GENOME-WIDE ASSOCIATION AND EXPRESSION STUDIES FOR IDENTIFICATION OF QTLS AND CANDIDATE GENES UNDERLYING ABIOTIC STRESS TOLERANCE DURING GERMINATION STAGE OF RICE

A Dissertation

by

RANJITA THAPA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Endang M. Septiningsih
Co-Chair of Committee,	Michael J. Thomson
Committee Members,	Rodante Tabien
	Scott Finlayson
Head of Department,	David Baltensperger

August 2019

Major Subject: Plant Breeding

Copyright 2019 Ranjita Thapa

ABSTRACT

Rice is one of the most important cereal crops feeding more than half of the world's population. Due to extreme climatic condition, different abiotic stresses like hypoxia stress and chilling stress have been the biggest threat to rice production. Direct sowing method is the most preferred way of planting in Asian countries and in U.S. due to the lower cost of planting and less labor requirement. The major challenge associated with direct sowing is flash flooding that can happen immediately after sowing due to unpredicted rainfall. In this study, we evaluated more than 250 rice accessions belonging to different groups of rice for various traits related to chilling tolerance and hypoxia tolerance during germination stage. Compressed Mixed Linear Model (CMLM) of GAPIT was used to conduct GWAS analysis for the identification of QTLs. From the GWAS study conducted for chilling stress tolerance, we identified 41 QTLs associated with different chilling indices like low temperature germinability, germination index, coleoptile growth under cold condition, plumule length at 4 d recovery, and plumule growth rate after cold germination. Out of 41 QTLs identified in the whole panel, 14 QTLs were potentially colocalized with known genes/QTLs and 27 QTLs were found to be novel.

From the GWAS analysis of hypoxia stress tolerance traits, there were 24 significant SNPs identified to be associated with different traits measured under hypoxia stress. Out of the 24 significant SNPs discovered in the whole panel, 11 QTLs were found to be potentially colocalized with previously identified candidate genes underlying flooding tolerance mechanism in rice. From the phenotypic evaluation of the whole panel for chilling stress tolerance and hypoxia stress tolerance, two lines with contrasting phenotypes under each stress condition were selected and used for global gene expression analysis. The results of these transcriptomics studies have

provided new insights of underlying biological processes, molecular functions and cellular components related to the phenotypic differences of the contrasting lines. The findings of our study will help in identification of promising candidate genes underlying hypoxia stress and chilling stress tolerance and would eventually assist rice breeding program to develop improved tolerant rice cultivars.

DEDICATION

To my mom and dad for their unlimited love, faith and support throughout my academic journey.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Endang Septiningsih, and my committee members, Dr. Michael Thomson, Dr. Scott Finlayson and Dr. Rodante Tabien, for their guidance and support throughout the course of this research. In addition, I would like to extend my sincerest gratitude to Dr. Michael Dickens, Dr. Shichen Wang and Dr. Noushin Ghaffari for their guidance on analysis of transcriptomics data.

I would also like to thank my labmates from Rice Genetics lab for their support during planting and harvesting of the crop in the field. Thanks also go to my friends, colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my parents and sisters for providing the love and guidance to pursue the most worthwhile things in my life.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Endang Septiningsih (chair), Dr. Michael Thomson (co-chair), Dr. Rodante Tabien, and Dr. Scott Finlayson of the Department of Soil and Crop Sciences. Dr. Septiningsih is also a faculty member of the Interdisciplinary Graduate Program in Genetics, while Dr. Thomson and Dr. Finlayson are also faculty members of the Molecular and Environmental Plant Sciences Interdisciplinary Graduate Program.

The RNA sequencing was conducted by the Texas A&M AgriLife Genomics and Bioinformatics Services. Dr. Endang Septiningsih designed the experiment and supervised the research and data analysis. All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was supported by a graduate research assistanship from Texas A&M University and a thesis research funding from Texas A&M AgriLife Research.

NOMENCLATURE

DEG	Differentially Expressed Gene
FC	Fold-Change
FDR	False Discovery Rate
GO	Gene Ontology
GWAS	Genome-wide Association Study
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MLM	Mixed Linear Model
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphism
TASSEL	Trait Analysis by Association Evolutionary Linkage
GAPIT	Genome Association and Prediction Integrated Tool

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Rice is a major crop grown in areas from 44°N latitude in North Korea to 35°S latitude in Australia. It is cultivated from 0 ft to 2700 ft above sea level. Rice is grown in at least 114 countries, and the annual production is around 0.1 million tonnes and above in more than 50 countries (FAO, 2010). Rice is the second most widely grown cereal crop and a staple food for more than half of the world's population, providing two-thirds of caloric intake for more than three billion people in Asia and one-third of caloric intake for 1.5 billion people in Africa and Latin America (Khan, Dar, et al., 2015). Rice fields cover around 155 million hectares more than any other crops, thus showing the importance of rice production ecosystems.

Rice belongs to the family Poaceae and is a semiaquatic and annual self-pollinating grass. There are two major cultivated rice species, *Oryza sativa* (grown worldwide) and *Oryza glaberrima* (grown in Africa). There are two major subpopulations within the *O. sativa* species: *O. sativa indica* and *O. sativa japonica*. The rice plant is believed to have originated in Southern India and then spread northwards to China (Khan, Dar, et al., 2015). It then reached Korea, the Philippines, Japan, and Indonesia by about 1000 B.C.(Khan, Dar, et al., 2015)

Rice is the single most important source of calories for humans. Among the cereal crops, rice is grown mainly for human consumption (Awika, 2011). Rice contributes about 21% of global per capita caloric intake and 27.1% of per capita caloric intake in developing countries (Awika, 2011). Rice production and consumption is highly localized. Asia makes up approximately 92% of global rice production (Awika, 2011), with China and India accounting for 50% of global rice

production in 2010. In Cambodia, 90% of agricultural land is under rice cultivation. In Vietnam, Cambodia and Myanmar, up to 80% of caloric intake comes from rice (Awika, 2011).

Rice planting is typically done following either the direct sowing method (DSR) or the traditional transplanting system (TPR). Direct seeding of rice refers to the process of establishing a rice crop from seeds sown in the field rather than by transplanting seedlings from the nursery. On the other hand, in the traditional transplanting system, rice seedlings are grown in the nursery and later transplanted in the main field. The major drawbacks of the transplanting system are huge water inputs, labor costs, and labor requirements. In recent years, there has been a shift from TPR to DSR in many countries of Southeast Asia (Pandey and Velasco, 1999). TPR has high labor requirements due to the need for uprooting nursery seedlings, puddling fields, and transplanting seedlings into fields. The adoption of the direct-seeded method for lowland rice significantly decreases the labor requirements and costs of rice production (Flinn and Mandac, 1986).

The dominant limiting factor for crop production is abiotic stress like excess or deficient water, high or low temperature, and high salinity (Boyer, 1982). Abiotic stresses induce diverse physiological and molecular responses in plants. These include changes in gene expression and metabolism, osmotic adjustment, induction of repair systems and chaperons, and the expression of late embryogenesis abundant (LEA) proteins (Ingram and Bartels, 1996). Therefore, improvement in abiotic stress tolerance is required to increase the crop yield in an extreme environmental condition.

Rice possesses atypical characteristics compared to other cereal crops. Rice is a tropical C3 grass that has evolved in a semi-aquatic, low-radiation habitat. In comparison with other cereal crops, rice exhibits peculiar characteristics of tolerance and susceptibility to abiotic stresses. Rice can thrive in waterlogged soil conditions and can tolerate submergence to an extent that would kill

other crops. Rice is moderately tolerant of salinity and soil acidity but highly sensitive to drought and cold. Even though these responses to stresses are generally superior to other crops, due to extreme weather conditions, many rice-growing regions demand cultivars with greater tolerance.

Unpredictable flood events are significant problems in the lowlands of South and Southeast Asia. Unpredictable floods after planting can cause poor seed germination and poor plant stand, ultimately leading to yield losses. Planting of improved cultivars having high flooding tolerance is considered to be a valuable and sustainable solution under such conditions. Direct seeding is emerging as a popular technology in both rainfed and irrigated rice growing areas, including flood prone areas. In such flood prone areas, heavy rainfall immediately after sowing causes waterlogging or even flooding especially with fields that are poorly drained or poorly levelled. As a consequence, the seeds drown and are unable to germinate resulting in poor crop establishment and low yield. Developing improved cultivars that can tolerate flooding during germination can provide some assurance of yield to the farmers. Moreover, this trait is also highly desirable in an irrigated rice system where direct seeding is increasingly being practiced because of shortage of labor and increasing costs of production. Also, in areas where direct seeding is practiced, shallow flooding right after sowing helps to suppress weed infestation. After screening more than 8000 accessions from the gene bank at IRRI, several genotypes tolerant of AG conditions have been identified (Angaji, Septiningsih, et al., 2010, Ismail, Ella, et al., 2008). Several studies have used these genotypes for QTL mapping. Several promising QTLs were derived and mapped from Khao Hlan On, a tolerant donor from Myanmar (Angaji, Septiningsih, et al., 2010). The largest QTL detected on chromosome 9 (*qAG-9-2* or *AG1*) has been fine mapped, and the gene underlying the QTL has been cloned (Septiningsih, Pamplona, et al., 2008). With the progress in molecular

breeding, a set of promising breeding lines have been developed using available best donors such as Khao Hlan On, Khaiyan and Ma-Zhan Red (Mackill, Ismail, et al., 2010).

However, there are various external conditions that could affect the expression of tolerance for flooding during germination, such as the condition of seeds, the seedbed, and flood water (Jagadish, Septiningsih, et al., 2012). Recent studies conducted by Jagadish et al. 2012 have shown that the use of fresh seeds and seeds stored under cooler conditions is important for survival. The improvement is because of the lower lipid peroxidation associated with higher activities of some of the enzymes involved in reactive oxygen scavenging, such as superoxide dismutase (SOD) and catalase (CAT) (Ella, Dionisio-Sese, et al., 2010). The optimum temperature of flood water ranges from 22°C to 28°C for most genotypes (Jagadish, Septiningsih, et al., 2012). The seed survivability decreases with algal growth in flood water and with increasing flood depth. Seed pre-soaking and priming increase the rate of seed germination.

