of selected sequences had been aligned by performing CLC Sequence Viewer 6.6.2 program for investigating a consistently conserved domain like HXD…H (Figure 3-4).

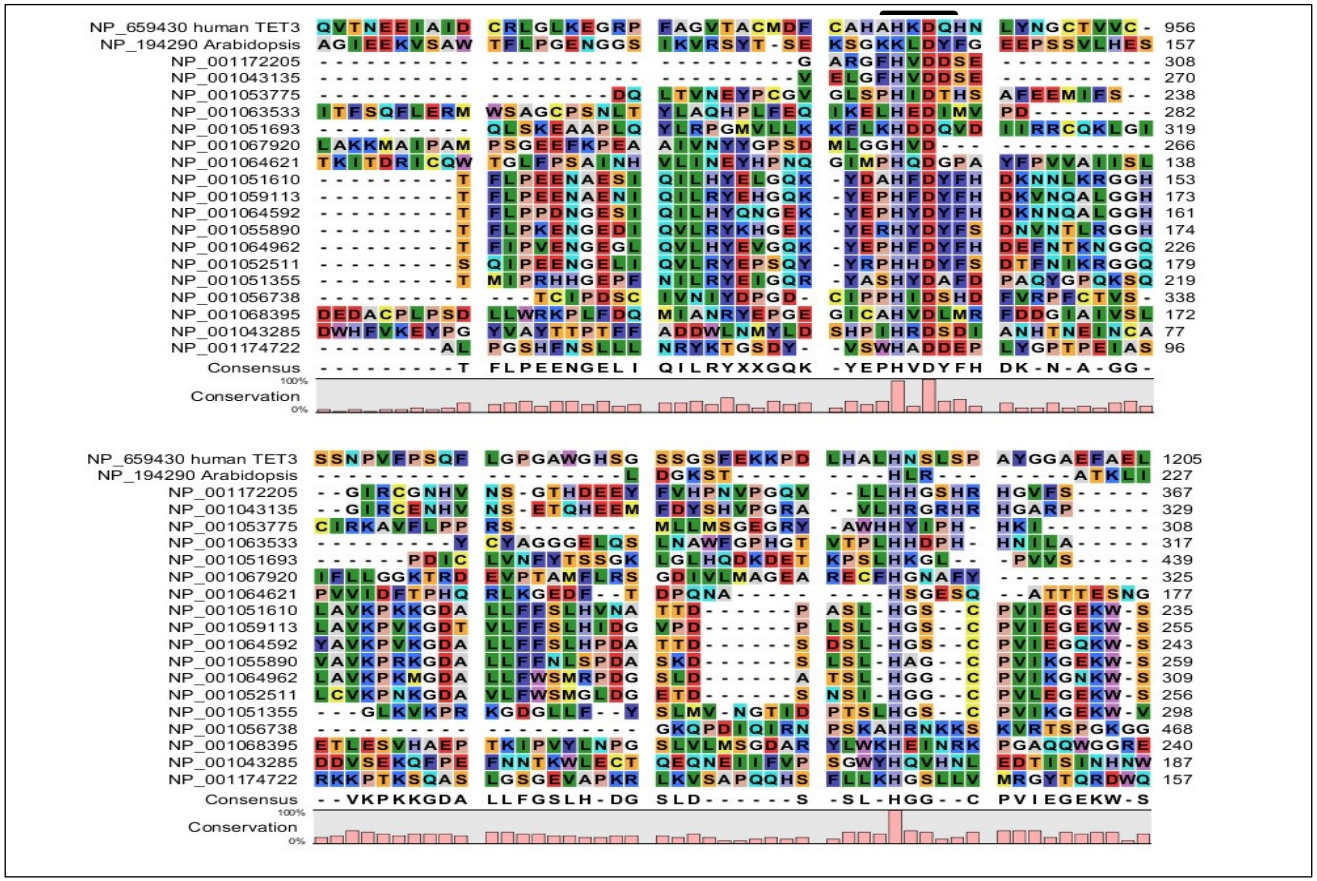


Figure 3-4 The multiple alignment of HXD…H motifs from 18 rice and Arabidopsis sequences together with Human TET protein by using ClustalW and CLC Sequence Viewer 6.6.2 program. A=Alanine, R=Arginine, N=Asparagine, D=Aspartic acid, B= Asparagine or Aspartic acid, C=Cysteine, Q=Glutamine, E=Glutamic acid, Z= Glutamine or Glutamic acid, G=Glycine, H=Histidine, I=Isoleucine, L=Leucine, K=Lysine, M=Methionine, F=Phenylalanine, P=Proline, S=Serine, T=Threonine, W=Tryptophan, Y=Tyrosine, V=Valine.

3.4.2 Protein analyses

3.4.2.1 Phylogenetic tree analysis

The relatedness of protein sequences was analysed by using neighbour joining algorithms performing through the CLC Sequence Viewer 6.6.2 programme. The phylogeny of 2ODD of 3 different species can be described that NP_659430 encoded by *tet3* is outgroup of plants dioxygenase whereas NP_194290 encoded by *AT4G25600* of *A. thaliana* shared a hypothetical common ancestor with dioxygenase proteins from *O. sativa* Japonica. Rice proteins encoded by differential TET3 gene homologues can be divided into 2 clades which individually shared a hypothetical common ancestor. The first clade comprises 5 proteins which are NP_001063533 (*Os09g0489200*), NP_001053775 (*Os04g0602700*), NP_001043135 (*Os01g0502400*), NP_001172205 (*Os01g0180900*) and NP_001051693 (Os03g0816500) whereas another clade contains 7 proteins which are NP_001064592 (*Os10g0413500*), NP_001051610 (*Os03g0803500*), NP_001059113 (*Os07g0194500*), NP_001055890 (*Os05g0489100*), NP_001064962 (*Os10g0497800*), NP_001052511 (*Os04g0346000*) and NP_001051355 (*Os03g0761900*). Moreover, within these two clades, there are three groups of protein which might be duplicated (Figure 3-5).

Similarity of human TET amino acid sequence to plants, and the identity of dioxygenase amino acid sequences within plants were identified by using the BlastP tool. As expected, amino acid sequence alignment results showed that 18 rice proteins and Arabidopsis proteins encoded by different 2ODD genes were less identical to TET3 (20 - 45%). In addition, only one rice protein (NP_001063533) showed high identical (83%) to Arabidopsis (NP_194290) whereas the other rice proteins showed less than 50% similarity to NP_194290.

Moreover, the similarity of duplicated rice protein present in the phylogenetic tree showed that the similarity of NP_001063533 to NP_001053775, NP_001043135 to NP_001172205 and NP_001064592 to NP_001051610 were 39%, 55% and 69%, respectively.

Based on these observations and results from phylogenetic analysis, it is shown that these proteins are homologues and some genes might have undergone duplication. For example, at a very early stage of the evolution, three groups of duplicated rice genes were formed. However, the proteins encoded by these duplicated genes do not show a very high degree of similarity. In fact, selection pressures during evolution acts at functional level; this means no strict overall similarity of sequences. In other words, as Koonin and Galperin, (2003) stated, secondary and tertiary protein structure is maintained better during evolution than is primary protein sequence. Furthermore, there are numerous examples of proteins that show slight sequence similarity but still contain identical or related amino acid residues in their active sites and have both similar catalytic mechanisms and similar structural adaptation. These shared features support the view that, such proteins are homologous though with low primary sequence similarity.

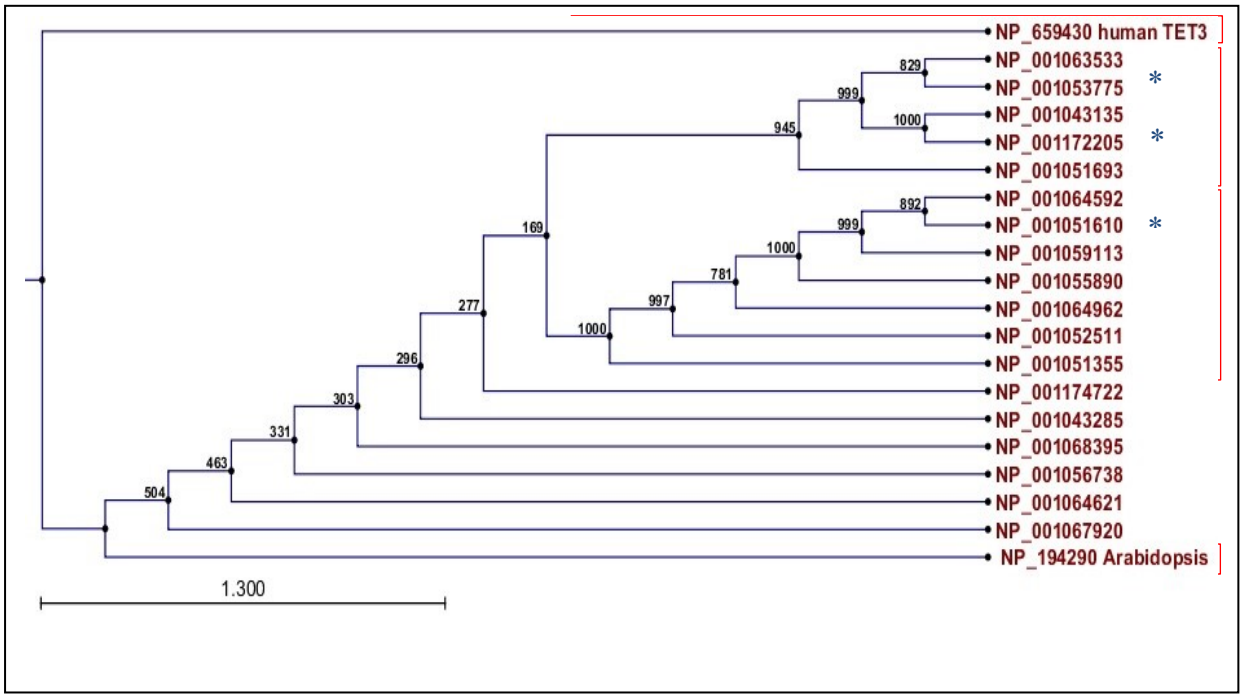


Figure 3-5 The phylogeny of 2ODD family. These proteins are encoded by 18 rice genes, one human gene and one Arabidopsis gene. The numbers at the branches are confidence values based on Felsenstein's bootstrap method (B = 1000 bootstrap replications). The horizontal lines are branches which represent evolutionary lineages changing over time and scale bar indicates1.300 is the length of branch that represents an amount genetic change of 1.300. Blue asterisks denote duplicated proteins. This phylogeny was obtained from "create tree" tool of CLC Sequence Viewer 6.7.1 program.

3.4.2.2 Protein interaction analysis

Protein interaction was investigated using STRING, which is a database designed to predict direct (physical) and indirect (functional) associations derived from four sources, such as genomic context, high-throughput experiment, co-expression and previous knowledge that would characterise a significant step towards a comprehensive description of cellular mechanisms and functions (Franceschini, et al., 2013). The results showed that there are 11 of 18 matched proteins from *O. sativa* Japonica appearing in this database. In addition, at the high confidence level (0.700), there are three genes encode dioxygenase proteins having interaction with other protein or proteins; these three are 4351051 (NP_001068395, *Os11g0657200*), 4340053 (NP_001056738, *Os06g0138200*) and 4339155 (NP_001055890, *Os05g0489100*). Protein 4339155 shows an interaction with five other proteins which are

4333246, 4335696, 4334010, 4349276 and OsJ_16680 (Figure 3-6), and its interactions are mentioned in the same publication saying that all these six proteins individually are putative target genes for the microRNA *OsmiR393*. However, it is confirmed later that gene encoding 4339155 is not the actual gene which is targeted by *OsmiR393* (Xia, et al., 2012).

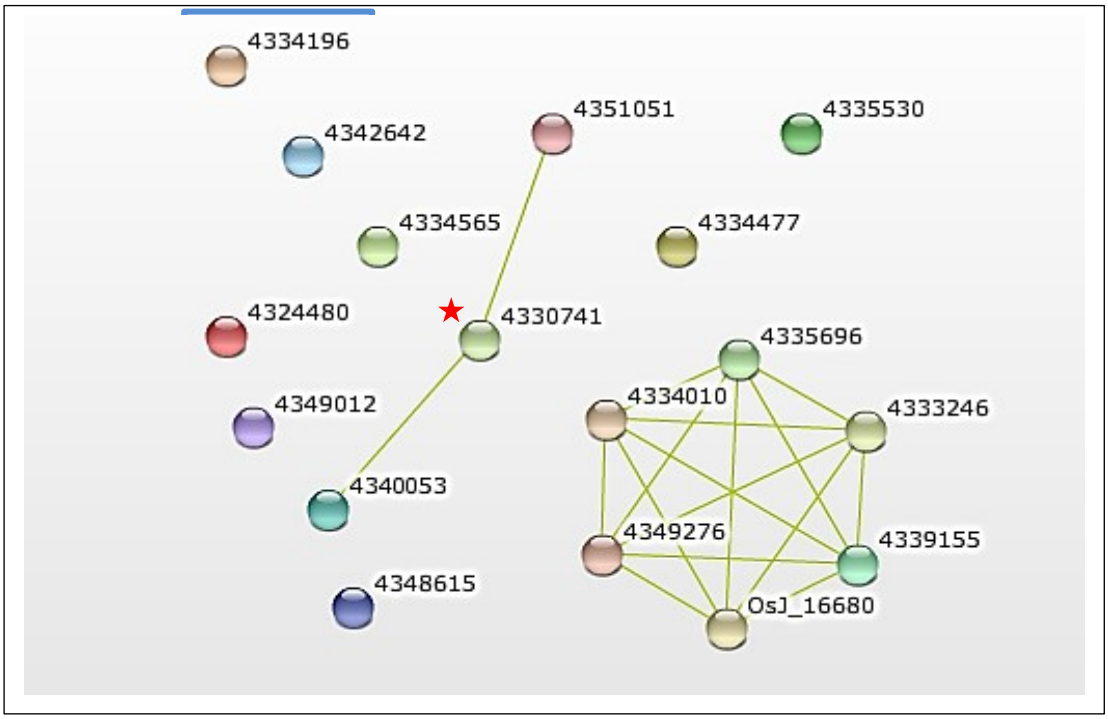


Figure 3-6 Interactions of dioxygenase proteins. Green line represents the interaction between proteins based on types of evidence for the association; Red boxes are dioxygenase proteins with interaction; blue boxes are dioxygenase proteins which have no interaction and red asterisks are other proteins. This diagram was retrieved by using STRING. 4334196, 4342642, 4349012, 4348615, 4335530, 4334477 and 4339155 are oxidoreductase which confidently predicted domain are prolyl 4-hydroxylase alpha subunit; 4334565, 4324480, 4340053 and 4351051 are oxidoreductase which confidently predicted domain are 2OG-Fe oxygenase family protein; 4330741 is methyltransferase; 4335696 is OsFBL16 - F-box domain and LRR containing protein; 4334010 is GRF-interacting factor 1; 4349276 is CorA-like magnesium transporter protein; 4333246 and OsJ_16680 are expressed protein.

3.4.3 Nuclear localisation sequences/signals (NLSs) prediction

NLSs are simple karyophilic clusters of arginines and lysines in nucleus-targeted proteins signal in which the anchoring of these proteins to specialized transporter molecules can be found on the pore complex or in the cytoplasm (Boulikas, 1993). In other words, NLSs are

the amino acid sequences or motifs which are recognised by the receptor of a carrier protein which can facilitate transfer from cytoplasm into the nucleus. NLSs can be transferred through at least two mechanisms. Firstly, small proteins which lack nuclear localization signals may transfer into the nucleus through nuclear pores, where they are subsequently sequestered by binding to intra-nuclear components. Secondly, larger proteins containing nuclear localisation signals are actively transported to the nucleus from the cytoplasm (Gao and Knipe, 1992). Different nuclear localised proteins may share the same NLS but a NLS has the opposite function of a nuclear export signal, which transfers proteins out of the nucleus. Hence, proteins which have a NLSs motif might be confirmed to have a function in the nucleus. To investigate such nuclear localisation of proteins encoded by the 2ODD gene family is a means to predict that these proteins might have a probability to bind with DNA in the nucleus and subsequently generate oxidative demethylation according to our hypothesis. After using three bioinformatics databases to predict NLSs, the results showed that from the ELM database, all 2ODDs showed NLS motifs whereas there are eight and five 2ODDs which have NLSs motif, according to AtsubP and RSLpred database, respectively (Table 3-2). Furthermore, Table 3-3 shows additional information of motifs specifically retrieved from ELM database. It can be seen that, for instance, *Os01g0180900* encodes 2ODD protein having LIG_FHA_1 motifs which are short phosphothreonine motifs binding a subset of FHA domains (a signal transduction module which recognizes phosphothreonine containing peptides on the ligand proteins) with large aliphatic amino acids at the pT+3 position. The consensus sequence of this motif is..(T)..[IVL].

Table 3-2 The prediction of subcellular localisation of 2ODD protein encoded by 18 rice genes, by using RSLpred, ELM and AtsubP databases

No.	Gene name	NLS prediction		
		RSLpred	ELM	AtsubP
1	*Os01g0180900*	mitochondria	nucleus	nucleus
2	*Os01g0502400*	mitochondria	nucleus	nucleus
3	*Os01g0546900*	mitochondria	nucleus	nucleus
4	*Os03g0761900*	chloroplast	nucleus	cell membrane
5	*Os03g0803500*	nucleus	nucleus	extra cellular
6	*Os03g0816500*	mitochondria	nucleus	nucleus
7	*Os04g0346000*	mitochondria	nucleus	unknown
8	*Os04g0602700*	mitochondria	nucleus	nucleus
9	*Os05g0489100*	nucleus	nucleus	extra cellular
10	*Os06g0138200*	mitochondria	nucleus	nucleus
11	*Os06g0286310*	mitochondria	nucleus	unknown
12	*Os07g0194500*	nucleus	nucleus	Extra cellular
13	*Os09g0489200*	mitochondria	nucleus	nucleus
14	*Os10g0413500*	nucleus	nucleus	Extra cellular
15	*Os10g0420000*	mitochondria	nucleus	unknown
16	*Os10g0497800*	mitochondria	nucleus	Cell membrane
17	*Os11g0488500*	mitochondria	nucleus	nucleus
18	*Os11g0657200*	mitochondria	nucleus	unknown

Table 3-3 Results of motif search retrieved from ELM database

No.	Gene name	ELM					
		Elm Name	**Matched Sequence**	**Positions**	**Elm Description**	**Pattern**	Retrieved more information
1	*Os01g0180900*	LIG_FHA_1	AATVRIR, GITPAVY, PSTFILP, ELTARIS	7-13, 100-106, 172-178, 416-422	Phosphothreonine motif binding a subset of FHA domains that show a preference for a large aliphatic amino acid at the pT+3 position.	..(T).. [ILV].	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS5GIWe&EXPECT_CUTOFF=100&r=1&bg=on
2	*Os01g0502400*	DOC_CYCLIN_1	RRLRL, RHLPP, KELY	43-47, 75-79, 128-131	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes.	[RK].L.{0,1}[FYLIVMP]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS535jt&EXPECT_CUTOFF=100&r=1&bg=on
3	*Os01g0546900*	MOD_PIKK_1	KDSSQKG	348-354	(ST)Q motif which is phosphorylated by PIKK family members.	...([ST])Q..	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS4k9QM&EXPECT_CUTOFF=100&r=1&bg=on

Table 3-3 Results of motif search retrieved from ELM database(continue)

No.	Gene name	ELM					
		Elm Name	**Matched Sequence**	**Positions**	**Elm Description**	**Pattern**	Retrieved more information
4	*Os03g0761900*	DOC_USP7_1	ANASF AAVSS ARESS PRSSW	35-39, 64-68, 70-74, 103-107	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions.	[PA][^P][^FYWIL]S[^P]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS4goVF&EXPECT_CUTOFF=100&r=1&bg=on
5	*Os03g0803500*	DOC_USP7_1	PDVSL	249-253	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions	[PA][^P][^FYWIL]S[^P]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2rYAU&EXPECT_CUTOFF=100&r=1&bg=on
6	*Os03g0816500*	MOD_CK1_1	SPGTPKA, SYQSVEM, SALTLAA	57-63 89-95 134-140	CK1 phosphorylation site	S..([ST])...	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2qO4x&EXPECT_CUTOFF=100&r=1&bg=on
7	*Os04g0346000*	LIG_FHA_2	NVTHLEN	41-47	Phosphothreonine motif binding a subset of FHA domains that have a preference for an acidic amino acid at the pT+3 position.	..(T)..[DE].	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2iYCT&EXPECT_CUTOFF=100&r=1&bg=on

Table 3-3 Results of motif search retrieved from ELM database (continue)

No.	Gene name	ELM					
		Elm Name	**Matched Sequence**	**Positions**	**Elm Description**	**Pattern**	Retrieved more information
8	*Os04g0602700*	DOC_USP7_1	PAPSA, PWKSL, PFVSK	21-25 151-155 185-189	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions.	[PA][^P][^FYWIL]S[^P]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2hVBh&EXPECT_CUTOFF=100&r=1&bg=on
9	*Os05g0489100*	DOC_USP7_1	PHHSR	55-59	The USP7 NTD domain binding motif variant based on the MDM2 and P53 interactions.	[PA][^P][^FYWIL]S[^P]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2eujC&EXPECT_CUTOFF=100&r=1&bg=on
10	*Os06g0138200*	LIG_FHA_1	FLTTWLP, GSTLKIA, EFTGSLP, VPTKRIS	33-39, 348-354, 358-364, 390-396	Phosphothreonine motif binding a subset of FHA domains that show a preference for a large aliphatic amino acid at the pT+3 position.	..(T)..[ILV].	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKUydRvpr&EXPECT_CUTOFF=100&r=1&bg=on
11	*Os06g0286310*	LIG_FHA_1	RDTCYVA	16-22	Phosphothreonine motif binding a subset of FHA domains that show a preference for a large aliphatic amino acid at the pT+3 position	..(T)..[ILV].	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2bmTg&EXPECT_CUTOFF=100&r=1&bg=on

Table 3-3 Results of motif search retrieved from ELM database (continue)

No.	Gene name	ELM					
		Elm Name	**Matched Sequence**	**Positions**	**Elm Description**	**Pattern**	Retrieved more information
12	*Os07g0194500*	LIG_14-3-3_3	RAVSWR	58-63	Consensus derived from reported natural interactors which do not match the Mode 1 and Mode 2 ligands.	[RHK][STALV].([ST]).[PESRDIFTQ]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2agGc&EXPECT_CUTOFF=100&r=1&bg=on
13	*Os09g0489200*	MOD_PIKK_1	ITFSQFL	241-247	(ST)Q motif which is phosphorylated by PIKK family members.	...([ST])Q..	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVS2VfE7&EXPECT_CUTOFF=100&r=1&bg=on
14	*Os10g0413500*	MOD_NEK2_1	FDISVK	253-258	NEK2 phosphorylation motif with preferred Phe, Leu or Met in the -3 position to compensate for less favorable residues in the +1 and +2 position.	[FLM][^P][^P]([ST])[^DEP][^DE]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVRxD7Xz&EXPECT_CUTOFF=100&r=1&bg=on

Table 3-3 Results of motif search retrieved from ELM database (continue)

No.	Gene name	ELM					
		Elm Name	**Matched Sequence**	**Positions**	**Elm Description**	**Pattern**	Retrieved more information
15	*Os10g0420000*	DOC_CYCLIN_1	KGLLP, KKLF	77-81, 231-234	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes.	[RK].L.{0,1}[FYLIVMP]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVRxGQIl&EXPECT_CUTOFF=100&r=1&bg=on
16	*Os10g0497800*	LIG_FHA_1	PWTEVLS	108-114	Phosphothreonine motif binding a subset of FHA domains that show a preference for a large aliphatic amino acid at the pT+3 position.	..(T)..[ILV].	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVRxCCoc&EXPECT_CUTOFF=100&r=1&bg=on
17	*Os11g0488500*	DOC_CYCLIN_1	KPLKP, KILV	28-32, 146-149	Substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes.	[RK].L.{0,1}[FYLIVMP]	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVRxAloh&EXPECT_CUTOFF=100&r=1&bg=on
18	*Os11g0657200*	MOD_CK1_1	SSDSEAD	24-30	CK1 phosphorylation site	S..([ST])...	http://elm.eu.org/cgimodel.py?fun=smartResult&userId=QiKVRx9skM&EXPECT_CUTOFF=100&r=1&bg=on

3.4.4 Mutant analysis

The identification of rice lines that carry mutations in the genes of interest showed that mutants are available for 14 of the 18 genes from different resources. The highest numbers of mutants have been found in Postech Rice T-DNA Insertion Sequence Database. Some databases such as UCD Rice Transposon Flanking Sequence Tag Database, Oryza Tag Line (OTL) Database, Taiwan Rice Insertional Mutants Database (TRIM) and Rice Mutant Database (RMD) also contain mutants in these genes. Focusing on genes having insertion mutation in Table 3-4, for example, in particular *Os01g0180900* can be found mutant lines in Postech and UCB database. The former, this gene was insert 315 bp of T-DNA (transfer DNA) on forward stand at the position of 4268366 to 4268678 whereas the latter, it was insert dSpm (defective suppressor-mutator) with 227 bp at the position of 4268307 to 4268533 on reverse strand (Table 3-4).

One means to identify the function of any gene is to generate mutant lines and subsequently, apply these lines in a study of phenotypic or biochemical change in the mutant and wild type. Practically, many researchers have developed methodology for studying the regulation of selected genes on plant development and adaptation through either comparison of gene expression or direct measurement of biochemical activity in both mutant and wild type. For example, the compatibility genes of *DMR6* (downy mildew resistance 6) encode 2OG-Fe(II) dioxygenase superfamily in Arabidopsis were identified by cloning and characterising both recessive Arabidopsis downy mildew resistance 6 mutant (*dmr6*) and *DMR6*. The result showed that, although the number of sporangiophores per seedling in mutant line was lower than wild type showing that mutation of *DMR6* leads to downy mildew resistance, expression of *DMR6* is induced during plant defence. Furthermore, *dmr6* mutants express enhanced levels of a subset of defence-associated genes, indicating that *DMR6* negatively affects plant

defences (Damme, et al., 2008). Another good example describing how to use a mutant for investigating the regulation of genes is conducting on rice *OsMET1-2 (closely related to* putative *MET1* genes of *Arabidopsis thaliana*) by comparing sensitivity of their genomic DNA of wild type and heterozygote mutant to methylation-sensitive endonuclease digestions, following by locus-specific assay of methylation status by bisulfite sequencing. The result found that compared with wild type and heterozygote mutant (*OsMET1-2*$^{-/-}$*)* was hypersensitive to digestion by HapI, implicating global loss of CG methylation at the 5′ CCGG sites in the mutant (Hu, et al., 2014).

Table 3-4 Gene mutant detail found in 5 database resources retrieved by using OryGenesDB tool

No.	Gene name	Mutant insertion detail					
		Source	Plant name	Orientation	Position	Mutagen	Sequence length (bp)
1	*Os01g0180900*	Postech	D11915	+ strand	OS01:4268366..4268678	T-DNA (RB border)	315
		UCD	RdSpm953A_3.1	- strand	Os01:4268307..4268533	dSpm	227
2	*Os01g0502400*	Postech	A45824	+ strand	Os01:17404322..17406034	T-DNA (RB boarder)	713
		TRIM	M0066742	- strand	Os01:17394482..17394830	T-DNA (N/A)	349
3	*Os01g0546900*	N/A	N/A	N/A	N/A	N/A	N/A
4	*Os03g0761900*	Postech	A09120	+ strand	Os03:32402597..32403425	T-DNA (RB boarder)	829
		TRIM	M0080339	+ strand	Os03:32403938..32404149	T-DNA (N/A)	212
		TRIM	M0117657	+ strand	Os03:32404747..32405008	T-DNA (N/A)	262
		OTL	AQCG11	- strand	Os03:32403118..32403652	Tos17 (RB boarder)	535
5	*Os03g0803500*	Postech	C11491	- strand	Os03:33524132..33524648	T-DNA (LB boarder)	517
		Postech	A31881	- strand	Os03:33523330..33525726	T-DNA(RB boarder)	397
		RMD	02Z15BH36	+ strand	Os03:33522635..33523198	T-DNA (LB boarder)	564
6	*Os03g0816500*	N/A	N/A	N/A	N/A	N/A	N/A
7	*Os04g0346000*	RMD	04Z11FF36	+ strand	Os04:16677580..16677873	T-DNA (RB boarder)	294
		Postech	A16714	- strand	Os04:16674100..16674556	T-DNA (RB boarder)	457
8	*Os04g0602700*	N/A	N/A	N/A	N/A	N/A	N/A
9	*Os05g0489100*	Postech	A03108	+ strand	Os05:24139603..24140108	T-DNA (RB boarder)	506
		TRIM	M0035834	- strand	Os05:24139149..24139495	T-DNA (N/A)	347

Table 3-4 Gene mutant detail found in 5 database resources retrieved by using OryGenesDB tool

No.	Gene name	Mutant insertion detail					
		Source	Plant name	Orientation	Position	Mutagen	Sequence length (bp)
10	*Os06g0138200*	OTL	AQHF02	+ strand	Os06:2031266..2031877	T-DNA (LB boarder)	612
		Postech	B14454	+ strand	Os06:2031927..2032226	T-DNA (LB boarder)	300
11	*Os06g0286310*	Postech	B08673	+ strand	Os06:10341176..10341723	T-DNA (LB boarder)	548
12	*Os07g0194500*	OTL	AMWG09	- strand	Os07:5150094..5150296	T-DNA (LB boarder)	203
		Postech	A05109	- strand	Os04:23856965..23857259	T-DNA (LB boarder)	295
13	*Os09g0489200*	RMD	03Z11CT15	- strand	Os09:19569583..19569985	T-DNA (LB boarder)	403
		RMD	04Z11JF50	+ strand	Os09:19570705..19571202	T-DNA (LB boarder)	498
14	*Os10g0413500*	Postech	D15180	+ strand	Os10:14429435..14429990	T-DNA (RB boarder)	556
		Postech	D13894	- strand	Os10:14430036..14430682	T-DNA (RB boarder)	647
15	*Os10g0420000*	TRIM	M0101007	+ strand	Os10:15273453..15273871	T-DNA (N/A)	419
16	*Os10g0497800*	N/A	N/A	N/A	N/A	N/A	N/A
17	*Os11g0488500*	Postech	B03212	+ strand	Os11:19353131..19353730	T-DNA (RB boarder)	600
		TRIM	M0027636	- strand	Os11:19355172..19355394	T-DNA (N/A)	223
18	*Os11g0657200*	N/A	N/A	N/A	N/A	N/A	N/A

T-DNA = Transfer DNA; dSpm = Defective-suppressor mutator; LB = Left border; RB = Right boarder

3.4.5 Gene annotation profiling

Details of 18 genes including gene name (RAP name), locus name (MSU name), chromosome location, gene position on chromosome, number of exon and coding exon, gene squence (nucleotide) length, mRNA (transcription) sequence length, coding sequence (CDS) length, conserved domain and function of these encoded proteins as well as gene overview have been annotated by using various database resources as mentioned in the materials and methods section of this chapter.

Table 3-5 shows that *Tet* gene homologues are located in all rice chromosomes except chromosomes 2, 8 and 12. Chromosomes 1, 3 and 10 have more (three) genes than other chromosomes. Compared, gene *Os04g0346000* contains the highest number of exons (11). Its exons caries 1072 bp of mRNA length, 804 bp of CDS length and 268 aa of protein sizes. (Table 3-5, Figure 3-7). Focusing on transcription and splice variants, it is found that *Os01g0502400* has two splice forms but only one form is transcripted to protein with 393 amino acids. Furthermore, there are three genes namely *Os06g0138200*, *Os10g0413500* and *Os11g0488500,* presenting that they have two splice forms as well as two transcripts (Table 3-5,Figure 3-7).

As regards the conserved domain, these genes encode proteins having a wide range of conserved domain which can be categorised into six groups. There are:- oxidoreductase conserved domain, 2OG-Fe(II) oxygenase domain containing protein, conserved domain which is similar to transcription factor jumonji (jmjC), conserved domain which is similar to Prolyl 4-hydroxylase, alpha-ketoglutarate-dependent dioxygenase AlkB-like conserved domain and conserved domain which is similar to Prolyl 4-hydroxylase. These conserved domains are associated with proteins that have several molecular functions and biological processes. For example, the conserved domain which is similar to oxidoreductase enzyme

encoded by *Os01g0180900* suggests this protein has activity for iron ion binding (GO:0005506) and oxidoreductase activity (GO:0016706) which acts on a paired donor with incorporation or reduction of molecular oxygen. One atom each of oxygen is incorporated in to both donors, namely 2- oxoglutarate (2OG) and an Fe^{2+} ion. Similarly, Welford and colleagues (2005) reported that 2ODDs catalyse two electron oxidation reactions by incorporating the oxidation of substrate to the oxidative decarboxylation of 2OG, resulting in succinate and carbon dioxide coproducts.

Table 3-5 Gene profiling retrieved from five databases resources

RAP Name/ MSU Name	Location	Position	Exon	Nucleotide (bp)	Transcript (bp)	CDS	Protein (aa)	Conserved Domain/ Type of region	Function
Os01g0180900/ LOC_Os01g08570.1	Chr01	4263181.. 4269134 (+ strand)	7	5954	1687	966	322	Similar to Oxidoreductase.	GO:0005506 GO:0016706 GO:0019538
Os01g0502400/ LOC_Os01g31770.1 LOC_Os01g31770.2	Chr01	17394305.. 17404420 (+ strand)	9, 8	10116	1642	1179	393	2OG-Fe(II) oxygenase domain containing protein	GO:0005506 GO:0016706 GO:0016491 GO:0019538
Os01g0546900/ LOC_Os01g36630.1	Chr01	20322698.. 20326374 (+ strand)	8	3677	1,305	1212	404	Similar to transcription factor jumonji (jmjC) domain-containing protein	GO:0000987 GO:0003700 GO:0005515 GO:0033746 GO:0033749 GO:0010030
Os03g0761900/ LOC_Os03g55380.1	Chr03	31503761.. 31507752 (+ strand)	8	3992	1357	933	311	Similar to Prolyl 4-hydroxylase	GO:0005506 GO:0016706 GO:0019538
Os03g0803500/ LOC_Os03g58890.1 LOC_Os03g58890.2	Chr03	33522573.. 33524525 (- strand)	7	1953	1114	900	300	Similar to Prolyl 4-hydroxylase	GO:0005506 GO:0016706 GO:0019538

Table 3-5 Gene profiling retrieved from five databases resources (continue)

RAP Name/ MSU Name	Location	Position	Exon	Nucleotide (bp)	Transcript (bp)	CDS	Protein (aa)	Conserved Domain/ Type of region	Function
Os03g0816500/ LOC_Os03g60190.1 LOC_Os03g60190.2	Chr03	35132995.. 35136955 (+ strand)	3	3961	2725	1035	345	Alpha-ketoglutarate-dependent dioxygenase AlkB-like	GO:0016706
Os04g0346000/ LOC_Os04g27850.1	Chr04	16672311.. 16679132 (+ strand)	11	6822	1072	804	268	2OG-Fe(II) oxygenase domain containing protein.	GO:0005506 GO:0016491 GO:0019538
Os04g0602700/ LOC_Os04g51360.1	Chr04	30912942 ..30914916 (- strand)	6	1975	1117	1047	349	2OG-Fe(II) oxygenase domain containing protein	GO:0016706
Os05g0489100/ LOC_Os05g41010.1	Chr05	24138032 ..24140913 (+ strand)	7	2882	1649	960	320	Similar to transcription factor jumonji (jmjC) domain-containing protein	GO:0005506 GO:0016491 GO:0016705 GO:0016706 GO:0031418 GO:0055114

Table 3-5 Gene profiling retrieved from five databases resources (continue)

RAP Name/ MSU Name	Location	Position	Exon	Nucleotide (bp)	Transcript (bp)	CDS	Protein (aa)	Conserved Domain/ Type of region	Function
Os06g0138200/ LOC_Os06g04660.1 LOC_Os06g04660.2	Chr06	2029472.. 2033740 (+ strand)	6, 2	4269	2218, 1050	1851, 723	616, 240	Similar to Oxidoreducta -se	GO:0016491 GO:0016706 GO:0055114
Os06g0286310/ LOC_Os06g17830.1	Chr06	10340159.. 10342467 (+ strand)	3	2309	1999	321	107	2OG-Fe(II) oxygenase domain containing protein.	GO:0016491
Os07g0194500/ LOC_Os10g27340.2	Chr07	5148833.. 5151305 (- strand)	6	2473	1125		319	Alpha-ketoglutarate -dependent dioxygenase AlkB-like	GO:0005506 GO:0016491 GO:0016705 GO:0016706 GO:0031418 GO:0055114
Os09g0489200/ LOC_Os09g31380.1	Chr09	19569239.. 19573497 (- strand)	6	4259	1496	1239	413	Similar to transcription factor jumonji (jmjC) domain-containing protein.	GO:0003700 GO:0006355

Table 3-5 Gene profiling retrieved from five databases resources (continue)

RAP Name/ MSU Name	Location	Position	Exon	Nucleotide (bp)	Transcript (bp)	CDS	Protein (aa)	Conserved Domain/ Type of region	Function
Os10g0413500/ LOC_Os10g27340.2	Chr10	14428234.. 14432381 (- strand)	7, 4	4148	1127, 1147	927, 609	308, 202	2OG-Fe(II) oxygenase domain containing protein	GO:0005506 GO:0016706 GO:0016491 GO:0019538
Os10g0420000/ LOC_Os10g28410.1 LOC_Os10g28410.2	Chr10	15273657.. 15276162 (- strand)	6	2506	989	711	237	Similar to transcription factor jumonji (jmjC) domain-containing protein	GO:0016491 GO:0016706
Os10g0497800/ LOC_Os10g35470.1	Chr10	19478522.. 19482368 (+ strand)	8	3847	1371	966	322	Similar to Prolyl 4-hydroxylase	GO:0005506 GO:0016706 GO:0019538 GO:0016491

Table 3-5 Gene profiling retrieved from five databases resources (continue)

RAP Name/ MSU Name	Location	Position	Exon	Nucleotide (bp)	Transcript (bp)	CDS	Protein (aa)	Conserved Domain/ Type of region	Function
Os11g0488500/ LOC_Os11g29690.1	Chr11	17206889.. 17210071 (+ strand)	6, 4	3183	1433	978, 696	326, 155	Alpha-ketoglutarate -dependent dioxygenase AlkB-like	GO:0016491 GO:0016706 GO:0055114
Os11g0657200/ LOC_Os11g43610.1	Chr11	28510461.. 28512375 (- strand)	4	1915	1020	780	260	Similar to Prolyl 4-hydroxylase	GO:0016491 GO:0016706 GO:0055114

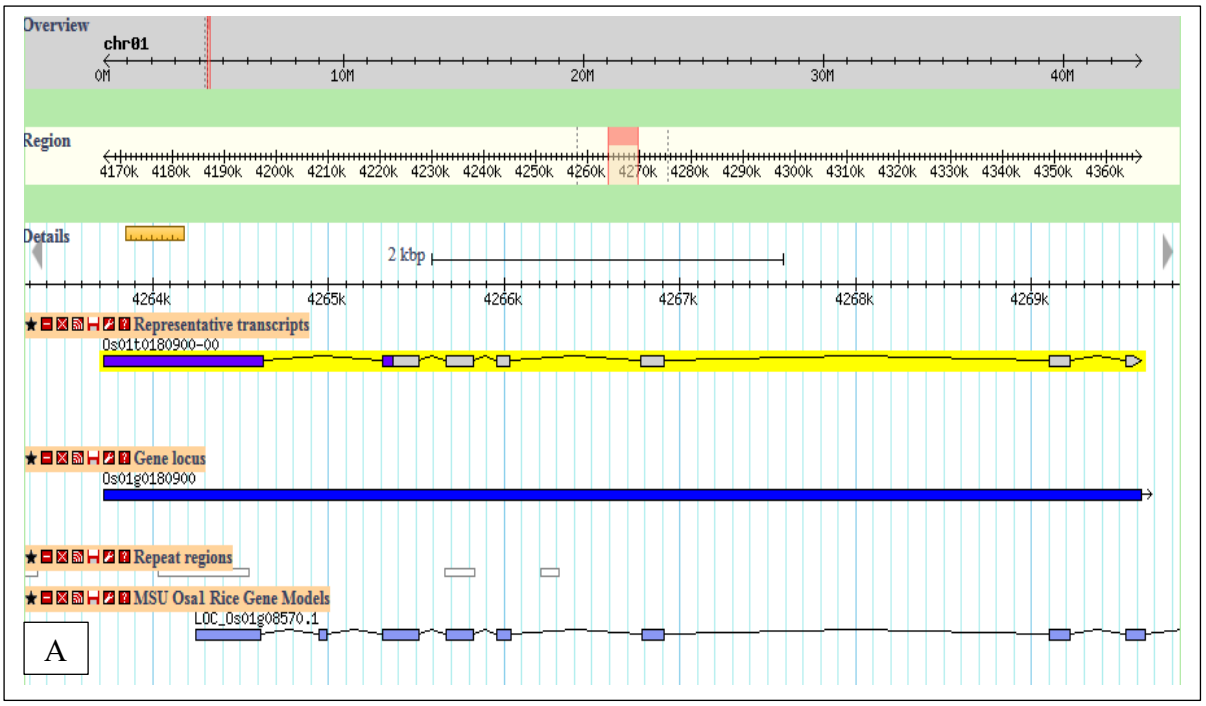
Go number definition is given below:

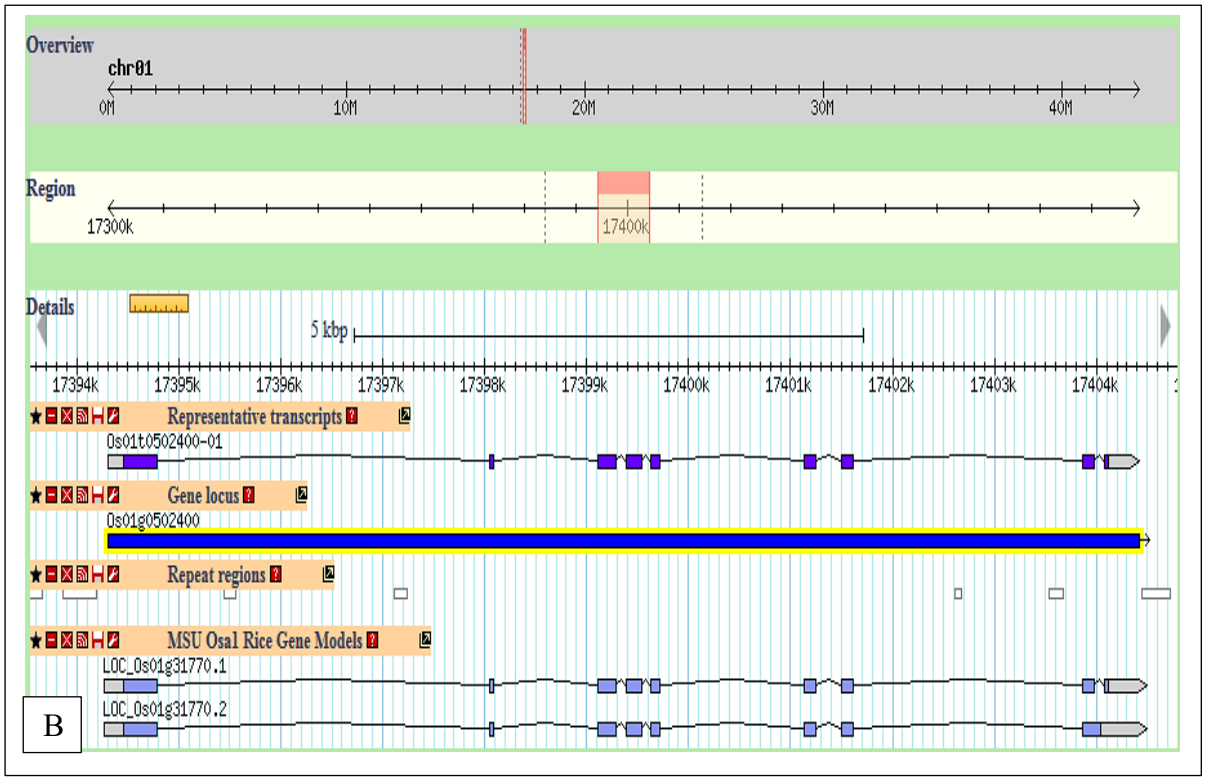
Molecular Function

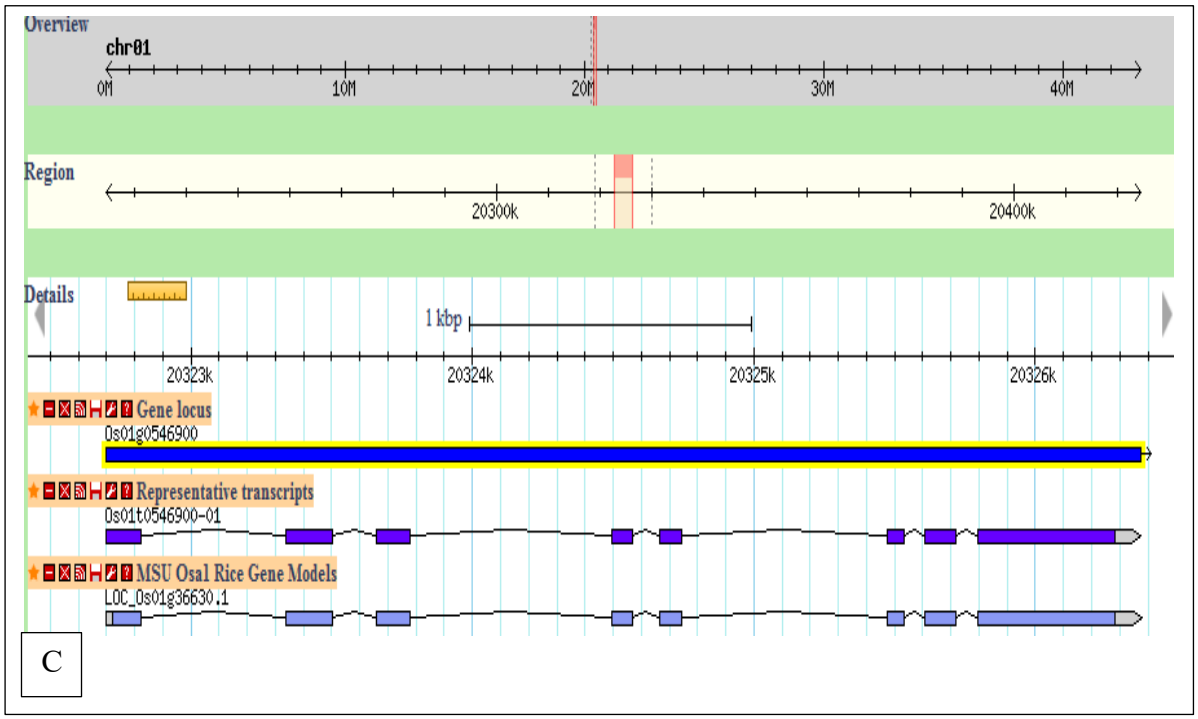
GO:0000987 = core promoter proximal region sequence-specific DNA binding
GO:0003700 = sequence-specific DNA binding transcription factor activity
GO:0005506 = iron ion
GO:0005515 = protein binding
GO:0016706 = oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2OG as one donor, and incorporation of one atom each of oxygen into both donors
GO:0016491 = oxidoreductase activity
GO:0031418 = L-ascorbic acid binding
GO:0033746 = histone demethylase activity (H3-R2 specific)
GO:0033749 = histone demethylase activity (H4-R3 specific)
GO:0003700 = sequence-specific DNA binding transcription factor activity

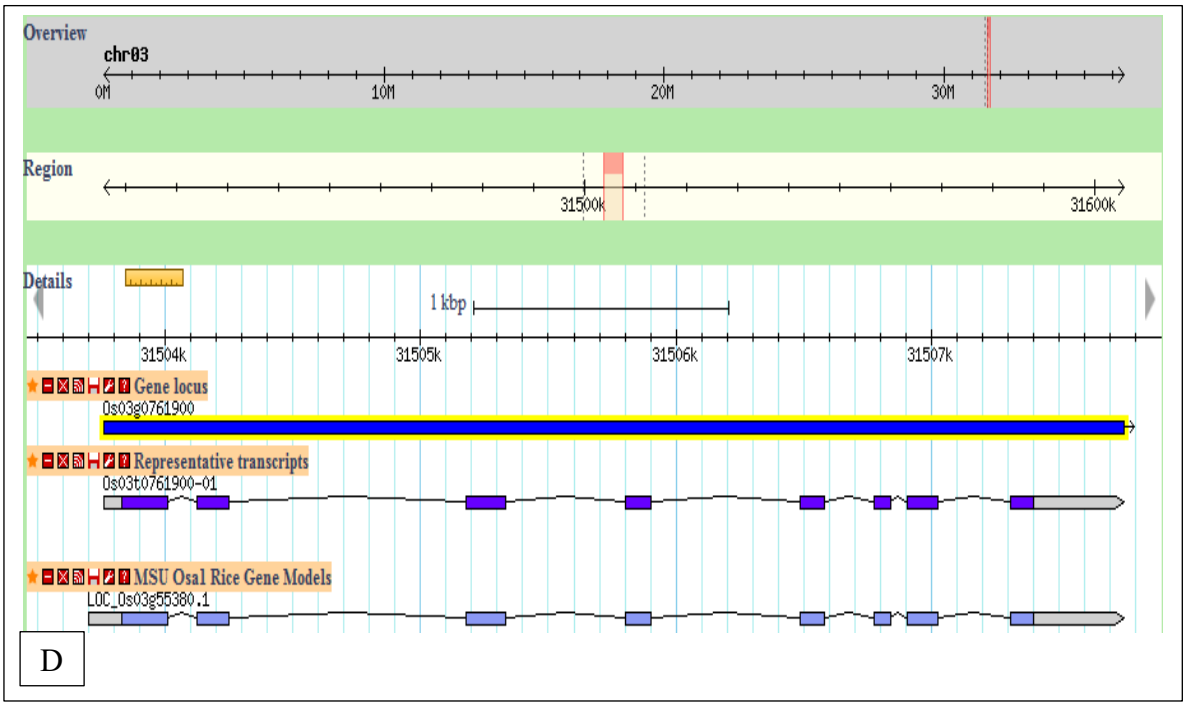
Biological Process

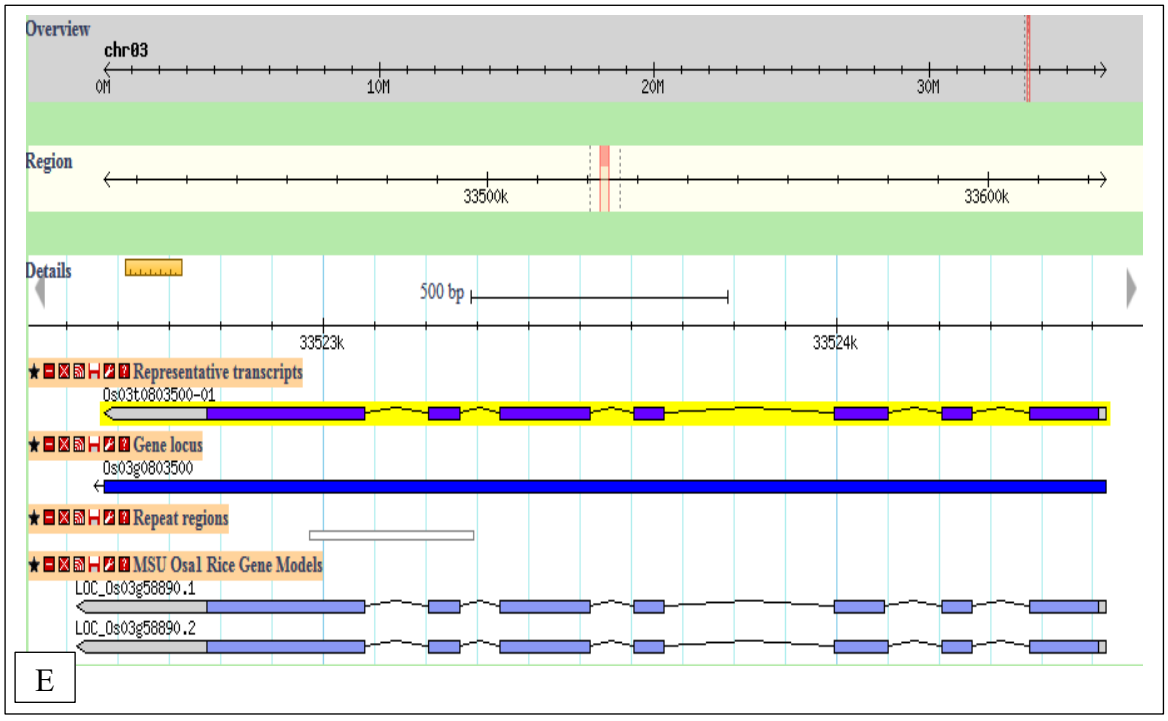
GO:0010030 = positive regulation of seed germination
GO:0019538 = protein metabolic process
GO:0055114 = oxidation-reduction process
GO:0006355 = regulation of transcription, DNA-template
RAP Name = Gene name obtained from Rice Annotation Project database
MSU Name = Gene name obtained from Michigan state University Rice Annotation Project database

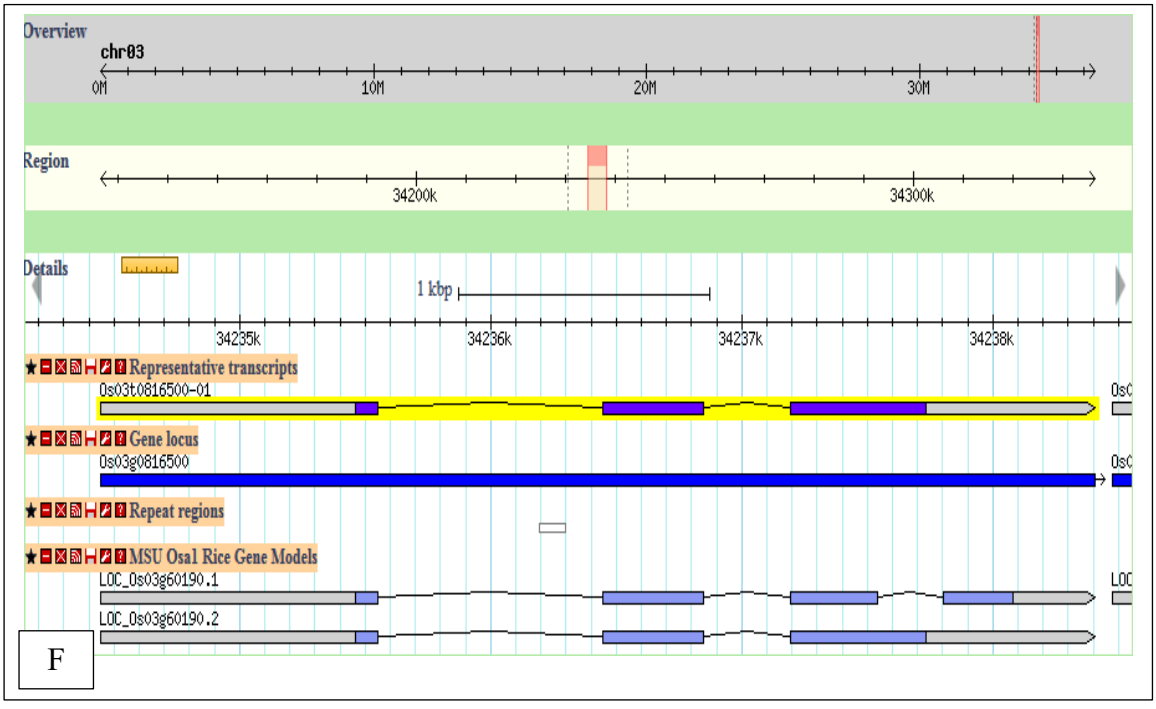

Overview
chr01
0M
10M
20M
30M
40M
Region
4170k 4180k 4190k 4200k 4210k 4220k 4230k 4240k 4250k 4260k 4270k 4280k 4290k 4300k 4310k 4320k 4330k 4340k 4350k 4360k
Details
2 kbp
4264k
4265k
4266k
4267k
4268k
4269k
Representative transcripts
Os01t0180900-00
Gene locus
Os01g0180900
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os01g08570.1
A

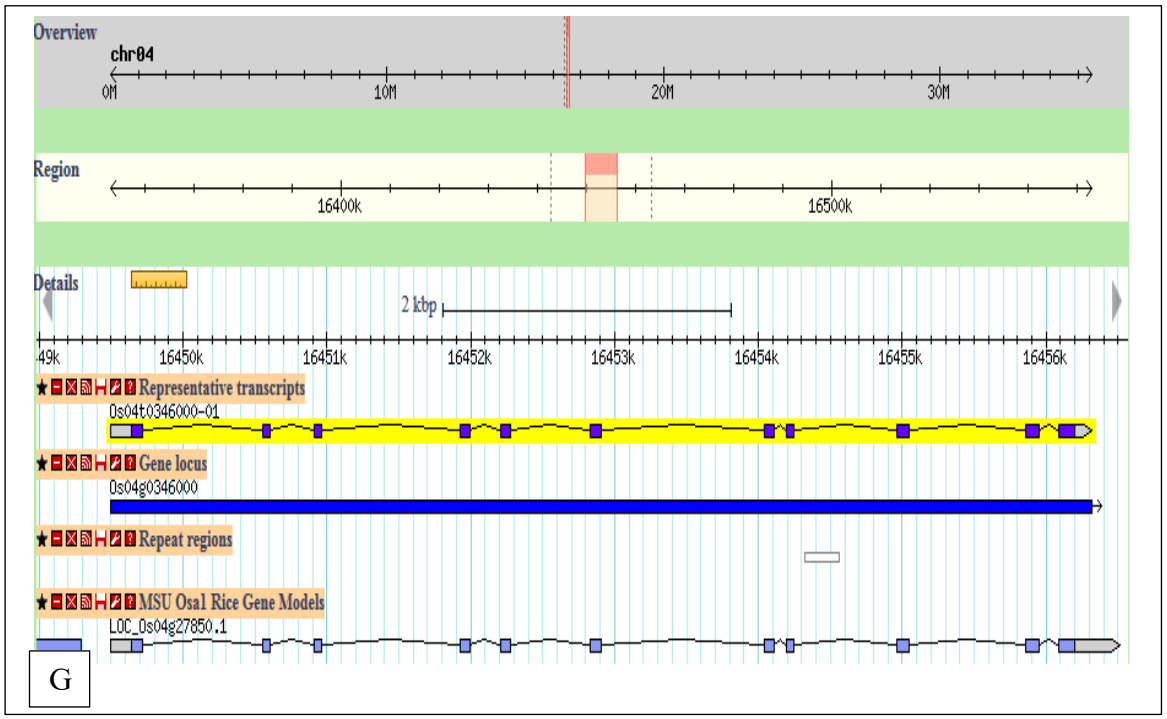

Overview
chr01
0M
10M
20M
30M
40M
Region
17300k
17400k
Details
5 kbp
17394k
17395k
17396k
17397k
17398k
17399k
17400k
17401k
17402k
17403k
17404k
Representative transcripts
Os01t0502400-01
Gene locus
Os01g0502400
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os01g31770.1
LOC_Os01g31770.2
B

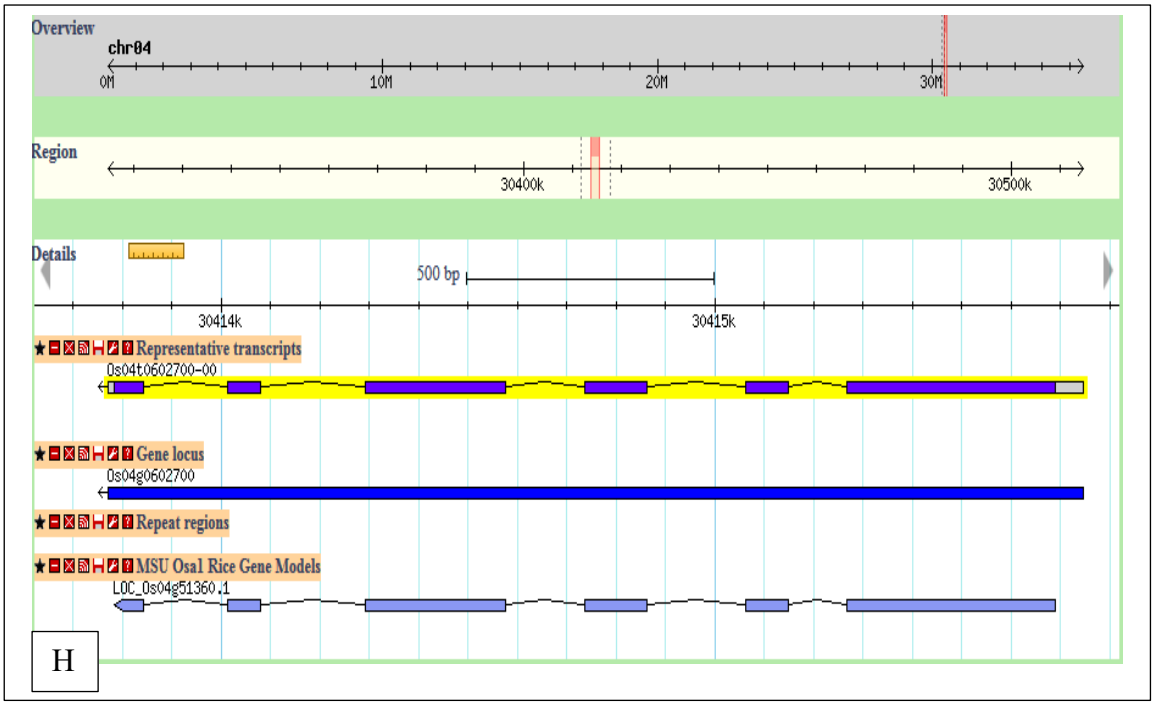

Overview
chr01
0M
10M
20M
30M
40M
Region
20300k
20400k
Details
1 kbp
20323k
20324k
20325k
20326k
Gene locus
Os01g0546900
Representative transcripts
Os01t0546900-01
MSU Osa1 Rice Gene Models
LOC_Os01g36630.1
C

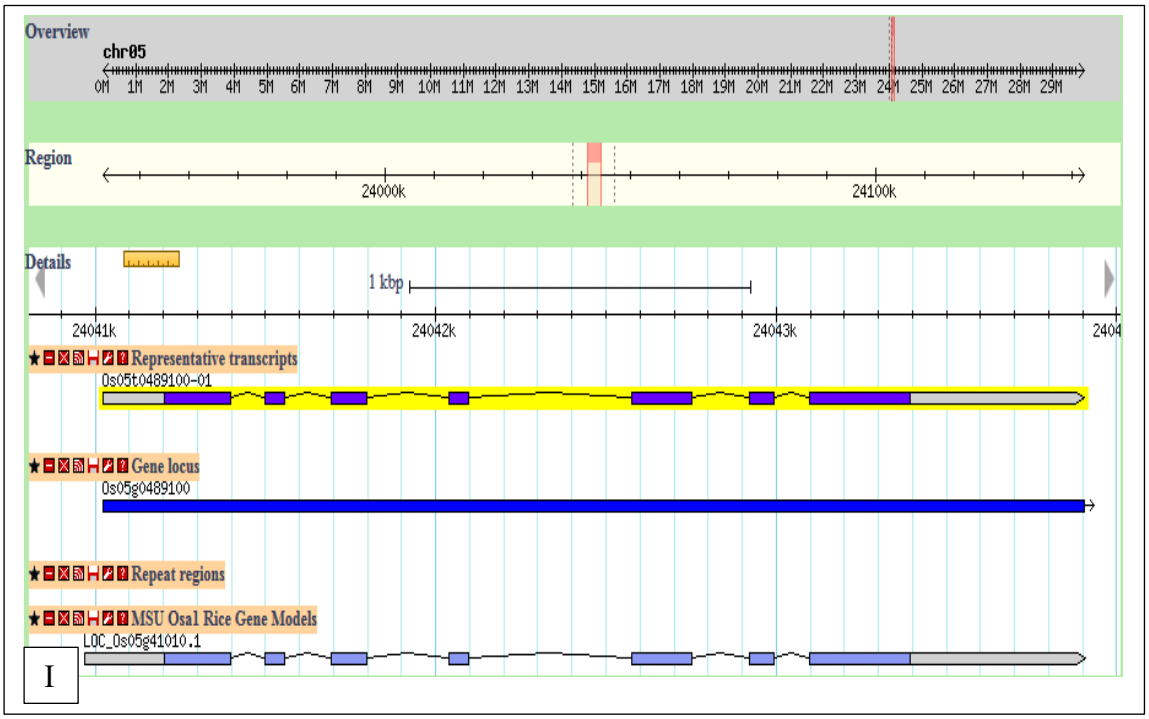

Overview
chr03
0M
10M
20M
30M
Region
31500k
31600k
Details
1 kbp
31504k
31505k
31506k
31507k
Gene locus
Os03g0761900
Representative transcripts
Os03t0761900-01
MSU Osa1 Rice Gene Models
LOC_Os03g55380.1
D

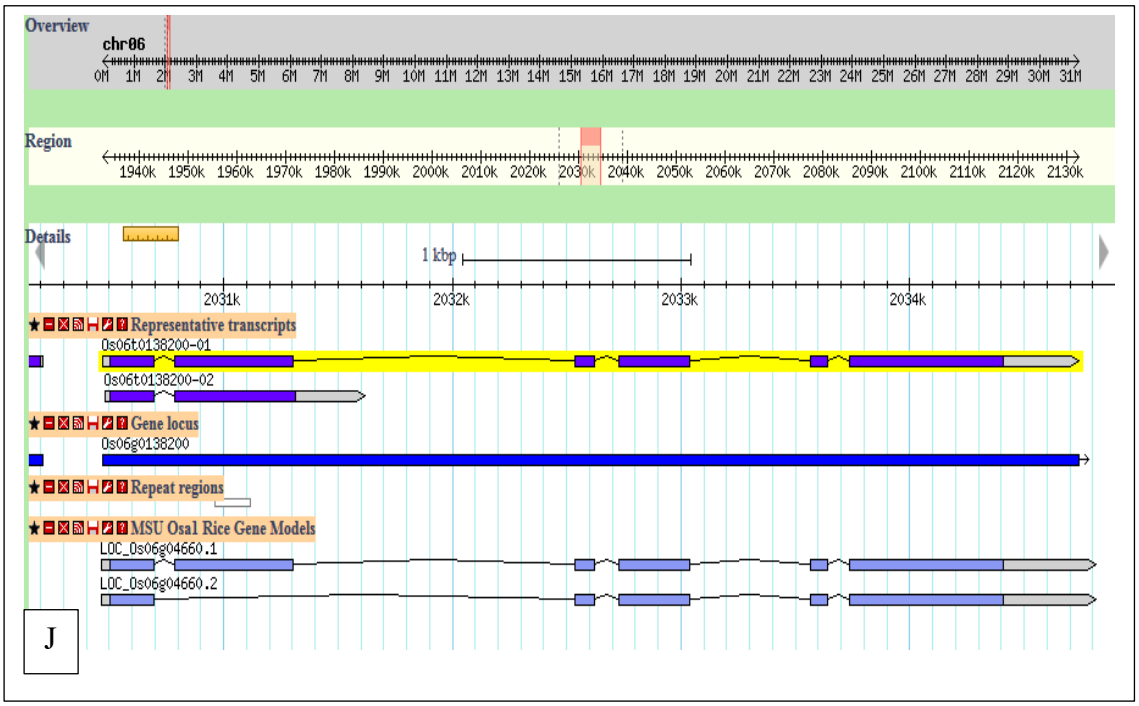

Overview
chr03
0M
10M
20M
30M
Region
33500k
33600k
Details
500 bp
33523k
33524k
Representative transcripts
Os03t0803500-01
Gene locus
Os03g0803500
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os03g58890.1
LOC_Os03g58890.2
E

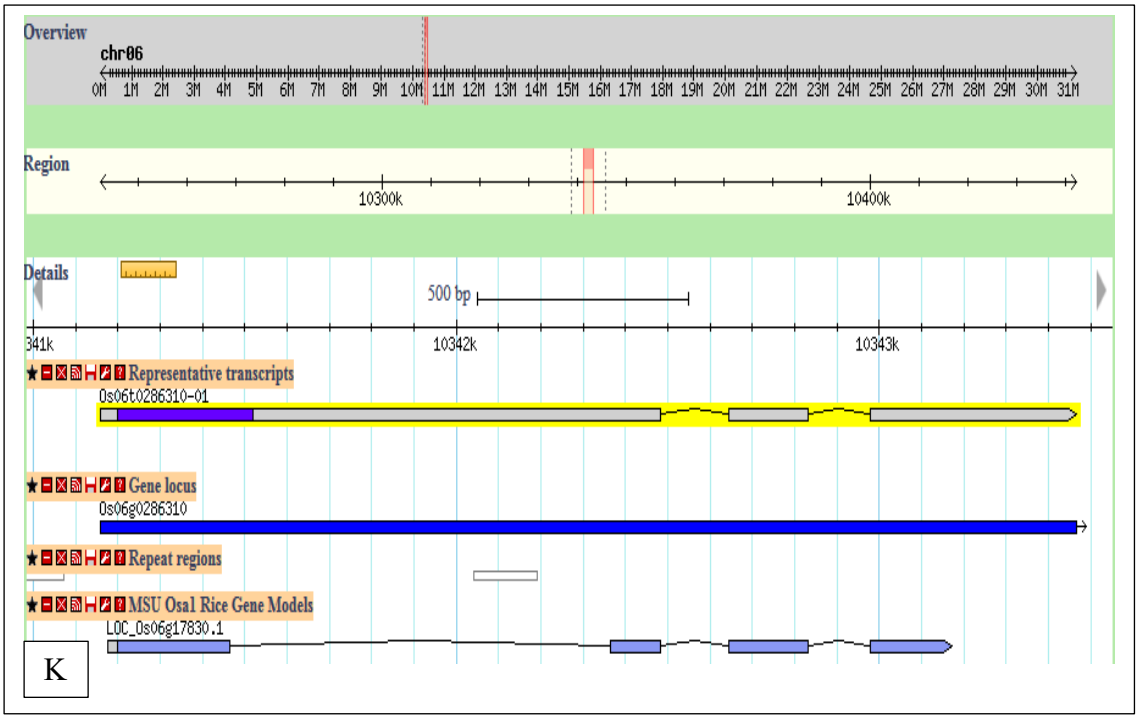

Overview
chr03
0M
10M
20M
30M
Region
34200k
34300k
Details
1 kbp
34235k
34236k
34237k
34238k
Representative transcripts
Os03t0816500-01
Gene locus
Os03g0816500
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os03g60190.1
LOC_Os03g60190.2
F


Overview
chr04
0M
10M
20M
30M
Region
16400k
16500k
Details
2 kbp
49k
16450k
16451k
16452k
16453k
16454k
16455k
16456k
Representative transcripts
Os04t0346000-01
Gene locus
Os04g0346000
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os04g27850.1
G

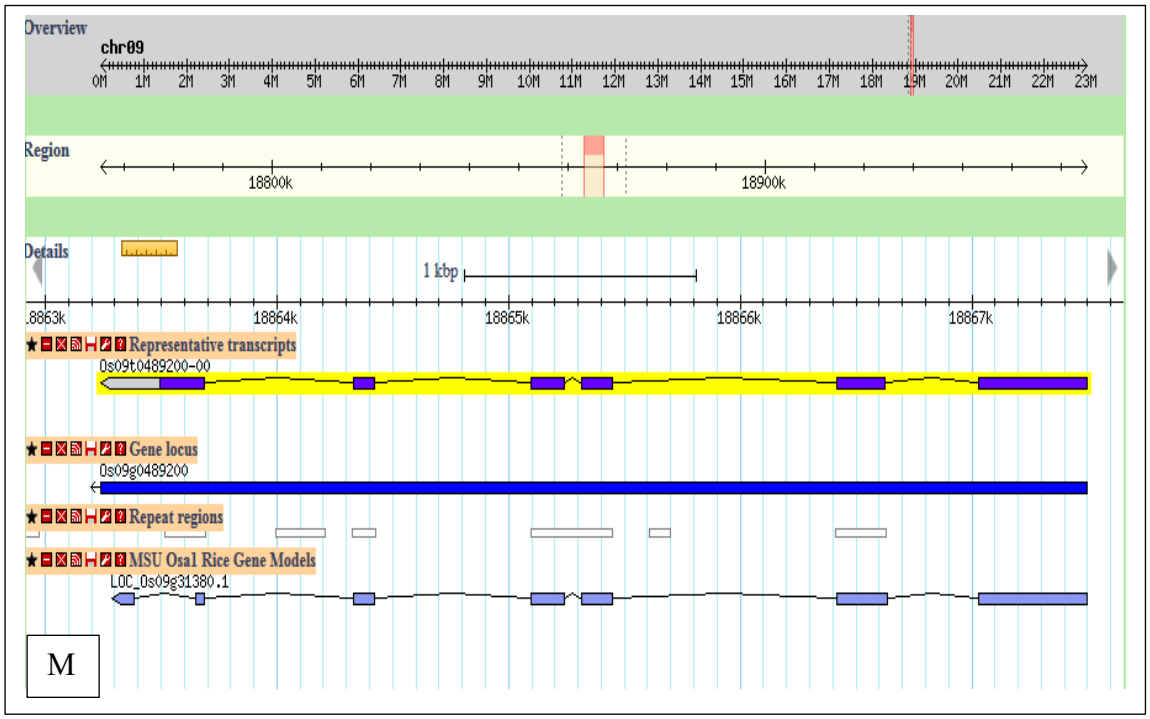

Overview
chr04
0M
10M
20M
30M
Region
30400k
30500k
Details
500 bp
30414k
30415k
Representative transcripts
Os04t0602700-00
Gene locus
Os04g0602700
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os04g51360.1
H

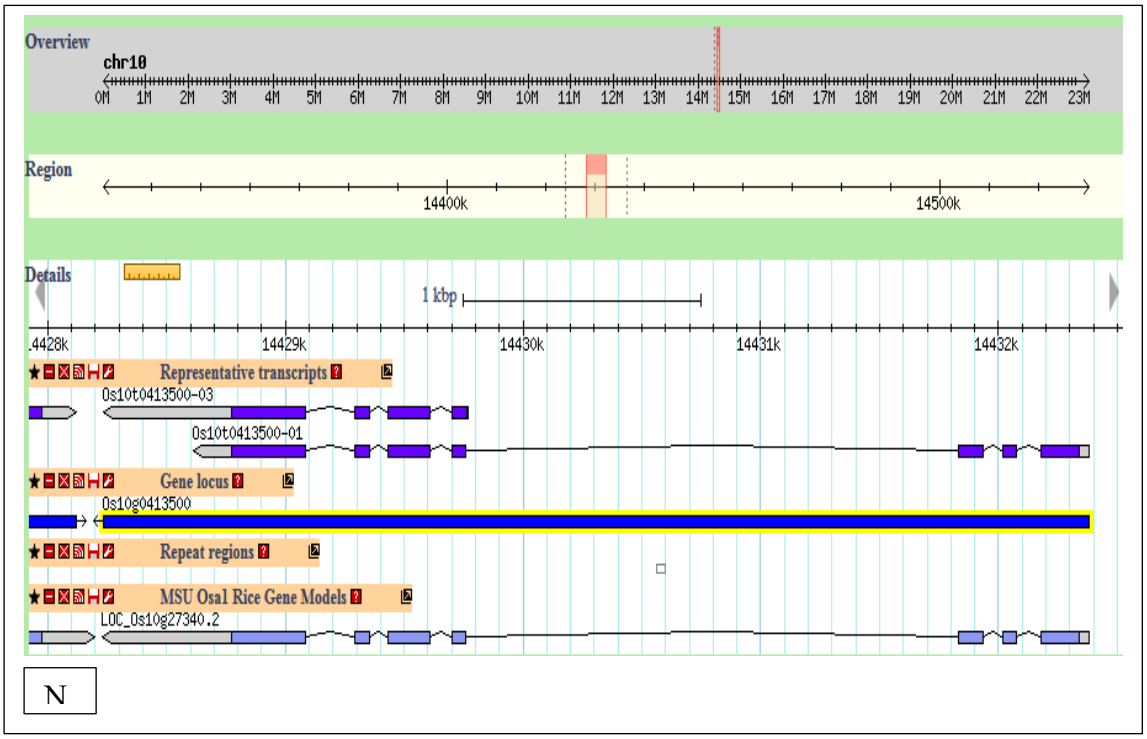

Overview
chr05
Region
24000k
24100k
Details
1 kbp
24041k
24042k
24043k
Representative transcripts
Os05t0489100-01
Gene locus
Os05g0489100
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os05g41010.1
I

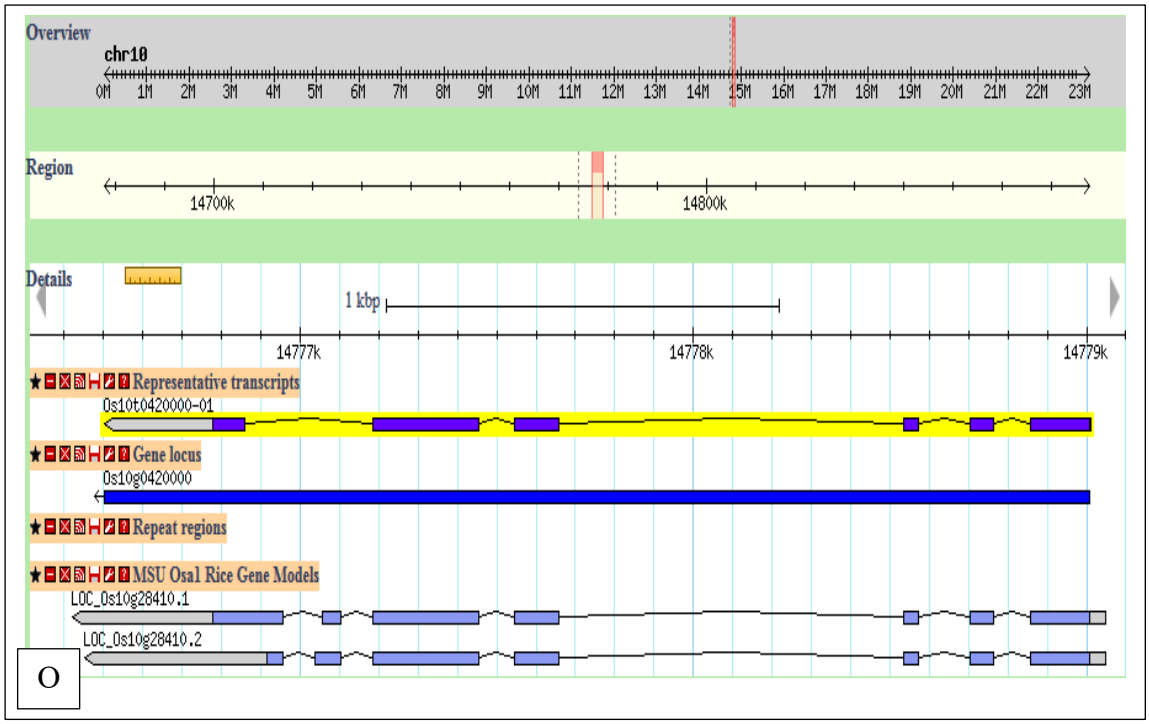

Overview
chr06
Region
Details
1 kbp
2031k
2032k
2033k
2034k
Representative transcripts
Os06t0138200-01
Os06t0138200-02
Gene locus
Os06g0138200
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os06g04660.1
LOC_Os06g04660.2
J

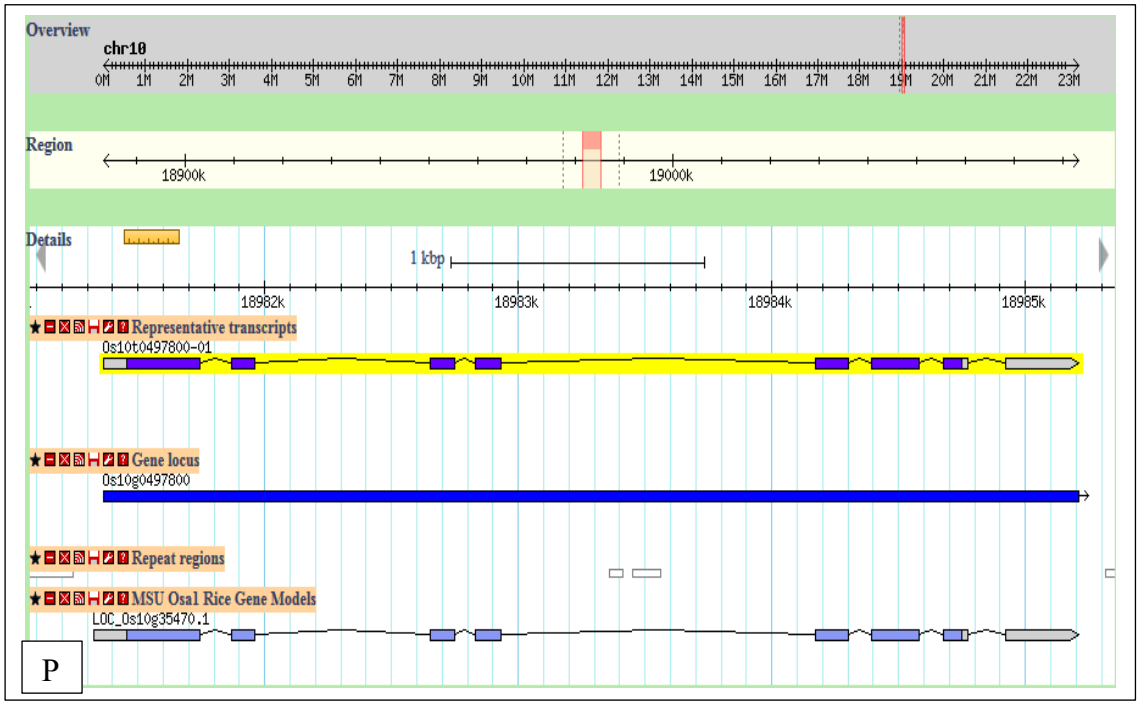

Overview
chr06
Region
10300k
10400k
Details
500 bp
341k
10342k
10343k
Representative transcripts
Os06t0286310-01
Gene locus
Os06g0286310
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os06g17830.1
K