Rice is often grown in high altitude mountain areas in countries like Nepal, China, India, Indonesia, and the Philippines. Rice grown over 1000 meters above sea level (masl) is generally affected by chilling injury, which results in spikelet sterility (Sthapit and Shrestha, 1991). Rice is extremely sensitive to chilling stress during germination stage (< 15°C) and at booting to heading stage (<18°C) (Jagadish, Septiningsih, et al., 2012). Cold stress during booting stage can lead to 100% spikelet sterility. However, cold tolerant cultivars like Chomrong showed only 30% spikelet sterility even at a mean minimum temperature of 15C between booting and heading in Lumle (1740 masl) of Kaski district in Nepal (Shrestha, Asch, et al., 2011), indicating that genetic diversity for this trait could be utilized to develop cold tolerant cultivars.

Researchers from all over the world have made significant efforts in understanding the mechanisms of responses to abiotic stresses in rice. A greater understanding of the physiology and

molecular biology of stress tolerance may provide a useful platform to improve stress tolerance varieties.

Rice is also a model species of monocot plants for genomic research. Tremendous advances have been made in the field of rice genetics and genomics after the success of genome sequencing of rice. These major achievements include the genetic mapping of a large number of genes/QTLs, controlling many important traits and the cloning and characterization of large number of rice genes (Jiang, Cai, et al., 2012). Having such a vast amount of genomic and genetic information, the main challenges lie in the development of efficient and effective breeding strategies that make use of the available information.

References

Angaji, S.A., E.M. Septiningsih, D. Mackill and A.M. Ismail. 2010. QTLs associated with tolerance of flooding during germination in rice (Oryza sativa L.). Euphytica 172: 159-168.

Awika, J.M. 2011. Major cereal grains production and use around the world. Advances in cereal science: implications to food processing and health promotion 1089.

Boyer, J.S. 1982. Plant productivity and environment. Science 218: 443-448.

Ella, E.S., M.L. Dionisio-Sese and A.M. Ismail. 2010. Proper management improves seedling survival and growth during early flooding in contrasting rice genotypes. Crop science 50: 1997-2008.

Flinn, J. and A. Mandac. 1986. Wet seeding of rice in less favored rainfed environments working paper. Agricultural Economics Department, International Rice Research Institute, Los Banos, Philippines.

Ingram, J. and D. Bartels. 1996. The molecular basis of dehydration tolerance in plants. Annual review of plant biology 47: 377-403.

Ismail, A.M., E.S. Ella, G.V. Vergara and D.J. Mackill. 2008. Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (Oryza sativa). Annals of Botany 103: 197-209.

Jagadish, S., E. Septiningsih, A. Kohli, M. Thomson, C. Ye, E. Redona, et al. 2012. Genetic advances in adapting rice to a rapidly changing climate. Journal of Agronomy and Crop Science 198: 360-373.

Jiang, Y., Z. Cai, W. Xie, T. Long, H. Yu and Q. Zhang. 2012. Rice functional genomics research: progress and implications for crop genetic improvement. Biotechnology advances 30: 1059-1070.

Khan, M.H., Z.A. Dar and S.A. Dar. 2015. Breeding strategies for improving rice yield—a review. Agricultural Sciences 6: 467.

Mackill, D.J., A.M. Ismail, A.M. Pamplona, D.L. Sanchez, J.J. Carandang and E.M. Septiningsih. 2010. Stress tolerant rice varieties for adaptation to a changing climate. Crop Environment & Bioinformatics 7: 250-259.

Pandey, S. and L. Velasco. 1999. Economics of alternative rice establishment methods in Asia: a strategic analysis. Social Sciences Division Discussion Paper, International Rice Research Institute, Los Banos, Philippines 1: 12-18.

Septiningsih, E.M., A.M. Pamplona, D.L. Sanchez, C.N. Neeraja, G.V. Vergara, S. Heuer, et al. 2008. Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Annals of Botany 103: 151-160.

Shrestha, S., F. Asch, M. Dingkuhn and M. Becker. 2011. Cropping calendar options for rice–wheat production systems at high-altitudes. Field Crops Research 121: 158-167.

Sthapit, B. and K. Shrestha. 1991. Breeding for cold tolerance at reproductive phase in the high hills of Nepal. International Rice Research Newsletter 16: 14.

CHAPTER II

GENOME-WIDE ASSOCIATION STUDY FOR TRAITS RELATED TO TOLERANCE TO FLOODING DURING GERMINATION

Synopsis

Rice is a unique cereal crop that possess the ability to germinate under submerged and hypoxic or anoxic stress conditions. The success of rice germination and survival under submerged condition is mostly determined by the rapid growth of coleoptile to timely escape from the water surface. Previous researches have shown the presence of genetic variability in the levels of tolerance to flooding during germination among rice accessions. Although multiple studies have been done focusing on the physiological metabolism of oxygen stress, not much researches have been done to explore a wealth of natural variations in coleoptile growth and related traits under stress conditions and only a limited number of work have been reported to unravel the candidate genes underlying various flooding tolerance mechanisms in rice. In this study, we screened the coleoptile lengths of a diverse rice panel consisting of 269 accessions under normal and flooded condition. Genome-wide association study was performed using a 7K SNP chip array and the phenotypic data of normal coleoptile length (NCL), flooded coleoptile length (FCL) and flooding tolerance index (FTI) of all accessions and the japonica subspecies only. We identified 6, 10 and 9 significant SNPs for FCL, NCL and FTI, respectively, in all accessions; while in a separate analysis of the japonica group, we detected 1, 2 and 26 significant SNPs for the same traits, respectively. Likewise in indica group, 10. 3 and 1 SNPs were found to be significantly associated with FCL, NCL and FTI respectively. Five loci, FTI1, FTI2, FTI3, FTI4 and FTI6 associated with FTI in all accessions analysis was identified to be associated with FTI of indica species too. Out

of the 24 significant SNPs discovered in ALL accessions, 11 QTLs were found to be colocalized with previously identified candidate genes underlying flooding tolerance mechanism in rice. From this study we identified 13, 15 and 11 novel QTLs from all accessions, indica group and japonica group respectively. The findings of our study may assist rice improvement program to develop rice cultivars with high tolerance under flooding stress during germination and very early seedling stage.

Introduction

Rice is cultivated by transplanting or direct-seeded systems. Farmers are moving towards direct seeded sowing in both rainfed and irrigated systems due to low labor requirement and more convenience [7]. In the United States, however, direct-seeded rice is the mainstay for rice production. The major hurdle associated with direct sowing is poor seed germination and poor seedling growth which ultimately affect the crop establishment. Unexpected heavy rainfall is one of the major climate change challenges that affects crop cultivation. Heavy rainfall often results in flooding in uneven agricultural land when there is no proper drainage system. The rainfed rice production areas of South Asia, Southeast Asia and tropical Africa are more prone to flooding due to excess rainfall. In some areas of South and South-east Asia, farmers suffer from crop loss due to periodic flooding events during monsoon season (Septiningsih, Pamplona, et al., 2008). With the escalating problems of climate change, farmers all over the world, including in the U.S. face similar problems.

Rice can survive submergence for days. The broad genetic diversity of rice landraces and traditional varieties has enabled its cultivation in very different agro-climatic zones and water regimes (Singh, Septiningsih, et al., 2017). Submergence caused by flooding creates hypoxia environment and restricts the germinating seed aerobic respiration leading to anaerobic respiration

which ultimately leads to low ATP production. In such a scenario, the ability to germinate and rapidly establish the seedling is a very essential trait. Therefore, breeding rice cultivars capable of surviving under prolonged flooded conditions during germination and very early seedling growth will help increase the success of direct-seeded cultivations (Pandey, 2002).

Unlike other cereal, rice has a very unique ability to germinate not only under submerged condition but also under anoxic conditions (Magneschi and Perata, 2008). The elongation of coleoptile at germination stage and highly developed aerenchyma in mature plants allow rice tissues to transport oxygen under water. The faster the coleoptile elongates, the sooner the seedling is able to avoid flooding stress which increases the chance of survival. During flooding and oxygen-deprived condition, various metabolic changes take place. Different enzymes associated with starch degradation, glycolysis and ethanol fermentation such as α-amylase, phosphofructokinase, fructose-6-phosphate-1-1-phosphotransferase, alcohol dehydrogenase and pyruvate dehydrogenase are highly active under submerged conditions (Gibbs, Morrell, et al., 2000, Kato-Noguchi and Morokuma, 2007, Lasanthi-Kudahettige, Magneschi, et al., 2007, Magneschi, Kudahettige, et al., 2009). However, not all rice accessions possess the ability to tolerate the stress under flooding during germination (Miro and Ismail, 2013).

One of the tolerance mechanisms in rice to germinate under submerged conditions is by the ability of rice seeds to rapidly degrade the starch reserves under hypoxia or anoxia conditions leading to rapid growth of coleoptile to timely escape from the stress conditions (Magneschi and Perata, 2008). The hydrolysis of starch to maltose for the production of glucose and ultimately sucrose is activated by low oxygen condition and sugar starvation (Lee, Chen, et al., 2009). The protein representing the hub of the mechanism is the *calcineurin b-like interacting protein kinase* (*CIPK15*) (Lee, Chen, et al., 2009, Nghi, Tondelli, et al., 2019). *CIPK15* activates the energy

sensor *sucrose non-fermenting 1-related protein kinase 1* (*SnRK1*), which regulates *myeloblastosis sugar response complex 1* (*MYSB1*) transcription factor (Nghi, Tondelli, et al., 2019). *MYSB1* migrates to the nucleus and binds to sugar-responsive elements and the promoters of *α-amylase 3* (Lu, Ho, et al., 2002, Lu, Lin, et al., 2007) whose transcription is promoted in the embryo and its activity is exerted in the starchy endosperm. The *α-amylase 3D* (*RAMY3D*) is the predominant *α-amylase* isoform expressed in anoxic rice (Lasanthi-Kudahettige, Magneschi, et al., 2007). An additional regulator of this pathway was recently identified by isolating the genetic determinant for a QTL of germination under flooding, *qAG-9-2*, derived from a tolerant donor Khao Hlan On (Kretzschmar, Pelayo, et al., 2015), which is an aromatic japonica from Myanmar (Angaji, Septiningsih, et al., 2010). The qAG-9-2 QTL has an indel of 20.9 Kb which is responsible for rice germinating vigor under anaerobic condition (Kretzschmar, Pelayo, et al., 2015). Development of near-isogenic line containing qAG-9-2 in IR-64 background which is sensitive to anaerobic germination (Miro and Ismail, 2013), highlighted that *trehalose-6-phosphate (T6P) phosphatase 7* (*TPP7*) is responsible for germination under anaerobic or submerged conditions. *TPP7* likely acts upstream of *SnRK1* and protects it from inactivating action of T6P through its conversion to trehalose (Kretzschmar, Pelayo, et al., 2015). T6P also acts as an indicator of sucrose availability via a homeostatic parameter in sugar signaling which regulates the source to sink resource allocation (Yu, Lo, et al., 2015).

Genetic studies on rice germination under oxygen deficiency focus mainly on identification of underlying genes and their functional mechanisms. QTL mapping and GWAS are two widely used tools to discover genetic control of complex traits. Recently the use of SNPs markers has been increasingly popular in both type of studies. The major difference between QTL mapping and GWAS study is the type of mapping population used in the study. QTL mapping is based on

the use of biparental population whereas GWAS uses diverse germplasm and identifies QTL based on the historic recombination via the presence of linkage disequilibrium between SNPs and QTLs. With the first report of unravelling 14 agronomic traits in rice landraces, GWAS has become a popular strategy to dissect complex traits in rice (Huang, Sang, et al., 2010). Compared with the linkage mapping approach, association mapping possesses the ability to uncover a large number of superior allele variations in broad natural population in a simple and rapid pattern.

A number of QTL studies on anaerobic germination based on bi-parental mapping populations have been reported and numerous QTLs have been detected (Angaji, Septiningsih, et al., 2010, Baltazar, Ignacio, et al., 2014, Baltazar, Ignacio, et al., 2019, Septiningsih, Ignacio, et al., 2013). Angaji et al. (2010) identified five QTLs on chromosome 1, 3, 5, 7 and 9 in BC_2F_2 lines derived from a cross between a tolerant line, Khao Hlan On and a susceptible line, IR64. Septiningsih et al. (2013) discovered six significant QTLs on chromosome 2, 5, 6, and 7 using an $F_{2:3}$ population derived from IR42, a susceptible parent, and Ma-Zhan Red, a tolerant parent from China. Jiang et al. (2004; 2006) conducted QTL analysis in 81 recombinant inbred lines (RILs) derived from the Kinmaze/DV85 cross and 148 F_2 progenies of USSR5/N22 cross grown under anoxic conditions and measured shoot length and identified seven QTLs on chromosome 1, 2, 5, 7 and 11. Recently, Hsu and Tung (2015) conducted a GWAS in a panel of indica, japonica and aus varieties and identified a total of 88 single-nucleotide polymorphisms (SNPs) associated with anaerobic response index (difference in coleoptile length between control and those submerged in water). In the same study, a biparental recombinant inbred lines population derived from a cross between Nipponbare and IR64 highlighted a QTL on the long arm of chromosome 1 which explained about 27% of the phenotypic variation (Hsu and Tung, 2015)

It appears that rice germplasm from Central Asia and South Asia has great potential for anaerobic or hypoxic germination tolerance and needs to be further explored (Septiningsih, personal comm.). Therefore, the aim of this study is to identify QTLs related to rice germination under hypoxia conditions using an enriched GWAS rice panel consisted of rice accessions mainly from those regions along with accessions from different parts of the world.

Materials and methods

Plant materials and germination experiment

The rice panel consisted of 269 diverse accessions collected from the United State Department of Agriculture Genetic Stocks-*Oryza* (USDA-GSOR) gene bank (USDA-ARS DBNRRC; Stuttgart, AR). Seed multiplication was performed in the Beaumont rice research station in summer, 2016 (March – August). The seeds from single panicle were sown for seed multiplication of all the accessions grown in 2017. After maturity, the seeds were harvested separately from panicles of single plants. The harvested seeds were dried in a heated air dryer at 37^{o}C for 5 days after harvest and then stored in freezer set at 4^{o}C. To maintain the seed source uniformity, we used seeds collected from single plants. Considering seed age effects on anaerobic germination, we only used seeds that were about nine months old. To break the seed dormancy, seeds were incubated at 50^{o}C for 5 days. Germination testing was done by following roll paper method to access the germination percentage of each accession. The detail of the method is found in http://www.knowledgebank.irri.org/step-by-step-production/pre-planting/seed-quality. Dormant seeds were excluded from this experiment. The oxygen concentration of the autoclaved distilled water used in the experiment was measured using Extech dissolved oxygen meter (Extech instruments, Nashua, NH 03063, U.S.A.).

Phenotyping for anaerobic germination

Seeds were surface sterilized in 20% diluted bleach (6-7% NaClO) for 15 min and then thoroughly rinsed with water. The experiment was conducted in three replications with 30 - 40 seeds in each replication. For control samples, 30 - 40 sterilized seeds were placed on water soaked filter paper placed inside the petri-dishes. For submerged samples, the sterilized seeds were submerged in 8 cm deep sterile water placed inside the 250 ml glass beakers. The glass beakers were wrapped in aluminum foil and placed in controlled growth chamber maintained at 30C for germination. The dark condition was provided to mimic the natural dark condition under soil during rice germination stage. After 4 days of germination, images of all germinating seeds were taken using a Pentax camera and the length of each germinating seedling from each replication was measured using ImageJ software. *The flooding tolerance index (FTI) will be calculated by the formula: Flooded coleoptile length (FCL)/normal coleoptile length (NCL).*

Statistical analysis of phenotype variation

Descriptive statistical analysis was performed in R version. The analysis of variation (ANOVA) was carried out to evaluate the effects of genotype (G) and genotype * environment (G*E) using general linear model in R software. Variation and correlation analysis were performed, respectively to evaluate the coefficient of variation and the relationship between NCL and FCL.

Genotyping rice varieties

An array of 7902 SNPs markers that covered the 12 rice chromosomes was developed by Morales et al. (nearly submitted). This SNP array was used to genotype the 269 rice accessions. After discarding the heterozygous markers and those with missing data > 20% and with minor allele frequency (MAF) < 5%, we obtained a set of 5805 SNPs for our association analysis.

Population structure and Kinship estimation

STRUCTURE program (Pritchard, Stephens, et al., 2000) was used to estimate the population structure by implementing the Bayesian model of Markov chain Monte Carlo (MCMC). The number of populations (k) was set to be 1 to 10 and five iterations were performed for each population. Burn-in value and number of replications of MCMC were set at 50,000 and 100,000 respectively. The k value was determined by the data log likelihood [LNP(D)] and an ad hoc statistics delta K, based on the change rate of [LNP(D)] between successive values of k. Structure harvester was used to perform these analysis (Earl, 2012). The coefficient of ancestry (Q) threshold was defined at 70% to consider an individual with its inferred ancestry from one single group and the accessions which were unable to be clearly assigned to only one group were determined as admixtures. To compare our results from STRUCTURE software, we also used Bayesian clustering program fastStructure (Raj, Stephens, et al., 2014) to calculate different levels of K (K = 1-10) where the command choose K.py was used to identify the model complexity that maximized the marginal likelihood (K=8).

Kinship coefficient

The control of spurious association is improved in association study when finer levels of relatedness are taken into account by fitting a marker-based kinship matrix in the models (Yu, Pressoir, et al., 2006). The coefficients of kinship between the pairs of accessions were determined using a set of 5805 SNPs. A matrix based on identity by descent was produced using TASSEL version 5.0 software (Bradbury, Zhang, et al., 2007) and used as a similarity matrix (K matrix) in the GWAS model.

Genome wide association study analysis

We conducted two separate GWAS analysis, i.e. using ALL 269 accessions and only japonica sub-species within the panel. By using population structure (Q matrix) and kinship relatedness data (K matrix), we used Mixed Linear statistical model (MLM) statistical model (Zhang, Ersoz, et al., 2010) to perform association analysis between the phenotypic traits and the SNP data. The MLM model used is:

$\boldsymbol{Y = \beta X + \gamma P + Zu + \varepsilon}$ $\boldsymbol{Y}$, where Y is the vector of phenotype data, X is the vector of genotype data, β represents the SNP effects, P is a vector of the Q matrix resolving population structure, γ is the effect of population structure, u refers to the random effect from kinship, Z is a Kinship matrix, and ε corresponds to random error. The expected p-values versus the observed p-values test statistics for the SNP markers were plotted (Q-Q plot) to assess the control of type I (false positives) errors under the multiple runs. The Q-Q plot helped us to visualize the distribution of the test statistics and to evaluate the inflation factor as well (λ). Based on the approximate significance values where the observed number of p-values exceeded the expected number of Q-Q- plots, a significance threshold of 0.001 was used to determine the marker-trait association. We also emphasized on reporting of significant SNPs observed above FDR corrected p-value < 0.05.