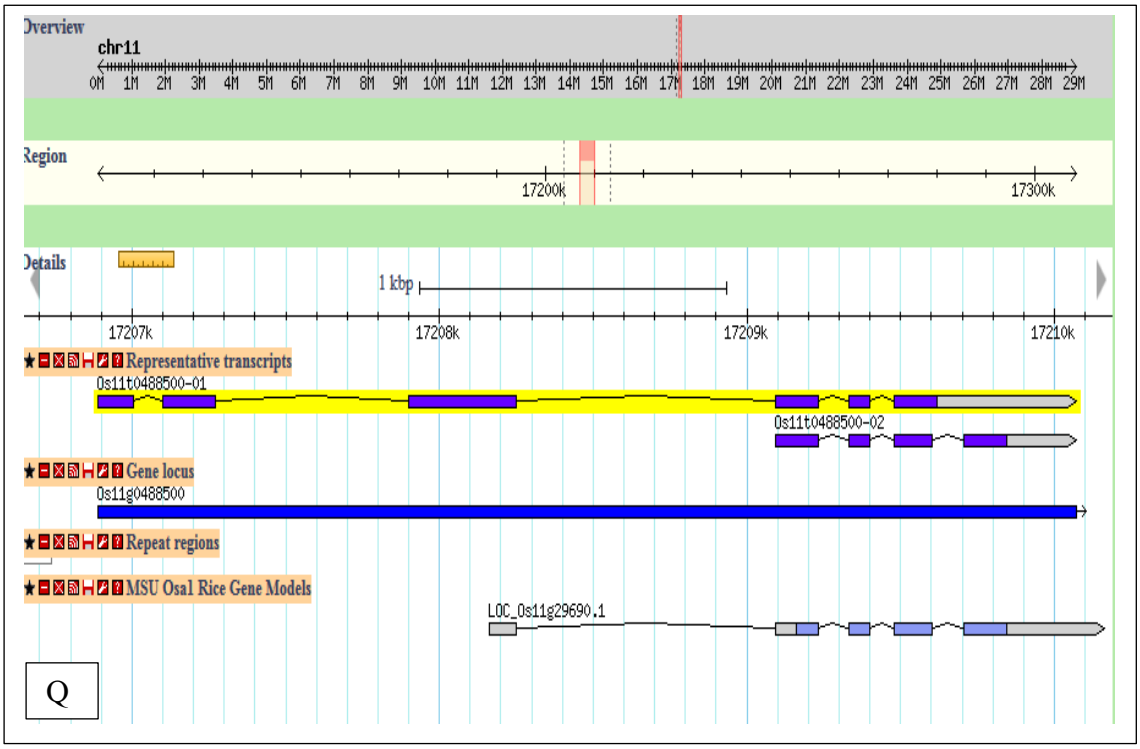

Overview
chr07
Region
Details
500 bp
5117k
5118k
5119k
Representative transcripts
Os07t0194500-01
Os07t0194550-00
Gene locus
Os07g0194500
Os07g0194550
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os07g09630.1
L


Overview
chr09
0M 1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M 16M 17M 18M 19M 20M 21M 22M 23M
Region
18800k
18900k
Details
1 kbp
.8863k
18864k
18865k
18866k
18867k
Representative transcripts
0s09t0489200-00
Gene locus
0s09g0489200
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os09g31380.1
M


Overview
chr10
0M 1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M 16M 17M 18M 19M 20M 21M 22M 23M
Region
14400k
14500k
Details
1 kbp
.4428k
14429k
14430k
14431k
14432k
Representative transcripts
0s10t0413500-03
0s10t0413500-01
Gene locus
0s10g0413500
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os10g27340.2
N


Overview
chr10
0M 1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M 16M 17M 18M 19M 20M 21M 22M 23M
Region
14700k
14800k
Details
1 kbp
14777k
14778k
14779k
Representative transcripts
Os10t0420000-01
Gene locus
Os10g0420000
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os10g28410.1
LOC_Os10g28410.2
O


Overview
chr10
0M 1M 2M 3M 4M 5M 6M 7M 8M 9M 10M 11M 12M 13M 14M 15M 16M 17M 18M 19M 20M 21M 22M 23M
Region
18900k
19000k
Details
1 kbp
18982k
18983k
18984k
18985k
Representative transcripts
Os10t0497800-01
Gene locus
Os10g0497800
Repeat regions
MSU Osa1 Rice Gene Models
LOC_Os10g35470.1
P



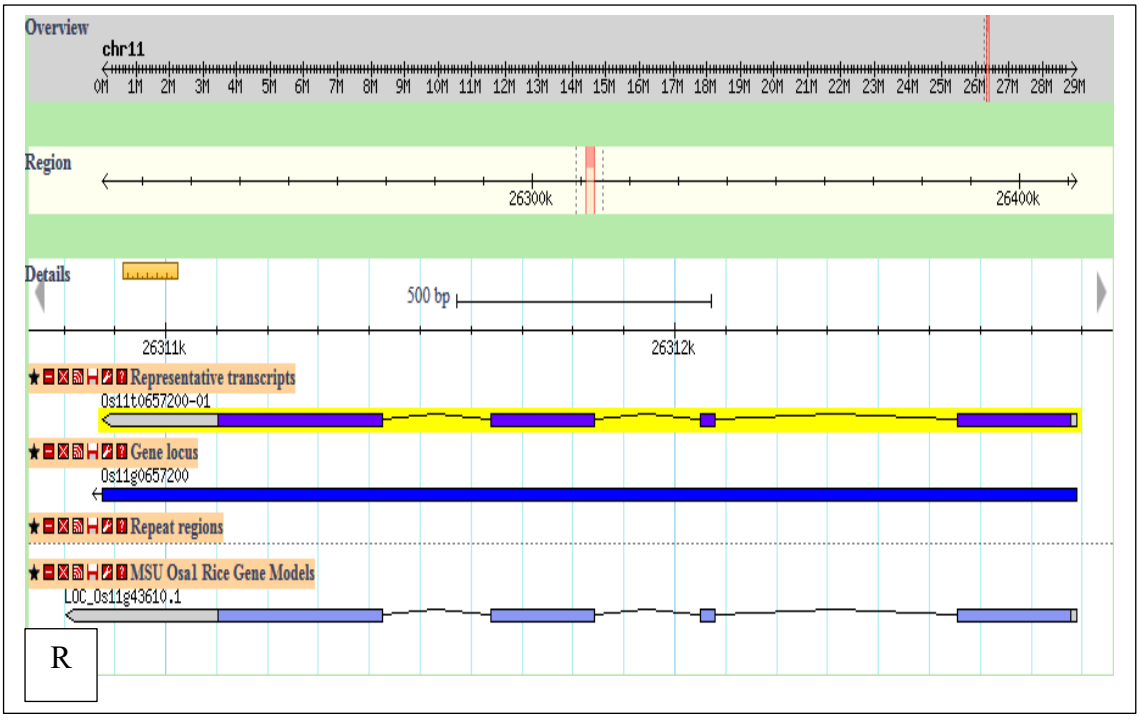


Figure 3-7 Panels A-R providing a gene over view showing genome region, encoded transcript and gene locus of 18 selected rice genes

3.5 Summary

This chapter has started with the investigation of *Tet* gene homologues in order to obtain gene name, and protein sequences by using various databases. Here, we have identified 18 genes which encode proteins having a 2ODD domain. Moreover, additional relevant information such as the relatedness of protein sequence, protein interaction, nuclear localisation prediction, mutant analysis as well as gene annotation also have been given with the aim to assemble and apply such gene information to aid both analysis and synthesis in the subsequent study.

Chapter 4

The expression of 2ODD family genes in various organs

4.1 Introduction

Recently, a study of proteins involved in processing of active DNA demethylation by oxidation has been widely described. The TET family of proteins, namely TET1, TET2 and TET3 are confirmed to be catalysts for this process. Mohr et al. (2011) reported that three members of TET family contain 1) three metal binding residues (TET1: amino acid (AA) 1672, 1674, 2028; TET2: 1382, 1384, 1884; TET3: 942, 944, 1538) which enable the protein to bind an Fe (II)-ion and thus form a catalytically active site, 2) a 2-oxoglutarate (2OG) binding site which is linked to a single conserved AA. Both of these sites form a double-strand β-helix (DSBH) (cupin) domain which folds with the characteristic of a 2ODD and is required for catalytic activities. 3) Moreover, TET 1 and TET3 also have a CXXC domain which binds the CpG island of DNA and subsequently causes a tight binding between TET1 and DNA (Figure 4-1). In addition, TET1 contains three nuclear localization signals (NLS) which ensure that it has a potential to function in the nucleus.

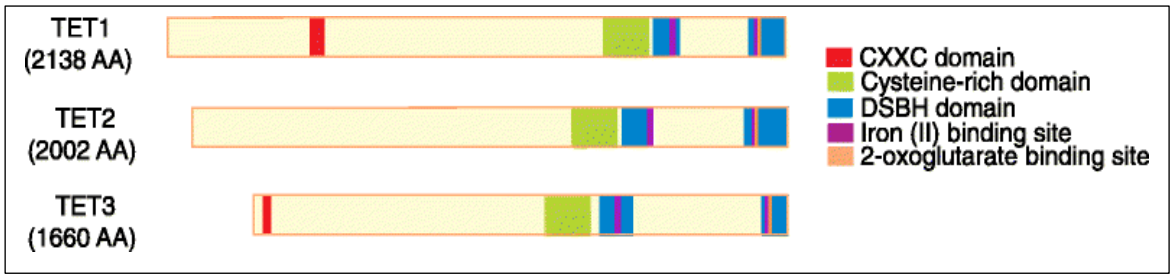


Figure 4-1 Primary structure and function of human TET proteins. A double-stranded β helix (DSBH), a cysteine-rich domain, and one 2-oxoglutarate and three iron (II) binding sites in the carboxyl terminus, which constitute their dioxygenase catalytic domain are presented in all TET protein whereas an amino-terminal CXXC zinc finger domain is only identified in TET1 and TET3, allowing these enzymes to bind DNA directly to CpG. (Scourzic, et., 2015)

The finding of TET genes in particular mammalian cells brings an understanding of the effect of TET family proteins in specific cells. Mohr et al. (2011) revealed that stem cells are able to perpetuate themselves through self-renewal and to generate functional mature cells by differentiation. The balance between self-renewal and differentiation is maintained by crucially coordinated gene expression. However, a variety of malignant disorders can be generated by disturbance of this balanced system. TET1 is implicated in the control of balanced gene expression. In murine embryonic stem cell, TET1 generates demethylation on the Nanog promoter and then leads to self-renewal or an undifferentiated state in the cell. Knockdown of *TET 1* causes TET 1 depletion which consequently brings about hypermethylation of the nanog promoter and then a decrease in *nanog* expression. Further, upregulation of markers (*Cdx2, Hand1, Gata6 and Gata4*) related to differentiation is observed and ES cells differentiate to specific cell types (Figure 4-2).

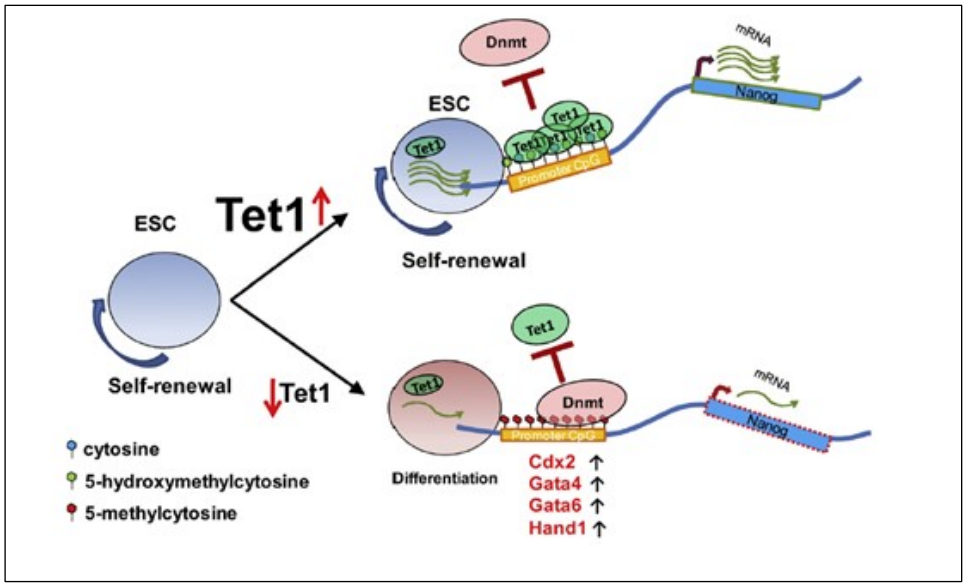


Figure 4-2 The regulation of murine embryonic stem cell (ESC) by TET1 (Mohr, et al., 2011)

The function of TET1 was also confirmed by Koh et al. (2011) who reported that *TET1*-depleted ESCs generate aggressive hemorrhagic teratomas with reduced neuroectoderm and increased endoderm and TET1 also functions to regulate the lineage differentiation potential of ESCs.

In addition, a well-known example of DNA demethylation is a loss of 5-mC in paternal DNA after fertilization of mammals whereas the erasure of the methylation mark in maternal DNA proceeds through passive DNA demethylation. Moreover, the increase of 5-hmC level in the fertilised egg also correlates with the decrease of 5-mC, while the female pronucleus remains methylated and contains low levels of 5-hmC (Kriukien, et al., 2012).

It was shown recently that a correlation exists between DNA methylation and *TET 2* by studying patients with chronic myelomonocytic leukemia (CMML). Such results show that *TET2* mutation is associated with 5-mC and 5-hmC in CMML presenting a higher degree of hypermethylation (Perez, et al., 2012).

In addition, Tet 3 also was confirmed to facilitate the loss of 5-mC, which proceeds before the first cell division and goes through the Tet3-dependent conversion of 5-mC to 5-hmC and later to both 5-fC and 5-caC. In particular, further 5-mC oxidative forms are relatively stable and persist to at least the four-cell stage.

Mounting evidence has been reported that TET1 and TET2 are also stable partners of O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT) in the nucleus of ESCs. OGT-TET1 complexes modulate chromatin conformation and gene expression at CpG-rich transcription start sites (TSSs) (Figure 4-3). Loss of TET1 significantly decreases OGT binding to chromatin, displaces OGT binding from its target gene and increases the expression of OGT-TET1 common target genes. However, the absence of OGT activity promotes 5-hmC accumulation (Balasubramani and Rao, 2013; Vella, et al., 2013).

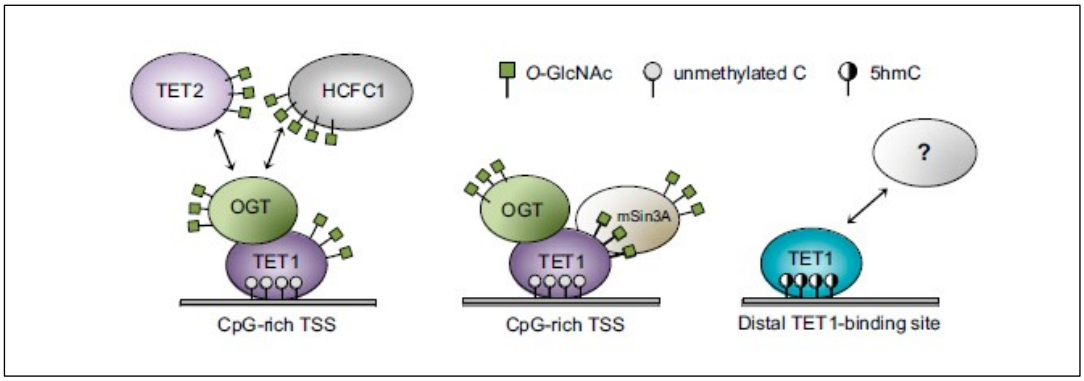


Figure 4-3 TET-OGT complexes modulate chromatin conformation and gene expression at CpG-rich TSSs in mouse ESCs (Balasubramani and Rao, 2013).

Very recent report studied in mouse also confirmed that the disturbance of TET2 leads to the increase of multipotent as well as myeloid progenitors, resulting in the accumulation of premalignant clones. Furthermore, TET proteins are also related to chromatin modifications and other cellular processes through the interaction with O-linked β-N-acetylglucosamine transferase (Nakajima and Kunimoto, 2014 ; Scourzic, et al., 2015)

It can be seen that all evidence mentioned in previous paragraphs suggests that TET family members are responsible to the reversible DNA methylation through oxidative demethylation in mammals and many metazoans, whereas this modification still has not been confirmed in plants. Although, there have been several investigations to determine whether or not oxidative demethylation is present in plants, no definitive conclusion has been reached.

Notably, some research groups suggest the presence in rice of 5-hmC, the oxidative product of 5-mC. For example, Wang and colleagues (2015) provide evidence of 5-hmC levels in DNA from different rice tissues using a dot-blot assay, followed by liquid chromatography-multi-stage mass spectrometry (LC-MS/MS/MS). The results showed that its level also has large inter-tissue variation. In addition the genome-wide profiles of 5-hmC modification in three different rice cultivars using a sensitive chemical labelling followed by next-generation

sequencing method was obtained. It found that thousands of 5-hmC peaks were identified, and a comparison of the distributions of 5-hmC among different cultivars shown the specificity and conservation of 5-hmC modification. Such evidence supports the need to continue the study of possible *Tet* gene homologues in rice.

4.2 Objectives

Here, we has started to identify the expression of selected genes from Chapter 3 in order to obtain a fundamental information of the expression of these genes in various tissues leading to subsequent investigation of their potential involvement in growth, development, and adaptation of plants.

4.3 Materials and methods

4.3.1 Rice organ collection

Fertile wild type seeds of IR64 and IR64.S1 varieties of *O. sativa* Indica and two varieties of *O. sativa* Japonica including Gleva and Nipponbare were grown following the protocol described in Chapter 3.

4.3.2 RNA isolation

RNA from various organs was extracted by using RNeasy Plant Mini Kit (Qiagen, UK) as mentioned in Chapter 3.

4.3.3 RNA quantification

Quantification of RNA samples was conducted by performing spectrophotometry according the instructions described in Chapter 3.

4.3.4 RNA qualification

In addition to using spectrophotometry to examine the purity of RNA, another way to ensure the purity of RNA is by running gel electrophoresis in which all macromolecules and their fragments can be separated based on size and charge. The steps applying for gel electrophoresis include gel preparation and gel electrophoresis, the deatils of which were described in Chapter 3 but used HyperLadder^{TM} I (200-10000 bp) instead of Hyperladder IV (100-1000 bp)

4.3.5 cDNA synthesis

RNA samples from different varieties and various organs were used to synthesise cDNA by using SuperScript® III First-Strand Synthesis SuperMix for qRT-PCR Kit (Invitrogen, UK). The protocol for synthesising cDNA is described in Chapter 3.

4.3.6 Primer designing for amplifying genes

For designing primers which would be specific for amplification from cDNA rather than genomic DNA, exon positions were identified by tracking mRNAs sequence and features through the NCBI database and used as a sequence source for selecting forward and reverse primers which are from different exons, via the primer3 online programme (http://www.ncbi.nlm.nih.gov/), in order to get the effective primers which can discriminate the amplicon from cDNA and gDNA by size (Table 4-1).

Table 4-1 Primer details

No.	Gene		Primer details									Product			
	Name	Number of Exon	Primer name	Sequence (5'->3')	Length	Start	Stop	Tm (°C)	Primer location		Annealing temperature (℃)	Estimated Product		PCR product	
									Exon No.	Exon Junction		Size on mRNA location	Size on gDNA location	Size	Location
1	*Os01g0180900*	9	1FP	CATGGTAGTCACAGGCATGG	20	46	65	59.98	7	-	55	176	492	176	mRNA
			1RP	TCCTGAAGATGTTGCACCTG	20	221	202	59.83	8	-					
2	*Os01g0502400*	9	2FP	GGGAAAGACAGGGATGTTGA	20	763	782	59.90	5	-	55	161	>1.2 kb	161	mRNA
			2RP	TCGAACATTTCCTCATGCTG	20	923	904	59.80	6,7	-					
3	*Os01g054690*	8	3FP	ATGTCGATGCGGGAGTTTGT	20	26	45	60.04	1	-	55	112	614	112	mRNA
			3RP	AACCAGGGTACTCCTTCACA	20	137	118	57.89	2	124/125					
4	*Os03g0761900*	6	4FP	TAGACATCATGGGGAGCCCT	20	618	637	59.73	4	633/634	55	123	895	123	mRNA
			4RP	AAAGACGCGACCCTTTGACT	20	740	721	59.89	6	729/730					
5	*Os03g0803500*	7	5FR	GCAGTGAAGCCAAAGAAAGG	20	233	252	59.99	6,7	-	55	170	295	170	mRNA
			5RP	AGGCTTACATCCGGAGGATT	20	402	383	59.92	7	-					
6	*Os03g0816500*	5	6FP	ACTTTCCAAGGAGGCAGCTC	20	1056	1075	59.96	3	-	55	101	400	101	mRNA
			6RP	AGCGTCTGATAATGTCAACCTGA	23	1156	1134	59.80	4	1137/1138					
7	*Os04g0346000*	11	7FP	GGTGAAACCCATTTTCCTCA	20	70	89	59.77	6	-	53	209	1155	209	mRNA
			7LP	TTTTGCCTCATCCATTTCGT	20	278	257	60.45	8	-					
8	*Os04g0602700*	6	8FP	GCCGATATGCTTGGCATCAC	20	884	903	59.76	4	-	55	119	500	119	mRNA
			8RP	AAGACCCATCCTCACCTTTCG	21	1002	982	59.72	6	987/988					
9	*Os05g0489100*	7	9FP	GCTGATAACCTGTCCGGCAA	20	465	484	60.39	1	843/844	55	105	653	105	mRNA
			9RP	ATTGGATCCTGGCTCTTGCG	20	547	528	60.47	3	-					

Table 4-1 Primer details

	Gene		Primer details								Annealing temperature (℃)	Products			
									Primer location			Estimated Product		PCR Product	
No.	Name	Number of Exon	Primer name	Sequence (5'->3')	Length	Start	Stop	Tm (°C)	Exon No.	Exon Junction		Size on mRNA location	Size on gDNA location	Size	Location
10	Os06g0138200	6	10FP	TCGTCAATATCCTCGCTGGC	20	627	646	59.97	3	-	55	128	128	128	mRNA
			10RP	CGTACGATCTCCAAGCTCCC	20	754	735	59.97	3	746/747					
11	Os06g0286310	3	11FP	GACGAAATCGCAAGCTTCACT	21	1498	1518	59.54	2	1511/1512	55	117	163	117	mRNA
			11RP	CCTCTCATCACAAGCAGCGA	20	1614	1595	60.11	3	-					
12	Os07g0194500	6	12FP	TGCTGAGGGGTGGGAAAATC	20	644	663	59.96		650/651	55	107	224	107	mRNA
			12RP	CACCGTGTCGCCTTTAACTG	20	737	718	59.49		-					
13	Os09g0489200	8	13FP	GGAACTGTTCCTGTTGAAGTTGG	23	559	578	59.61	2	-	55	127	543	127	mRNA
			13RP	AAACAAGGGATGCTGAGCCA	22	685	664	58.04	3	678/679					
14	Os10g0413500	7	14FP	GGAGTGCGAGCACCTGATCT	20	233	252	61.95	2	-	55	139	212	139	mRNA
			14RP	AACTTCGTCCTGCTTCTTCTCG	22	371	350	60.61	3	362/363					
15	Os10g0420000	6	17LP	GAATCTGCCAATGGACTGGT	20	281	300	59.93	4	-	53	235	325	235	mRNA
			17RP	GTAGTTGCCTGCGACTCTCC	20	515	494	60.02	5	-					
16	Os10g0497800	7	18LP	GGGCAGCGTATAGCAACTCT	20	673	692	59.50	5	-	53	161	229	160	mRNA
			18RP	TCTCCCATCTTGGGTTTGAC	20	833	812	59.90	6	-					
17	Os11g0488500	6	19LP	GGAGTACCCTTGGTCTGCAA	20	312	331	60.11	3	-	53	211	1154	211	mRNA
			19RP	GCTTCCATGTCATCAACGTG	20	522	501	60.12	5	-					

Table 4-1 Primer details

	Gene		Primer details								Annealing temperature (℃)	Products			
									Primer location			Estimated Product		PCR Product	
No.	Name	Number of Exon	Primer name	Sequence (5'->3')	Length	Start	Stop	Tm (°C)	Exon No.	Exon Junction		Size on mRNA location	Size on gDNA location	Size	Location
18	Os11g0657200	4	20FP	CAGGTGAGGGCATCTGTGC	19	462	480	60.75	4	469/470	55	106	346	106	mRNA
			20RP	TGTTCTGCCCGACTGAAGTG	20	567	548	60.25	4	-					

4.3.7 Gene amplification

4.3.7.1 To amplify genes by polymerase chain reaction (PCR)

cDNA samples were used to amplify the gene of interest with the specific primer following the protocol as described in Chapter 2. Each amplification was carried out in a GeneAmp PCR system 9700 (Applied Biosystems) using the following programme: 5 min at 94 °C for preheating followed by annealing cycle for 35 cycles of 30 s at 94 °C, 30 s at 55 °C, and 30 s at 72 °C, and finally, 7 min at 72 °C for the final elongation.

4.3.7.2 Agarose gel electrophoresis analysing

The method is the same as that in Chapter 3.

4.3.7.3 PCR product sequencing and editing

The PCR products were sent to Source Bioscience for sequencing and the derived sequence were edited manually and using FinchTV 1.4.0 and BioEdit version 7.0.9.0 software to ensure the correct trimming and produce a high quality full length nucleotide sequence for each amplified product.

4.3.8 Microarray data analysis of predicted *Tet* gene homologues in various organs and different varieties

According to previous work (Jain, et al., 2007), microarray analysis performed by employing one-cycle target labeling and control reagents (Affymetrix). Affymetrix GeneChip® Rice Genome Arrays (Gene Expression Omnibus platform accession no. GPL2025) was used to prepare a collection of transcriptome profiles of 22 stages of vegetative and reproductive development and stress response in rice (Kapoor, et al., 2008). This array contains probes to query 51,279 transcripts representing two rice cultivars, with approximately 49, 824 rice

transcripts comprising 48,564 japonica transcripts and 1,260 transcripts representing the indica cultivar. The microarray analysis data were deposited in the Gene Expression Omnibus database at the NCBI under the series accession numbers GSE6893.

In this present investigation, four more stages which are seedling root, mature leaf, young leaf and 3-4 cm length of panicle were included to analyse gene expression profile during development stages. Sixty-six cell intensity (CEL) files were retrieved from previous accession number GSE6893 and further analysed using R version 3.2.0/Bioconductor version 3.1 (Gentlemen, 2004; Patel, et al., 2012) wherein the RMA algorithm were used for normalisation and $\log_2$ transformation. The Multi expression viewer (MeV) version.4 tool was applied both to generate heat maps by applying normalised and $\log_2$ transformed data and to subsequently cluster on rows of expression value using Euclidean Distance metrics. Consequently, expression data for the predicted *Tet* gene homologues were extracted using the locus identification (locus name) mentioned in Chapter 3. A differential expression analysis was performed by taking a shoot apical meristem as the reference to identify genes expressing at more than the two-fold level (Jain, et al., 2007) in various stages, with $P \leq 0.05$ (Sharma, et al., 2009).

Moreover, for comparison among different analysis, gene expression data were retrieved from two different databases through Genevestigator V3 software and the efP rice browser tool.

4.3.9 qPCR identification of predicted *Tet* gene homologues in various organs and different varieties

This qPCR analysis was conducted by using a Rotor Gene^{TM} 6000 instrument. All reaction components were mixed following the protocol mentioned in section 2.3.2.3 in Chapter 2. The mastermix (Bioline), namely SensiMix SYBR No-ROX, contains the SYBR® Green I

dye, dNTPs, stabilisers and enhancers which are all the components necessary for real-time PCR. For the template, two types of cDNA were prepared 1) cDNA which was used as a control template and for creating a standard curve of both reference gene and the gene of interest, was prepared by pooling all of cDNA sample and then conducting five successive dilutions which are aliquot five times to make the control template dilution and 2) cDNA samples derived from cDNA synthesis which were used as samples for studying the expression of the gene of interest. The expression level was quantified by qPCR through Relative Standard Curve Method. All reactions were placed into the qPCR machine and were run following the thermal cycle: 15 sec at 95 °C for preheating followed by 30 cycles of 15 sec at 55 °C, and 30 sec at 72 °C.

4.4 Results and discussion

Samples were collected from different organs at the required state (Figure 4-4). Before commencing RT-qPCR, an RNA sample was extracted, its concentration determined and it was purified by using a Nano Drop instrument and gel electrophoresis. RNAs having the 260/280 and the 260/230 ratios around 2.0-2.10 and 2.2-2.4, respectively, with no genomic DNA contamination were selected (Figure 4-5). In addition, the purified RNAs were used to synthesise cDNA which was subsequently used for quantification with the specific primer. PCR products were sequencing in order to identify primer efficiency. Finally, RT-qPCR was conducted for studying the expression of selected genes in numerous organs and varieties.

(a) (b) (c) (d)

Figure 4-4 Selected samples from rice: (a) the 3-4 cm of panicles; (b) the mature leaves; (c) young leaves and (d) young roots

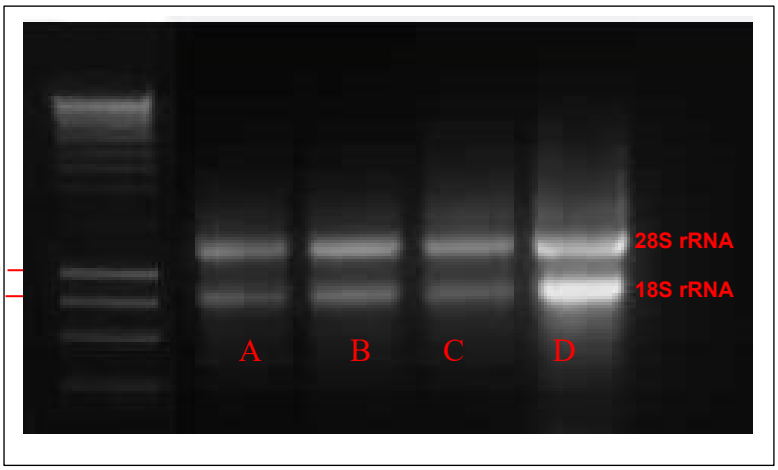


Figure 4-5 No genomic DNA contamination of RNAs from four different tissue samples; A: panicles, B: mature leaves, C: young leaves, D: young roots

4.4.1 Expression analysis of predicted *Tet* gene homologues in vegetative and reproductive organ

To gain an understanding of the developmental stage during which the predicted *Tet* gene homologues are expressed, spatial and temporal expression patterns of these genes were analysed within seedling root, mature leaf, young leaf and young panicle. In this context, evidence from an in-house generated microarray expression dataset prepared by using 57 k Affymetrix GeneChip® rice genome arrays (Affymetrix Inc., Santa Clara, CA, USA) was employed, as mentioned previously. In that dataset, the average intensity values of the non-rice control probe sets were found to be < 10. Thus, the value '10' was taken as the cut-off to discriminate between expressed and non-expressed genes in those particular organs (Sharma, et al., 2009; Kapoor, et al., 2008; Jain, et al., 2007).

On the basis of signal intensities achieved for rice transcripts, there is one gene, *Os01g0180900*, did not express in any of the four analysed organs and only one gene, *Os09g0489200*, did not express in vegetative organs but was expressed at a very low level in young reproductive organs (Table 4-2). The other five genes (*Os04g0346000*, *Os03g0816500*, *Os06g0286310*, *Os04g0602700* and *Os11g0488500*) showed a low expression level in almost organs except for *Os03g0816500* and *Os06g0286310*, which had a moderate expression level in panicle and *Os11g0488500*, which expressed moderately in leaf (both young and mature). The next two genes, *Os03g0761900* and *Os07g0194500*, showed low to moderate expression level in vegetative organs but a high expression level in young panicle. The remaining nine genes showed specific / preferential enhancement in transcript abundance of both reproductive and vegetative organs. In particular, *Os01g0502400, Os10g0420000*, *Os06g0138200* and *Os10g0413500* all showed a higher expression level in all organs than did the other genes (*Os03g0803500*, *Os05g0489100*, *Os10g0497800*, *OS01g0546900* and *Os11g0657200*) (Figure 4-6A)

This analysed expression data of those genes gained from modifying the microarray expression dataset using the RMA algorithm (Robust Multichip Average via quantile adjustment) for normalisation showed clustering similar to that of the gene expression value gained from Genevestigator database (Hruz, et al., 2008) but was slightly different to the clustering derived from the eFP database (Patel, et al., 2012). The clustering obtained from three different methods of normalisation mainly showed that gene expression can be divided into two main groups. The first group included genes which did not express (Figure 4-6 B) or showed the expression profile from no expression to moderate expression (Figure 4-6 A,C) throughout organs. Another group exhibited the group of genes which had an expression profile from low to high (Figure 4-6 B) or showed an expression profile from moderate to high (Figure 4-6 A, C) in all over organs. Moreover, the expression pattern of each gene in various organs obtained from those two different methods of normalisation did not show significant differences when normalising by using RMA algorithm but they were different in terms of the exact value of expression despite using the same microarray dataset. Moreover, there was no probe found for gene *Os04g0346000* and *Os06g0286310* in either the Genevestigator or eFP database, respectively (Figure 4-6 B, C). Such differences between these three results may be caused by the way that normalisation method was used. Genevestigator database applied the MAS 5.0 algorithm (Microarray Suite software 5.0) which is frequently used to carry out the probe-pairs adjustment and to alter the images into text files as intensity information, for normalisation, but the eFP database validated the microarray dataset using MAS5.0 together with the RMA algorithm and in the present study, only RMA was used as the normalisation method. Similarly, one study compared algorithms for the analysis of Affymetrix microarray dataset (Harr and Schlotterer, 2006). Several algorithms such as RMA and MAS 5.0 were applied, leading to different expression values. Consequently, they also concluded that different normalisation methods should be performed depending on the aim of studies which are either to detect differential gene expression

between independent samples or to identify co-expressed genes. In addition, Su, et al., (2011) also confirmed that the normalisation is designed to account for the systematic differences through different array chips and for the first step in the normalization processes, the difference of selected image analysis algorithms ordinarily produce different analysis results. Both of available MAS5 and RMA are popular image analysis algorithms which are used with the Affymetrix array system.

Table 4-2 RMA normalized expression values obtained for the predicted *Tet* gene homologues by using rice microarray data.

No	Probe Set ID	Gene name	Seedling root	Mature leaf	Young leaf	Young panicle
1	OsAffx.23162.1.	*Os01g0180900*	5.6	5.4	6.0	6.2
2	OsAffx.30109.1.	*Os09g0489200*	5.8	5.7	6.4	11.4
3	Os.52052.1.S1_at	*Os04g0346000*	16.4	24.8	19.4	19.2
4	Os.27965.1.S1_at	*Os03g0816500*	35.3	28.3	26.7	74.7
5	Os.55102.1.S1_at	*Os06g0286310*	37.4	38.7	35.0	55.2
6	OsAffx.4146.1.S	*Os04g0602700*	29.1	34.4	30.6	36.9
7	Os.51298.1.S1_at	*Os11g0488500*	38.0	53.6	53.5	36.0
8	Os.55956.1.S1_at	*Os03g0761900*	67.5	28.4	60.1	142.9
9	Os.50554.1.S1_at	*Os07g0194500*	30.1	16.3	58.3	415.0
10	Os.50961.1.S1_at	*Os03g0803500*	199.7	55.8	220.0	488.7
11	Os.25020.1.S1_at	*Os05g0489100*	343.2	161.5	199.0	243.8
12	Os.15848.1.A1_a	*Os10g0497800*	317.5	218.5	207.6	123.1
13	Os.33852.1.S1_at	*OS01g0546900*	97.5	123.9	164.6	183.1
14	Os.11416.1.S1_at	*Os11g0657200*	117.4	112.2	111.4	87.6
15	Os.37540.1.S1_x	*Os10g0420000*	215.2	217.2	384.8	396.2
16	Os.11678.2.S1_at	*Os06g0138200*	235.9	400.5	522.2	490.5
17	OsAffx.19184.1.	*Os01g0502400*	322.9	405.5	404.7	402.6
18	Os.18412.1.S1_at	*Os10g0413500*	592.8	383.6	517.4	481.1

Rows with values less than 10 (cut-off value) defined as no expression are highlighted in yellow

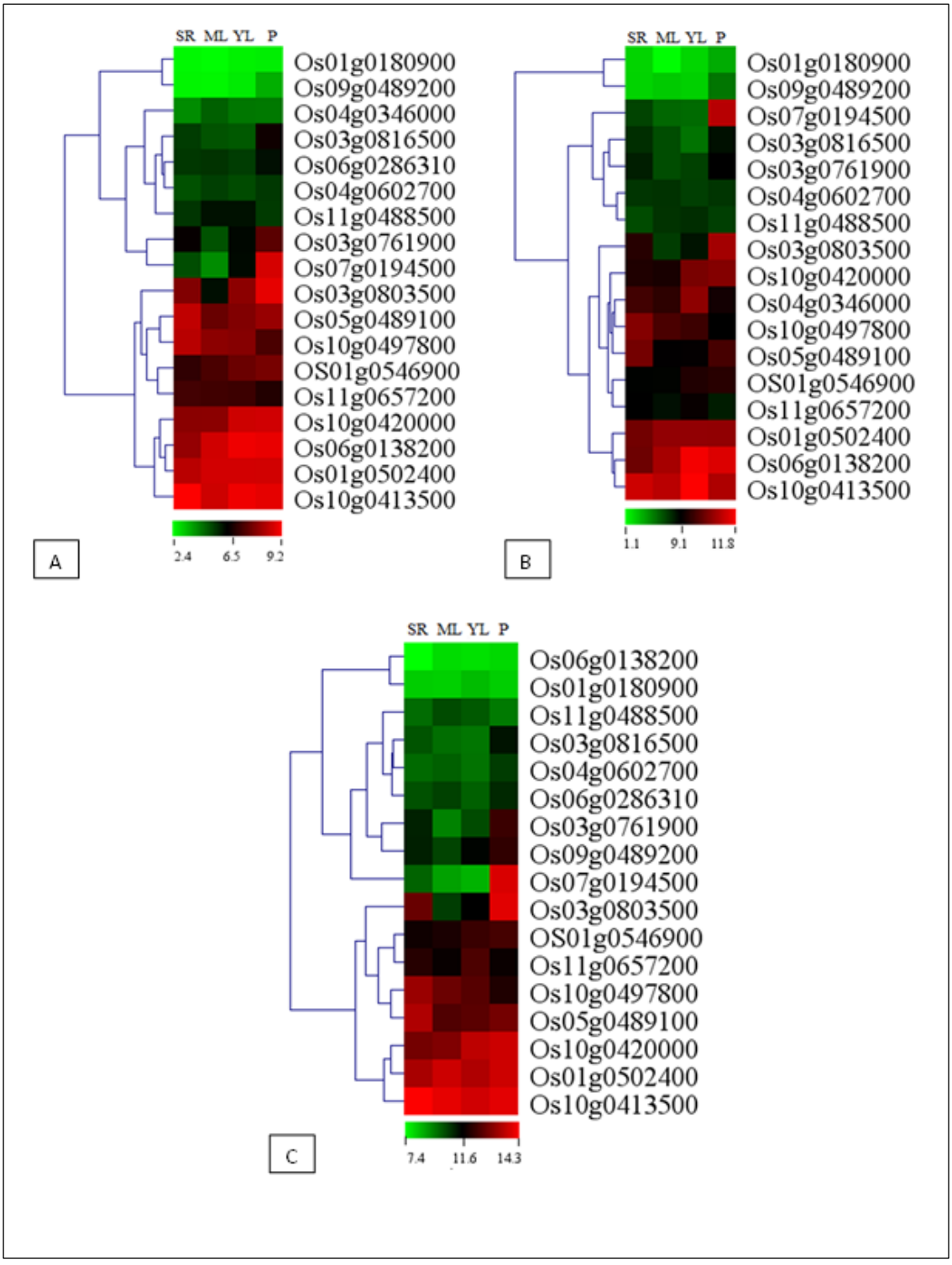


Figure 4-6 Microarray-based expression analysis of predicted *Tet* gene homologues of rice. The expression profiles were analysed in vegetative tissues (seedling roots, mature leaf and young leaf), and reproductive tissue (panicle). The colour bar represents Log2 expression values. A cluster dendrogram derived by grouping on rows of expression value using Pearson correction efficient is shown on the left-hand side of the expression maps. Derived data which was normalised by using A) RMA algorithm, B) RMA together with MAS 5.0 methods and C) MAS 5.0 algorithm.

In comparison with seedling root, the transcript levels of *Os05g0489100*, *Os03g0803500* and *Os03g0761900* were significantly lower in mature leaf and *Os10g0497800* was significantly lower in panicle (Figure 4-6 A; Table 4-3; more than two-fold downregulation at $P \leq 0.05$). In contrast, *Os06g0138200* and *Os03g0803500* had significantly different expression in young leaf and panicle, respectively (Figure 4-6 A; Table 4-3; more than two-fold upregulation at $P \leq 0.05$).

Statistically significant and differential expression profiles of all genes were further analysed by quantitative PCR (qPCR). The transcript accumulation pattern for all genes observed by qPCR was different to those observed from the microarray analysis (Figure 4-6, Figure 4-7). In qPCR analysis, the expression profiling can be divided in to four main patterns. Firstly, there are a majority of genes (10 genes) showing significantly preferential expression in mature leaf. Within this group, five genes showed significantly highest expression in mature leaf; in contrast three genes and two genes showed moderate expression and low expression, respectively. However, it can be seen that in the other three organs, these genes showed very low expression. For the next group, only one gene, *Os01g0546900*, showed significantly highest expression in young leaf, but in seedling root and mature leaf, this gene showed moderate expression and very low expression in panicle. The other group of two genes, were expressed at a low level in all four organs but in seedling root, were more significantly expressed in mature leaf, young leaf and panicle. The last five genes had no significantly differential expression within the four individual organs and the expression value also was low (Figure 4-8).

Comparison of the expression patterns of qPCR and microarray analysis showed there are four genes which had the same expression profile with qPCR analysis: specifically, *Os01g0546900* and *Os04g0346000* had high expression in young leaf and mature leaf respectively, whereas *Os09g0489200* and *Os11g0488500* had low expression in all organs. In

contrast, genes *Os10g0413500* and *Os10g0497800* exhibited low expression in all organs analysed by qPCR whereas in microarray analysis these genes were highly expressed in all organs. For the remaining genes, the expression patterns from the two different analyses showed considerable variation within organs. For example, both *Os03g0761900* and *Os03g0803500* showed high expression in panicle when analysed by microarray but analysed by qPCR, they were highly expressed in seedling root. The next gene, *Os07g0194500*, showed high expression in panicle but low expression in all organs when analysed by microarray and qPCR, respectively. Moreover, the four genes *Os01g0180900*, *Os04g0602700*, *Os06g0286310* and *Os03g0816500* were expressed at a low level in all organs when analysed by microarray data but were highly expressed in mature leaf when analysed by qPCR. Finally, the last five genes *Os01g0502400*, *Os11g0657200*, *Os10g0420000*, *Os06g0138200* and *Os05g0489100* all showed high expression in mature leaf from both analyses.

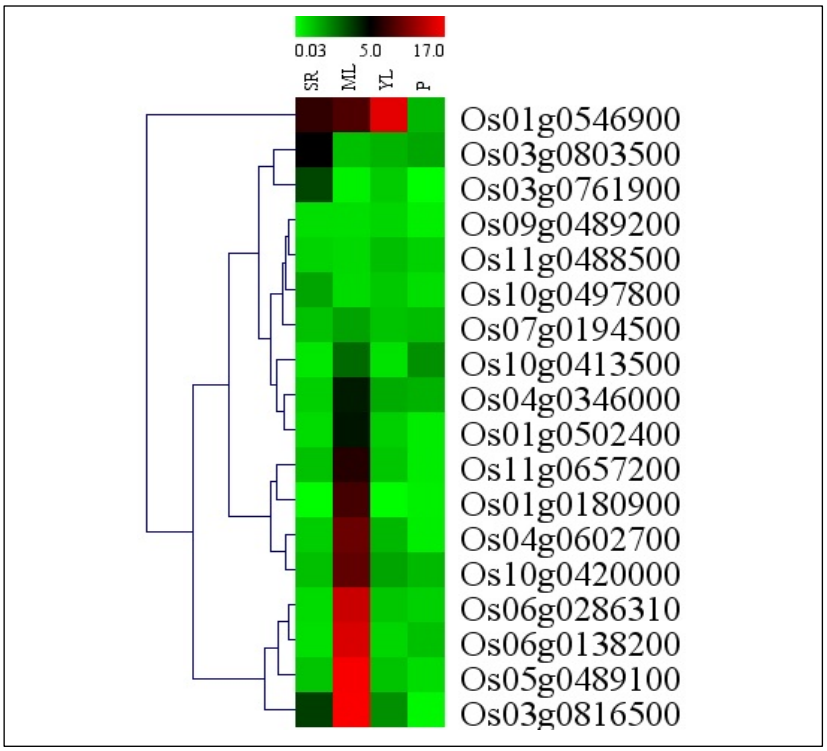


Figure 4-7 qPCR expression analysis of predicted *Tet* gene homologues in rice. The expression profiles of genes were analysed in vegetative tissues (seedling roots, mature leaf and young leaf), and reproductive tissue (panicle). The colour bar represents relative expression values normalised with respect to eEF-1α gene. A cluster dendrogram derived by grouping on rows of expression value using Pearson correction efficient is shown on the left-hand side of the expression maps.

Table 4-3 Differential expression analysis of predicted Tet gene homologues (obtaining data from Affymetrix GeneChip® Rice Genome Array)

No.	Probe Set ID	Gene Name	[Mature leaf] / [Seedling root]		
			Corrected P Value	Fold Change	Regulation
1	Os.19184.1.S1_at	*Os01g0502400*	0.004	1.26	Up
2	Os.33852.1.S1_at	*OS01g0546900*	0.004	1.27	Up
3	Os.52050.1.S1_at	*Os04g0346000*	0.004	1.51	Up
4	OsAffx.4146.1.S1_at	*Os04g0602700*	0.004	1.18	Up
5	Os.25020.1.S1_at	*Os05g0489100*	0.003	2.13	Down
6	Os.55102.1.S1_at	*Os06g0286310*	0.003	1.03	Up
7	OsAffx.30109.1.A1_at	*Os09g0489200*	0.003	1.02	Down
8	Os.18412.1.S1_at	*Os10g0413500*	0.003	1.54	Down
9	Os.37540.1.S1_x_at	*Os10g0420000*	0.002	1.01	Up
10	Os.15848.1.A1_a_at	*Os10g0497800*	0.002	1.45	Down
11	Os.50961.1.S1_at	*Os03g0803500*	0.007	3.57	Down
12	Os.27965.1.S1_at	*Os03g0816500*	0.006	1.25	Down
13	Os.55956.1.S1_at	*Os03g0761900*	0.009	2.38	Down
14	Os.51298.1.S1_at	*Os11g0488500*	0.013	1.41	Up
15	Os.11416.1.S1_at	*Os11g0657200*	0.013	1.04	Down
16	Os.11678.2.S1_at	*Os06g0138200*	0.019	1.70	Up
17	OsAffx.23162.1.S1_at	*Os01g0180900*	1.068	1.05	Down
18	Os.50554.1.S1_at	*Os07g0194500*	0.789	1.85	Down

Genes with fold change ≥ 2.0 and P-value ≤ 0.05 are highlighted and defined as differential expression compared to seedling root.

Table 4-3 Differential expression analysis of predicted Tet gene homologues (obtaining data from Affymetrix GeneChip® Rice Genome Array)

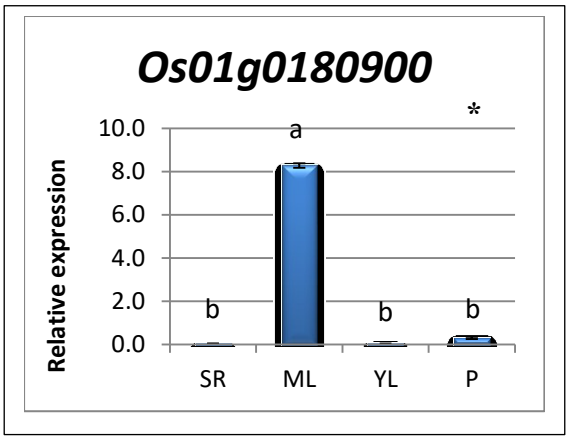
No.	Probe Set ID	Gene Name	[Young leaf] / [Seedling root]		
			Corrected P Value	Fold Change	Regulation
1	Os.19184.1.S1_at	*Os01g0502400*	0.004	1.25	Up
2	Os.52050.1.S1_at	*Os04g0346000*	0.004	1.19	Up
3	Os.25020.1.S1_at	*Os05g0489100*	0.004	1.72	Down
4	Os.18412.1.S1_at	*Os10g0413500*	0.004	1.15	Down
5	Os.15848.1.A1_a_at	*Os10g0497800*	0.003	1.54	Down
6	Os.33852.1.S1_at	*OS01g0546900*	0.005	1.69	Up
7	Os.37540.1.S1_x_at	*Os10g0420000*	0.005	1.79	Up
8	Os.11678.2.S1_at	*Os06g0138200*	0.013	2.21	Up
9	Os.50961.1.S1_at	*Os03g0803500*	0.011	1.1	Up
10	Os.55102.1.S1_at	*Os06g0286310*	0.021	1.06	Down
11	Os.51298.1.S1_at	*Os11g0488500*	0.021	1.41	Up
12	Os.11416.1.S1_at	*Os11g0657200*	0.030	1.05	Down
13	OsAffx.4146.1.S1_at	*Os04g0602700*	0.028	1.05	Up
14	Os.27965.1.S1_at	*Os03g0816500*	0.051	1.32	Down
15	Os.55956.1.S1_at	*Os03g0761900*	0.160	1.12	Down
16	Os.50554.1.S1_at	*Os07g0194500*	0.238	1.94	Up
17	OsAffx.23162.1.S1_at	*Os01g0180900*	1.613	1.06	Up
18	OsAffx.30109.1.A1_at	*Os09g0489200*	0.846	1.09	Up

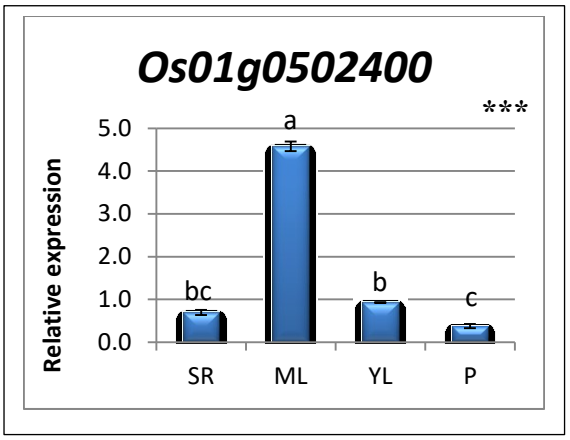
Gene with fold change ≥ 2.0 and P-value ≤ 0.05 are highlighted and defined as differential expression compared to seedling root.

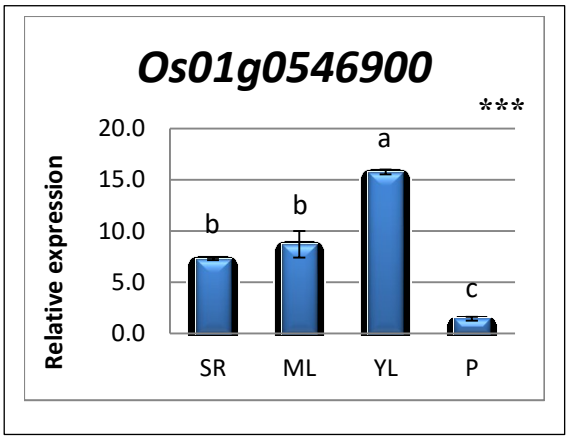
Table 4-3 Differential expression analysis of predicted Tet gene homologues (obtaining data from Affymetrix GeneChip® Rice Genome Array)

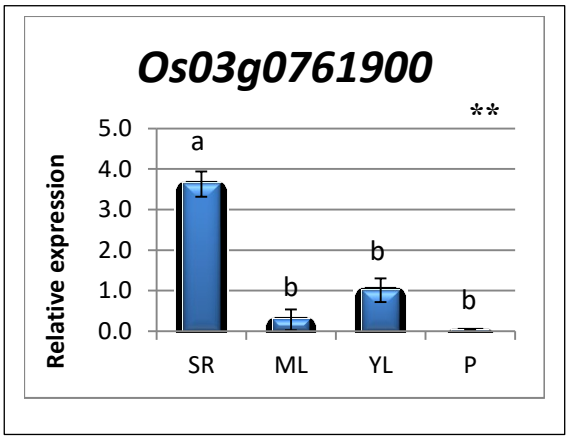
S. No.	Probe Set ID	Gene Name	[Panicle] / [Seedling root]		
			Corrected P Value	Fold Change	Regulation
1	Os.19184.1.S1_at	*Os01g0502400*	0.004	1.25	Up
2	Os.33852.1.S1_at	*OS01g0546900*	0.004	1.88	Up
3	Os.52050.1.S1_at	*Os04g0346000*	0.004	1.17	Up
4	OsAffx.30109.1.A1_at	*Os09g0489200*	0.004	1.96	Up
5	Os.18412.1.S1_at	*Os10g0413500*	0.003	1.23	Down
6	Os.37540.1.S1_x_at	*Os10g0420000*	0.003	1.84	Up
7	Os.15848.1.A1_a_at	*Os10g0497800*	0.003	2.56	Down
8	Os.25020.1.S1_at	*Os05g0489100*	0.010	1.41	Down
9	Os.11416.1.S1_at	*Os11g0657200*	0.020	1.33	Down
10	Os.51298.1.S1_at	*Os11g0488500*	0.032	1.05	Down
11	Os.50961.1.S1_at	*Os03g0803500*	0.030	2.45	Up
12	Os.55102.1.S1_at	*Os06g0286310*	0.026	1.48	Up
13	Os.55956.1.S1_at	*Os03g0761900*	0.114	2.12	Up
14	OsAffx.4146.1.S1_at	*Os04g0602700*	0.104	1.27	Up
15	Os.27965.1.S1_at	*Os03g0816500*	0.518	2.12	Up
16	Os.11678.2.S1_at	*Os06g0138200*	0.894	2.08	Up
17	OsAffx.23162.1.S1_at	*Os01g0180900*	1.532	1.10	Up
18	Os.50554.1.S1_at	*Os07g0194500*	0.786	13.79	Up

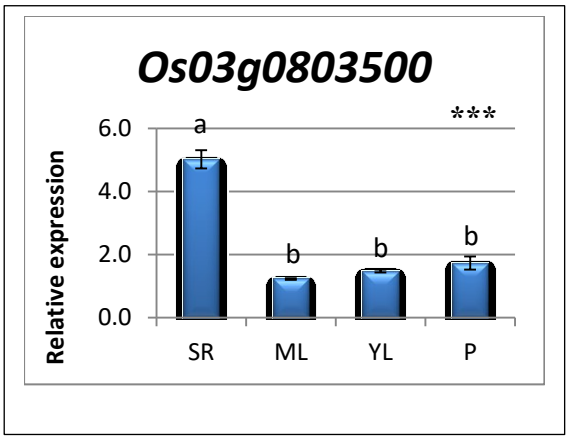
Genes with fold change ≥ 2.0 and P-value ≤ 0.05 are highlighted and defined as differential expression compared to seedling root.

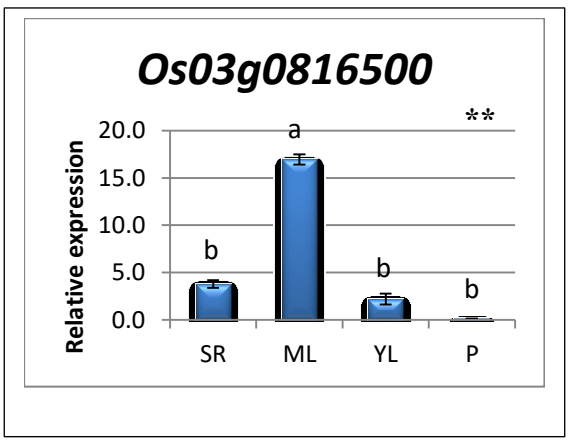

Os01g0180900
Relative expression
10.0
8.0
6.0
4.0
2.0
0.0
a
b
b
b
*
SR
ML
YL
P

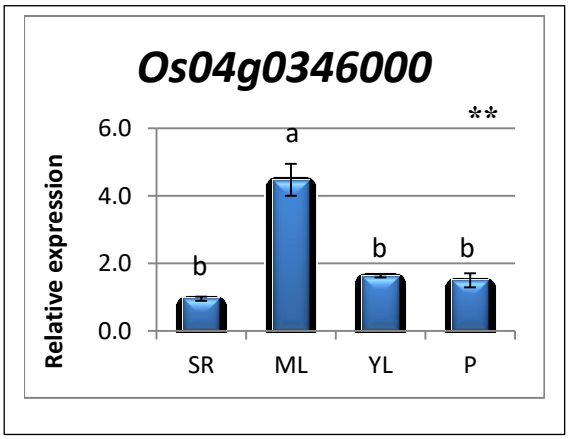

Os01g0502400
Relative expression
5.0
4.0
3.0
2.0
1.0
0.0
a
bc
b
c
***
SR
ML
YL
P

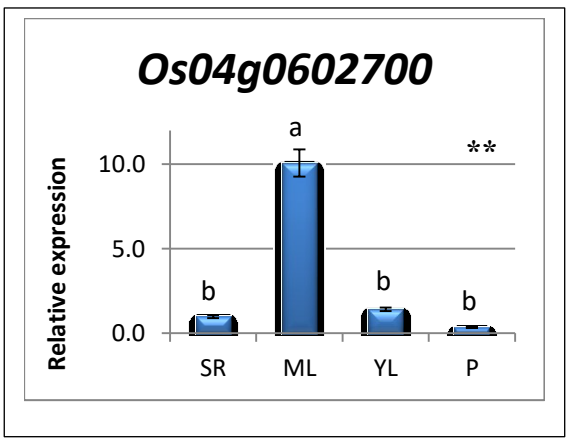

Os01g0546900
Relative expression
20.0
15.0
10.0
5.0
0.0
b
b
a
c
***
SR
ML
YL
P

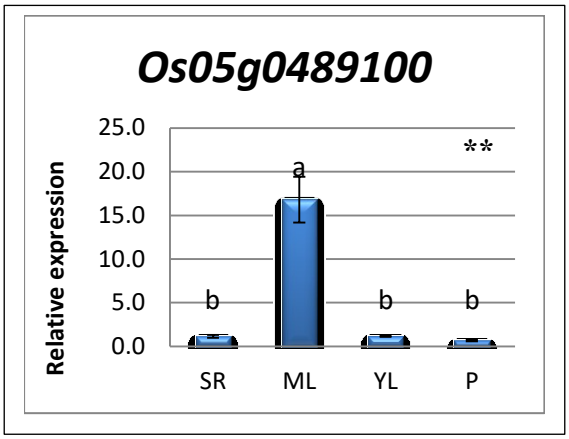

Os03g0761900
Relative expression
5.0
4.0
3.0
2.0
1.0
0.0
a
b
b
b
**
SR
ML
YL
P

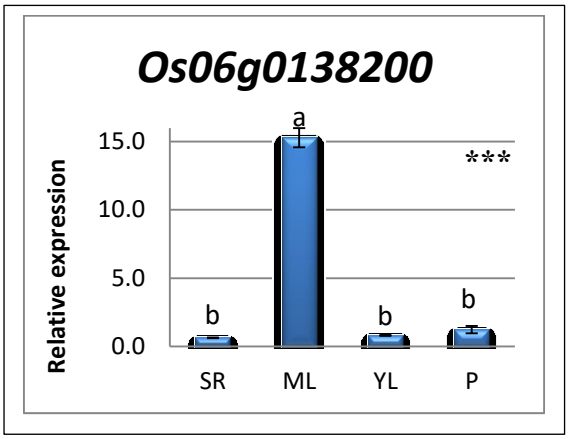

Os03g0803500
Relative expression
6.0
4.0
2.0
0.0
a
b
b
b
***
SR
ML
YL
P

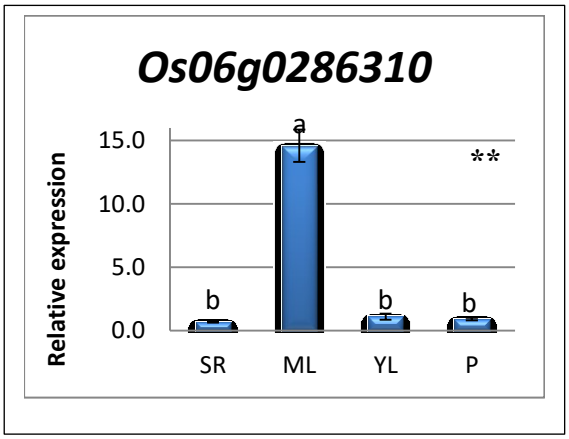

Os03g0816500
Relative expression
20.0
15.0
10.0
5.0
0.0
b
a
b
b
**
SR
ML
YL
P

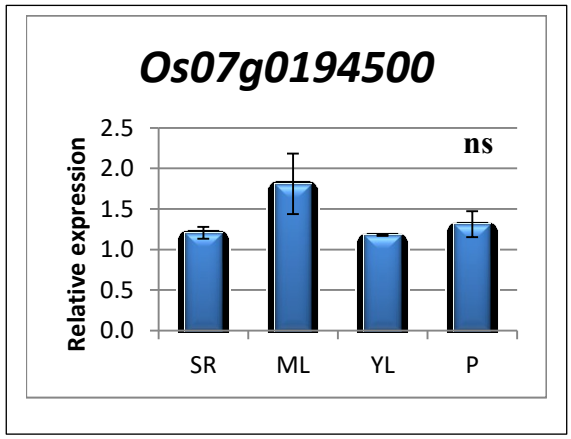

Os04g0346000
Relative expression
6.0
4.0
2.0
0.0
b
a
b
b
**
SR
ML
YL
P

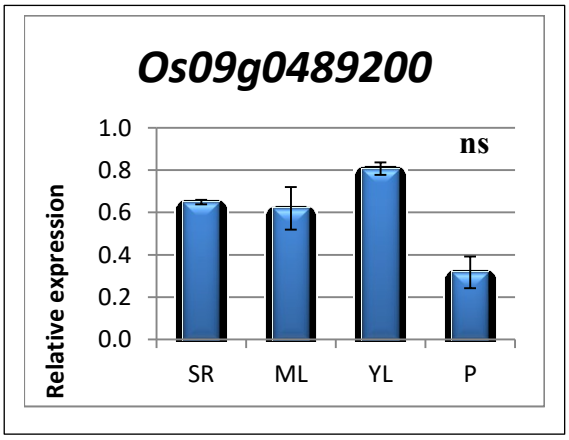

Os04g0602700
Relative expression
10.0
5.0
0.0
b
a
b
b
**
SR
ML
YL
P

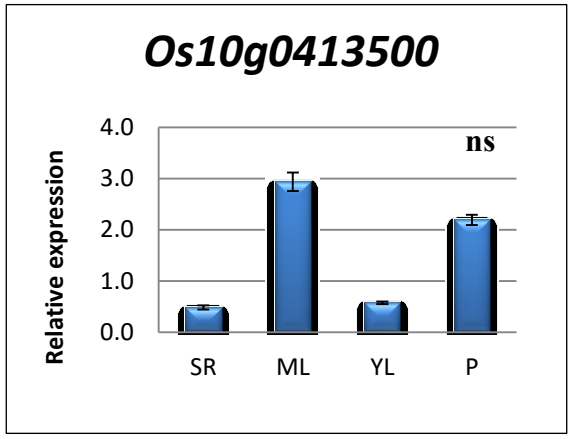

Os05g0489100
Relative expression
25.0
20.0
15.0
10.0
5.0
0.0
a
b
b
b
**
SR
ML
YL
P

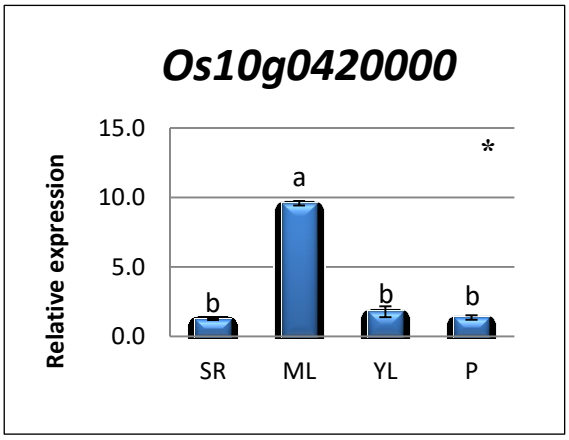

Os06g0138200
Relative expression
15.0
10.0
5.0
0.0
a
b
b
b
***
SR
ML
YL
P

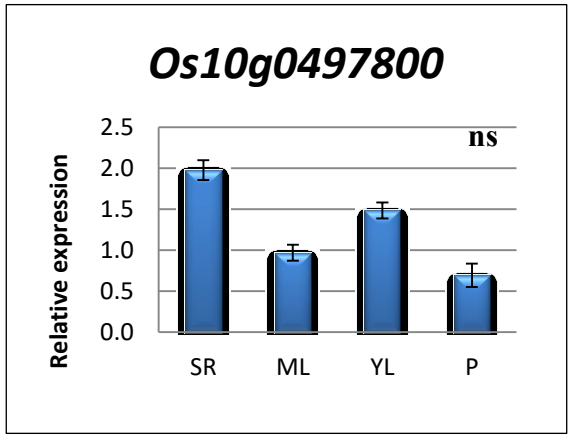

Os06g0286310
Relative expression
15.0
10.0
5.0
0.0
a
b
b
b
**
SR
ML
YL
P


Os07g0194500
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
ns
SR
ML
YL
P


Os09g0489200
Relative expression
1.0
0.8
0.6
0.4
0.2
0.0
ns
SR
ML
YL
P


Os10g0413500
Relative expression
4.0
3.0
2.0
1.0
0.0
ns
SR
ML
YL
P


Os10g0420000
Relative expression
15.0
10.0
5.0
0.0
a
b
b
b
*
SR
ML
YL
P


Os10g0497800
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
ns
SR
ML
YL
P

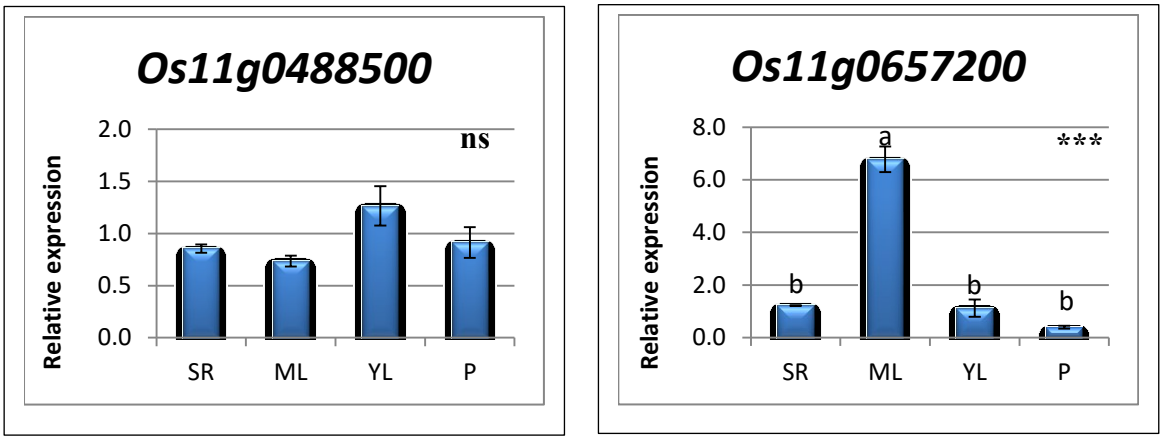


Figure 4-8 Bar graph presenting qRT-PCR data for the relative expression of 18 predicted *Tet* gene homologues in four different organs normalised with respect to the *eEF-1α* gene. Values are the mean ± standard error of three biological replicates. Data were analysed using an ANOVA with Student-Newman-Keuls test (* is P<0.05, ** is P<0.01, *** is P<0.001). Different letters represent significant. Value 0.1-5.0 = low expression, 5.1 -10.00 = moderate expression and 10.1 onward = high expression

The difference of expression profiles of many genes analysed from both methods can generally occur due to many factors. First, the rice varieties used in both methodologies are different those used in the microarray analysis, *O. sativa* Indica domesticated in eastern India, was applied but in the qPCR analysis, it was a *O. sativa* Japonica variety from somewhere in South China (Liu, et al., 2010). Because these two subspecies have been subject to different extracellular cues or different environmental condition they may respond differently from the control in terms of phenotypic patterns, biological behaviour as well as biochemical components. Moreover, O'Conno and Adams (2010) also stated that different cells have dissimilar sets of transcription regulators which affect the diversity of gene expression profiles characteristic of different cell types. For example, the regulation of gene expression is predominantly controlled by a transcriptional step which can be regulated by proteins called transcription regulators (Alberts, et al., 2008; Liu, et al., 2010) including repressors which are able to turn genes off, and activators which are able to turn genes on. Therefore, different varieties may have such different regulators and may lead to different factors controlling gene regulation as a result of the divergence of gene expression patterns. There is

also evidence reveal that genes themselves have an interaction among them. A certain gene generates a protein, possibly an enzyme that activates a second gene which may produce a protein as a transcription regulator or other components such as small RNA that subsequently regulates the next downstream gene. As the same reason given from first example, different varieties also have different proteins or small RNA (Sanchez, et al., 2006) generated so it is possible that either each variety may produce different proteins/small RNA affecting other individual genes, or one may not produce and another one may produce both of them leading to the different components between of two varieties. Likewise, this production also may have downstream effect resulting in the different pattern of gene expression.

Secondly, another factor that may cause differences in results from microarray and qPCR methods is the different platform and procedures (Morey, et al., 2006). For example, an accurate result is gained from good quality RNA, as gene expression can be affected by contamination with other components such as different tissues, salts, alcohols, and phenol, which can affect reverse transcriptases used in both qPCR and RNA amplification procedures for microarray labelling (Freeman, et al., 1999). Furthermore, the effects of dye biases (Yang, et al., 2002) and non-specific and/or cross hybridizations of labelled targets to array probes (Chuaqui, et al., 2002) are distinctive to microarray procedures and also qPCR which has its own sources of error such as amplification biases (Chuaqui, et al., 2002), the exponential amplification of errors (Freeman, et al., 1999), and mispriming or primer dimer formation (Bustin, 2002). In addition, data normalization between microarray analysis and qPCR have a fundamental difference, the former requires global normalization, while the latter generally applies the expression of one or more reference genes against which all other gene expression for calibration. Therefore, appropriate selection and application of normalisation criteria may also have an impact on the correlations found between these methods.

Gene expression is dynamic in that the same gene may have a different action under different circumstances. The quantities of gene transcripts in a particular cell are measured by observing mRNA molecules which may directly reveal the function of that cell (O'Conno and Adams, 2010). Therefore , it reasonably straight forward to say that if the expression patterns of genes from both methods show similar pattern in whatever of observed samples or treatments, it can be concluded that those genes may demonstrate high transcription in those observed samples or under that treatment. Focusing on genes which showed significantly the highest expression value in both analysis methods, there are seven genes showing the same pattern of expression in leaf. Only *Os01g0546900* showed high expression in young leaf, whereas six genes *Os01g0502400*, *Os04g0346000*, *Os05g0489100*, *Os06g0138200*, *Os10g0420000*, and *Os11g0657200* showed high expression in mature leaf.

Seemingly, the majority of predicted *Tet* gene homologues have higher expression in leaf than in other organs. This indicates that these genes may be significantly involved in critical metabolism in the leaf which one of the most important organs of the plant.

4.5 Summary

In many metazoa Tet family enzymes have been revealed to play a major role in demethylation by removing methyl groups on cytosine through an oxidative pathway leading to the activation of silenced gene and also by generating oxidative derivative products including 5-hmC and 5-caC. Fascinatingly, these oxidative products have also been detected in rice and Arabidopsis. This finding prompted us to investigate the expression of predicted *Tet* gene homologues in order to gain fundamental information and in the future to study the possible link of these genes with demethylation in some cellular process.

To obtain an understanding of the developmental period during which the predicted *Tet* gene homologues are expressed, spatial and temporal expression patterns of these genes were

analysed within seedling root, mature leaf, young leaf and young panicle by using data from both microarray and qPCR.

Genes which showed similarly high expression profile from both data analysed by microarray and qPCR were selected. There are seven of 18 genes showed highest expression in leaf more than in seedling root and panicle; these included *Os01g0546900* which had highest expression in young leaf and six genes, *Os01g0502400*, *Os04g0346000*, *Os05g0489100*, *Os06g0138200*, *Os10g0420000*, and *Os11g0657200*, which have highest expression in mature leaf. The remaining genes showed lower expression that varied between organs. Those genes showing highest expression in leaf are considered to be selected for further study in order to investigate the functional link with demethylation in cellular process in the leaf.

Chapter 5
Molecular approaches to *Sub1A* regulation under flooding and drought

5.1 Introduction

5.1.1 Global warming impacts

For decades, global warming caused primarily by increases in "greenhouse" gases such as carbon dioxide (CO_2) had led to a warming planet and subsequent climate change which has contributed to the severity of the weather with a consequent impact including severe flooding and rigorous drought (Hartmann, et al., 2013 ;Shah, 2015). For example, during the period from mid-December through until early February in 2014, the UK was affected very severely by an extraordinary sequence of winter storms, resulting in suffering from extensive flooding, which was a combination of fluvial, pluvial and groundwater flooding with an element of coastal flooding in some areas. Beyond the problems of inundated land, neighbouring areas and areas of poor drainage or high water tables were also subjected to waterlogging of soils (Slingo, et al., 2014). Likewise, beginning on December 2014, a series of massive floods from the northeast monsoon severely hit Indonesia, West Malaysia, and Southern Thailand.

Furthermore, the frequency of global drought has also been increasing. By the end of April 2015, there was a general increase in drought conditions in many locations. Europe is facing dryness, generally through the central part of the region and in Asia (Vilsack and Harden, 2015), drought continues to be focused in the eastern and particularly China is faced with a persistent drought which was suggested as being perhaps the worst in this region in more than 100 years (Shah, 2015).

Both flooding and drought also subsequently affected various human's activities on the Earth, particularly the agriculture sector. These circumstances had led to reduction of agricultural productivity across the continents. In particular rice production, for example, in Southeast Asia, it was clearly demonstrated that many of the region's rice systems are vulnerable to typhoons and floods. In late 2011, a series of typhoons ripped across Southeast Asia, initiating floods that destroyed about 12.5% rice farmlands in Thailand, along with 12% in Cambodia, 6% in the Philippines, 7.5% in Lao PDR, and 0.4% in Viet Nam (Redfern, et al., 2012). In addition, floods and strong winds brought on by the typhoons (GIEWS, 2012) also consequently had cut Thailand's rough-rice production by about 7.2 million tonnes while the Philippines lost some 600,000 tonnes of milled rice. Another example was in the report of Agricultural Development and Advisory Service, UK (ADAS, UK) (Twining, 2014); this stated that some regions of the UK were subject to unexpectedly severe storms, which caused huge flooding and lead to submergence of arable plants and grassland for several weeks. Estimates of yield losses based on research and expert opinion, suggested winter wheat and winter oilseed rape had 15% and 20% of average yield loss when submerged less than 15 d, whereas with more than 15 d of submergence, these plants had 100% of average yield loss.

Not only flooding but drought caused by El Niño is also likely to reduce crop yield. In 2014, the Northeast of Thailand where a majority of rice is grown is facing serious drought due to below average rainfall. This phenomenon led to the depletion of the aggregate rice crop. It was estimated that rice production yield was put at 7.2 million tonnes, 26% below the yiled in 2013. During 2015, Thailand is still facing drought, which causes significant loss in the quantity and quality of the rice crop. Likewise, Cambodia is also facing dry conditions leading to reduction of rice production; due to drought in the year 2014, the rice yield declined to 44.8 million tonnes which is smaller than in 2013 (GIEWS, 2015).

5.1.2 Abiotic stresses response in rice

5.1.2.1 Flooding stress response

Abiotic stress is a major factor inhibiting plant growth and subsequent limiting the yield of rice crop. It is necessary for plants to cope with any stress for surviving and reproduction otherwise it will either be damaged or die. Unlike animals, plants cannot move to avoid such stresses so they have to adapt by develop numerous mechanisms to dealing with stress tolerance. Naturally, rice can tolerate shallow flooding but most rice varieties will die if completely submerged for more than a few days.

Deep-water rice is able to survive in deep-water flood using the escape strategy (Figure 5-1) which is regulated by two ethylene-responsive factors (ERFs), SNORKEL1 (SK1) and SNORKEL2 (SK2), that during flooding generate substantial internode elongation via GA synthesis (Hattori, et al.,2009; Nagai et al., 2010). SK1 and SK2 are present in some wild rice genotypes that show a deep-water response (Hattori, et al., 2009) but are absent in non-deep-water rice varieties evaluated to date (Niroula, et al., 2012).

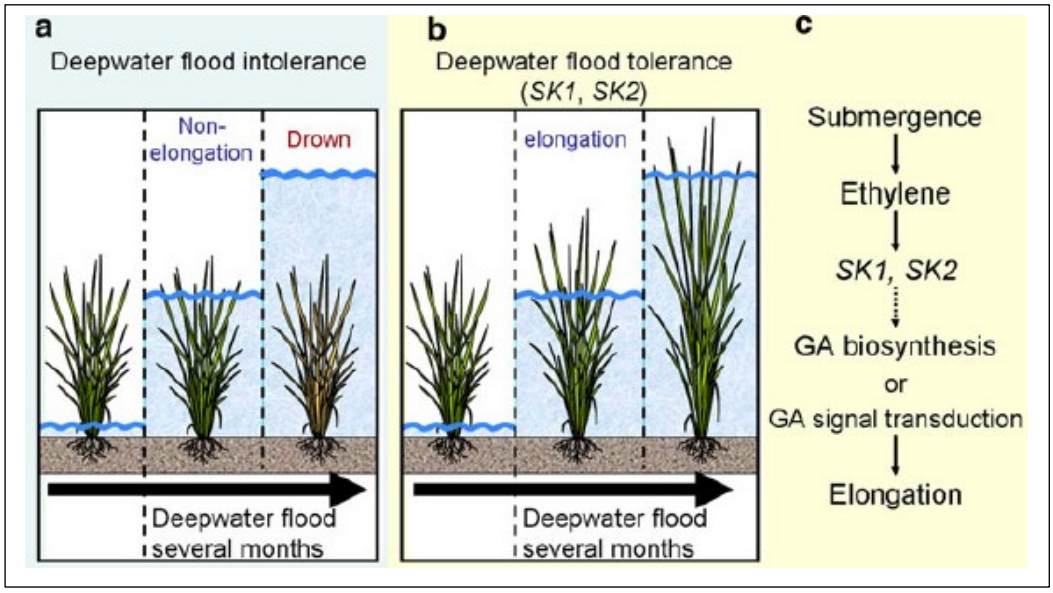


Figure 5-1 Deepwater response of intolerant and tolerant deepwater rice. a Non-deepwater rice strategy in a deepwater flood. Non-deepwater rice is unable to elongate the internodes, so it dies from oxygen shortage. b The adaptation to deepwater flood of deepwater rice by an escape strategy. Deepwater rice elongates its leaves and internodes to rise above the water level for survival. c The regulation of deepwater tolerance in deepwater rice genotype. When submerged, the ethylene content in plants increases. This accumulated ethylene induces the expression of *SK1* and *SK2*. Then, SK1 and SK2 have an direct or indirect function to promote the accumulation of GA which positively regulates leaves and internodes elongation or GA signal transduction (Nagai, et al., 2010).

Lowland rice genotypes also have this adaptive response, but it is only beneficial if floodwaters are shallow or rise gradually (Voesenek, et al., 2004). This is because shoot elongation is a favourable trait only when the associated costs in terms of energy and carbohydrates for cell division and elongation are outweighed by being able to reach the water surface before carbohydrate starvation occurs.

Although lowland rice is also considered a flood tolerant crop, most rice cultivars die within 7 d of complete submergence (Xu, et al., 2006). However, there are only a few rice varieties which are able to resist complete submergence for more than two weeks. Some regional rice varieties are able to resist prolong submergence; for example, the ancient Indian rice variety, Flood Resistant 13A (FR13A) with poor grain and yield qualities has an unusual ability to

survive when completely submerged for over 14 d (Xiong, et al, 2012) and still recover to regrow (Mackill, et al., 1993). This persistent flood tolerant variety has been used as a donor parent to introduce the submergence tolerant trait into other intolerant varieties which may have a good production and grain quality but lack submergence tolerance.