Results

Oxygen concentration and phenotypic variation

The oxygen concentration of the autoclaved distilled water used in the experiment was measured to be 5.2 mg/liter. A total of 269 accessions were screened for coleoptile length under control and submerged germination conditions, and a wide range of variations were observed among the genotypes (Figure 1). Coleoptile length ranged from 0.06 cm to 4.29 cm with a mean

value of 1.56 cm under normal condition (NCL) (Table 1). While under flooded germination condition (FCL), the length of coleoptile ranged from 0 cm to 4.54 cm with a mean value of 1.51 cm. The analysis of variance test (ANOVA) indicated that the effect of genotype and genotype*treatment interaction had significant effects on the length of the coleoptiles (Table 3). The CV value of coleoptile length was 0.52 for FCL and 0.45 for NCL showing abundant amount of variations among the accessions under flooding stress as well as in normal germination conditions. The anaerobic germination index, FTI was calculated as the ratio of coleoptile length under the two germination conditions (FTI = FCL/NCL). The range of FTI was observed to be 0 cm to 8.02 cm with a mean value of 1.15 cm. The CV value of FTI was estimated to be 0.79 reflecting a wide range of variation among accessions. We found positive correlation between FCL and NCL ($R^2 = 0.29$), and negative correlation between NCL and FTI ($R^2 = 0.45$; Table 2).

Table 1. Phenotypic variation among the accessions in the rice panel for traits under hypoxia stress

trait	mean (cm) ± SE	range (cm)	CV
FCL	1.51 ± 0.04	0 ~ 4.54	0.52
NCL	1.56 ± 0.04	0.06 ~ 4.29	0.45
FTI	1.15 ± 0.05	0 ~ 8.02	0.79

Figure 1 Phenotypic variation among all accessions for traits under hypoxia stress

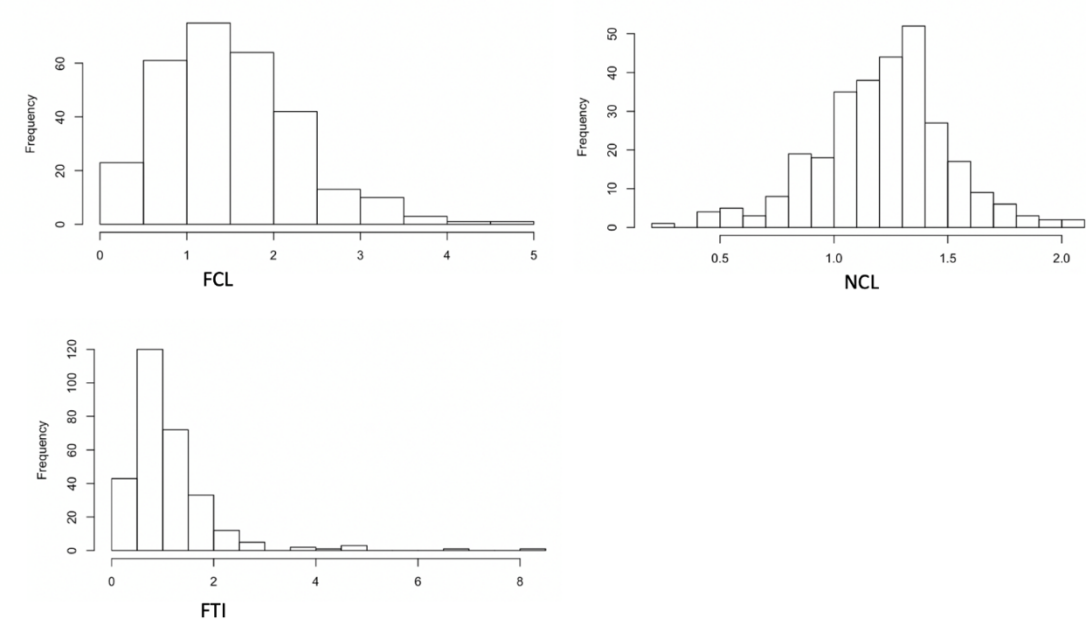


Table 2 Correlation among different traits under hypoxia stress

	AG	control	index
FCL	1	0.29***	0.45***
NCL	0.29***	1	-0.45***
FTI	0.45***	-0.45***	1

Table 3 Treatment effect and treatment-genotype interaction by analysis of variance (ANOVA)

	Df	Sum of squares	Mean sum of squares	F value	Pr(>F)	Significance
genotypes	316	605.5	1.92	16.098	<2e-16 ***	S
treatment	1	0.1	0.14	1.17	0.28	NS
genotypes*treatment	288	266.4	0.93	7.78	<2e-16 ***	S
Residuals	1187	141.3	0.12			

Population Structure

The STRUCTURE V2.3.1 software showed presence of 8 populations in our germplasm collection which sufficiently explained genetic structure among 269 accessions. Using the Bayesian clustering software fastStructure22, we calculated varying levels of K means. The optimal number of subpopulations was predicted to be K=8, based on model complexity and model component analysis as computed by fastStructure22. The number of subpopulations at K=8 clearly defined variation within and between *indica, aus, aromatic, tropical japonica, temperate japonica, glaberrima*, *hybridus* and admixed accessions.

Genome-wide association study of whole panel

GWAS analysis was performed for ALL accessions, indica accessions and japonica accessions separately. Association mapping was performed under compressed MLM by using both population structure matrix and kinship matrix as covariates for ALL accessions. The Q-Q plot showed that compressed MLM model was significantly reducing false positives by accounting for the confounding effects of subpopulation structure and relatedness in the population. An over-correction for the population structure was observed when both subpopulation structure and kinship matrix were considered in the mixed model for indica and japonica populations. To address

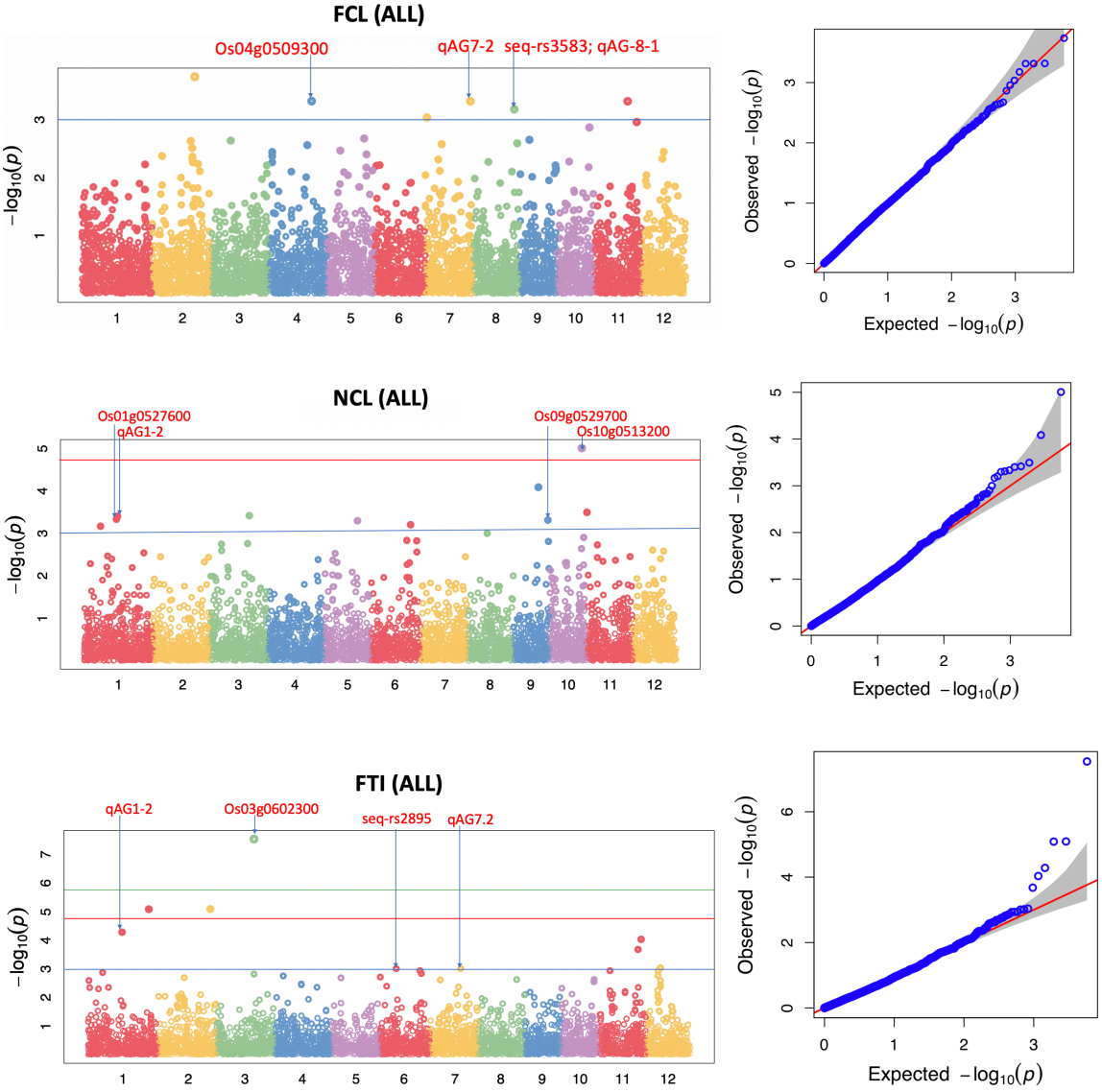
the overcorrection, we used MLM model of GAPIT with 3 PCs in japonica varietal groups and 2 PCs in indica varietal groups.

At a threshold of $p < 0.001$, we discovered 6, 10 and 9 significant SNPs associated with FCL, NCL and FTI respectively (Table 4, Figure 2). Using a more stringent threshold of FDR corrected p-value ($q<0.05$), 1 SNP (NCL1) was found to be associated with FCL, and 3 SNPs associated with FTI (FTI1, FTI2 and FTI3). The comparison of identified significant SNPs among FCL, NCL and FTI showed that one significant SNP (NCL5) was common in between NCL and FTI. This result shows some degree of correlation existed between the coleoptile length under control germination condition and flooded germination condition. Out of 25 significant SNPs, 1 SNP was found to be associated with both NCL and FTI. Among, 24 significant SNPs, all the significant SNPs were found to have occupied only one location (one SNP per locus). Therefore, from the ALL accessions analysis, we identified 24 QTLs associated with FCL, NCL and FTI. The amount of phenotypic variance explained by each SNP was estimated as R^2. The R^2 value of the identified significant SNPs ranged from for 3.78% to 4.26% for FCL, 2.56 to 5.14% for NCL, and 4.0 to 11.49 % for FTI (Table 4).