FR13A adopts a quiescence strategy (Figure 5-2, Figure 5-4) by inhibiting the elongation growth and carbohydrate consumption for extending survival when submerged, and accumulating carbohydrate for recovery when desubmerged. To date it has been found that such quiescent strategy, which lowland rice use to cope with particular abiotic stress as flooding, is modulated through *Sub1A* gene (Xu, et al., 2000; Xu, et al.,2006; Niroula, et al., 2012; Xiong, et al., 2012; Schmitz, et al., 2013) which is mediated by SLENDER RICE 1 (SLR1) and SLR1 Like 1 (SLRL1) protein (Fukao and Bailey-Serres, 2008; Dongen and Licausi, 2014) and consequently restrict elongation growth.

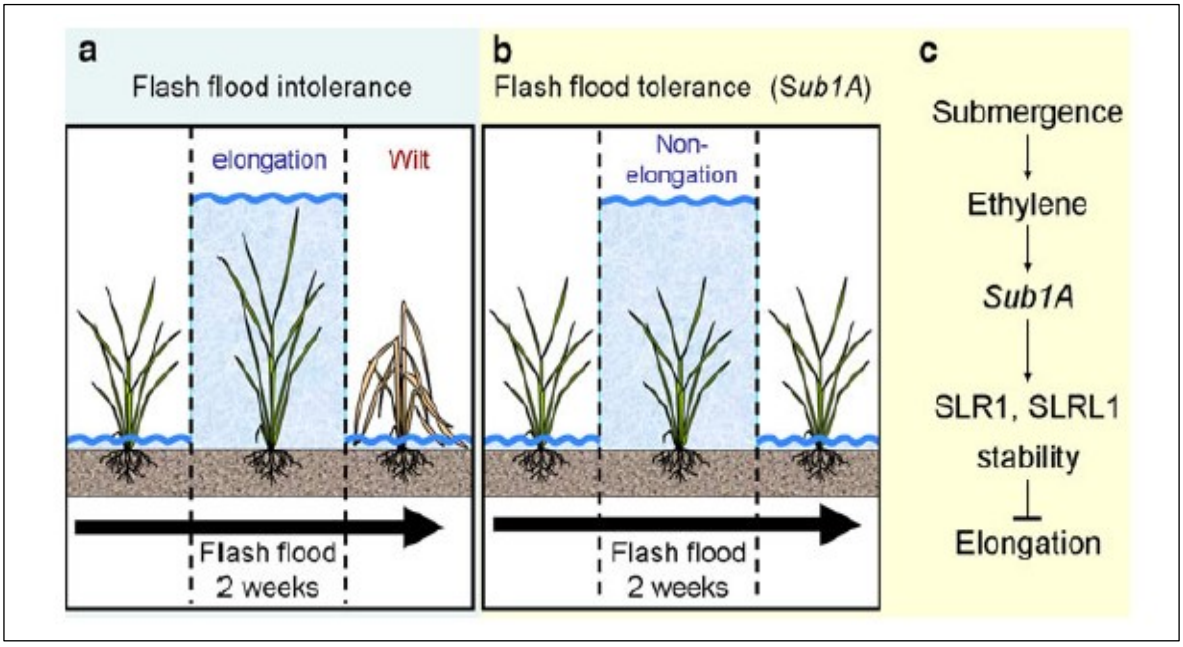


Figure 5-2 Flash flood response model of intolerant and tolerant lowland rice. a The flash flood intolerant rice elongates its leaves and internode to avoid sinking. The emerged leaves are unable to breach the surface leading to prolong complete submergence. Consequently, it wilts and dies after the water reduces. b The quiescence strategy of lowland tolerance rice facing with flash flood at the rice seedling stage. Flash flood tolerant rice stops growing because it does not need to consume energy while submerged. After the reduction of water,

the growth is restarted using accumulated energy. c The regulation of flash flood tolerance. When submerged, the accumulation of ethylene induces the expression of *Sub1A* which consequently promotes the accumulation of SLR1 and SLRL1 which are GA signaling repressors and inhibit internode elongation (Nagai, et al., 2010).

5.1.2.2 SUB1A-mediated submergence tolerance response in rice

The quiescence strategy is modulated by the presence of the major quantitative trait locus (QTL) *Submergence 1* (*Sub 1*) located on rice chromosome 9 that encodes a variable cluster of two or three Ethylene Response Factor (EFR) genes named *Sub1A*, *Sub1B* and *Sub1C* (Fukao, et al., 2006; Xu et al., 2006). Studies to date have shown that *Sub1B* and *Sub1C* are invariably present in all *Indica* and *Japonica* varieties, whereas the presence of *Sub1A* is variable and also exclusively exists in a subset of *Indica* varieties (Figure 5-3) (Xu, et al., 2006)

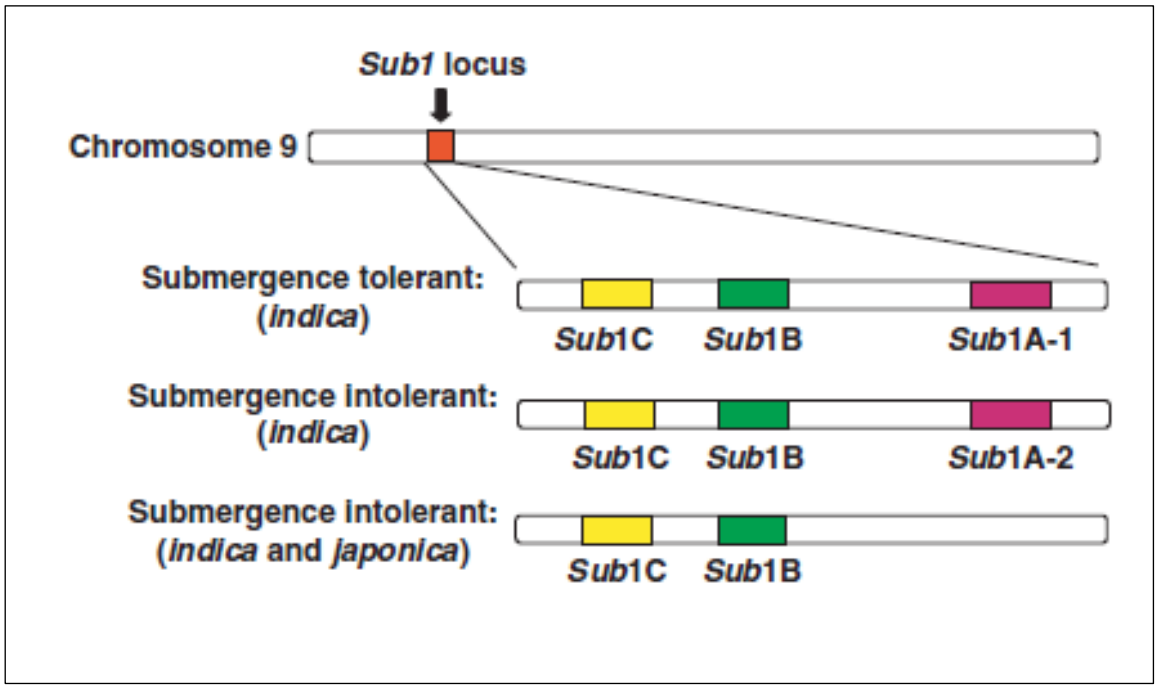


Figure 5-3 *Sub1* locus of *O. sativa*. This *Sub1* locus comprises up to three ethylene responsive factors which are *Sub1A*, *Sub1B* and *Sub1C*. *Sub1B* and *Sub1C* is invariable in both *indica* and *japonica* rice. *Sub1A-1* allele is present in tolerance submergence accession only whereas the accessions possess *Sub1A-2* or even no *Sub1A* conferring intolerance submergence (Fukao et al., 2006; Xu et al., 2006; Fukao, et al., 2009).

Sub1A encodes a full length transcript of 1,312 nucleotides carrying 846 nucleotides of an open reading frame (ORF) which encodes a protein of 281 amino acids, and also comprises two alleles including *Sub1A-1* and *Sub1A-2*. These two alleles are discriminated by a single amino acid substitution within the coding region. In the tolerant *Sub1A-1* allele, a single nucleotide polymorphism (SNP) at position 556 is responsible for amino acid at position 186 being Serine whereas the intolerant *Sub1A-2* allele, the SNP is responsible for Proline (Pro) (Xu, et al., 2006).

Under submergence, the expression levels of *Sub1A-1* and *Sub1A-2* are high and low, respectively (Fukao, et al., 2006). Investigation of *Sub1A* expression in selected *Indica* varieties revealed that the possession of the strongly submergence-induced *Sub1A-1* was correlated with submergence tolerance whereas the poorly submergence-induced *Sub1A-2* was associated with intolerance. This indicates that rice varieties which possess *Sub1A-1* allele are submergence tolerant whereas possession of the *Sub1A-2* allele confers submergence-intolerance. Furthermore, the presence of *Sub1A* is also variable in different rice varieties; it was confirmed that *Sub1A* is absent from all *Japonica* and some *Indica* accessions resulting in submergence intolerance. Support for *Sub1A-1* as the determinant allele of submergence tolerance came from the finding that the over-expression of this allele in the intolerant *Japonica* Liaogeng and M202 leads to a distinct improvement in submergence tolerance (Xu, et al., 2006).

5.1.2.3 *Sub1A* is involved in drought tolerance

In general, rice is sensitive to drought because it is naturally requires a great deal of water but different rice varieties range in their tolerance. Moreover, cultivation of rice in rain-fed areas which are prone to flooding and drought due to lack of water management also suffers from frequent reduction of production yield. To survive in such stress conditions, rice plants

modulate responses of adaptability through complex signalling pathways, which are incorporated at various levels (Lanceras, et al., 2004 ; Yue, et al., 2006; Venuprasad, et al., 2009). As some rice plants with genotypes containing *Sub1A* are able to recover after desubmergence by forming new leaves (Fukao, et al., 2006; Fukao and Bailey-Serres, 2008), this indicates that *Sub1A* contributes to prevention of meristematic cells from submergence and reoxygenation stress. This finding led to the hypothesis that *Sub1A* may play a protective role to support endurance of other abiotic stresses, such as drought. Thus, the contribution of *Sub1A* to water deficit tolerance was evaluated. The result showed that under dehydration, *Sub1A* expression was highly enhanced and consequently upregulated other transcripts encoding *ERFs* associated with acclimation to drought such as *DEHYDRATION RESPONSIVE ELEMENT BINDING PROTEIN 1s* (*DREB1s*) including *DREB1A* and *DREB1E*, and other *ERFs* comprising *AP59* and *AP37* (all of which encode transcription factor containing an APETELA2 (AP2) domain). Additionally, enhancement of *Sub1A* expression also increased ABA responsiveness which leads to consequent expression of *LEA* (LATE EMBRYOGENESIS ABUNDANT) and subsequently reduced the leaf water loss (Fukao, et al., 2011) (Figure 5-4). Finally, such *Sub1A* intervention lead to drought tolerance of rice containing *Sub1A*.

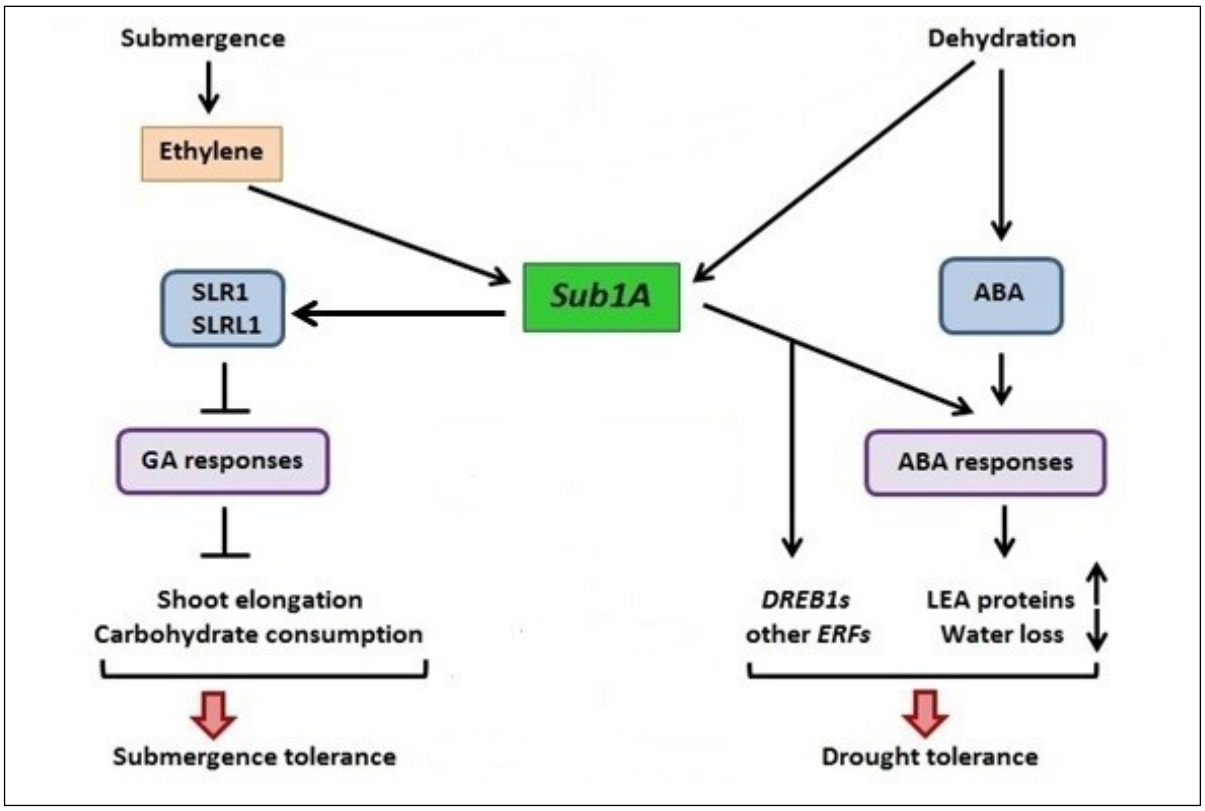


Figure 5-4 *Sub1A*-mediated abiotic stress responses in rice. Under submergence, *Sub1A* expression promotes the accumulation of GA signaling repressors, SLR1 and SLRL1 which inhibit the GA responses and consequently restrict shoot elongation as well as carbohydrate consumption resulting in submergence tolerance. Under dehydration, *Sub1A* expression upregulated transcription of *ERFs* associated with acclimation to drought (e.g., *DREB1s* and *AP59*) and also increased ABA responsiveness which induces expression of *LEA* mRNAs and suppresses leaf water loss and consequently leading to drought tolerance (modified from Fukao, et al., 2011).

5.1.2.4 The involvement of gases in plant acclimation to flooding and drought stress

Plant life relies on the production of carbohydrates via photosynthesis by using light energy for CO_2 fixation. These carbohydrates are used to create various plant structures and fuel energy production as adenosine triphosphate (ATP) through respiration, which requires a sufficient source of O_2. Furthermore, a large number of metabolic pathways are driven through enzymatic reaction catalysed by several oxygen-dependent enzymes called monooxygenase and dioxygenase. All metabolic reactions catalysed by oxygen are driven as usual to support a plant's life cycle when the oxygen level is normal, which can be defined as normoxia and typically is ~ 21% of oxygen concentration in the atmosphere (McKeown,

2014; Silver and Erecińska, 1998). However, plants metabolic mechanisms are restricted when plants encounter particular abiotic stress. Flash flooding and drought may significantly alter the concentration of available oxygen. Those stresses can lead to a shortage of oxygen with the plants facing hypoxia (2-5% of oxygen concentration), anoxia (0% of oxygen concentration) and reoxygenation (Blokhina, et al, 2003) . Consequently, plants cells would suffer from a severe shortage of energy and carbohydrates (Bailey-Serres and Voesenek, 2008; Licausi and Perata 2009) resulting in lack of energy used in cell activities and finally may lead to the death of plants (Akhtar and Nazir, 2013; Vasellati, et al, 2011).

Interestingly, rice plants, particularly low land rice, has evolved adaptive processes to survive low O_2 triggered by those stress conditions, through the regulation of *Sub1A* which serves as a conjunction point between flooding and drought tolerance. During submergence, the ethylene and submergence-induced *Sub1A-1* allele of *Sub1A* allows endurance of complete submergence by restricting GA responsiveness and suppressing genes associated with cell wall loosening, starch and sucrose catabolism and flowering (Fukao and Bailey-Serres, 2008; Fukao, et al., 2006). Throughout drought stress, this allele generates dehydration tolerance by increasing ABA responsiveness inducing expression of *LEA* mRNAs and consequent suppression of leaf water loss and also increasing the expression of *ERFs* genes associated with acclimation of drought tolerance (Fukao, et at., 2011).

However, beside the adaptability to abiotic stress responses through the regulation of *Sub1A* initiated by ethylene, there are other mechanisms involved in plant tolerance to such stresses. Plants use change in ethylene, CO_2 and O_2, all of which are major importance substances occurring in various specific cellular concentrations and according to the plant organ, they act as signals to adapt their survival strategy when faced with abiotic stress (Voesenek and Sasidharan, 2013). These internal signals of increased levels of ethylene and/or decreasing O_2 concentrations have different kinetics and the dynamics of both gases also varies strongly

between plant organs such as roots and shoots (Voesenek and Blom, 1999) that consequently may lead to different means of regulation of adaptive morphological and metabolic changes during flooding in both organs.

Importantly, for plants during flooding, oxygen diffuses slower in water than in air, generating an oxygen shortage. This results in biochemical adaptation of plants from an aerobic to an anaerobic respiration. Additionally, ethylene also accumulates and serves as a signal to activate hypoxic responsive-genes. Unlike plants, animals have an oxygen sensing mechanism to induce abiotic stress responsiveness that is regulated by heme proteins which directly plays a central role as an oxygen sensor (Voesenek and Sasidharan, 2013) and then leads to the heterodimerization of hypoxia-inducible factor 1 (HIF1) constructed from HIF-1α and HIF-1β (Bruick, 2003; Haddad, 2004; Brahimi-Horn, et al., 2005; Hemelaers, 2012; Pucciariello and Perata, 2012) (Figure 5-5) which induce downstream hypoxia responsive elements of such as erythropoietin (EPO), phosphoglycerate kinase (PGK) and lactate dehydrogenase (LDH). These consequently result in an animal acclimatizes to the oxygen deficiency (Hemelaers, 2012).

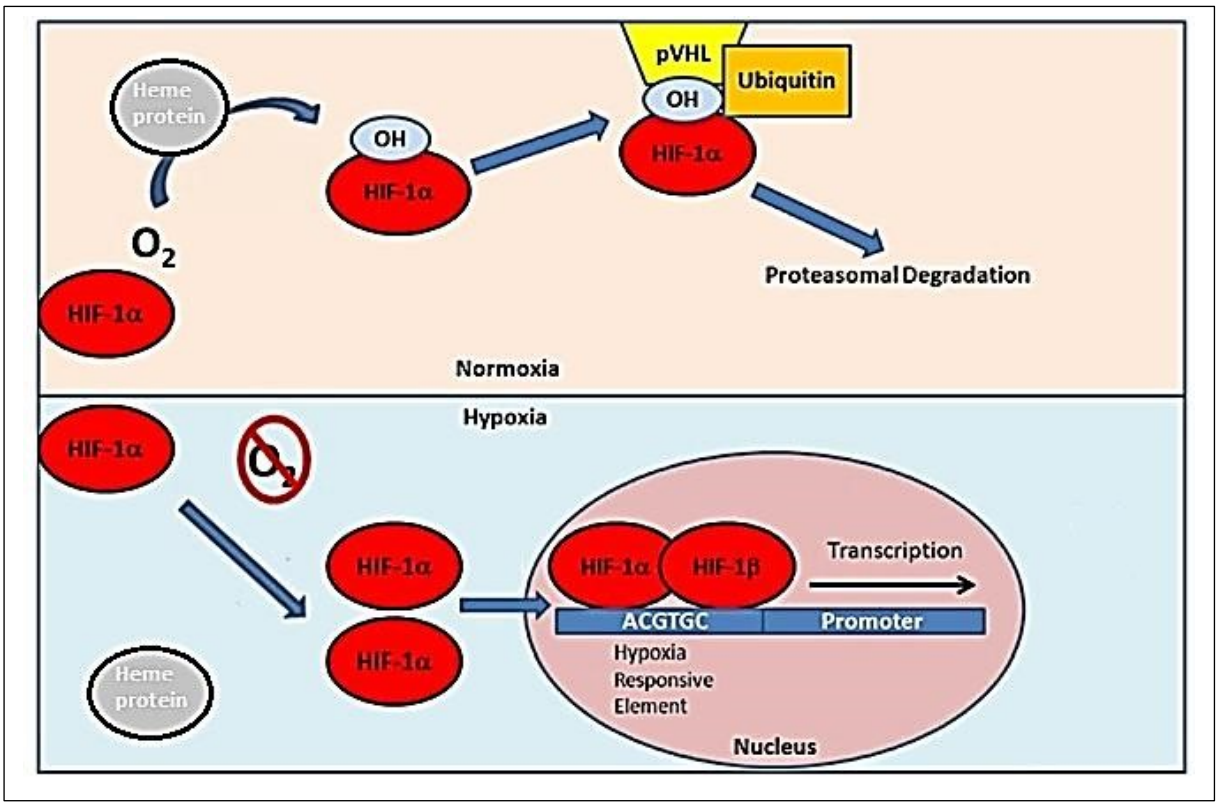


Figure 5-5 Oxygen sensing mechanism regulated through HIF-1α in animal cells. In normoxia, heme protein containing an iron molecule and a flavin group (FL) that is formed to ROS (peroxide, OH), through the proton gift of NADPH. ROS acts to degrade the hypoxia-inducible factor-1α (HIF-1α). During hypoxia, there is no formation of ROS so that HIF-1α is not degraded and then is transported to the nucleus, forming a heterodimer with HIF-1β. Consequently, this heterodimer (HIF-1) binds to DNA and then induces the downstream hypoxia-inducible gene (Modified from Goggins, et al, 2013)

Mounting evidence revealed that dicot plants, specifically *A. thaliana*, also have oxygen sensing mechanisms but these are different from those present in animals. In *A. thaliana*, the oxygen sensing process is driven by posttranslational regulation of key hypoxia response transcription factors through the N-end rule pathway (NERP). NERP is an evolutionarily conserved pathway for protein degradation associated with the fate of a protein substrates of the group VII ERF transcription factors having individual N-terminal residues with the sequence 'NH2-Met-Cys'. This protein substrate is classified as stabilising and destabilising depending on the fate of the protein (Gibbs, et al., 2011; Sasidharana and Mustrophc, 2011). The Cys residue at the N-terminus of these proteins is oxidized in the presence of O_2 and therefore leads to post-translational modification which finally is degraded. Though, the N-

terminus is stabilized under low O_2. In *A. thaliana*, only destabilising residues including RAP2.12, RAP2.2 and RAP2.3 which are members of group VII ERF transcription factors (Papdi, et al., 2015) are targeted for proteasomal degradation due to the fact it contains hydrophobic and basic N termini, respectively. Moreover, it has been confirmed that NERP and those substrates are the main key for regulating hypoxia-responsive gene expression in *A. thaliana* under hypoxia (Gibbs, et al., 2011; Sasidharana and Mustrophc, 2011; Papdi, et al., 2015) (Figure 5-6).

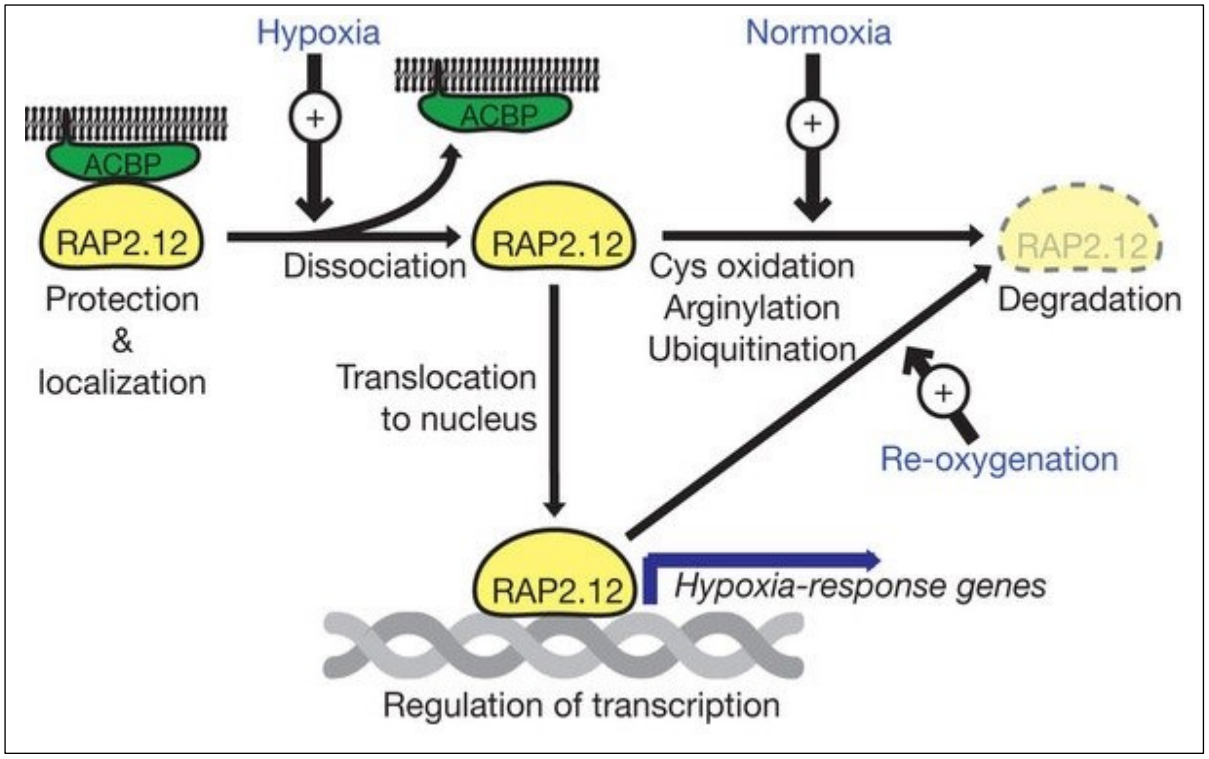


Figure 5-6 Model of the oxygen sensing mechanism in *A. thaliana*. The transcription factor RAP2.12 is constitutively upregulated under aerobic conditions. Consequently, RAP2.12 protein is always present and is associated with membrane-localized acyl-CoA-binding proteins (ACBP) to prevent transferring of RAP2.12 into the nucleus and to protect it against proteasomal degradation. During hypoxia, RAP2.12 transfers into the nucleus due to no association with ACBP, and then it activates anaerobic-gene expression. Upon reoxygenation, RAP2.12 is quickly degraded via the N-end rule pathway and proteasome-mediated proteolysis to downregulate the hypoxic response (Licausi, et al., 2011)

Focussing on rice, the means by which low land rice uses to adapt their survival under low oxygen condition is the well-known quiescence strategy regulated through *Sub1A*. However, the regulation via *Sub1A* under oxygen depletion has not been determined to date. There is a report revealing uncertain regulation of Sub1A (Xiong, et., 2012). It was found that despite the fact that some *Indica* cultivars such as FR13A and Goda Heenati carry the same *Sub1A-1* allele, they exhibited differential tolerance to submergence. This leads to an interest in additional influences on Sub1A regulation. Additionally, the crosstalk between flooding and drought mediated by *Sub1A* has been first reported but no more research has supported this finding. With those unclear events mentioned above, the regulation of *Sub1A* under abiotic stress still requires more investigation.

Furthermore, no evidence has yet been found that suggests monocots also possess an oxygen sensing mechanisms like those in animals and *A. thaliana*. There has been an attempt to examine oxygen sensing mechanism(s) in rice (Gibbs, et al., 2011) but it is still elusive. As mentioned above, in animals, the transcription factor HIF-1α is a key role in oxygen sensing mechanism regulated by heme protein which acts as oxygen sensor but unfortunately, it has been found that no HIF-1 α orthologue is present in rice. Furthermore in *A. thaliana*, the association of NERP and transcription factors of group VII ERF including RAP2.12, RAP2.2 and RAP2.3 play a role in oxygen sensing mechanism. Although *Sub1A* of rice also is a member of group VII ERF and contains N-terminal residues with 'NH2-Met-Cys, its regulation has been shown to be uncoupled from the N-end rule pathway, possibly because of either the absence of a significantly positioned lysine downstream of the N-end or the differences of tertiary structure. Therefore, oxygen sensing strategy in rice, which is normally used for a model of monocot, is still mysterious.

With the unclear regulation of *Sub1A* under low oxygen levels, the uncertain regulation of *Sub1A* has been studied by investigating the expression of some genes such as *SLR1* that

already has shown some interaction with *sub1A*. However, further investigation is required. Moreover, 2ODD genes, which encode proteins that typically catalyse the oxidation of an organic substrate by using a dioxygen molecule, have also been selected to be used in an investigation of the possible link between them and *Sub1A.*

5.2 Objectives

The main aim of this study is to evaluate the molecular approaches on *Sub1A* regulation under flooding and drought. Here, two groups of genes were selected to investigate their influences under both stress tolerance.

First, the *SLR* gene was chosen for the following reason. Under submergence, *Sub1A* has an influence on *SLR* which functions to inhibit the GA response that finally causes the restriction of rice shoot elongation. Moreover, it is found that the crosstalk between drought and flooding tolerance is mediated through the *Sub1A* regulator. This prompted us to examine the influence of *SLR* on acclimation response to water deficit as well.

Additionally, 2ODD genes were selected with the reason that under flooding and drought, reduction of O_2 level will occur and finally lead to hypoxic condition in cells. In addition, 2ODD genes encode proteins containing a domain that catalyses a substrate by using dioxygen. The alteration of O_2 level may have a direct impact on transcription of those genes under O_2 depletion.

With the two premises mentioned above, this study aims to investigate the expression of the *SLR* gene and selected 2ODD genes in tolerant and intolerant rice under both flooding and drought in order to evaluate the molecular regulation of *Sub1A* and to investigate the expression details of 2ODD genes under abiotic stress.

5.3 Material and methods

5.3.1 Submergence and dehydration treatment

The rice seeds were germinated and grown with the same protocol and conditions mentioned in section 2.2.1. For submergence treatment, fourteen-day-old seedlings in soil-containing pots (15 seedlings/pot) were completely submerged for up to 21 d in a black plastic tank (55 x 110 x 100 cm) which was filled with 90 cm of water and left standing for 1 d before placement of the potted plants in the tank. The tank water was not circulated or refreshed during the treatment. The turbidity of the tank water did not visibly increase during the submergence period. Air O_2, Dissolved O_2 and temperature were measured every specified day using a dissolved oxygen meter (Lutron, UK). After that, pots were moved to normal conditions by placing them in a shallow tray filled with water. For drought treatment, seedlings pots were exposed to dehydration by withholding water for up to 14 d and after that were placed in a shallow tray filled with water for complete soil rehydration. All submergence and dehydration treatments were replicated in three independent biological experiments. Seedling height was recorded before submergence and during submergence. Leaves from both treatments were harvested at noon, on the day of treatment specified, immediately frozen in liquid nitrogen, and stored at -80°C until use. Whole plant and leaf viability was evaluated on the day of treatment specified.

5.3.2 Genotyping

Genomic DNA Sample from IR64, IR64.S1, Gleva and Nipponbare were extracted following the protocol described in section 2.3.1.1 and then were amplified by performing PCR as the protocol described in 2.3.2.2 with forward primer “GAAAGCCCGGACAAACTTC” and reverse primer “AGGCTCCAGATGTCCATGTC”. Consequently, PCR products from IR64 and IR64.S1 were applied to discriminate the SNP by sequencing and restriction analysis; all

protocols have been described in sections 2.3.3 and 2.3.4, respectively. For restriction analysis, BseNI was used as restriction endonuclease enzyme shown specific recognition site at the single nucleotide polymorphism (SNP) of each nucleotide sequence as shown below.

5'...A C T G G N↓...3'
3'...T G A C↑C N ...5'

5.3.3 qPCR identification

qPCR was performed using the Applied Biosystems StepOne™ (Life technology, UK) system with 2x qPCRBIO SyGreen Mix Hi-ROX kit and then the mixed reaction components were used to conduct real time PCR amplification following three stages as described in section 2.3.2.3 with the specific primer pairs as shown in Table 5-1. After performing a relative standard curve method to evaluate the qPCR efficiency of each gene, the expression level of all genes was quantified through the comparative Ct method. Output data were processed using the software provided with this instrument.

Table 5-1 List of primer pairs used in quantification of selected genes involved in abiotic stress study

Gene name	Forward primer	Reverse primer
Sub1A	CTGCTTCTTGCTCAACGACA	CGAGCAGCACTCCAGTTGTA
SLR1	GATCGTCACCGTGGTAGAGC	GAGGGAATCGAACATGGTGG
eEF-1α	GCACGCTCTTCTTGCTTTCA	GAGTACTTGGGAGTGGTGGC
Os03g0803500	The same primer pairs used in Chapter 4	
Os07g0194500		
Os09g0489200		

5.4 Results and discussion

5.4.1 Morphological and molecular observations

Four varieties of rice seeds including IR64.S1, IR64, Gleva and Nipponbare were geminated in potted soil. 14 d old seedlings were submerged in a tank containing 90 cm of water for 21 d. Leaves were collected at 0, 1, 3, 5, 7, 10, 14 and 21 d during submergence as well as 1 and 7 d after recovery. The shoot length and water conditions and temperature were monitored at the specified times.

Plant survival depends on various aspects of the floodwater, including depth and duration of submergence, turbulence, dissolved gases, particularly O_2 and CO_2 (Ram et al., 2002; Das et al., 2009). This variable condition of floodwater may have an impact on rice survival so that the water conditions, particularly O_2 were measured to evaluate the responses of the plants.

The normal O_2 concentration in the air ranges 20.7-21.9% throughout the day whereas at the surface and bottom of the water, the O_2 concentration was slightly reduced (Table 5-2). During submergence, the O_2 concentration measured at noon was slightly decreased but it did not lead to hypoxic conditions in which the O_2 concentration is 2-5% (Blokhina, et al, 2003). This indicated that O_2 level in water surrounding rices was still high, an observation that is similar to that from Singh et al. (2009) who reported that the concentration of O_2 was higher in the afternoon throughout the submergence water profile. Moreover, Das et al. (2005) also detected high O_2 concentrations at or shortly after midday and near the water surface. In contrast to the outside cells of an organ surrounded by water, the intracellular O_2 concentration significantly decreased leading to cellular hypoxic condition due to reduced gas exchange between the plant and environment. This arises because O_2 diffusion in water is slower than in air, an approximate 10^4-fold reduction, leading to greatly reduced gas exchange that consequently reduces cellular O_2 resulting in restriction of aerobic respiration

(Nishiuchi, et al., 2012; Bailey-Serres and Voesenek, 2008; Mommer and Visser, 2005). Moreover, in air, stromata open allowing the exchange of gases such as O_2, CO_2 but under submergence, the stomata are hypothesized to close (Winkel, et al., 2013; Mommer and Visser, 2005), therefore gas exchange is restricted and subsequenty this limits aerobic respiration and results in morphological and metabolic alteration.

Table 5-2 O_2 Concentration and temperature at specified times

Time point	O_2 Concentration (%)			Temperature (C°)		
	Room	Surface water	Bottom water	Room	Surface water	Bottom water
0 d submerged (Plant free)	21.1	22.4	22.4	25.4	24.8	24.8
1 d submerged	21.9	21.0	20.0	24.6	24.1	24.3
3 d submerged	20.8	20.1	20.9	25.2	24.3	24.4
5 d submerged	20.8	20.5	20.4	24.3	25.3	25.2
7 d submerged	20.9	20.6	20.5	25.4	25.5	25.4
10 d submerged	20.8	20.4	20.2	25.1	25.6	25.6
14 d submerged	20.7	19.8	19.3	24.1	25.0	25.0
21 d submerged	20.7	19.3	19.0	24.7	25.1	25.3
1 d recovery	20.8	-	-	-	-	-
7 d recovery	20.8	-	-	-	-	-

Observation of shoot elongation, which is a key phenotype (Schmitz, et al., 2013), under submergence for 21 d showed that in Indica rice, IR64.S1 demonstrated limited elongation up to 21 d of submergence whereas IR64 still gradually extended its shoot and at the last day of submergence it had around two-fold higher than the first day of submergence. Similar to IR64, Gleva and Japonica also showed the elongation of their shoots but the rate of shoot extension was more than in IR64. It was found that at 5 d of submergence, both Gleva and Japonica had almost two-fold higher than at 1 d of submergence and at the last day of submergence, the shoot length was more than three-fold greater than on the first day of submergence (Figure 5-7). Examination of the controls of individual varieties, showed similar growth in that the shoot length gradually increased and at 21 d, it was almost two-fold compared to the first day. Compared to control of individual varieties, submerged IR64.S1 stopped growing, so that its shoot length was shorter than control, whereas shoot length of submerged IR64 was not significantly different to that of its control. For submerged Gleva and Nipponbare, both had a significant increase of shoot length with almost three-fold at 21 d (Figure 5-8).

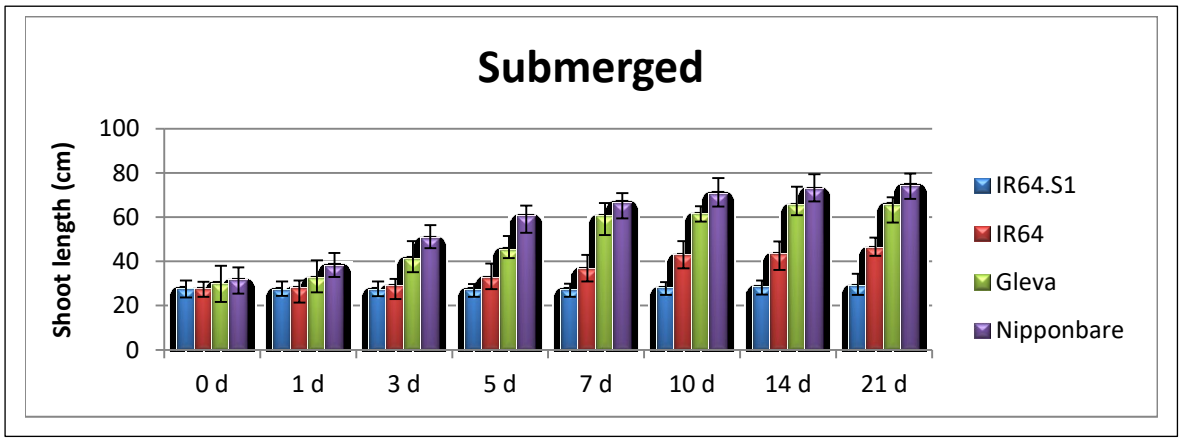
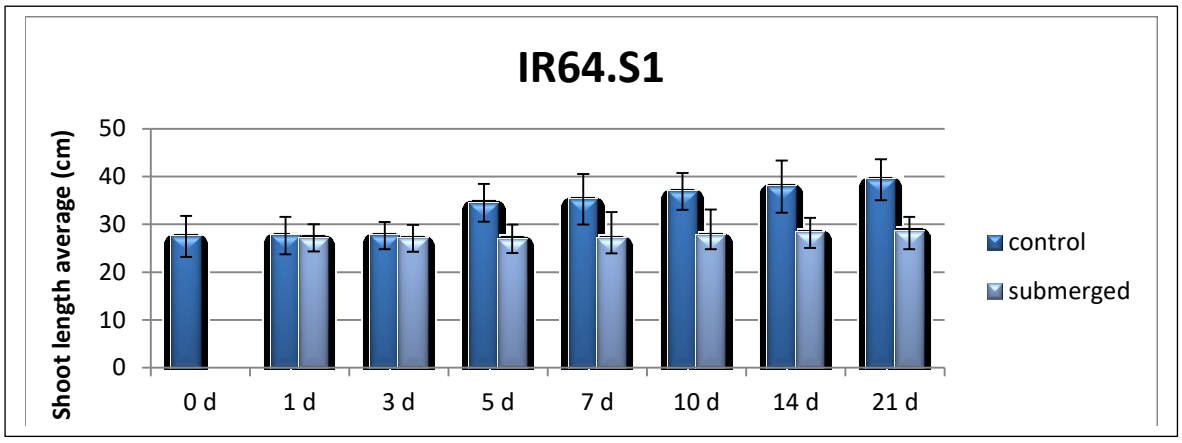
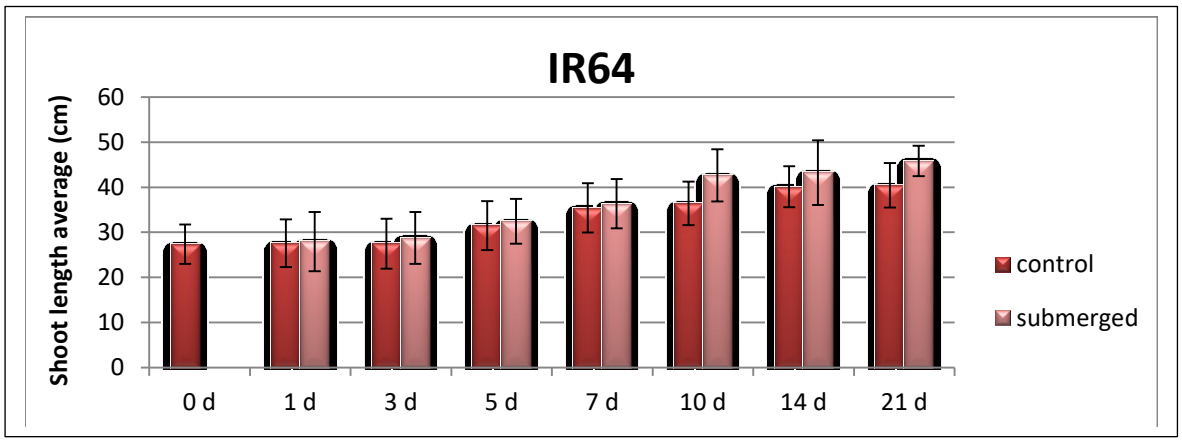
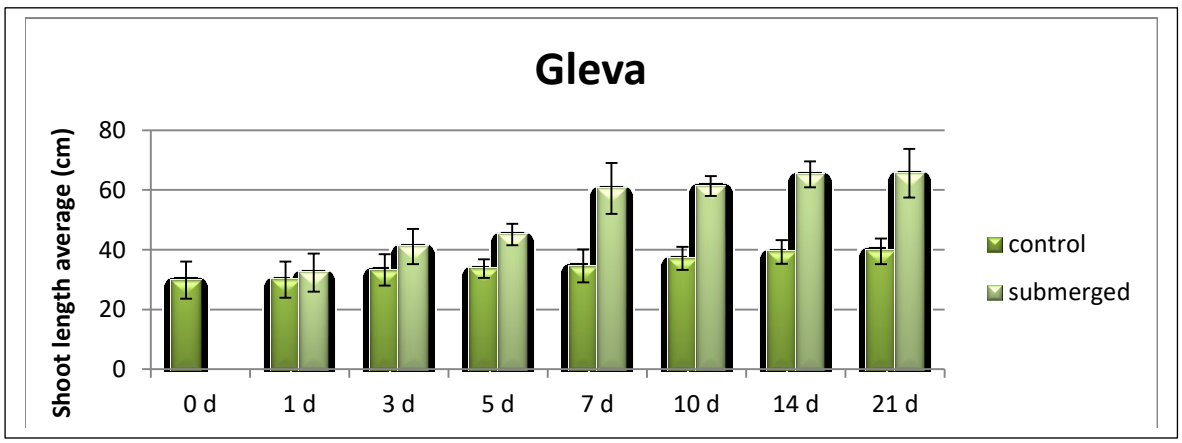


Figure 5-7 Bar graph indicates shoot length average of IR64.S1, IR64, Gleva and Nipponbare under control and submergence for three weeks. Error bars denote the standard error from three biological replicates.