Table 4 Significant SNPs from GWAS analysis of All accessions

	SNP	Chr.	Position	P.value	R2	FDR P-values	effect
FCL	FCL1	2	26231409	0.0001831	4	0.70325381	-0.0367842
	FCL2	4	25930219	0.0004815	4.05	0.70325381	0.03079102
	FCL3	7	27806332	0.0004846	4.26	0.70325381	-0.0296509
	FCL4	11	20647651	0.0004847	4.2	0.70325381	0.08718391
	FCL5	8	25091497	0.0006659	4.01	0.77298017	0.12078237
	FCL6	7	1066928	0.0009219	3.78	0.89182733	-0.1042113
NCL	NCL1	10	20272810	9.86E-06	5.15	0.05720261	-0.0101271
	NCL2	9	15752720	8.21E-05	4.17	0.23828037	0.0998868
	NCL3	11	430255	0.0003205	3.81	0.36617344	-0.0392573
	NCL4	3	24681102	0.0003842	3.5	0.36617344	-0.064255
	NCL5	1	21106769	0.0003965	3.59	0.36617344	0.11431886
	NCL6	1	20418175	0.0004648	3.39	0.36617344	0.10465117
	NCL7	9	21729544	0.0004876	3	0.36617344	0.05491551
	NCL8	5	21206702	0.0005047	3	0.36617344	-0.0282856
	NCL9	6	24873284	0.0006257	2.74	0.39524737	-0.0173751
	NCL10	1	10501214	0.000681	2.56	0.39524737	-0.1217055
FTI	FTI1	3	23419698	2.90E-08	11.5	0.00016806	0.02061248
	FTI2	2	32323814	8.10E-06	7.57	0.01582775	0.15624803
	FTI3	1	37615383	8.18E-06	8.45	0.01582775	-0.0040043
	NCL5	1	21106769	5.20E-05	6.32	0.07552096	-0.122634
	FTI4	11	26212757	9.24E-05	5.96	0.10728259	-0.0100172
	FTI5	11	24303688	0.0002095	5.27	0.2026635	0.18434248
	FTI6	12	9074224	0.0009198	4.19	0.56358967	-0.1036928
	FTI7	7	18718153	0.0009639	4.01	0.56358967	-0.0268567
	FTI8	6	10199497	0.0009807	4.07	0.56358967	-0.1888272

Figure 2 Manhattan and QQ plots of ALL accessions for traits under hypoxia stress



GWAS analysis for indica sub-species

At a cutoff value of $p < 0.001$ and FDR < 0.05, we identified 29 and 20 significant SNPs associated with different phenotypic traits (Table 5, Figure 3). Among the 29 significant SNPs ($p < 0.001$), 1 2, and 26 GWAS sites were found to be associated with FCL, NCL and FTI, respectively (Table 5). Out of the39 significant SNPs, some of them were found to have shared

genomic region. The significant GWAS sites, InFTI5, InFTI6 and InFTI15 associated FTI were found to have shared site at chromosome 10; other 8 significant SNPs associated with FTI, InFTI1, InFTI3, InFTI8, InFTI9, InFTI14, InFTI17, InFTI19 and InFTI21, were found to be located in the same genomic region in chromosome 12; 4 significant SNPs, FTI6, InFTI4, InFTI10 and InFTI12 associated with FTI were found to be located in same genomic region at chromosome 12, and two significant SNPs, InNCL1 and InNCL12 associated with NCL shared genomic region at chromosome 11. Therefore, in total, we identified 16 distinct QTLs associated with FCL, NCL and FTI in indica species. From GWAS analysis of indica species, we didn't find any SNPs associated with all three measured traits. However, we found 5 significant SNP, FTI1, FTI2, FTI3, FTI4 and FTI6 associated with FTI of ALL accessions analysis appeared in FTI of indica analysis. The phenotypic variance explained by the significant SNPs (R^2) were estimated to be high compared to the phenotypic variance explained by the significant SNPs in GWAS analysis of ALL accessions. The R^2 of identified significant SNPs associated with NCL, FCL and FTI were ranging from 13.42 % to 13.76 %, 12.93 % and 13.14 % to 53.52 %, respectively (Table 5).

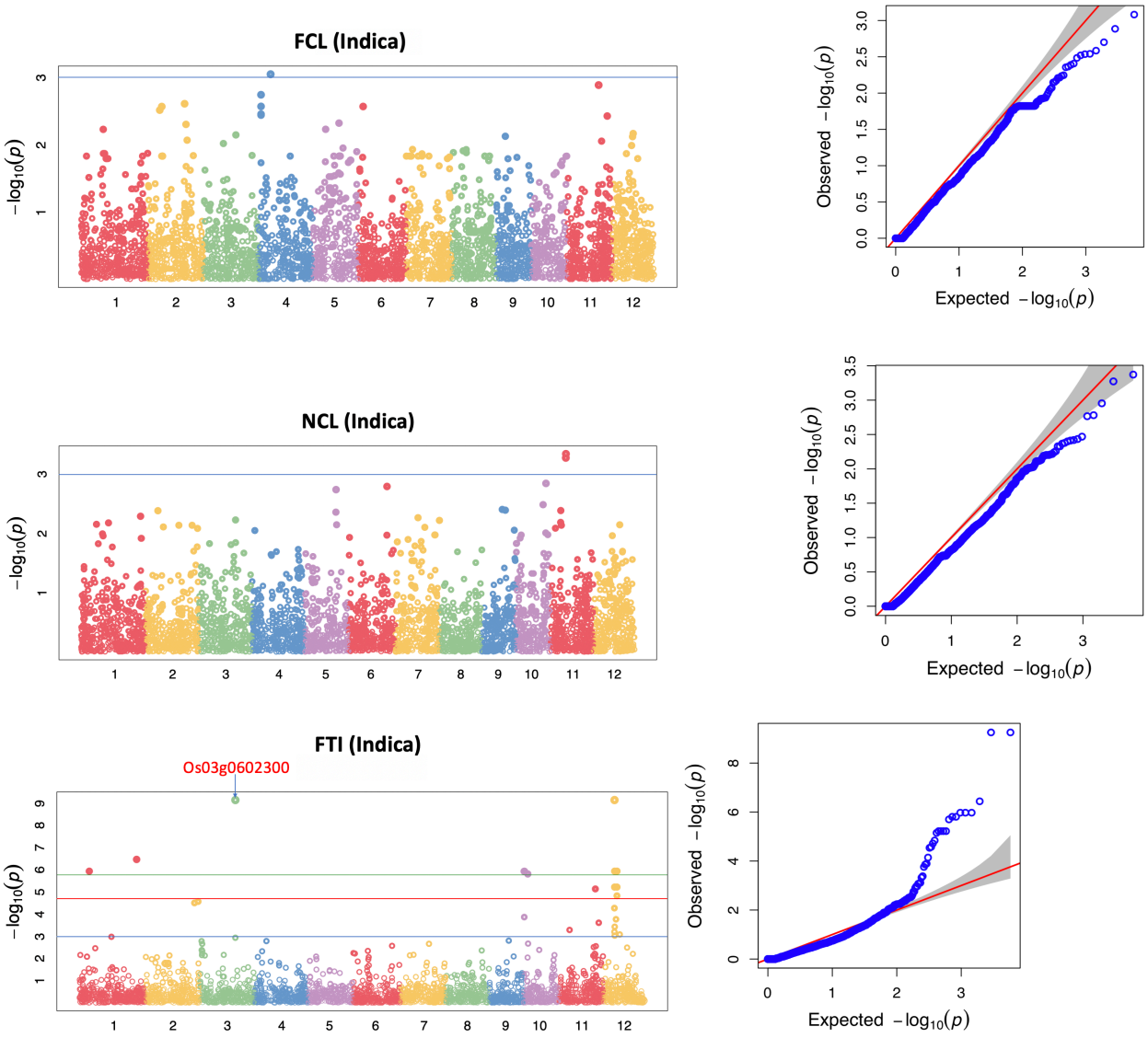
Table 5 Significant SNPs from GWAS analysis of indica subspecies

	SNP	Chr.	Position	P.value	R2	FDR P-values	effect
FCL	InFCL1	4	7915413	0.0008887	12.94	1	-0.3583993
NCL	InNCL1	11	10551942	0.0004547	13.77	1	NA
	InNCL2	11	10479143	0.0005311	13.42	1	0.06385978
FTI	FTI3	1	37615383	3.44E-07	34.08	0.000665	-0.0611721
	InFTI2	1	6606552	1.17E-06	30.56	0.000856	0.19845836
	InFTI13	2	34692243	2.71E-05	21.98	0.009838	0.01587893

Table 5 (continued)

	SNP	Chr.	Position	P.value	R2	FDR P-values	effect
	FTI2	2	32323814	3.14E-05	21.59	0.010718	0.21223339
	FTI1	3	23419698	7.23E-10	53.53	2.10E-06	0.25852879
	InFTI5	10	650031	1.18E-06	30.53	0.000856	NA
	InFTI6	10	407551	1.18E-06	30.53	0.000856	-0.3641423
	InFTI7	10	2792487	1.56E-06	29.75	0.001003	NA
	InFTI15	10	469475	0.0001368	17.83	0.041773	-0.0332702
	InFTI11	11	23993552	7.34E-06	25.46	0.003045	-0.2524297
	FTI4	11	26212757	0.0002462	16.37	0.068054	-0.0217327
	InFTI18	11	7105792	0.0005195	14.55	0.131096	0.23184569
	InFTI1	12	7538167	7.23E-10	53.53	2.10E-06	NA
	InFTI3	12	7663131	1.17E-06	30.56	0.000856	0.05695475
	InFTI4	12	9127499	1.17E-06	30.56	0.000856	NA
	InFTI8	12	7599203	6.07E-06	25.98	0.00271	NA
	InFTI9	12	7599728	6.07E-06	25.98	0.00271	-0.3503684
	InFTI10	12	9181124	6.07E-06	25.98	0.00271	-0.0563619
	FTI6	12	9074224	6.07E-06	25.98	0.00271	-0.0634215
	InFTI12	12	9129374	1.49E-05	23.56	0.005769	NA
	InFTI14	12	7462305	5.38E-05	20.2	0.017352	NA
	InFTI16	12	8506147	0.0001712	17.27	0.049692	NA
	InFTI17	12	7672269	0.0003737	15.35	0.098582	NA
	InFTI19	12	7478430	0.000633	14.07	0.153078	NA
	InFTI20	12	10965554	0.0008296	13.43	0.192609	0.03744419
	InFTI21	12	7667404	0.000934	13.15	0.208507	NA

Figure 3 Manhattan and QQ plots of indica group for traits under hypoxia stress



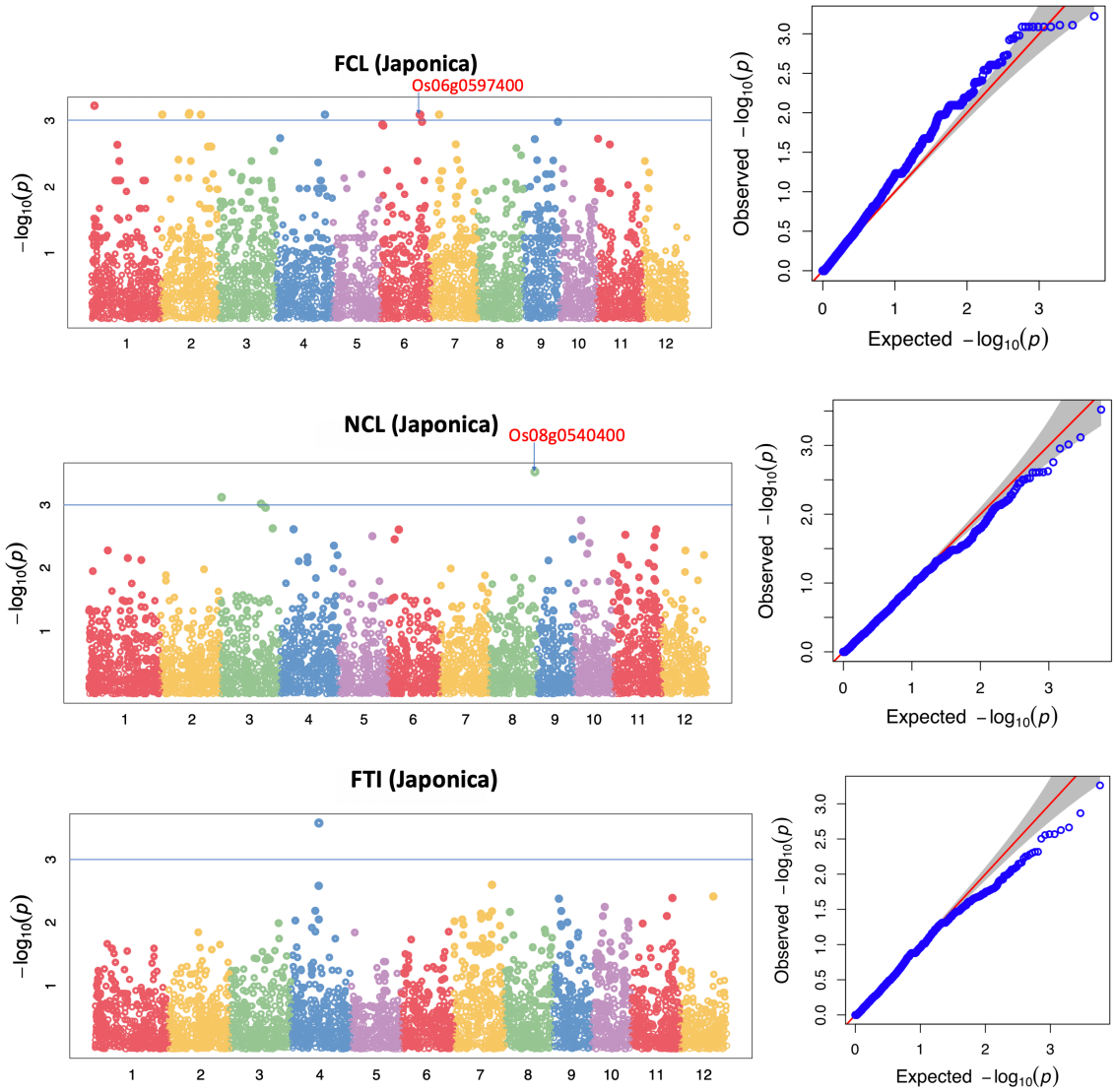
GWAS analysis for japonica sub-species

At a cutoff value of $p < 0.001$, we identified 14 significant SNPs associated with different phenotypic traits (Table 6, Figure 4). Among the 14 significant SNPs, 10, 3, and 1 GWAS sites were found to be associated with FCL, NCL and FTI, respectively (Table 6, Figure 4). Out of the14 significant SNPs, some of them were found to have shared genomic region. The significant GWAS sites, JaFCL2 and JaFCL3 were found to have shared site at chromosome 2. Therefore, in total, we identified 13 distinct QTLs associated with FCL, NCL and FTI in japonica species. The phenotypic variance explained by the significant SNPs (R^2) were estimated to be high compared to the phenotypic variance explained by the significant SNPs in GWAS analysis of ALL accessions. The R^2 of identified significant SNPs associated with NCL, FCL and FTI were ranging from 11.45 % to 13.32 %, 13.96 % to 14.7 % and 16.52 %, respectively (Table 6).

Table 6 Significant SNPs from GWAS analysis of japonica subspecies

	SNP	Chro	Position	P.value	R2	FDR P-value	effect
FCL	JaFCL1	1	2014954	0.0005991	14.74	0.4424231	0.12951506
	JaFCL2	2	17911664	0.0007773	14.09	0.4424231	-0.0583426
	JaFCL3	2	17929072	0.0007773	14.09	0.4424231	-0.0211654
	JaFCL4	2	1000481	0.0008166	13.97	0.4424231	0.15928076
	JaFCL5	2	25103588	0.0008166	13.97	0.4424231	0.08035501
	JaFCL6	4	30174083	0.0008166	13.97	0.4424231	-0.3225388
	JaFCL7	6	24628911	0.0008166	13.97	0.4424231	0.01151134
	JaFCL8	7	5526898	0.0008166	13.97	0.4424231	0.15191592
	JaFCL9	2	17578247	0.0008166	13.97	0.4424231	-0.0202231
	JaFCL10	6	25026144	0.0008166	13.97	0.4424231	0.12714062
NCL	JaNCL1	8	27300242	0.0004381	13.32	1	0.01034174
	JaNCL2	3	942403	0.0006721	12.4	1	-0.0416726
	JaNCL3	3	27346396	0.0010477	11.46	1	-0.0777238
FTI	JaFTI1	4	16285717	0.0002694	16.53	1	0.08415291

Figure 4. Manhattan and QQ plots of japonica group for traits under hypoxia stress



Co-localization of GWAS sites with known Genes/QTLs/SNPs

The significant SNPs discovered in this study were compared with reported QTLs, genes or SNPs previously detected using linkage and association mapping approaches. We found that some of the significant associated loci identified in our study were co-localized with previously identified genes/QTLs identified to be controlling for flooding stress tolerance in rice (Table 7). Among 24 GWAS sites discovered in ALL accessions analysis, 11 GWAS sites were found to be

colocalized with previously identified genes/QTLs/SNPs. Among 16 and 13 significant GWAS sites identified in indica and japonica subpopulation, 2 and 2 GWAS sites were found to be colocalized with previously identified genes/QTLs respectively.

We found 3, 4 and 5 significant SNPs associated with FCL, NCL and FTI to be co-localized with previously identified genes/QTLs in ALL accessions. The significant SNP, FCL2 associated with FCL was found to be located at a distance of 57.58 Kb away from a gene, *sho1* previously discovered to be responsible for shoot apical meristem formation in seedling (Nagasaki, Itoh, et al., 2007). Other significant SNP, FCL3 was found to be located at a distance of 85.25 Kb away from the QTLs, qAG7 and qAG7-2 previously identified to be controlling for flooding tolerance in rice (Angaji, Septiningsih, et al., 2010, Jiang, Liu, et al., 2006). We also discovered that a significant SNP, FCL5 associated with FCL located at a distance of 163.66 Kb away from a SNP, seq-rs3583 previously reported by Zhang et al., 2017 as significant SNP associated with flooding tolerance in rice (Zhang, Lu, et al., 2017).

The significant GWAS site, NCL6 associated with NCL was identified to be located at a distance of 179.05 Kb from a gene, *shl2* controlling shoot apical meristem formation(Nagasaki, Itoh, et al., 2007). The significant SNP, NCL7 associated with NCL was found to be located at 153.06 Kb distance away from *OsVPS22*, a gene controlling seedling lethality (Zhang, Hou, et al., 2013). A significant GWAS site, NCL1 was identified to be located only at a 38 Kb distance away from *Nod26-like intrinsic protein 3-1* which was previously reported to be maintaining shoot growth under boron deficient condition (Hanaoka, Uraguchi, et al., 2014). The significant SNP, NCL5 on chromosome 1 in our study was found to be located closely to a significant SNP (seq-rs3583), associated with coleoptile elongation under flooding stress detected in another GWAS study (Hsu and Tung, 2015). On chromosome 1, the same significant SNP, NCL5 associated with

NCL and FTI, was identified to be located in the genomic interval of *qAG1-2*. This locus was previously detected using both linkage mapping and association mapping approach (Angaji, Septiningsih, et al., 2010, Hsu and Tung, 2015).