IR64.S1
Shoot length average (cm)
50
40
30
20
10
0
0 d
1 d
3 d
5 d
7 d
10 d
14 d
21 d
control
submerged


IR64
Shoot length average (cm)
60
50
40
30
20
10
0
0 d
1 d
3 d
5 d
7 d
10 d
14 d
21 d
control
submerged


Gleva
Shoot length average (cm)
80
60
40
20
0
0 d
1 d
3 d
5 d
7 d
10 d
14 d
21 d
control
submerged

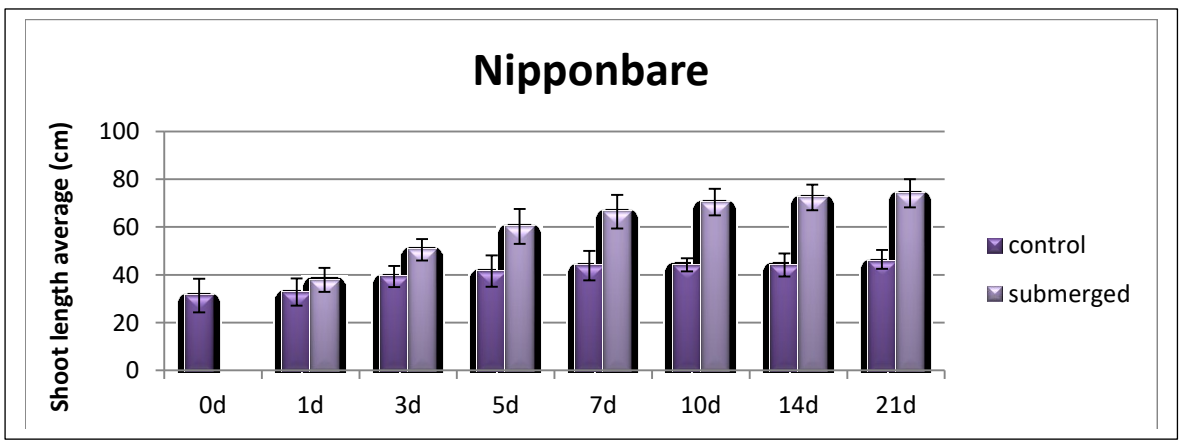

Nipponbare
Shoot length average (cm)
100
80
60
40
20
0
0d
1d
3d
5d
7d
10d
14d
21d
control
submerged

Figure 5-8 Comparison of shoot length average of IR64.S1, IR64, Gleva and Nipponbare under control and submerged conditions. Error bars denote the standard error from three biological replicates.

The variable growth of those four varieties is influenced by their genetic composition, specifically by the ERFs, Sub1A found in lowland rice and SK1/SK2 found in Deep-water rice (Hattori, et al.,2009; Nagai et al., 2010). In Indica rice, Sub1A is confirmed to be a regulator to control the adaptability of response to abiotic stress particularly flooding whereas Japonica does not contain Sub1A (Fukao, et al., 2006; Xu et al., 2006). To discriminate between two varieties of Indica, a restriction enzyme method and PCR product sequencing were applied. The result showed that both IR64.S1 and IR64 possess Sub1A but there are different alleles. IR64.S1 carries Sub1A-1 which allows flood tolerance whereas IR64 carries Sub1A-2 which gives flood intolerance (Figure 5-9, Figure 5-10). These contrasting genotypes in the two varieties resulted in the different responses to submergence as described above (Figure 5-7, Figure 5-8). This finding is supported by Singh et al. (2010) and Septiningsih et al. (2009) confirming that Sub1A-1 exists in IR64.S1 allowing tolerance to flooding and Sub1A-2 is present in IR64 and determines intolerance to flooding (Xu, et al., 2006).

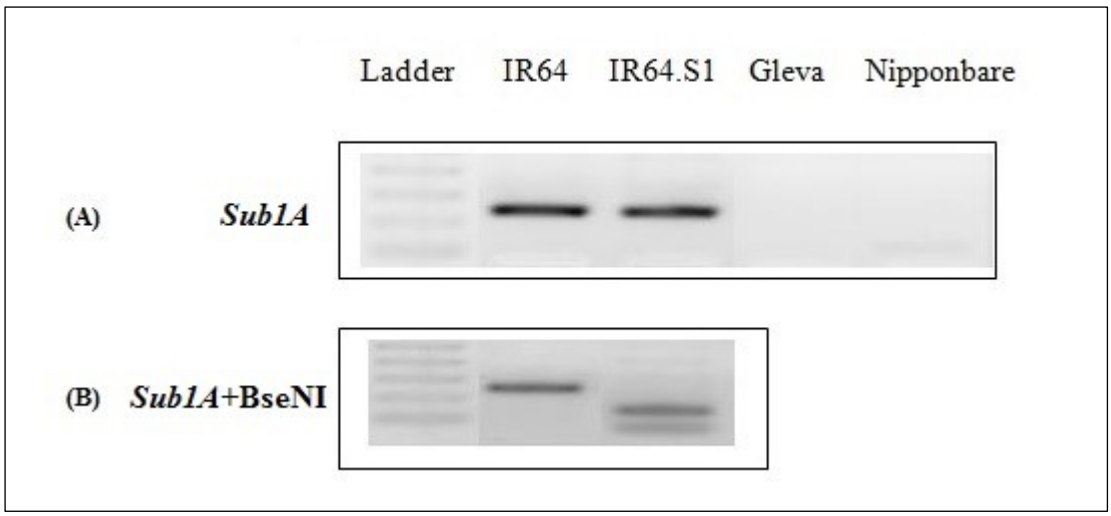


Figure 5-9 Genotyping of four varieties. (A) *Sub1A* is present in *Indica* rice (IR64 and IR64.S1) whereas it is not present in *Japonica* rice (Gleva and Nipponbare). (B) *Sub1A* amplicon was cut by enzyme restriction BseNI recognizing a specific SNP (described in Materials and methods) resulting in the cleavage of the *Sub1A* amplicon indicating that IR64.S1 carries *Sub1A-1* whereas in IR64, this enzyme does not recognize the SNP so that its amplicon was not cleaved indicating that IR64 carries *Sub1A-2*.

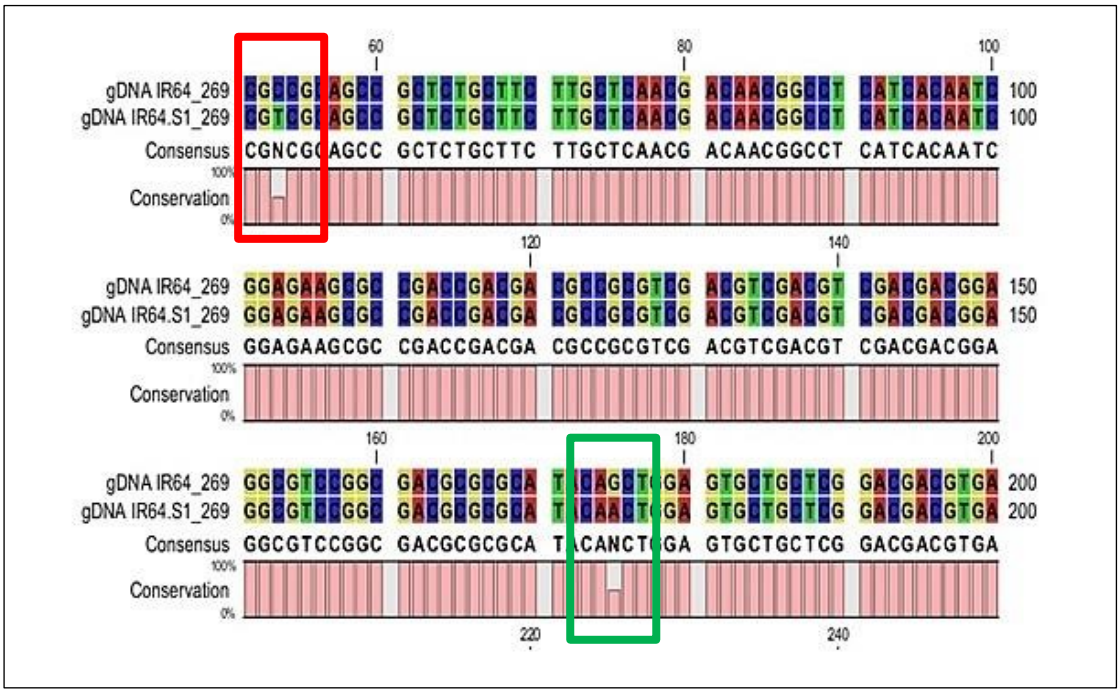


Figure 5-10 Sequencing of IR64 and IR64.S1 amplicon. Red box shows functional SNP that distinguishes the two varieties having different alleles, which are *Sub1A-2* in IR64 and *Sub1A-1* in IR64.S1 generating different amino acids proline and serine leading to intolerance and tolerance, respectively. Green box indicates non-functional SNP which leads to no difference in amino acid encoded (Xu, et al., 2006).

Moreover, it seemed IR64 containing Sub1A had a response to submergence as similar as other two Japonica, Gleva and Nipponbare, which do not contain Sub1A (Figure 5-9) and this was also confirmed by Fukao et al. (2006) and Xu et al. (2006) revealing Sub1A is absent from all Japonica therefore both do not survive prolonged submergence. For IR64, after submergence, the shoot length increased steadily resulting in leaves that could not break through water surface and this elongation was not significantly different with its control (Figure 5-7 and Figure 5-8). In addition, its leaves rolled in, turned to yellow and decayed (Figure 5-11, Figure 5-12). Gleva and Nipponbare also showed the same changes in leaf phenotype as IR64 but their submerged shoots were less damaged than IR64 (Figure 5-11, Figure 5-12). This occurred because their growth rate was faster than IR64 allowing leaves to emerge through the water surface. This circumstance leads to photosynthesis and increases gaseous exchange in leaves which were over the water and subsequently generates the synthesis of carbohydrate which is transferred to submerged tissue for consumption. Furthermore, within 5 d of submergence, the shoot length of Gleva and Nipponbare significantly increased almost two-fold compared to their control. The elongation increased until at 21 d of submergence, the shoot length was almost four-fold more than their control. Similarly, Singh et al. (2010) stated that intolerant rice varieties and deep-water rice exhibit rapid growth under submerged conditions (escape response). That excessive response in Gleva and Nipponbare firstly caused us to consider the possible influence of *SK1*/*SK2* regulators which deep-water rice use to cope with flooding by an escape strategy which leads to significant elongatione of the shoot, which finally emerges from the water surface and survives. However, this hypothesis rejected because unlike deep-water rice, Gleva and Nipponbare are lowland rice which definitely do not contain *SK1*/*SK2* (Hattori, et al.,2009; Nagai et al., 2010). Therefore, the possible reason for such response to submergence of both varieties is weighted towards the regulation by ethylene. Generally, submerged tissue initially has a response to submergence by rapidly accumulating the gaseous phytohormone, ethylene,

(Fukao, et al., 2008; Voesenek, et al., 2004) due to physical entrapment and active biosynthesis during the stress. This primary response subsequently triggers a range of acclimation responses including shoot elongation, adventitious root formation and carbohydrate metabolism. Moreover, ethylene affects the balance of gibberellic acid (GA) and abscisic acid (ABA) contents, which facilitates GA-promoted elongation of shoots during submergence (Fukao, et al., 2008). Beside the effect of ethylene and GA (Fukao, et al., 2006; Fukao and Bailey-Serres, 2008), previous studies have confirmed that *Sub1C* also has an involvement in the regulation of shoot elongation. It is suggested that, in the absence of *Sub1A-1*, *Sub1C* enables shoot elongation during submergence, through a GA-dependent mechanism (Niroula, et al., 2012; Fukao and Bailey-Serres, 2008) whereas in the presence of *Sub1A-1, a* feedback regulation of ethylene allows an increase in GA responsiveness and consequent *Sub1C* mRNA accumulation, resulting in a restriction of shoot elongation (Niroula, et al., 2012; Fukao and Bailey-Serres, 2008).

In addition, it was found that intolerant rice (IR64, Gleva and Nipponbare) similarly adapted to submergence by elongating their shoot. This observation also confirmed the findings of Schmitz et al. (2013) who reported that flood intolerant rice varieties typically exhibit rapid elongation upon prolonged submergence. However, they observed different growth rates, in that shoot length in *Japonica* increased more rapidly than that in IR64. This is possibly a result of the influence of *Sub1A*. Although IR64 carries *Sub1A-2* alleles which encode an intolerant phenotype, *Sub1A-2* may still have an influence on the repression of shoot elongation and carbohydrate consumption under submergence stress due to reduced ethylene-mediated GA biosynthesis (Septiningsih, et al., 2009; Perata and Voesenek, 2007; Fukao, et al., 2006). However, this repression of *Sub1A-2* is significantly weaker compared to *Sub1A-1*. Gleva and Nipponbare with no *Sub1A* are abundantly regulated by ethylene and *Sub1C* therefore allowing fastest elongation (Figure 5-13). Tolerant rice can withstand up to two

weeks of complete submergence (Niroula, et al., 2012; Xu, et al., 2006). Here, all varieties were subjected to prolonged submergence for more than two weeks (21 d). The result showed that shoots and leaves collapsed and decayed in all of the varieties even IR64.S1 which contain *Sub1A-1* (Figure 5-15). This finding confirmed previous studies that *Sub1A-1* is capable of enhancing the tolerance of submergence only up to two weeks.

Moreover, to investigate the tolerance of the four rice varieties, a recovery stage was monitored. Fourteen-day-old seedlings of all genotypes were completely submerged for a longer duration (14 d). A significant difference in the appearance of the plants was observed immediately after desubmergence and at 7 d of recovery. As shown in Figure 5-12 and Figure 5-14, immediately after desubmergence, the shoots of *Japonica* and IR64 collapsed and decayed, but similar damage was observed only in fewer leaves in IR64.S1. At 7 d of recovery, it is clearly shown that most of IR64.S1 shoots were green, strong as well as able to stand up. For IR64 and Gleva, the shoots were very weak but there was some new green tillers emerging from the submerged tiller sheath. Similarly, Nipponbare also showed new tillers emerging from submerged tiller sheaths but its submerged tillers were stronger than the other two cultivars (Figure 5-17 and Figure 5-18). It seems that all varieties are able to recover. This is because after desubmergence, shoots are exposed to the air that allows gasous exchange and generates the normal processes such as photosynthesis and respiration. Consequently, the growth involves typical developments such as the emergence of new tillers, self-supporting shoots, and the production of new leaves leading to subsequent survival. However, it was found that the presence of *Sub1A* in IR64.S1 allowed this variety to to recover more rapidly than other varities. This difference may possible be a result of ROS scavenging (Fukao, et al., 2011). There is evidence revealing that ROS are generated not only during submergence but also during the recovery time; this is associated with the increase in concentration of O_2. After desubmergence, reoxygenation occurs allowing fast increase of O_2

concentration and subsequent oxidative stress. This rapid change triggers the over accumulation of ROS leading to cell damage. In tolerant rice, *Sub1A* is a key factor to enhance ROS amelioration resulting in limitation of oxidative damage and chlorophyll degradation during reoxygenation. Fukao et al. (2011) observed that *Sub1A* steadily increased the abundance of transcripts which encode ROS scavenging enzymes, generating greater tolerance to oxidative stress.

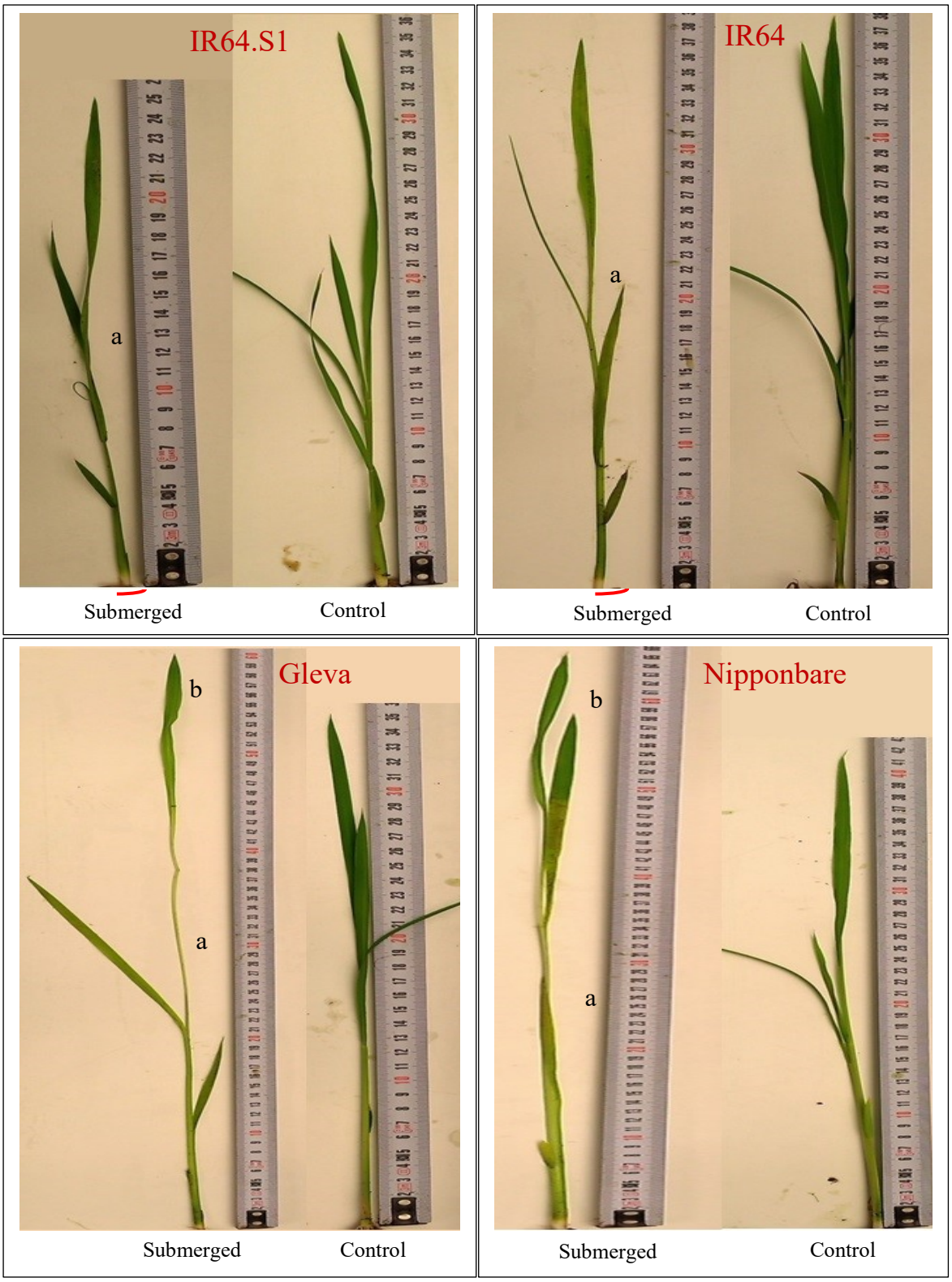


Figure 5-11 Phenotypes of four rice varieties when submerged at 7 d. IR64.S1 stopped growing but submerged shoot and leaves were still green whereas IR64, Gleva and Nipponbare still had shoot elongation. IR64 shoots did not breach the water surface but Gleva and Nipponbare shoot emerged through the water surface allowing some leaves to be above the water. a) Submerged leaves folded to thin and become light green. b) Non-submerged leaves were green.

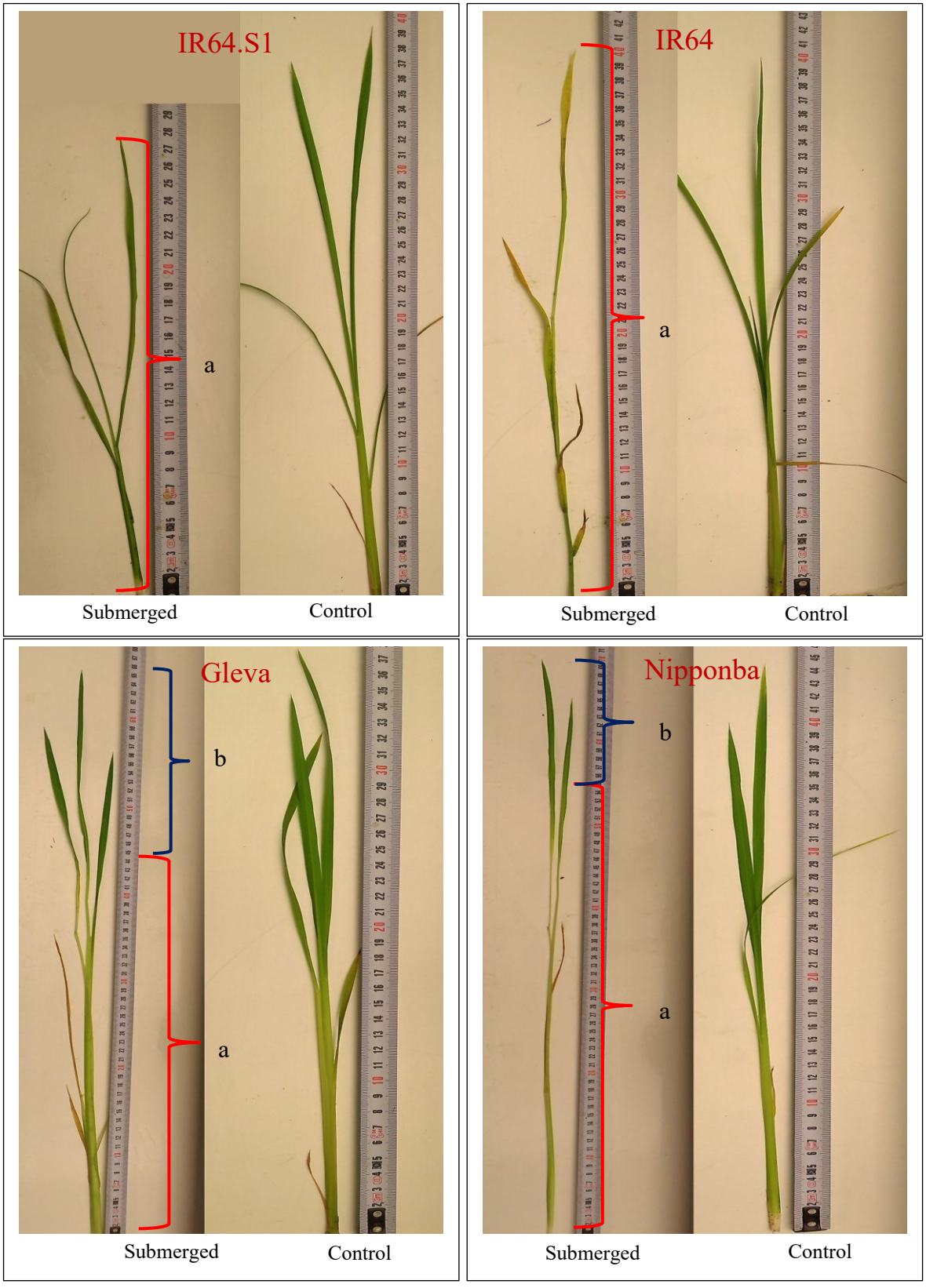


Figure 5-12 Phenotypes of four rice varieties when submerged for 14 d. IR64.S1 stopped growing but submerged shoot and leaves had less damage whereas IR64, Gleva and Nipponbare still had shoot elongation. However, submerged tissue of these varieties was severely damaged. a) and b) indicates the same appearance described in Figure 5-11

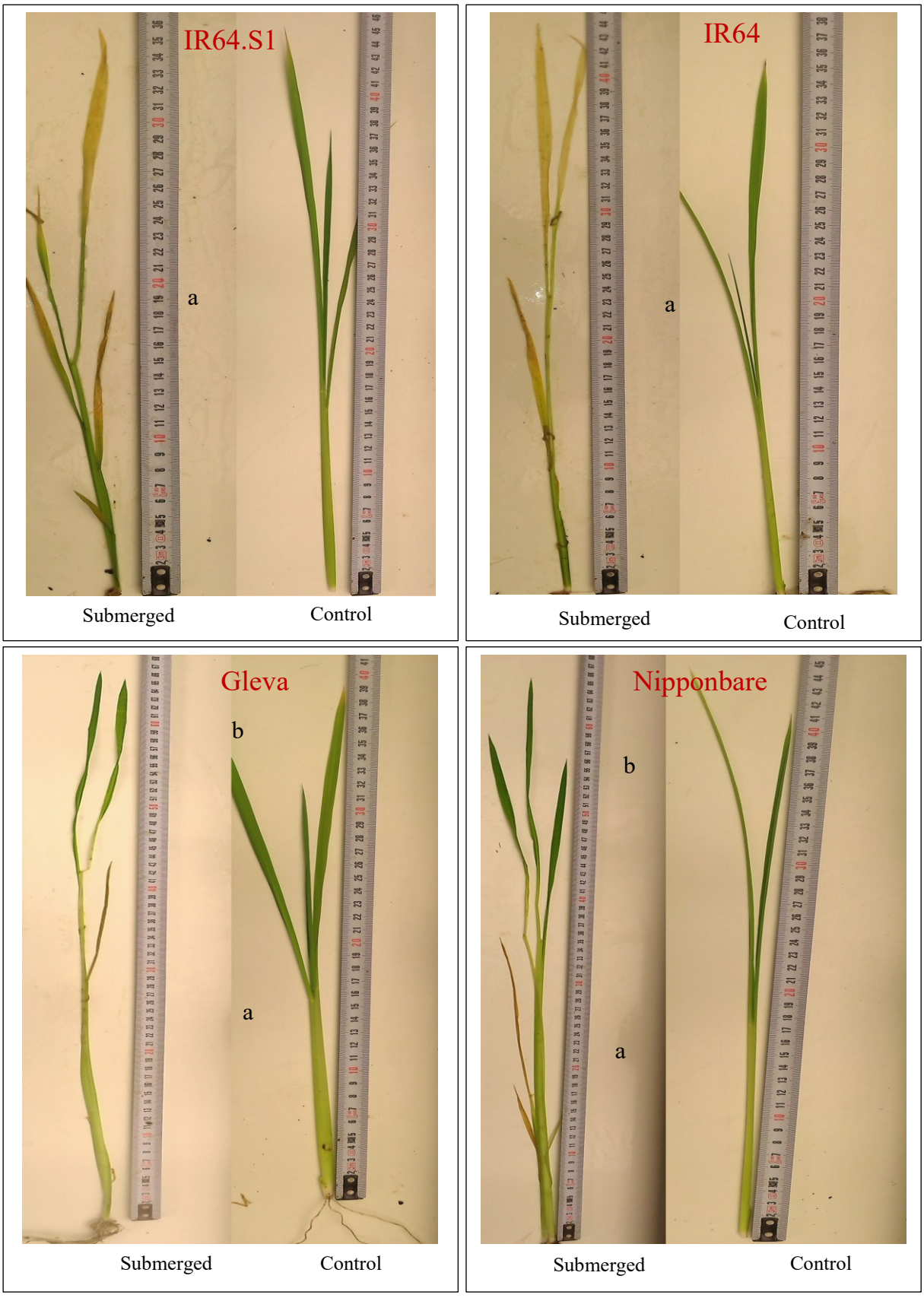


Figure 5-13 Phenotypes of four rice varieties when submerged at 21 d. Submerged tissue of all varieties had severe damage and decayed. a) and b) indicates the same appearance described in Figure 5-11 and Figure 5-12.

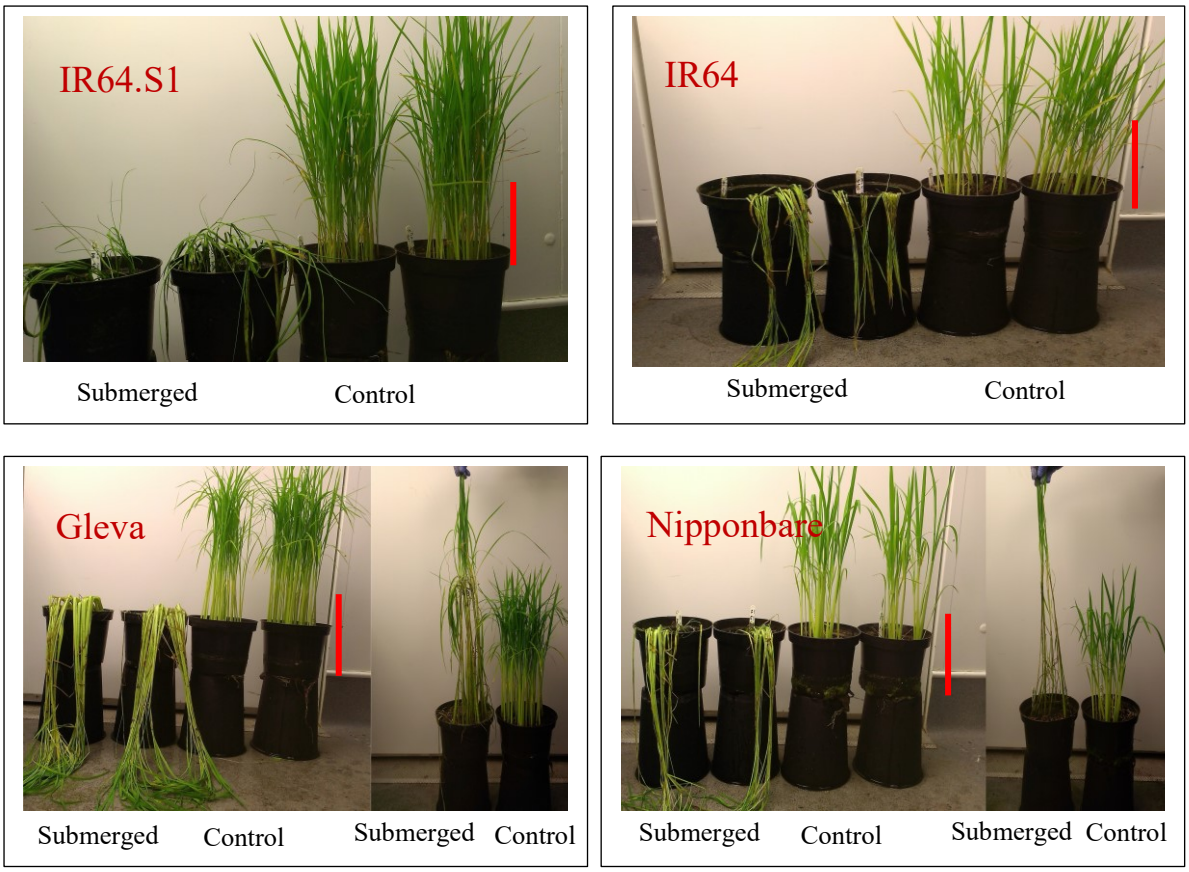


Figure 5-14 Phenotype appearance of four rice varieties immediately after desubmergence compared to control. All varieties were submerged up to 14 d in a tank filled with water to a depth of 90 cm. IR64.S1 stunts but shoots are able to stand up. IR64, Gleva and Nipponbare collapse but the shoot length varied. For IR64, the shoot length increased gradually and was not significantly different to control but for both Gleva and Nipponbare, the shoot length increased rapidly leading to an almost three-fold increase more than their control.
Red bar = 20 cm.

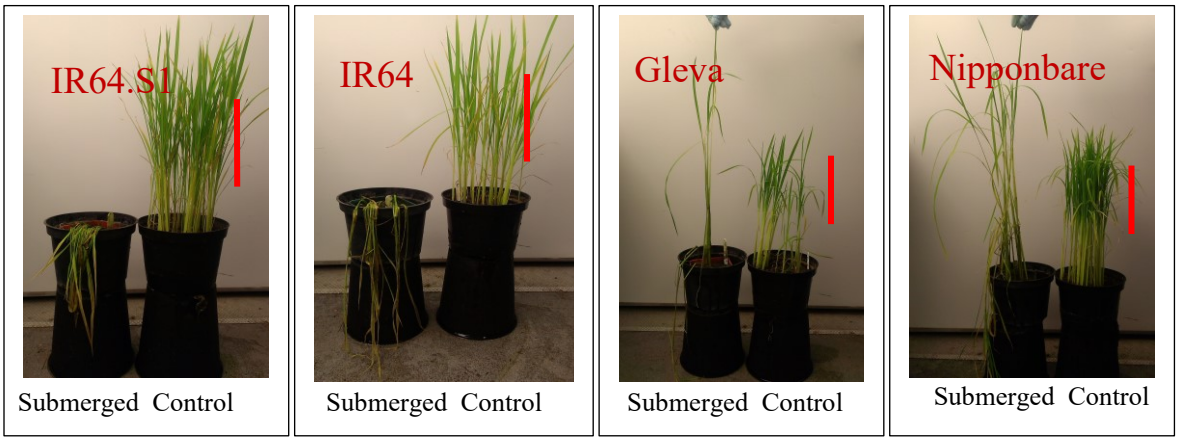


Figure 5-15 Phenotype appearance of four rice varieties at 21 d of submergence compared to its own control. All varieties had severe damages even tolerant rice, IR64.S1. Red bar = 20 cm.

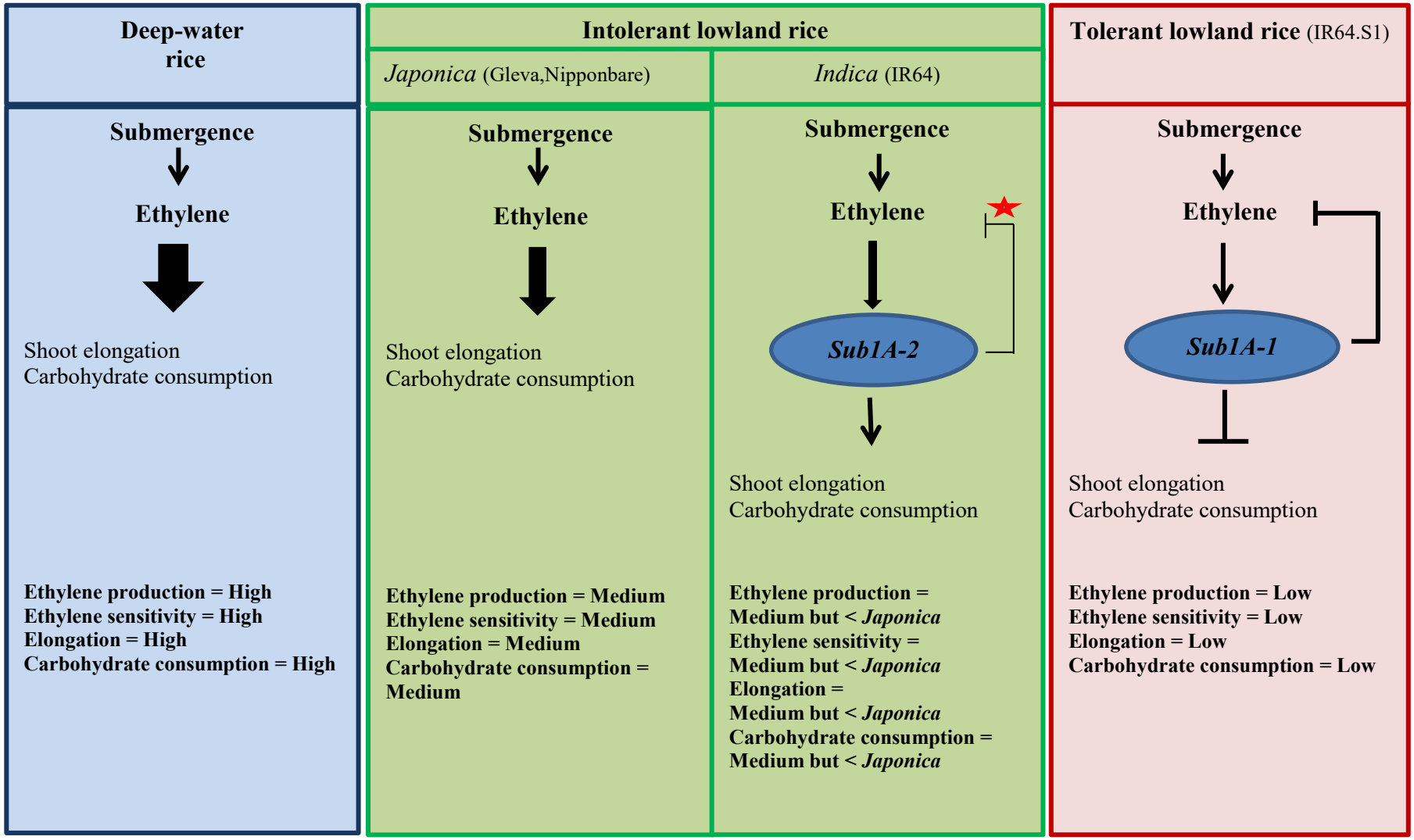

Deep-water rice
Intolerant lowland rice
Japonica (Gleva,Nipponbare)
Indica (IR64)
Tolerant lowland rice (IR64.S1)
Submergence
Ethylene
Shoot elongation
Carbohydrate consumption
Sub1A-2
Sub1A-1
Ethylene production = High
Ethylene sensitivity = High
Elongation = High
Carbohydrate consumption = High
Ethylene production = Medium
Ethylene sensitivity = Medium
Elongation = Medium
Carbohydrate consumption = Medium
Ethylene production = Medium but < Japonica
Ethylene sensitivity = Medium but < Japonica
Elongation = Medium but < Japonica
Carbohydrate consumption = Medium but < Japonica
Ethylene production = Low
Ethylene sensitivity = Low
Elongation = Low
Carbohydrate consumption = Low

Figure 5-16 The interaction of ethylene and Sub1A regulating rice response to submergence. When submerged, ethylene rapidly accumulates but its production is unequal in different varieties. Ethylene is most synthesized and detected in deep-water rice resulting in the fastest growth rate and subsequent emergence of the shoot through the water surface allowing an escape strategy. Intolerant lowland rice, *Japonica* and IR64 had a similar response to submergence but the growth rate of IR64 was slower than that of *Japonica*. This is because the presence of *Sub1A-2* in IR64 might also promote the feedback regulation of ethylene synthesis but this signal may be extremely poor compared to tolerant rice. On the other hand, tolerant rice (IR64.S1) contains *Sub1A-1* which inhibits the production of ethylene through feedback regulation. Finally, restriction of shoot elongation arises. Red asterisk indicates very weak depression of ethylene synthesis influenced by *Sub1A-2*. The figure is modified from Fukao, et al., 2008.