The significant GWAS site, NCL6 associated with NCL was identified to be located at a distance of 179.05 Kb from a gene, *shl2* controlling shoot apical meristem formation(Nagasaki, Itoh, et al., 2007). The significant SNP, NCL7 associated with NCL was found to be located at 153.06 Kb distance away from *OsVPS22*, a gene controlling seedling lethality (Zhang, Hou, et al., 2013). A significant GWAS site, NCL1 was identified to be located only at a 38 Kb distance away from *Nod26-like intrinsic protein 3-1* which was previously reported to be maintaining shoot growth under boron deficient condition (Hanaoka, Uraguchi, et al., 2014). The significant SNP, NCL5 on chromosome 1 in our study was found to be located closely to a significant SNP (seq-rs3583), associated with coleoptile elongation under flooding stress detected in another GWAS study (Hsu and Tung, 2015). On chromosome 1, the same significant SNP, NCL5 associated with NCL and FTI, was identified to be located in the genomic interval of *qAG1-2*. This locus was previously detected using both linkage mapping and association mapping approach (Angaji, Septiningsih, et al., 2010, Hsu and Tung, 2015).

The four significant SNPs, associated with FTI found to be colocalized with previously identified QTL/genes were FTI1, FTI3, FTI5, FTI8. One significant SNP, FTI1 associated with FTI of ALL accessions and indica subpopulation was found to be located at a distance of 120.46 Kb away from a gene, *CYP85A1* which was responsible for rice seedling growth where the gene was specifically controlling rice lamina bending and leaf unrolling by promoting castesterone (Asahina, Tamaki, et al., 2014). Another significant SNP, FTI3 associated with associated with FTI of ALL accessions and indica subpopulation, was found to be closely located to a genetic loci,

AQEP001 detected under submergence stress during germination stage in rice by Nghi et al. (2019) (Nghi, Tondelli, et al., 2019). A significant SNP, FTI5 was detected to be at a distance of 19.66 Kb away from a QTL identified to be controlling for submergence tolerance by Toojinda et al., 2003 (Toojinda, Siangliw, et al., 2003). We also detected the significant SNP, FTI8 located at a distance of 2.17 Kb away from a significant SNP, seq-rs2895 previously detected by Zhang et al., 2017 for flooding tolerance in rice (Zhang, Lu, et al., 2017).

The two significant SNPs potentially colocalized with previously identified genes/QTLs in japonica subspecies were JaFCL7 (associated with FCL) and JaNCL1(associated with NCL). The significant SNP, JaFCL7 was found to be located at a distance of 205.04 Kb distance away from a gene, OsPNH1 previously reported to be controlling for maintenance of shoot apical meristem and leaf development in rice (Nishimura, Ito, et al., 2002). The other significant SNP, JaNCL1 was found to be located at a distance of 157.14 Kb away from a gene OsCPK21, previously reported to be responsible for rice seedling growth (Asano, Hakata, et al., 2011).

Table 7 Colocalized genes identified in the GWAS sites

SNPs id	Chr	position	traits associated	colocalized genes/QTLs	function	distance away from the colocalized genes	
FCL2	4	25930219	FCL (ALL)	Os04g0509300	Shoot apical meristem formation.	57.588	Nagasaki et al., 2007
FCL3	7	27806332	FCL (ALL)	qAG7; qAG7-2	Flooding tolerance	85.25	Jiang et al., 2004; Angaji et al., 2010
FCL5	8	25091497	FCL (ALL)	seq-rs3583	Flooding tolerance	163.66	Zhang et al., 2017
NCL1	10	20272810	NCL (ALL)	Os10g0513200	maintaining growth under boron-deficient conditions seedling stage	38.8	Hanaoka et al., 2014
NCL5	1	21106769	NCL (ALL); FTI (ALL)	qAG1-2 ; id1012150	Anaerobic germination tolerance	87.23	Hsu et al., 2015
NCL6	1	20418175	NCL (ALL)	Os01g0527600	Shoot apical meristem formation.	179.05	Nagasaki et al., 2007
NCL7	9	21729544	NCL (ALL)	Os09g0529700	Chalky kernel. Seedling lethality.	153.06	Zhang et al., 2013
FTI1	3	23419698	FTI (ALL)	Os03g0602300	rice lamina bending and leaf unrolling by promoting castasterone(CS)	120.46	Asahina et al., 2014
FTI5	11	24303688	FTI (ALL)	RM206	Submergency tolerance	19.66	Toojinda et al., 2003
FTI8	6	10199497	FTI (ALL)	seq-rs2895	Flooding tolerance	2.17	Zhang et al., 2017
FTI1	3	23419698	FTI (Indica)	Os03g0602300	rice lamina bending and leaf unrolling by promoting castasterone(CS)	120.46	Asahina et al., 2014

Table 7 (continued)

SNPs id	Chr	position	traits associated	colocalized genes/QTLs	function	distance away from the colocalized genes	
JaFCL7	6	24628911	FCL (Japonica)	Os06g0597400	Maintenance of shoot apical meristem. Leaf development.	205.04	Nishimura et al., 2002
JaNCL1	8	27300242	NCL (Japonica)	Os08g0540400	seedling growth	157.14	Asano et al., 2011

Discussion

One of the most important traits for the success of a direct-sowing system is good germination rate and high seedling survival. Development of cultivars with high tolerance to flooding enables successful cultivation of rice in both rainfed condition, where flooding is anticipated any time after sowing, and under irrigated condition where lands are not well leveled or if any unpredicted rains occur. Enhancing the tolerance of rice cultivars to flooding during germination allows intentional flooding with a shallow depth of water after sowing which subsequently helps in weed control and reduces the need of manual weeding and herbicide application. During germination under flooded condition, rice seeds develop a shoot which is limited to white coleoptile (Kordan, 1976). Coleoptile growth under anoxia condition is due to the enlargement of cells as cell division is restricted in initial phase of flooding (ÖPIK, 1973). The growing rice coleoptiles assimilates nutrients from endosperm for growth and maintenance of high turgor pressure (Atwell, Waters, et al., 1982). Cell enlargement is directly influenced by cell wall plasticity and turgor pressure.

During the flooded condition, the genotype that can grow faster in hypoxia condition can thrive well. The faster the growth of coleoptile, the sooner it can avoid hypoxia and higher the chance of survival. More than 8000 accessions from IRRI gene bank have been previously screened for flooding tolerance during germination stage under submergence condition for 3 weeks to identify the donor for breeding rice cultivars with improved tolerance to hypoxic stress (Angaji, Septiningsih, et al., 2010). But surprisingly, only 19 accessions (0.23%) exhibited a survival rate higher than 70%. Subsequent screening in replicated trial reduced the number to further i.e. only 5 accessions (0.06%) were observed to be tolerant to flooding during germination (Ismail, Ella, et al., 2008). This finding indicates that the flooding tolerance is not present in majority of rice

accessions although rice is identified to be more submergence tolerance than other cereal crops belonging to the Poaceae family. It can be argued, however, the very low percentage of tolerant accession could be attributed to the very stringent selection. Shorter duration of flooding may have identified additional tolerant accessions. Previously, several researches have performed QTL mapping associated with anaerobic germination using highly tolerant genotypes such as Khao Hlan On (Angaji, Septiningsih, et al., 2010), Ma-Zhan Red (Septiningsih, Ignacio, et al., 2013), Nanhi and Karshu (Baltazar, Ignacio, et al., 2014, Baltazar, Ignacio, et al., 2019). All of these studies used the percent seedling survival after 21 DAS in anaerobic condition. Other studies done by Hsu et al. (2015) and Zhang et al. (2017) were focused on evaluating the length of the coleoptile under submerged conditions (Hsu and Tung, 2015, Zhang, Lu, et al., 2017). Both the survival rate and longer coleoptile growth are important traits contributing to the success of direct seeded sowing under flooding stress.

In our study we measured coleoptile length under hypoxia stress and estimated the anaerobic germination index by dividing the length of the coleoptile growth under anaerobic stress by the length of coleoptile under normal condition. Higher index means longer coleoptile in anaerobic stress than in normal condition, and the reverse suggests poor coleoptile growth in anaerobic condition. Although the stress was only for four days, the evaluation of coleoptile length under submerged condition and anaerobic germination index enabled us to identify the most tolerant accessions under flooding stress. A wide range of phenotypic variation was observed among the accessions under submerged condition. The accessions belonging to japonica sub-species, especially tropical japonica, were observed to have faster coleoptile elongation rate (higher index) than other sub-populations showing that japonica sub-species are more adapted to germination under submerged condition. Similar results of higher tolerance ability of japonica sub-

species have been previously reported by Hsu et al. (2015). A number of QTL studies using japonica varieties as donor parents in bi-parental mapping populations have been reported, and they found that the japonica alleles enhance anaerobic germination in the segregating populations (Angaji, 2008, Jiang, Liu, et al., 2006, Septiningsih, Ignacio, et al., 2013).

In this study, many of the accessions showing highly vigorous growth under normal condition showed suppressed coleoptile growth under flooding stress and these were the susceptible entries. The analysis of variance also showed that there is a significant interaction between flooding stress and genotypes. The flooding stress has significant effect on coleoptile growth.

From our GWAS analysis, we discovered 24, 16 and 13 distinct QTLs in ALL accessions, indica sub species and japonica sub-species respectively. The phenotypic variance explained by the identified SNPs from ALL accessions analysis were in the range of 2.56% to 11.49%. Unlike the major Sub1 QTL reported for submergence tolerance during vegetative stage by previous study, no such major QTLs explaining sufficient phenotypic variance have been reported for tolerance to hypoxia stress in rice (Septiningsih, Ignacio, et al., 2013). All the QTLs previously identified for tolerance to flooding stress during germination have much smaller effects than Sub1 (Angaji, Septiningsih, et al., 2010, Septiningsih, Ignacio, et al., 2013). Therefore, an appropriate breeding strategy like gene pyramiding could be used to combine the best alleles into one genetic background.