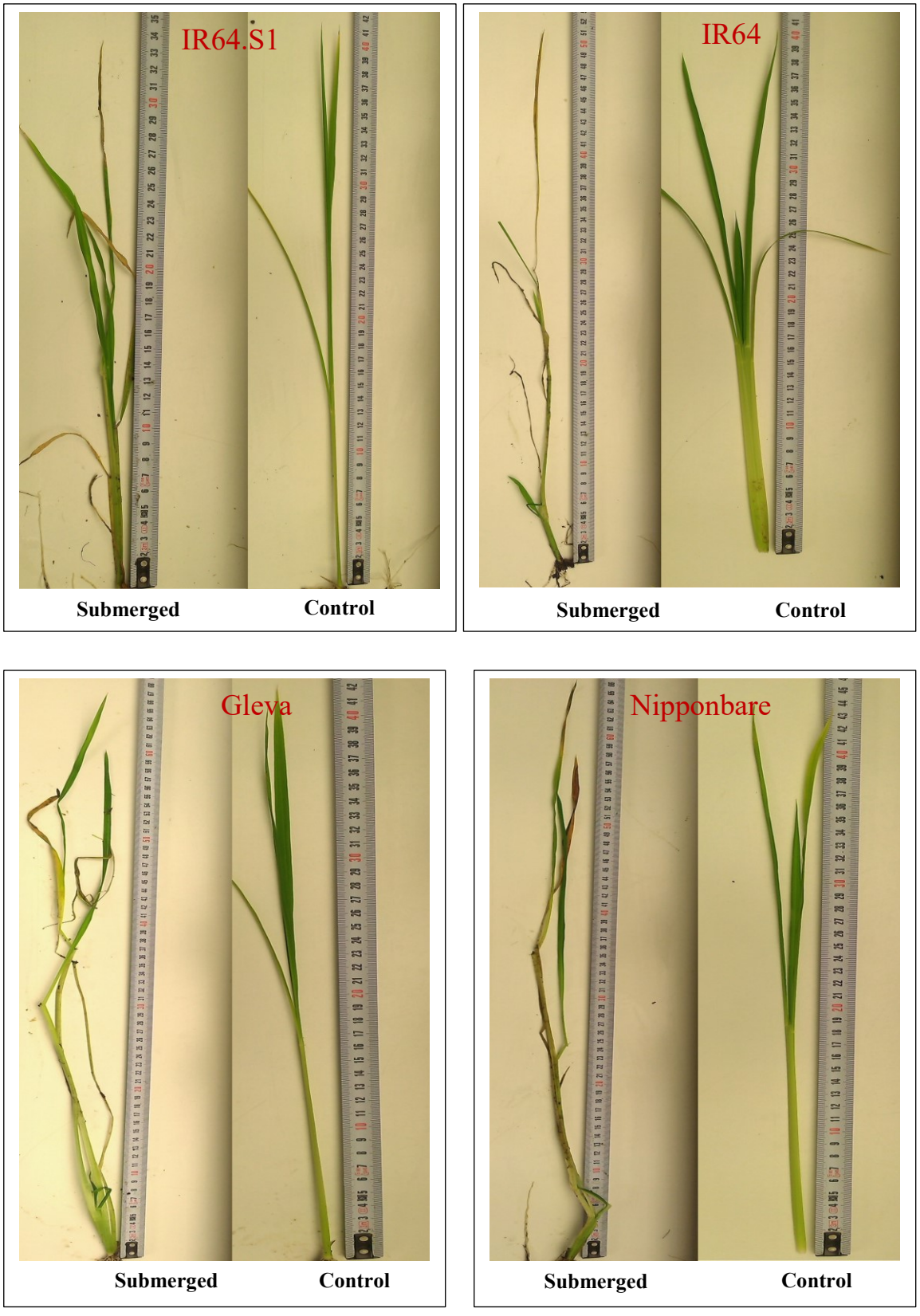


Figure 5-17 Phenotype of four rice varieties after 7 d of recovery. Shoot is able to arise in IR64.S1 whereas a new tiller emerged from damaged sheath in IR64, Gleva and Nipponbare.

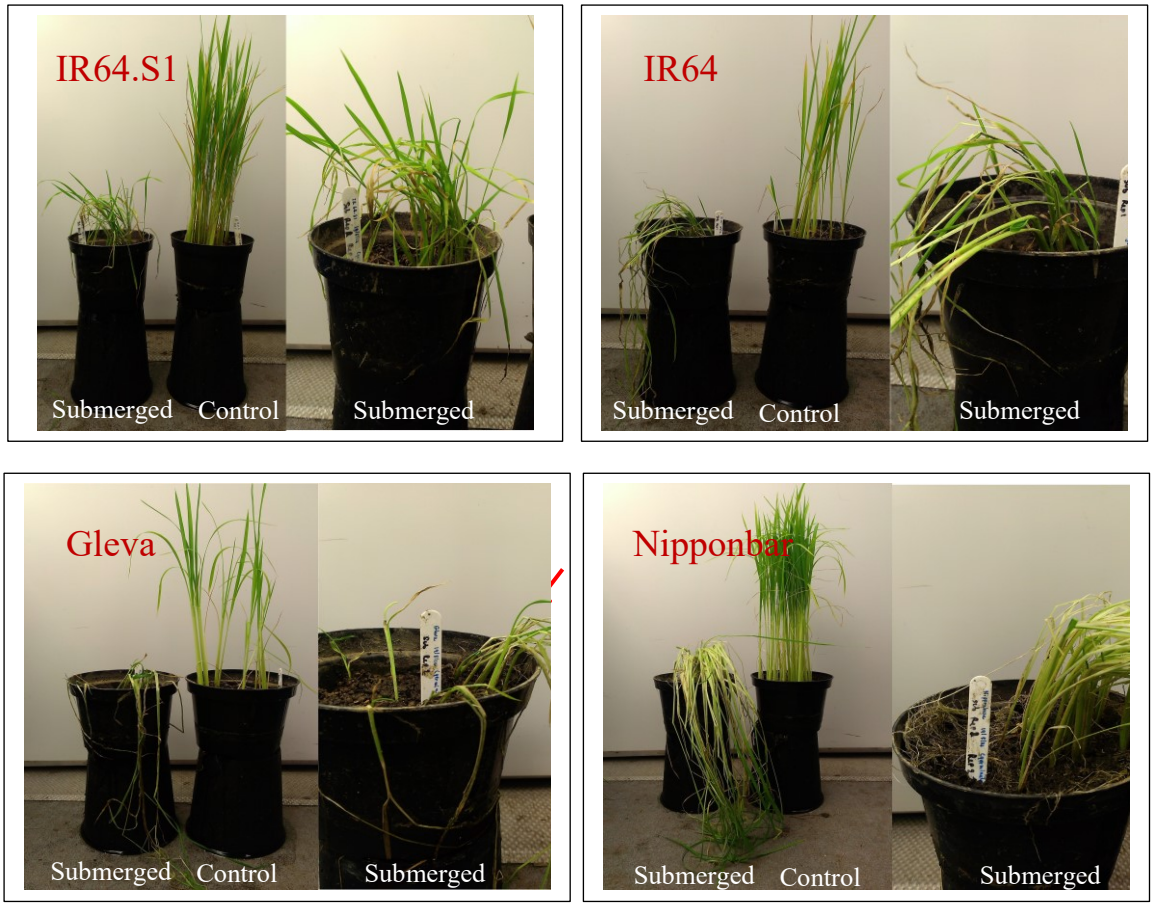


Figure 5-18 Phenotype appearance of four rice varieties after 7 d of recovery compared to the control**.** Red arrow represents the emerging leaf from submerged leaf sheath. Red bar = 20 cm.

Interestingly, it was reported that *Sub1A* also mediates acclimation to drought. To evaluate this discovery, the same varieties of rice previously used in the study of submergence tolerance were tested. Fourteen-day-old seedlings in soil-containing pots grown in a 2.5-liter pot for 14 d were exposed to dehydration stress by withholding water for up to 14 d. Next, pots were placed in a shallow tray filled with water for complete soil rehydration. The results showed that at 7 d and 14 d of dehydration, there is no significant difference in phenotype appearance of all varieties but shoots and leaves were severely wilted in all varieties at 14 d of dehydration (Figure 5-19). Similarly, at 7 d of rehydration, leaves of all varieties recommenced leaf development. However, a small difference of phenotype change in all varieties was noticed. IR64 and Gleva showed better recovery from dehydration than IR64.S1

and Nipponbare (Figure 5-19). All wilted leaves of the former group turned to green and looked fresh. For the latter group, fewer dried leaves became fresh in appearance and some still wilted (Figure 5-19). This finding is similar to the previous report of Fukao et al., (2011) who found that at 8 d of dehydration, leaves of *Japonica* inbred line, M202, and a near isogenic SUB1 introgression line, M202(Sub1) severely wilted and at 14 d of recovery both genotypes formed new leaves (Figure 5-20). Nevertheless, the contrast beetween this study and Fukao et al. (2011) is about the potential recovery beetween varieties. In this study it was found that *Sub1A* did not influence drought tolerance whereas Fukau et al. (2011) observed that there was a significant difference in the formation of new leaves between the two genotypes. The results showed 71.7% of M202 (Sub1) plants restarted leaf development as compared with only 11.7% of the M202 plants. The different results from these two studies may result from several reasons. First, the different periods of recovery, namely 7 d and 14 d. may lead to diverse results. This study collected data at 7 d of recovery and found all varieties restarted leaf development leading leaves to be fresh but there was no significant difference between rice with or without *Sub1A,* whereas Fukao et al. (2011) noticed the significant difference in the formation of leaves at 14 d of recovery. Next, different varieties lead to different acclimation to stress due to genetic differences. Fukau et al. (2011) studies a *Japonica* near isogenic *Sub1A* introgression line (M202(Sub1)) whereas this study was conducted in an *Indica* near isogenic *Sub1A* introgression line (IR64.S1). Unfortunately, the Fukau et al. (2011) observation only stated that withholding water was for a period of 8 d and viability was evaluated after initiation of recovery. For this initiation of recovery, the recovery day was not initiated immedietely after dehydration for 8 d and it seemed dehydration exposure had extended continuously without confirming the exact number for drought treatment as shown in Figure 5-21. So that, this recovery platform from Fukao et al. (2011) may not allow an effective comparison with this study. However, to clarify this, the molecular investigation was conducted and would be explained in next following section.

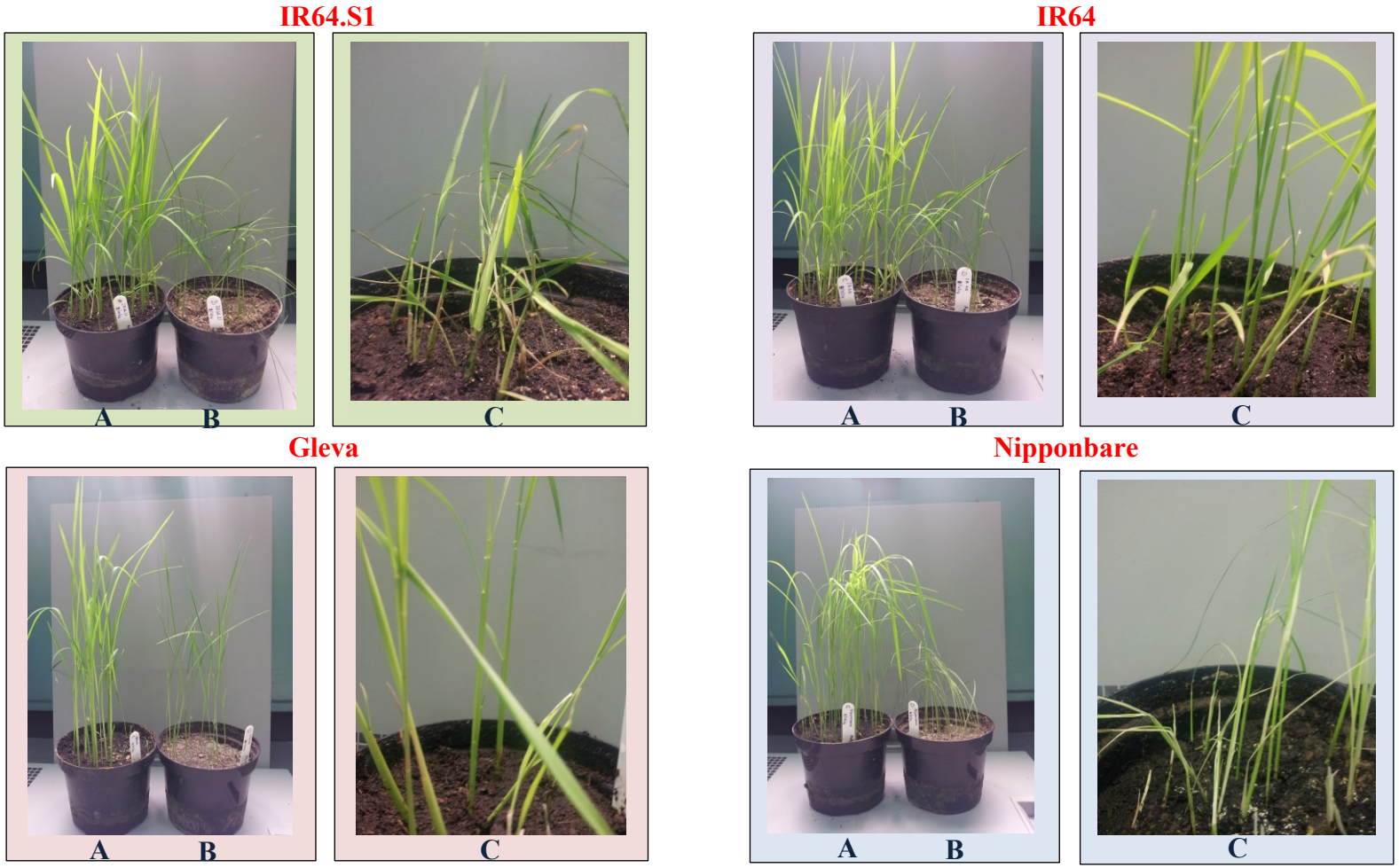


Figure 5-19 Phenotype appearance of four rice varieties subject to dehydration and rehydration. A) Control B) Withholding water for 14 d C) 7 d of rehydration; Red bar = 20 cm.

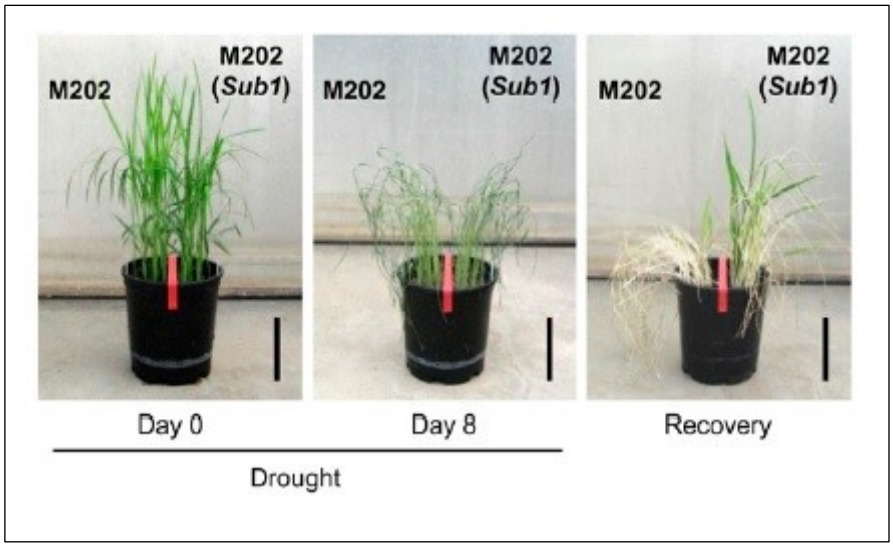


Figure 5-20 M202 and M202(Sub1) lines planted in the same pot were subjected to drought treatment for 8 d and were allowed to recover under regular watering conditions for 14 d. Bars = 10 cm. This picture retrieved from Fukao et al. (2011) that showed the difference of phenotypes between Day 8 of drought treatment and Recovery day. At Day 8 of drought treatment, leaves just wilted but at Recovery, leaves greatly dried and damaged. This indicated that the initiation of recovery may not take place immediately after 8 day of dehydration.

5.4.2 Analysis of gene expression in response to abiotic stresses

Beside the comparative analysis at phenotypical level of the four rice varieties described above, an *Indica* inbred line, IR64, a near isogenic SUB1 introgression line, IR64.S1, *Japonica* susceptible varieties, Gleva and Nipponbare, were also subjected to comparative molecular analysis. Under submergence, fourteen-day-old plants were completely submerged for up to 14 d and leaf tissue was collected at six specified times. The abundance level of *Sub1A* mRNA increased rapidly after 1 d of submergence in IR64.S1 leaves and remained at this high level for up to 14 d of submergence stress, whereas the accumulation of *Sub1A* mRNA gradually reduced for up to 7 d of desubmergence. In contrast to IR64.S1, the level of *Sub1A* in submerged and non-submerged IR64 showed the same pattern, that it very poor expression at every time point of both treatment. This finding confirmed the morphological analysis which was regulated by *Sub1A* that was described previously. Under submergence,

rice containing *Sub1A* is restricted in growth development particularly shoot elongation associated with avoiding the energy consumption (Nagai, et al., 2010) leading to the limitation of shoot length (Figure 5-7). For *Japonica* rice, *Sub1A* is not present in Gleva and Nipponbare, and therefore *Sub1A* expression in both varieties was not reported.

The regulation of flood tolerance response is conferred by *Sub1A*. This regulation is associated with many other factors. Slender Rice-1 (SLR1) is also one of the key genes associated with flooding tolerance adaptation that was associated with *Sub1A*. SLR1 is a gibberellin (GA) signalling repressor functioning concomitantly reduced GA-inducible gene expression under submerged conditions. In the *Sub1A* overexpression line, it was confirmed that *Sub1A* promotes accumulation of *SLR1* transcripts in aerial tissue under submerged conditions (Fukao et al., 2008). To confirm the role of *Sub1A* which promotes the accumulation of *SLR1* transcript, this study also investigated *SLR1* expression under submergence. Interestingly, the results showed that without submergence, *SLR1* was more highly expressed in *Indica* than in *Japonica*. When submerged, in *Indica*, *SLR1* expression was reduced up to 14 d of submergence and particularly in the *Sub1A* overexpression variety (IR64.S1), *SLR1* transcript level was significantly low at 7 d of submergence. In other words, *SLR1* was downregulated under submergence. In addition, it slightly increased at 1 d of recovery and at 7 d of recovery it increased to the same level as the control. These results are in contrast to the previous report revealing that *Sub1A* promotes the accumulation of *SLR1* which acts to repress GA inducible gene expression (Fukao, et al., 2008). In this study, it was found that *Sub1A* did not have an influence on accumulation of *SLR1* transcript under submergence stress. The distinction of *SLR1* accumulation in these two studies may be due to two main factors. First, the varieties of Sub1 introgression line are dissimilar. This study used *Indica* variety as Sub1 introgession line (IR64, IR64.S1) whereas Fukao et al. (2008) used *Japonica* variety (M202, M202(Sub1)). It is true that the submergence tolerance response is

mediated by *Sub1A*; however, it was revealed that beside *Sub1A*, the Sub1 locus haplotype typically comprises of two more ERFs, namely *Sub1B* and *Sub1C*. For classically intolerant rice, IR64 contains *Sub1B-1* and *Sub1C-3* alleles, whereas M202 possesses *Sub1B-2* and *Sub1C-2* alleles (Schatz, et al., 2014; Fukao, et al., 2006; Xu, et al., 2006). Wild *Indica* tolerant varieties possess only the specific *Sub1C-1* allele (Fukao, et al., 2006; Xu, et al., 2006). This distinction it may be responsible for the difference of *SLR1* expression. Another factor, tissue type is different in these two studies. This study used leaves as a sample whereas Fukao et al. (2008) used roots, therefore cellular differentiation may reflect to the regulation of *SLR1*. For submerged and non-submerged *Japonica* rice, *SLR1* expression was not significantly different at almost every time point. The constant levels of *SLR1* transcripts in *Japonica* may be caused by a requirement for specific amounts of the SLR1 protein for survival, so this gene does not show any substantial change in expression.

In addition to the investigation of *Sub1A* and *SLR1* expression under submergence, 2ODD genes were also examined. As O_2 is important to a plant's life, without oxygen, a plant cannot undertake the respiration required for various metabolic processes. Therefore, all mechanisms driven by O_2 will be affected if oxygen is absent or reduced in amount. Under submergence, cellular oxygen is depleted. This directly affects particular 2ODD enzymes which are able to oxidise an organic substrate by using a dioxygen molecule. The hypothesis that 2ODD genes may have an influence on the regulation of tolerance response to submergence by acting as O_2 sensing was established. As reported by Mustroph et al. (2010) aerial tissue of fourteen-day-old seedlings from M202 and M202 (Sub1) were submerged for 24 h and subjected to RNA extraction and hybridization on Affymetrix microarrays. Three genes including *Os03g0803500*, *Os07g0194500* and *Os09g0489200* (Figure 5-21) showing dramatic change/difference in submerged or non-submerged samples of both varieties were selected for subsequent study.

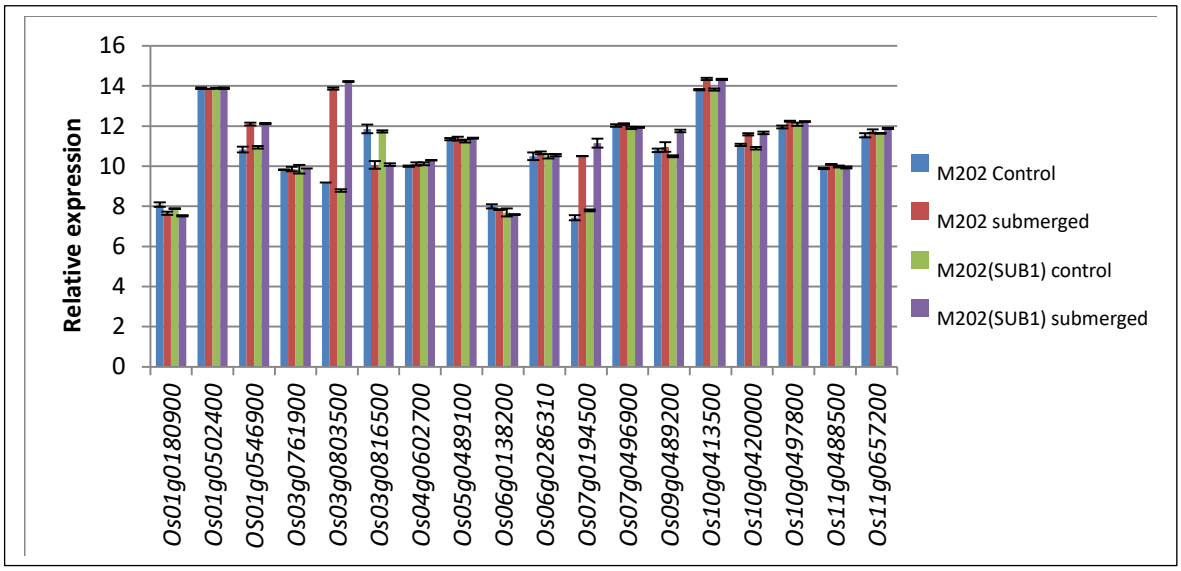
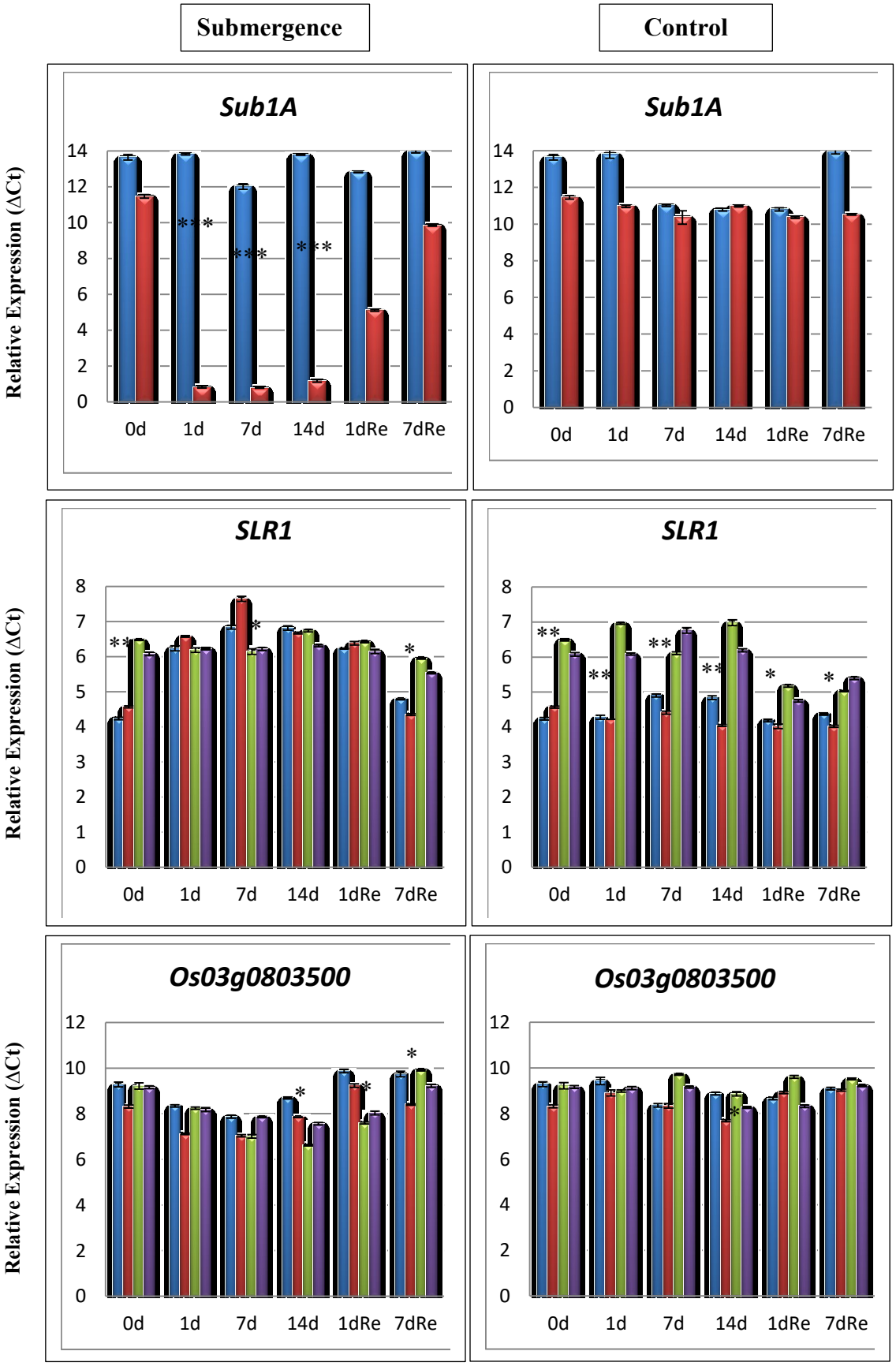


Figure 5-21 Micro array data of 2 OG (II) dioxygenase genes expression from aerial tissue of submergence tolerant M202 (Sub1) and intolerant M202 lines after submergence for 24 h (Mustroph et al, 2010).

Here, the alteration of 2ODD genes expressions under submergence in tolerant and intolerant varieties was investigated. Leaf samples of four varieties (IR64, IR64.S1, Gleva and Nipponbare) were collected at the specified times mentioned above. Non-submerged leaves were used as control. Three independent biological replicates were analysed for each treatment/varieties. The result shows that the expression of three of the 2ODD genes in the four varieties did not show any significant difference in either submergence or non-submergence. The expression of these genes was considerably lower compared to *Sub1A* and very low even compared to *SLR1*. Only *Os03g0803500* showed slight difference between varieties (Figure 5-22). It was found that this gene was downregulated when submerged up to 7 d in IR64.S1 and up to 14 d in Gleva. However, its transcript level was significantly low. This finding indicates that these three selected genes may not be implicated in the regulation of flood tolerance response and they are not likely to be involved in O_2 sensing during flooding.


Submergence
Control
Sub1A
SLR1
Os03g0803500
Relative Expression (ΔCt)
0d
1d
7d
14d
1dRe
7dRe

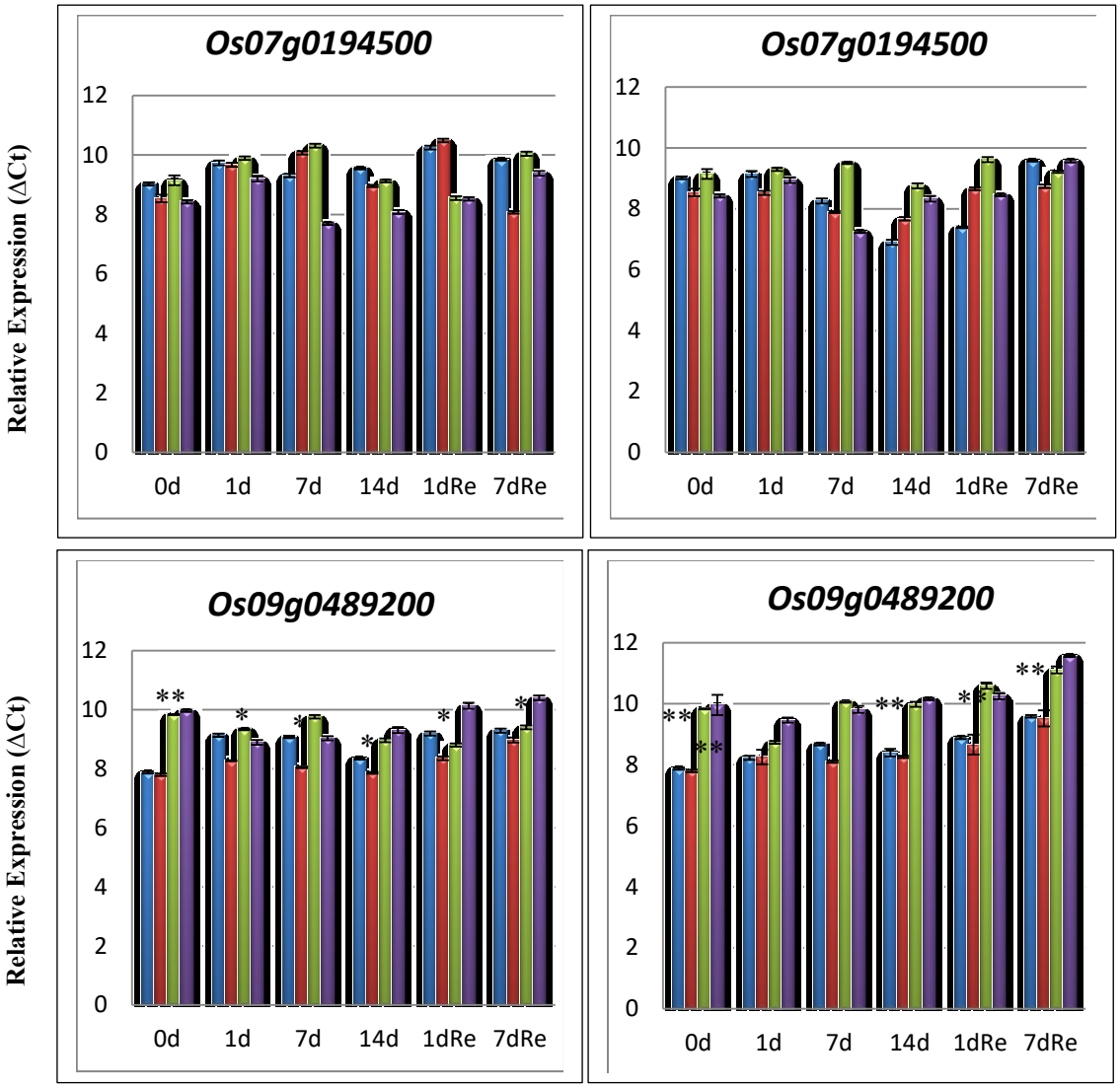


Figure 5-22 The expression analysis of *Sub1A*, *SLR1* and three selected 2ODD genes in four verities. Submerged and non-submerged leaves were collected at six specified time points and RNA samples were subjected to quantitative RT-PCR analysis. Relative levels of individual transcripts were calculated by normalisation with respect to the *eEF-1α* gene. Values are the mean ± standard error of three biological replicates. Data was analysed using an ANOVA with Student-Newman-Keuls test (* is P<0.05, ** is P<0.01, *** is P<0.001)

Recently, only Fukao et al. (2011) revealed that the presence of *Sub1A* improves the survival of tolerant rice during drought stress. Under submergence, there are many genes involved in the regulation of flood tolerance response. *SLR1* is also one of the genes that was triggered by Sub1A protein and consequently depresses GA inducible genes leading to a limitation of

shoot length and it also mediates the interaction between GA and ABA. Under dehydration, the main key hormone that regulates drought stress is ABA. In drought tolerant varieties, *Sub1A* increases ABA responsiveness leading to suppression of leaf water loss and consequent drought tolerance. Therefore, to evaluate drought tolerance adaptability, *SLR1* expression under dehydration also was examined. Furthermore, drought stress leads to the depletion of O_2 in cells so that 2ODD genes were also examined in order to confirm and evaluate the regulation of *Sub1A* under drought tolerance. This may lead to a discovery of a link between those genes in a regulation of drought tolerance response as well.

As shown in Figure 5-23, under dehydration, almost all genes showed significantly low expression except *SLR1* showed moderate expression. Comparing between with and without *Sub1A* varieties, *Sub1A* transcripts was significantly higher in IR64.S1 than in IR64. However, the level of *Sub1A* transcripts both in IR64.S1 and IR64 under dehydration was lower when compared to submergence. In both IR64 and IR64.S1, *SLR1* was upregulated during exposure to 7 d of dehydration whereas it was downregulated after 7 d of submergence. Regarding the 2ODD genes, all genes did not show any significant difference in IR64 and IR64.S1 except at 14 d of dehydration, when *Os09g0489200* moderately expressed in IR64.S1 and its transcript was also significantly higher than in IR64. Those results suggested that expression of *Sub1A* was not triggered by drought and that *SLR1* may be slightly implicated in the interaction of GA and ABA in the same manner as occurs during submergence treatment. Moreover, *Sub1A* also may not have an interaction with the three genes of 2ODD genes, indicating these dioxygenase genes may not function in O_2 sensing.

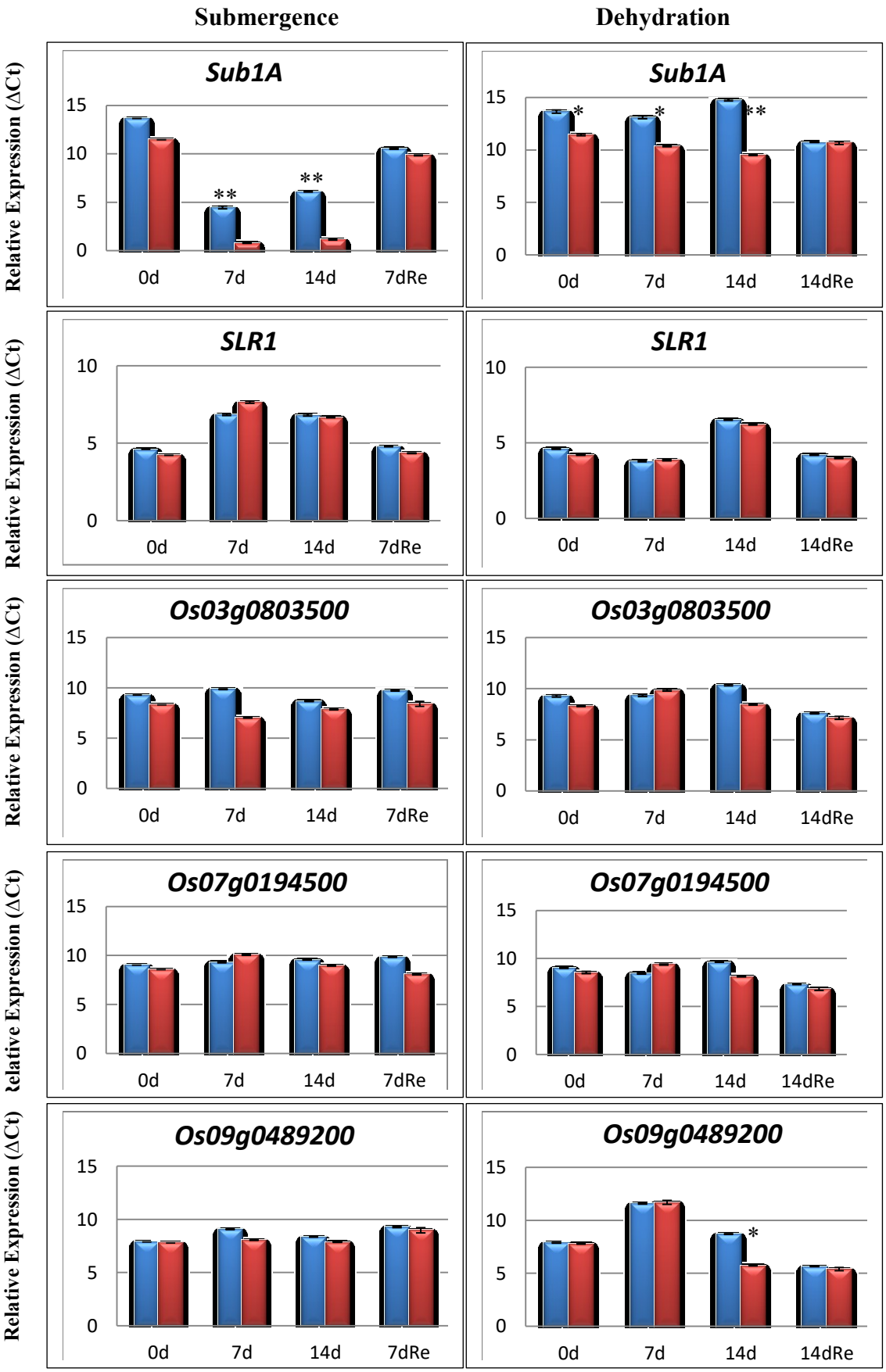

Submergence
Dehydration
Sub1A
SLR1
Os03g0803500
Os07g0194500
Os09g0489200
Relative Expression (ΔCt)
0d
7d
14d
7dRe
14dRe

Figure 5-23 The expression analysis of *Sub1A*, *SLR1* and three selected 2 OG (II) dioxygenase genes in IR64 and IR64.S1 under submergence and dehydration at four specified time points. RNA samples were subjected to quantitative RT-PCR analysis. Relative levels of individual transcripts were calculated by normalisation with respect to the *eEF-1α* gene. Values are the mean ± standard error of three biological replicates. Data was analysed using an ANOVA with Student-Newman-Keuls test (* is P<0.05, ** is P<0.01)

5.5 Summary

The comparative analysis of the phenotypic response of four rice varieties showed that IR64.S1, which contains *Sub1A*, showed limited shoot length when submerged up to 14 d and it was able to recover after water decrease. IR64 is able to elongate its shoot gradually and this elongation is not significantly different to its control at every specified time points. In contrast, Japonica, Gleva and Nipponbare all showed dramatic growth under water. Up to 14 d of submergence, the shoot length increased four-fold compared to the control.

Comparative molecular analysis confirmed that IR64.S1 possesses the *Sub1A-1* allele and IR64 possesses the *Sub1A-2* allele, whereas Gleva and Nipponbare did not have a *Sub1A* gene.

Moreover, the expression profiling of *Sub1A* and some genes in tolerant and intolerant rice under submergence and dehydration were also examined in order to confirm and to evaluate the regulation of *Sub1A*. The results showed that *Sub1A* of IR64.S1 is highly expressed under submergence but was not expressed under drought. Moreover, the expression of *SLR1* and three of 2OOD genes including *Os03g0803500*, *Os07g0194500* and *Os09g0489200* did not show any significant difference when plants were either submerged or dehydrated. In addition, the transcript levels of these genes were also not significantly different in either IR64 and IR64.S1 and even in *Japonicas* which were used as the susceptible check. This

indicates that *SLR1*, *Os03g0803500*, *Os07g0194500* and *Os09g0489200* may not have an interaction with *Sub1A* under abiotic stress in these varieties.

Chapter 6 General discussion

Plants are able to adapt to environmental change for survival by using various strategy depending on the factors affecting on plants and the response of plants to these factors. Biological and physical factors are counted to have an influence on plant mechanism regulation.

Epigenetic regulations play crucial roles in plant development and adaptation to environmental stress and it can be triggered by biological, chemical and physical factors. Recently, there has been evidence revealing 2ODD family particularly TET family proteins play a crucial role in DNA demethylation which is a type of epigenetic regulations in mammalian and other metazoan systems. These TETs biologically generate various oxidative products of 5-mC including 5-hmC, 5-caC and 5-fC as found in certain cancer cells. It has been challenging that these oxidised derivatives of 5-mC also were found in DNA of rice and Arabidopsis leaves but the particular protein responsible for the oxidation of 5-mC leading to its oxidative products in plants is still elusive.

For the first part of this thesis, I primarily investigated the expression pattern of predicted *Tet* gene homologues in rice tissues and I found that rice leaves presented highest expression of these homologues. Initially, it was the intention to investigate the involvement of predicted *Tet* gene homologues in leaf development by examining the relative level of oxidised 5-mC in wild type and mutant rice. To detect the level of these oxidised derivative of 5-mC, I were going to use precise mass spectrometry method. Unfortunately, I got practical problem that I could not obtain the mutant seeds and were unable to access an appropriate mass spectrometry method. Besides, during that time UK was facing with severe flooding destroying a huge number of plants. This situation brought us change our study objectives to mainly focus on physical factors as abiotic stresses particularly, flooding and drought stress. I

hypothesized that 2ODD may be triggered by these abiotic stresses leading to the alteration in epigenetic regulation.

6.1 Investigation of the correlation of predicted *Tet* gene homologues and the level of oxidised derivatives of 5-mC in wild type and mutant rice

Although I could not go further to identify the correlation of 2ODD gene and oxidised derivative of 5-mC in rice development of both mutant and wild type due to practical problems, a study of this topic is still of considerable interest and its results would be used as an additional evidence to support the suggestion that the presence of these oxidised products in plantsmay lead to consequent revelation of the mystery of DNA demethylation controlled by plant enzymes.

Generally, there are two ways to study enzyme activity, all of which are observing the disappearance of substrate and the appearance of product. For measuring the appearance of product is typically more precise because it is easier to measure the small change in product than small change in substrate. To study oxidative demethylation in plants, there are several researchers continuously revealing the possibility of its occurrence by monitoring generated derivative bases of 5-mC. With reference to 5-hmC, there has been a debate about its existence in the plant genome as there are considered to be no enzymes in plants with an equivalent function to that of the TET family enzymes (Jang, et al., 2014). The first report of the presence of 5-hmC in plants was revealed by Terragni, et al., (2012). They examined the genomes of Arabidopsis, soybean, and rice using recombinant β-GT (β-glucosyltransferase), and found very low levels of 5-hmC (<0.07% of the total nucleotides). Subsequent study using only the antibody-based dot blot method revealed the presence of 5-hmC in Arabidopsis (Yao, et al., 2012) but it was no evidence informing the detected amount of this oxidised base. Futhermore, Liu, et al. (2013) applied a liquid chromatography coupled with

tandem mass spectrometry (LC-MS/MS/MS) method. They found the relatively low levels (0.8 modifications per 10^6 nucleosides) of the 5-hydroxymethyl-2′-deoxycytidine (5-hmdC) and subsequently suggested that unlike observations made from mammals, these oxidative products may not be formed from enzyme-mediated oxidation reactions but rather from passive oxidation damage. Erdmann, et al. (2015) attempted to confirm the presence of 5-hmC in Arabidopsis by using various methods including thin-layer chromatography, immunoprecipitation-chip, ELISA, enzymatic radiolabeling, and mass spectrometry. In contrast to two previous research groups, it was concluded that 5-hmC is not present in Arabidopsis. Given the consideration on this conflicting conclusion, I have noticed that the presence or absence of bases oxidation is limited by the precise methodology used to detect those bases. Moreover I also observed that the level of detected bases were very low. These difficulties caused us to reconsider the hypothesis supporting the occurrence of DNA demethylation in plant modulated by enzymatic oxidation forming oxidised derivatives of 5-mC. Based on the conclusion of absence or very low level presence of oxidative products of 5-mC, the enzymatic oxidation of DNA demethylation may impossibly occur in Arabidopsis as concluded by Jang, et al., (2014).

Unlike *Arabidopsis*, consensus reports has promoted the presence of 5-hmC in rice. Ketsuwan (2013) detected 5-hmC in leaf, panicle, seedling and root by using immuno dot blot assay. Next, Wang et al. (2015) detected 5-hmC modification by employing a dot-blot assay and consequently quantified the level of that base in different rice tissues using liquid chromatography-multistage mass spectrometry (LC-MS/MS/MS). The results showed 5-hmC levels largely varied between tissues and thousands of 5-hmC peaks were identified whereas their levels were very low. Although it seems the amount of detected 5-hmC was very low, the presence of this base in rice leads one to question how this base is generated. Is it mediated from oxidative enzyme(s) as in mammals or is it possibly generated from ROS as is

likely in Arabidopsis? This finding in Arabidopsis suggests a prospective study of the initiation of 5-hmC in rice by ROS as well. If the results show ROS tend to be involved with the 5-hmC production, then an attempt to investigate oxidative enzyme mediating DNA methylation would be stopped. However, the detection of low levels of 5-hmC and its derivatives in DNA (and RNA)is still extremely challenging and will motivate us to further investigate its roles in plant development.

Although the initial aims of our hypothesis did not proceed as hoped, the discovery of predicted *Tet* genes homologues led to an investigation of the expression of these genes in various rice tissues. As these genes encode protein which are members of the 2ODD family, these genes might have an involvement in all mechanisms that would be modulated by O_2, particularly the mechanisms regulated in leaves. I found that at least seven of the 18 genes presented higher expression in leaves so that theses seven genes are of interest in any future investigation of their function in plant development.

Before continuing to study deeply about the correlation of Tet genes homologues expression and oxidative bases of 5-mC abundance in wild type and mutant, I had turned to other objectives relating with physical factor as flooding and drought stress which also might associate to the regulation of 2ODD gene. With the hypothesis that the alteration of O_2 level under flooding and drought in rice cells may cause the regulation of 2ODD gene leading to a response to such abiotic stresses, I established this study in order to evaluate the regulation system of rice adaptation to flooding stress and drought.

6.2 Evaluation of abiotic stresses response in rice

To investigate the correlation of other 2ODD genes and *Sub1A* / *SLR1* was conducted by identifying the expression of those gene under flooding and drought stress. In this study, I hypothesised that flooding and drought stresses lead to the limitation of the infusion of gases

into cells resulting in anoxic or hypoxic condition within the cell. Genes encoding proteins responsible for oxidation mechanisms might be influenced by this alteration in O_2 availability, leading to repression or expression of these genes, and eventually resulting in downstream regulation of genes involved in flood or drought tolerance.

Sub1A has been confirmed to confer flood tolerance in rice as well as to improve survival of tolerant rice during drought stress. Furthermore, *SLR1* has also been confirmed to be activated by SUB1A. In order to evaluate the resistance of rice under stresses, I investigated the correlation of three 2ODD genes (selected by considering the expression changing as mentioned in Chapter 5 including *Os03g0803500*, *Os07g0194500* and *Os09g0489200)*, *Sub1A* and *SLR,* by primarily starting with identifying the expression of these genes.

I also found that *Sub1A* was highly expressed under submergence whereas *SLR1* and three of the 2ODD genes did not show any significant difference between submerged and control treatments. Moreover, regarding the recovery state of submerged rice, several reports suggested that after being completely submerged for 14 d, tolerant rice was able to recover and would produce new tillers and leaves whereas intolerant rice was unable to recovery and decayed. In this study, my results were in contrast with previous reports. I found that both tolerant and intolerant rice could recover. In addition, for 2ODD genes, their expression was not significant different between both treated and control of all varieties of rice.

Considering the expression of those genes under drought, besides the objective stated above, I conducted this study because of the need to confirm the role of *Sub1A* to improve survival of tolerant rice during drought stress as first reported by Fukao et al. (2011). Unfortunately, I also could not discover the significant difference of genes expression of those genes under drought as well.

Focusing on the recovery state, it was surprising that even intolerant rice varieties were able to recover despite the fact that they do not possess *Sub1A-1*. This arose because there were many factors affected to be assessed while conducting the experiments. The major factor that I thought may have a considerable effect on our result is the experimental design, in particular the stress condition which was applied. Here, the tank was placed close to fluorescent light, and the level of water (from the top of leaves to water surface) was only 50 cm at the first day of submergence. Moreover, the age of the rice seedlings was also relevant. Here, I used fourteen-day-old seedlings which were at the stage of growing rapidly so that all intolerant rice varieties grew very rapidly at this stage except the tolerant rice with its quiescent strategy. Finally, ethylene initiated under stress condition may also have promoted a GA response; hence intolerant rice without *Sub1A-1* would be promoted by ethylene leading to rapid elongation. With these reasons, the shoots of intolerant rice were able to emerge above the water surface and lay at the surface. Eventually, rice would have enough light and gas exchange leading to survival after water receded.

Considering the possible correlation between *Sub1*A and three selected 2ODD genes, there seems to be no relation between them because no significant differences were found in 2ODD gene expression under both stress and drought, in tolerant rice. This finding does not allow any further conclusions to be drawn about oxygen sensing in this cereal.

6.3 Future work

- To deeply identify the functional roles of 2ODD involving either in epigenetic regulation or O_2 sensing under oxidative stress, further studies could be conducted as follows:-

- Investigation of the expression of 2ODD genes under stress by
    - using samples grown in very complete design flooding or drought condition as similar as the condition occurring in nature such as level of flooding water, the

wavelength of light, turbulence water, etc. and sample facing with real flooding and drought in nature.

- Using additional root sample as well as leaves
- Observing the expression of 18 genes of 2ODD in roots and leaves (present work used only three genes) in both tolerant and intolerant rice grown in previous conditions mentioned above.

➢ Identify the correlation of oxidative base abundance in mutant lines of those 18 genes and wild type grown under stress by using various methods such as reversed-phase HPLC coupled with tandem mass spectrometry method, stable isotope-labeled standards method and liquid chromatography-multistage mass spectrometry (LC-MS/MS/MS) method.

- To identify the contribution of ROS to the formation of oxidised bases by measuring of 5-hmC, 5-fC and 5-caC as well as the reduction of ROS in wild type and mutant line grown under oxidative stress.

References

Aik, W., McDonough, M., Thalhammer, A., Chowdhury, R., & Schofield, C. (2012). Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol, 22*, 691–700.

Akhtar, I., & Nazir, N. (2013). Effect of waterlogging and drought stress in plants. *Inter J Water Resour Environ Sci, 2*(2), 34-40.

Alberts, B., Johnson, A., Lewis, J., Martin Raff, M., Roberts, K., & Walter, P. (2008). *Molecular biology of the cell.* New York: Garland Science.

Aravin, L., & Koonin, E. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, research0007.1–0007.8.

Bailey-Serres, J., & Voesenek, L. (2008). Flooding stress: acclimations and genetic diversity. *Ann Rev Plant Biol, 59*, 313–339.

Balasubramani, A., & Rao, A. (2013). O-glcNAcylation and 5-methylcytosine oxidation:an unexpected association between OGT and TETs. *Molec Cell, 49*, 618-619.

Baubec, T., Pecinka, A., & Rozhon, W. (2009). Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. *Plant J, 57*, 542-554.

Bilitewski, U. (2009). DNA microarrays: an introduction to the technology. *Methods Mol Biol, 509*, 1-14.

Bjørnstad, L., Zoppellaro, G., Tomter, A., Falnes, P., & Andersson, K. (2011). Spectroscopic and magnetic studies of wild-type and mutant forms of the Fe(II)- and 2-oxoglutarate-dependent decarboxylase ALKBH4. *Biochem J , 434*, 391–398.

Blokhina, O., Virolainen, E., & Fagerstedt, K. (2003). Antioxidants, oxidative damage and oxygen deprivation stress: a review. *Ann Bot, 91*, 179-194.

Borsani, O., Zhu, J., Verslues, P., Sunkar, R., & Zhu, J. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell, 123*, 1279-1291.

Boulikas, T. (1993). Nuclear localization signals (NLS). *Crit Rev Eukaryot Gene Exp, 3*(3), 193-227.

Brahimi-Horn, C., Mazure, N., & Pouysségur, J. (2005). Signalling via the hypoxia-inducible factor-1alpha requires multiple posttranslational modifications. *Cell Signal, 17*, 1–9.

Brown, T., & Brown, D. (2016). *Epigenetics.* Southampton: ATDBio.

Bruick, R. (2003). Oxygen sensing in the hypoxic response pathway: Regulation of the hypoxia-inducible transcription factor. *Genes Dev, 17*, 2614–2623.

Burn, J., Bagnall, D., Metzger, J., Dennis, E., & Peacock, W. (1993). DNA methylation, vernalization, and the initiation of flowering. *PNAS, 90*(1), 287–291.

Bustin, S. (2002). Invited review: quantification of mRNA using real-time reverse transcription PCR (RT-PCR):trends and problems. *J Mol Endocrinol, 29*, 23-39.

Bustin, S. (2004). *A-Z of Quantitative PCR.* California: IUL Biotechnology Series, International University Line, La Jolla,.

Bustin, S., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real time PCR experiments. *Am Assoc Clin Chem, 55* (4), 611-622.

Cao, X., & Jacobsen, S. (2002b). Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol, 12*, 1138-1144.

Changchui, H. (2004). International Rice Commission Newsletter. *International Year of Rice 2004* (p. 46). Rome: Food and Agriculture Organization of the United Nations (FAO).

Chen, T., & Li, E. (2004). Structure and function of eukaryotic DNA methyltransferases. *Curr Top Dev Biol, 60*, 55-89.

Cho, J., Ryu, J., Jeong,YM, Park, J., Song, J., & Amasino, R. (2012). Control of seed germination by light-induced histone arginine demethylation activity. *Dev Cell, 22*, 736–748.

Choudhury, S., Panda, P., Sahoo, L., & Panda, S. (2013). Reactive oxygen species signaling in plants under abiotic stress. *Plant Signal Behav, 8*(4), e23681.

Chuaqui, R., Bonner, R., Best, C., Gillespie, J., Flaig, M., Hewitt, S., et al. (2002). Post-analysis follow-up and validation of microarray experiments. *Nat Genet, 32*, 509-514.

Clifton IJ, M., Ehrismann, D., Kershaw, N., Granatino, N., & Schofield, C. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem, 100*, 644–669.

Cubas, P., Vincent, C., & Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature, 401*, 157-161.

Damme, v., Huibers, R., J Elberse, J., & and Ackerveken, G. (2008). Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *The Plant J, 54*, 785–793.

Das, K., Panda, D., Sarkar, R., Reddy, J., & Ismail, A. (2009). Submergence tolerance in relation to variable floodwater conditions in rice. *Environ Exp Botany, 66*, 425–434.

Das, K., Sarkar, R., & Ismail, A. (2005). Elongation ability and non-structural carbohydrate levels in relation to submergence tolerance in rice. *Plant Sci, 168*, 131–136.

Dong, C., Zhang, H., Xu, C., Arrowsmith, C., & Min, J. (2014). Structure and function of dioxygenases in histone demethylation and DNA/RNA demethylation. *IUCrJ , 1*, 540–549.

Dongen, J., & Licausi, F. (2014). *Low-oxygen stress in plants: Oxygen sensing and adaptive responses to hypoxia.* Aachen, German and Pisa, Italy: Springer Science and Business Media.

Dowena, R., Pelizzolaa, M., Schmitza, R., Listera, R., Dowenc, J., Nerya, J., et al. (2012). Widespread dynamic DNA methylation in response to biotic stress. *PNAS, 109*(32), E2183–E2191.

Easwaran, P. H. (2003). *A complex interplay of regulatory domains controls cell cycle dependent subnuclear localization of DNMT1 and is required for the maintenance of epigenetic information.* Berlin: Mathematisch-Naturwissenschaftlichen Fakultät.

Epigenetik. (2016). *Exploring epigenetics.* Farmingdale: Epigentek Group Inc.

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D., & Basra, S. (2009). Plant drought stress: effects mechanisms and management. *Agron Sustainable Dev, 29*(1), 185-212.

Farrow, S., & Facchini, P. (2014). Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front Plant Sci, 5*(524), 1-15.

Federuk, M. (2006, August). *Decodeing Iceland's dna.* British Columbia: The science creative quaterly.

Feil, R., & Fraga, M. (2012). Epigenetics and the environment:emerging patterns and implications. *Nat Rev Genet, 13*, 97-109.

Feng, S., & Jacobsen, S. E. (2011). Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol, 14*(2), 179-186.

Finnegan, E., Genger, R., Peacock, W., & Dennis, E. (1998). DNA methylation in plants. *Annu Rev Plant Biol, 49*(1), 223–247.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res, 41(Database issue)*, D808-815.

Freeman, W., Walker, S., & Vrana, K. (1999). Quantitative RT-PCR: pitfalls and potential. *BioTechniques, 26*, 112-125.

Fukao, T., & Bailey-Serres, J. (2008). Ethylene—A key regulator of submergence responses in rice. *Plant Sci, 175*, 43-51.

Fukao, T., & Bailey-Serres, J. (2008). Submergence tolerance conferred by *Sub1A* is mediated by SLR1 and SLRL1 restriction of gibberellin responses in rice. *PNAS, 105*(43), 16814–16819.

Fukao, T., Bailey-Serres, J., & Yeung, E. (2011). The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *Plant Cell, 23*(1), 412-427.

Fukao, T., Harris, T., & Bailey-Serres, J. (2009). Evolutionary analysis of the *Sub1* gene cluster that confers submergence tolerance to domesticated rice. *Ann Bot, 103*, 143–150.

Fukao, T., Xu, K., Ronald, P., & Bailey-Serres, J. (2006). A variable cluster of ethylene responsive-like factors regulates metabolic and developmental acclimation responses to submergence in rice. *Plant Cell, 18*, 2021–2034.

Furne, I. J., & Matzke, M. (2011). Methylation and demethylation of the *Arabidopsis* genome. *Curr Opin Plant Biol, 14*(2), 137–141.

Gibbs, D., Lee, S., MdIsa, N., Gramuglia, S., Fukao, T., Basse, G., et al. (2011). Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. *Nature, 479*, 415-419.

GIEWS. (2012). *Global Information and Early Warning System on Food and Agriculture.* Rome: Global Watch, FAO.

GIEWS. (2015). *Global Information and Early Warning System on Food and Agriculture.* Rome: Global Watch, FAO.

Gillam, P. (2015, October 13). *IGCSE biology posts.* Retrieved January 4, 2016, from PMG biology: https://pmgbiology.com/tag/chromosome/

Goff, S., Ricke, D., Lan, T., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica). *Science, 296*(5565), 92-100.

Goggins, B., Chaney, C., Graham L. Radford-Smith, G., Horvat, J., & Keely, S. (2013). Hypoxia and integrin-mediated epithelial restitution during mucosal inflammation. *Front. Immunol., 4*(272), 1-9.

GRiSP. (2013). *Rice almanac* (4th ed.). Los Baños (Philippines): International Rice Research Institute.

Guimarães, E. (2009). *Rice breeding.* New York: Springer.

Gutzat, R., & Scheid, O. (2012). Epigenetic responses to stress: triple defense. *Curr Opin Plant Biol, 15*(5), 568–573.

Haddad, J. (2004). Oxygen sensing and oxidant/redox-related pathways. *Biochem Biophys Res, 316*, 969–977.

Hajkova, P. (2003). *Epigenetic reprogramming in mouse germ cells.* Berlin-Dahlem: Max-Planck Institut für Molekulare Genetik.

Harr, B., & Schlotterer, C. (2006). Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons. *Nucleic Acids Res, 34*, e8.

Hartmann, D. L., Klein Tank, A. M., & Rusticucci, M. (2013). *Observations:Atmosphere and Surface.* Cambridge, United Kingdom and New York, USA.: Cambridge University Press.

Hashimoto, H., Liu, Y., Upadhyay, A., Chang, Y., Howerton, S., Vertino, P., et al. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydreoxymethylation. *Nucleic Acid Res, 40*(11), 4841-4849.

Hattori, Y., Nagai, K., Furukawa, S., & al, e. (2009). The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature, 460*, 1026–1030.

Hemelaers, V. (2012). *Oxygen sensing mechanisms in plants.* Biology . Utrecht: Utrecht University.

Herrera, C., & Bazaga, P. (2010). Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. *New Phytol, 187*, 867-876.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator V3: A reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*, Article ID 420747, 5 pages.

Hu, L., Li, N., Zhong, S., Yang, J., Zhou, T., Wu, Y., et al. (2014). Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *PNAS, 21*, 10642–10647.

Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., et al. (2007). F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol, 143*, 1467–1483.

Jang, H., Shin, H., Eichman, B., Huh, J., & et, a. (2014). Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem Biophysl Res Commun, 446* , 1067–1072.

Jin, C., You, G., He, Y., Tang, C., Wu, P., & and Zheng, S. (2007). Iron deficiency-induced secretion of phenolics facilitates the reutilization of root apoplastic iron in red clover. *Plant Physiol , 144*, 278–285.

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J., Tyagi, A., et al. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics, 9*, 451.

Kaundal, R., & Raghava, G. (2009). RSLpred: an integrative system for predicting subcellular localization of rice proteins combining compositional and evolutionary information. *Proteomics, 9*(9), 2324 - 2342.

Kawahara, Y., de la Bastide, M., Hamilton, J., Kanamor, i. H., McCombie, W., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice, 6*, 4.

Kellinger, M. W., Song, C. X., & J, C. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol, 19*(8), 831-833.

Kershaw, N., Caines, M., Sleeman, M., & Schofield, C. (2005). The enzymology of clavam and carbapenem biosynthesis. *Chem Commun (Camb), 14*(34), 4251–4263.

Ketsuwan, K. (2013). *Molecular aspect of gene expression in rice.* Reading: University of Reading.

Kevil, C., Walsh, L., Laroux, F., Kalogeris, T., Grisham, M., & Alexander, J. (1997). An improved, rapid northern protocol. *Biochem Biophys Res Commun, 238*, 277–279.

Koh, K. P., Yabuuuchi, A., Rao, S., Huang, Y., Cunnif, K., Nardone, J., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineages specification in mouse embryonic stem cells. *Cell Stem Cell, 8*, 200–213.

Koonin, E., & Galperin, M. (2003). *Sequence - Evolution - Function: Computational approaches in comparative genomics.* Boston: Kluwer Academic.

Korlach, J., & Turner, S. W. (2012). Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol, 22*(3), 251-261.

Kriukien, E., Liutkeviciute, Z., & Klimasauskas, S. (2012). 5-Hydroxymethylcytosine – the elusive epigenetic mark in mammalian DNA. *Chem Soc Rev, 41*, 6916-6930.

Laborte, A., Nelson, A., Jagadish, K., Aunario, J., Sparks, A., Ye, C., et al. (2012). *Rice feels the heat.* Los Baños: International Rice Research Institute.

Lan, P., Li, W., Wen, T., Shiau, J., Wu, Y., Lin, W., et al. (2011). iTRAQ protein profile analysis of Arabidopsis roots reveals new aspects critical for iron homeostasis. *Plant Physiol, 155*, 821–834.

Lanceras, J., Pantuwan, G., Jongdee, B., & Toojinda, T. (2004). Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiol, 135*, 384–399.

Lauria, M., & Rossi, V. (2011). Epigenetic control of gene regulation in plant. *Biochim Biophys Acta, 1809*(8), 369-378.

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet, 11*, 204–220.

Licausi, F., & Perata, P. (2009). Low oxygen signaling and tolerance in plants. *Adv Bot Res, 50*, 139-198.

Licausi, F., Kosmacz, M., Weits, D., Giuntoli, B., & Giorgi, F. (2011). Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. *Nature, 479*, 419-422.

Lira-Medeiros, C., Parisod, C., Fernandes, R., Mata, C., Cardoso, M., & Ferreira, P. (2010). Epigenetic variation in mangrove plants occurring in contrasting natural environment. *PLoS ONE, 5*(4), e10326.

Liu, F., Xu, W., We, Q., Zhang, Z., Xing, Z., Tan, L., et al. (2010). Gene expression profiles deciphering rice phenotypic variation between Nipponbare (*Japonica*) and 93-11 (*Indica*) during oxidative stress. *PLoS ONE, 5*(1), e8632.

Liu, S., Dunwell, T. L., Pfeifer, G. P., Dunwell, J. M., Ullah, I., & al, e. (2013). Detection of oxidation products of 5-methyl-2'-deoxycytidine in *Arabidopsis* DNA. *PLoS ONE, 8*(12), e84620.

Loewe, L. (2008). Genetic mutation. *Nat Educat, 1*(1), 113.

Lorsbach, R., Moore, J., Mathew, S., Raimondi, S., Mukatira, S., & Downing, J. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *17*(3), 637-641.

Mackill, D., Amante, M., Vergara, B., & Sarkarung, S. (1993). Improved semidwarf rice lines with tolerance to submergence of seedlings. *Crop Sci, 33*, 749–753.

Maclean, J., Dawe, D., Hardy, B., & Hettel, G. (2002). *Rice Almanac.* (3rd, Ed.) OXON: CABI.

Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A., King, G., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet*, *38*(8), 948-952.

Martin, A., Troadec, C., Boualem, A., Rajab, M., Fernandez, R., Morin, H., et al. (2009). A transposon-induced epigenetic change leads to sex determination in melon. *Nature, 461*, 1135-1138.

Martinez, N., & Walhout, A. (2009). The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *BioEssays, 31*, 435–445.

McDonough, M., Loenarz, C., Chowdhury, R., Clifton, I., & Schofield, C. (2010). Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struc Biol, 20*, 659–672.

McKeown, S. (2014). Defining normoxia, physoxia and hypoxia in tumours—implications for treatment response. *Br J Radiol, 87*(1035), 20130676.

McLean, C., Gluckmanb, P., & Sheppard, A. (2012). Phenotypic diversity and epigenomic variation – The utility of mass spectrometric analysis of DNA methylation. *J Proteomics, 75*, 3400-3409.

Meza, T., Moen, M., Vågbø, C., Krokan, H., Klungland, A., Grini, E., et al. (2012). The DNA dioxygenase ALKBH2 protects *Arabidopsis thaliana* against methylation damage. *Nucleic Acids Res , 40*, 6620–6631.

Mohanthy, S. (2013, Jan 29). Rice facts: Trends in global rice consumption. *Rice Today, 12*(1), pp. 44-45.

Mohanthy, S. (2014, May 14). Rice facts: The global rice market winks at El Niño and Thai problems. *Rice today, 13*(3), pp. 38-39.

Mohr, F., Dohnerb, K., Buskea, C., & Rawata, V. P. (2011). *TET* genes: new players in DNA demethylation and important determinants for stemness. *Exp Hematol, 39*, 272–281.

Moldenhauer, K., Charles, E., Wilson, J., Counce, P., & Hardke, J. (2013). Rice growth and development. In J. Hardke, *Arkansas rice production handbook* (pp. 9-20). Little Rock: University of Arkansas Division of Agriculture Cooperative Extension Service.

Mommer, L., & Visser, E. (2005). Underwater photosynthesis in flooded terrestrial plants: a matter of leaf plasticity. *Ann Bot, 96*, 581–589.

Morey, J., Ryan, J., & Van Dolah, F. (2006). Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biol Proced Online, 8*(1), 175-193.

Murrey, K. K. (1996). DNA sequencing by mass spectrometry. *J Mass Spectrom, 31*(11), 1203-1215.

Mustroph, A., Lee, S., Oosumi, T., Zanetti, M., & et, a. (2010). Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. *Plant Physiol, 152*(3), 1484-500.

Nagai, K., Hattori, Y., & Ashikari, M. (2010). Stunt or elongate? Two opposite strategies by which rice adapts to floods. *J Plant Res, 123*, 303–309.

Nakajima, H., & Kunimoto, H. (2014). TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci, 105*, 1093–1099.

Niroula, R., Pucciariello, C., The Ho, V., Giacomo Novi, G., Fukao, T., & Perata, P. (2012). SUB1A-dependent and -independent mechanisms are involved in the flooding tolerance of wild rice species. *Plant J, 72*, 282–293.

Nishiuchi, S., Yamauchi, T., Takahashi, H., Kotula, L., & Nakazono, M. (2012). Mechanisms for coping with submergence and waterlogging in rice. *Rice, 5*(2), 1-14.

O'Conno, C., & Adams, J. (2010). *Essentials of cell biology.* Cambridge, MA: NPG Education.

Papdi, C., Perez-Salamo, I., Joseph, M., Giuntoli, B., Bogre, L., Koncz, C., et al. (2015). The low oxygen, oxidative and osmotic stress responses synergistically act through the ethylene response factor VII genes *RAP2.12*, *RAP2.2* and *RAP2.3*. *Plant J, 82*(5), 772–784.

Patel, R., Nahal, H., Breit, R., & Provart, N. (2012). BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. *Plant J, 71*, 1038–1050.

Paun, O., Bateman, R., Fay, M., Hedrén, M., Civeyrel, L., & Chase, M. (2010). Stable epigenetic effects impact adaptation in allopolyploid orchids (*Dactylorhiza*: Orchidaceae). *Mol Biol Evol, 27*, 2465-2473.

Pavet, V., Quintero, C., Cecchini, N., Rosa, A., & Alvarez, M. (2006). *Arabidopsis* displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by Pseudomonas syringae. *Mol Plant-Microbe Interact, 19*, 577–587.

Perata, P., & Voesenek, L. (2009). Submergence tolerance in rice requires Sub1A, an ethylene-response-factor-like gene. *Trends Plant Sci, 12*, 43–46.

Perez, C., Martınez-Calle, N., Martın-Subero, J., Segura, V., Delabesse, E., FernandezMercado, M., et al. (2012). *TET2* mutations are associated with specific 5-methylcytosine and 5-hydroxymethylcytosine profiles in patients with chronic myelomonocytic leukemia. *PLoS ONE, 7*(2), e31605.

Promega. (2016). *Epigenetics.* Madison: Promega Corporation.

Pucciariello, C., & Perata, P. (2012). How plants sense low oxygen. *Plant Signal Behav, 7*(7), 813–816.

Ram, P., Singh, B., Singh, A., Singh, P., Singh, H., Boamfa, I., et al. (2002). Submergence tolerance in rainfed lowland rice: physiological basis and prospects for cultivar improvement through marker-aided breeding. *Field Crops Res, 76*, 131–152.

Razin, A., & H, C. (1977). Distribution of 5-methylcytosine in chromatin. *PNAS, 74*, 2725-2728.

Redfern, S., Azzu, N., & Binamira, J. (2012). *Rice in Southeast Asia: facing risks and vulnerabilities to respond to climate change.* Rome: Food and Agriculture Organization of the United Nations (FAO).

Robert-Seilaniantz, A., Grant, M., & Jones, J. (2011). Hormone crosstalk in plant disease and defense: more than just jasmonate salicylate antagonism. *Annu Rev Phytopathol, 49*, 317-343.

Robertson, A. B., Dahl, J. A., Ougland, R., & Klunglang, A. (2012). Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. *Nat Protoc, 7*(2), 340-350.

Ruiz, T. M., Galisteo, A. P., Marin, M. I., Macias, M. I., Ariza, R. R., & Arjona, T. R. (2006). *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *PNAS, 103*(18), 6853–6858.

Russell, N. (2014). *UN awards big data project for confronting crazy weather in rice production.* Cali: International Center for Tropical Agriculture (CIAT).

Sasidharan, S., & Mustroph, A. (2011). Plant oxygen sensing is mediated by the N-end rule pathway: A milestone in plant anaerobiosis. *The Plant Cell, 23*(12), 4173–4183.

Schatz, M., Maron, L., Stein, J., Wences, A., Gurtowski, J., Biggers, E., et al. (2014). Whole genome de novo assemblies of three divergent strains of rice, Oryza sativa, document novel gene space of aus and indica. *Genome Biol, 15*(11), 506.

Schmitz, A., Folsom, J., Yusuke Jikamaru, Y., Ronald, P., & Walia, H. (2013). SUB1A-mediated submergence tolerance response in rice involves differential regulation of the brassinosteroid pathway. *New Phytol, 198*(4), 1060–1070.

Scourzic, L., Mouly, E., & Bernard, O. (2015). TET proteins and the control of cytosine demethylation in cancer. *Genome Med, 7*(1), 9.

Septiningsih, E., Pamplona, A., Sanchez, D., & Neeraja, C. (2009). Development of submergence-tolerant rice cultivars: the *Sub1* locus and beyond. *Ann Bot, 103*(2), 151–160.

Shah, A. (2015). *Climate Change and Global Warming Introduction.* The UK: Global Issues.

Sharma, R., Mohan, R., Singh, M., Garima Malik, G., Deveshwar, P., & Tyagi, A. (2009). Rice cytosine DNA methyltransferases – gene expression profiling during reproductive development and abiotic stress. *FEBS Journal, 276*(21), 6301–6311.

Silver, I., & Erecińska, M. (1998). Oxygen and ion concentrations in normoxic and hypoxic brain cells. *Adv Exp Med Biol, 454*, 7-16.

Singh, N., Dang, T., Vergara, G., Pandey, D., Sanchez, D., Neeraja, C., et al. (2010). Molecular marker survey and expression analyses of the rice submergence-tolerance gene *SUB1A*. *Theor Appl Genet, 121*(8), 1441–1453.

Singh, S., Mackill, D., & Ismai, A. (2009). Responses of *SUB1* rice introgression lines to submergence in the field:Yield and grain quality. *Field Crop Res, 113*(1), 12–23.

Slingo, J., Belcher, S., Scaife, A., McCarthy, M., Saulter, A., & McBeath, K. (2014). *The Recent Storms and Floods in the UK.* Devon, the UK: Met Office.

Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N., & Sano, H. (2002). Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J Biol Chem, 277*(40), 37741–37746.

Su, L., Wu, L., & Tseng, A. (2011). Comparison of mRNA Gene Expression Analysis between Quantitative-real time PCR Assay and DNA Microarray Based Assay. *2011 International Conference on Agricultural and Natural Resources Engineering. 3-5*, pp. 158-167. Singapore: Advances in Biomedical Engineering.

Taft, R., Pang, K., Mercer, T., Dinger, M., & Mattick, J. (2010). Non-coding RNAs: regulators of disease. *J Pathol, 220*(2), 126–139.

Tammen, S., Friso, S., & Choi, S. (2012). Epigenetics: the link between nature and nurture. *Mol Aspects Med, 34*(4), 753-764.

Tan, L., & Shi, Y. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development, 139*(11), 1895-1902.

Tomilin, N. (2008). Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *BioEssays , 30*(4), 338–348.

Tripathi, K., Warrier, R., Govila, O., & Ahuja, V. (2011). *Series of crop specific biology documents.* New Delhi: Department of Biotechnology of Ministry of Science & Technology and Ministry of Environment and Forests.

Twining, S. (2014). *Impact of 2014 Winter Floods on Agriculture in England.* Wolverhampton, UK: ADAS UK Ltd.

Valencia-Sanchez, M., Jidong Liu, J., Gregory J. Hannon, G., & Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev, 20*(5), 515-524.

Vanyushin, B. F., & Ashapkin, V. V. (2011). DNA methylation in higher plants: Past, present and future. *Biochim Biophys Acta, 1809*(8), 360–368.

Vasellati, V., Oesterheld, M., Medan, D., & Loreti, J. (2011). Effects of flooding and drought on the anatomy of Paspalum dilatatum. *Ann Bot, 88*(3), 355-360.

Vella, P. S. (2013). Tet proteins connect thre O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Molec Cell, 49*(4), 645-656.

Venuprasad, R., Dalid, C., Del Valle, M., Zhao, D., Espiritu, M., StaCruz, M., et al. (2009). Identification and characterization of large-effect quantitative trait loci for grain yield under lowland drought stress in rice using bulk-segregant analysis. *Theor Appl Genet, 120*(1), 177–190.

Verhoeven, K., Jansen, J., van Dijk, P., & Biere, A. (2010). Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol, 185*(4), 1108–1118.

Vilsack, T., & Harden, K. (2015). *Beyond drought.* USA: United state department of Agriculture.

Voesenek, L., & Blom, C. (1999). Stimulated shoot elongation: a mechanism of semiaquatic plants to avoid submergence stress. In H. Lerner, *Plant responses to environmental stresses: from phytohormones to genome reorganization* (pp. 431–448). New York, USA: Marcel Dekker.

Voesenek, L., & Sasidharan, R. (2013). Ethylene — and oxygen signalling – drive plant survival during flooding. *Plant Biol (Stuttg), 15*(3), 426-435.

Voesenek, L., Rijnders, J., Peeters, A., & van de Steeg, H. (2004). Plant hormones regulate fast shoot elongation under water: from genes to communities. *Ecology, 85*(1), 16–27.

Wada, Y., Miyamoto, K., Kusano, T., & Sano, H. (2004). Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. *Mol Genet Genomics, 271*(6), 658-666.

Wang, W., Zhao, X., Pan, Y., Zhu, L., Fu, B., & Zhikang, L. (2011). DNA methylation changes detected by methylation-sensitive amplified. *J Genet Genomics, 38*(9), 419-424.

Wang, X., Song, S., Wu, Y., Li, Y., Chen, T., Huang, Z., et al. (2015). Genome-wide mapping of 5-hydroxymethylcytosine in three rice cultivars reveals its preferential localization in transcriptionally silent transposable element genes. *J Exp Bot, 66*(21), 6651-6663.

Welford, R., Kirkpatrick, W., McNeill, L., Pur, M., Oldham, N., & Schofield, C. (2005). Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate

dependent oxygenases from bacteria, plants and humans. *FEBS Letters, 579*(23), 5170–5174.

Winkel, A., Colmer, T., Ismail, A., & Pedersen, O. (2013). Internal aeration of paddy field rice (Oryza sativa) during complete submergence – importance of light and floodwater O2. *New Phytol, 197*(4), 1193–1203.

Woon, E. C., Demetriades, M., Bagg, E. A., Aik, W., Krylova, S. M., Ma, J. H., et al. (2012). Dynamic combinatorial mass spectrometry leads to inhibitors of a 2-oxoglutarate-dependent nucleic acid demethylase. *Medic Chem*, *55*(5), 2173-2184.

Xia, K., Wang, R., Ou, x., Fang, Z., Tian, C., Duan, J., et al. (2012). OsTIR1 and OsAFB2 downregulation via OsmiR393 overexpression leads to more tillers, early flowering and less tolerance to salt and drought in rice. *PLoS One, 7*(1), e30039.

Xiong, H., Yang, J., & Li, Y. (2012). Identification of submergence-responsive genes in two Indica rice genotypes carrying *SUB1A-1* but exhibiting differential tolerance. *J Plant Biol, 55*(3), 233-241.

Xu, K., Xu, X., Fukao2, T., Maghirang-Rodriguez, R., Heuer, S., & al, e. (2006). Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature, 442*, 705-708.

Xu, K., Xu, X., Ronald, P., & Mackil, D. (2000). A high-resolution linkage map of the vicinity of the rice submergence tolerance locus *Sub1*. *Mol Gen Genet, 263*(4), 681-689.

Yang, Y., Dudoit, S., Luu, P., Lin, D., Peng, V., Ngai, J., et al. (2002). Normalization for cDNA microarray data:a robust composite method addressing single and multiple slide systematic variation. *Nucl Acids Res, 30*(4), e15.

Yao, Q., Song, C., He, C., Kumaran, D., & Dunn, J. (2012). Heterologous expression and purification of *Arabidopsis thaliana* VIM1 protein: In vitro evidence for its inability to recognize hydroxymethylcytosine, a rare base in *Arabidopsis* DNA. *Protein Express Purif, 83*(1), 104–111.

Yue, B., Xue, W., Xiong, L., Yu, X., Luo, L., Cui, K., et al. (2006). Genetic basis of drought resistance at reproductive stage in rice: Separation of drought tolerance from drought avoidance. *Genetics, 172*(2), 1213–1228.

Zhang, M., Kimatu, J. N., Xu, K., & Liu, B. (2010). DNA cytosine methylation in plant development. *J Genet Genomics, 37*(1), 1-12.

Zhang, X., & Bruice, T. (2006). The mechanism of M.Hhal DNA C5 cytosine methyltransferase enzyme: A quantum mechanics molecular mechanics approach. *PNAS, 103*(16), 6148-6153.

Zhao, W., Wang, J., He, X., Huang, X., Jiao, Y., Dai, M., et al. (2010). *Oryza sativa Indica Assembly and Gene Annotation.* Shenzhen: Beijing Genomics Institute (BGI).

Zhao, Z., Zhang, Y., Liu, X., Zhang, X., Liu, S., Yu, X., et al. (2013). A role for a dioxygenase in auxin metabolism and reproductive development in rice. *Dev Cell, 27*(1), 113-122.

Zhu, J. (2009). Active DNA demethylation mediated by DNA glycosylases. *Ann Rev Genet, 43*, 143–166.