The phenotypic screening of this study showed that japonica sub-population used in our study is more tolerant than other sub-population. In consent with this, the estimated phenotypic variation (R^2) explained by the significant SNPs discovered in japonica species is higher than the phenotypic variation explained by the significant alleles discovered from GWAS analysis of ALL

accessions. The phenotypic variation explained by the significant SNPs identified in japonica species were in the range of 14.7% to 17.8% whereas the phenotypic variance explained by the significant SNPs from ALL accessions analysis were in the range of 11.45 % to 16.52 %. This could be because many of the japonica accessions are highly tolerant to anaerobic germination condition.

Out of 25 significant SNPs discovered from GWAS analysis of ALL accessions, 5 significant SNP associated with AG index was rediscovered in GWAS analysis of indica subspecies. This finding suggest that this SNP may come from indica sub-species. Among the 24 significant QTLs identified in ALL accessions, 11 QTLs were found to be co-localized with previously reported QTLs/genes. Five significant SNPs, FTI1 and FTI3, associated FTI were found to be located in LD with previously identified genes/SNPs, CYP85A1 and AQEP001, respectively (Angaji, Septiningsih, et al., 2010, Asahina, Tamaki, et al., 2014, Hsu and Tung, 2015, Nghi, Tondelli, et al., 2019, Redona and Mackill, 1996). CYP85A1 was found to be controlling shoot growth by causing lamina bending and leaf unrolling by regulating castasterone (Asahina, Tamaki, et al., 2014). A significant SNP, 37615382 associated with coleoptile elongation was identified under flooding condition [Nghi et al. (2019)]. Four significant SNPs, NCL1, NCL5, NCL6 and NCL7 associated with coleoptile growth under normal condition were found to be co-localized with genes, OsNIP3;1, qAG1-2, sh12, and OsVP22, Node26-like intrinsic protein 3-1 respectively (Angaji, Septiningsih, et al., 2010, Hsu and Tung, 2015). These genes were previously reported to be controlling shoot growth during seedling stage.

Conclusion

GWAS has been widely used in rice genetic research for the identification of candidate genes underlying complex quantitative traits. Recently, association mapping has been developed to isolate genes controlling different agronomic traits like awn distribution, grain size, heading date, panicle number per plant (Si, Chen, et al., 2016, Yano, Yamamoto, et al., 2016). The efficiency of GWAS study in novel gene discovery has already been proved to be powerful (Yano, Yamamoto, et al., 2016). In our study, we used 7K SNP chip to identify chromosomal regions controlling coleoptile growth under submerged condition. The candidate genes identified in our study will be functionally characterized in the near future. Further, these findings would help in designing markers to be used in markers assisted breeding program for rice improvement against flooding during germination stress.

References

Angaji, S.A. 2008. Mapping QTLs for submergence tolerance during germination in rice. African Journal of Biotechnology 7.

Angaji, S.A., E.M. Septiningsih, D. Mackill and A.M. Ismail. 2010. QTLs associated with tolerance of flooding during germination in rice (Oryza sativa L.). Euphytica 172: 159-168.

Asahina, M., Y. Tamaki, T. Sakamoto, K. Shibata, T. Nomura and T. Yokota. 2014. Blue light-promoted rice leaf bending and unrolling are due to up-regulated brassinosteroid biosynthesis genes accompanied by accumulation of castasterone. Phytochemistry 104: 21-29.

Asano, T., M. Hakata, H. Nakamura, N. Aoki, S. Komatsu, H. Ichikawa, et al. 2011. Functional characterisation of OsCPK21, a calcium-dependent protein kinase that confers salt tolerance in rice. Plant molecular biology 75: 179-191.

Atwell, B., I. Waters and H. Greenway. 1982. The effect of oxygen and turbulence on elongation of coleoptiles of submergence-tolerant and-intolerant rice cultivars. Journal of Experimental Botany 33: 1030-1044.

Baltazar, M.D., J.C.I. Ignacio, M.J. Thomson, A.M. Ismail, M.S. Mendioro and E.M. Septiningsih. 2014. QTL mapping for tolerance of anaerobic germination from IR64 and the aus landrace Nanhi using SNP genotyping. Euphytica 197: 251-260.

Baltazar, M.D., J.C.I. Ignacio, M.J. Thomson, A.M. Ismail, M.S. Mendioro and E.M. Septiningsih. 2019. QTL mapping for tolerance to anaerobic germination in rice from IR64 and the aus landrace Kharsu 80A. Breeding Science: 18159.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Earl, D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources 4: 359-361.

Gibbs, J., S. Morrell, A. Valdez, T. Setter and H. Greenway. 2000. Regulation of alcoholic fermentation in coleoptiles of two rice cultivars differing in tolerance to anoxia. Journal of Experimental Botany 51: 785-796.

Hanaoka, H., S. Uraguchi, J. Takano, M. Tanaka and T. Fujiwara. 2014. O s NIP 3; 1, a rice boric acid channel, regulates boron distribution and is essential for growth under boron-deficient conditions. The Plant Journal 78: 890-902.

Hsu, S.-K. and C.-W. Tung. 2015. Genetic Mapping of Anaerobic Germination-Associated QTLs Controlling Coleoptile Elongation in Rice. Rice 8: 38. doi:10.1186/s12284-015-0072-3.

Huang, X., T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C. Li, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nature genetics 42: 961.

Ismail, A.M., E.S. Ella, G.V. Vergara and D.J. Mackill. 2008. Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (Oryza sativa). Annals of Botany 103: 197-209.

Jiang, L., S. Liu, M. Hou, J. Tang, L. Chen, H. Zhai, et al. 2006. Analysis of QTLs for seed low temperature germinability and anoxia germinability in rice (Oryza sativa L.). Field crops research 98: 68-75.

Kato-Noguchi, H. and M. Morokuma. 2007. Ethanolic fermentation and anoxia tolerance in four rice cultivars. Journal of plant physiology 164: 168-173.

Kordan, H. 1976. Oxygen as an Enviromnental Factor in Influencing Normal Morphogenetic Development in Germinating Rice Seedlings. Journal of Experimental Botany 27: 947-952.

Kretzschmar, T., M.A.F. Pelayo, K.R. Trijatmiko, L.F.M. Gabunada, R. Alam, R. Jimenez, et al. 2015. A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. Nature Plants 1: 15124.

Lasanthi-Kudahettige, R., L. Magneschi, E. Loreti, S. Gonzali, F. Licausi, G. Novi, et al. 2007. Transcript profiling of the anoxic rice coleoptile. Plant Physiol 144: 218-231. doi:10.1104/pp.106.093997.

Lee, K.-W., P.-W. Chen, C.-A. Lu, S. Chen, T.-H.D. Ho and S.-M. Yu. 2009. Coordinated responses to oxygen and sugar deficiency allow rice seedlings to tolerate flooding. Sci. Signal. 2: ra61-ra61.

Lu, C.-A., T.-h.D. Ho, S.-L. Ho and S.-M. Yu. 2002. Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of α-amylase gene expression. The Plant Cell 14: 1963-1980.

Lu, C.-A., C.-C. Lin, K.-W. Lee, J.-L. Chen, L.-F. Huang, S.-L. Ho, et al. 2007. The SnRK1A protein kinase plays a key role in sugar signaling during germination and seedling growth of rice. The Plant Cell 19: 2484-2499.

Magneschi, L., R. Kudahettige, A. Alpi and P. Perata. 2009. Comparative analysis of anoxic coleoptile elongation in rice varieties: relationship between coleoptile length and

carbohydrate levels, fermentative metabolism and anaerobic gene expression. Plant Biology 11: 561-573.

Magneschi, L. and P. Perata. 2008. Rice germination and seedling growth in the absence of oxygen. Annals of Botany 103: 181-196.

Miro, B. and A.M. Ismail. 2013. Tolerance of anaerobic conditions caused by flooding during germination and early growth in rice (Oryza sativa L.). Frontiers in plant science 4: 269.

Nagasaki, H., J.-i. Itoh, K. Hayashi, K.-i. Hibara, N. Satoh-Nagasawa, M. Nosaka, et al. 2007. The small interfering RNA production pathway is required for shoot meristem initiation in rice. Proceedings of the National Academy of Sciences 104: 14867-14871.

Nghi, K.N., A. Tondelli, G. Valè, A. Tagliani, C. Marè, P. Perata, et al. 2019. Dissection of coleoptile elongation in japonica rice under submergence through integrated genome wide association mapping and transcriptional analyses. Plant, cell & environment.

Nishimura, A., M. Ito, N. Kamiya, Y. Sato and M. Matsuoka. 2002. OsPNH1 regulates leaf development and maintenance of the shoot apical meristem in rice. The Plant Journal 30: 189-201.

ÖPIK, H. 1973. Effect of anaerobiosis on respiratory rate, cytochrome oxidase activity and mitochondrial structures in coleoptiles of rice (Oryza sativa L.). Journal of Cell Science 12: 725-739.

Pandey, S. 2002. Direct seeding: research strategies and opportunitiesInt. Rice Res. Inst.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Raj, A., M. Stephens and J.K. Pritchard. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197: 573-589.

Redona, E. and D. Mackill. 1996. Mapping quantitative trait loci for seedling vigor in rice using RFLPs. Theoretical and Applied Genetics 92: 395-402.

Septiningsih, E.M., J.C.I. Ignacio, P.M. Sendon, D.L. Sanchez, A.M. Ismail and D.J. Mackill. 2013. QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. Theoretical and Applied Genetics 126: 1357-1366.

Septiningsih, E.M., A.M. Pamplona, D.L. Sanchez, C.N. Neeraja, G.V. Vergara, S. Heuer, et al. 2008. Development of submergence-tolerant rice cultivars: the Sub1 locus and beyond. Annals of Botany 103: 151-160.

Si, L., J. Chen, X. Huang, H. Gong, J. Luo, Q. Hou, et al. 2016. OsSPL13 controls grain size in cultivated rice. Nature genetics 48: 447.

Singh, A., E.M. Septiningsih, H.S. Balyan, N.K. Singh and V. Rai. 2017. Genetics, Physiological Mechanisms and Breeding of Flood-Tolerant Rice (Oryza sativa L.). Plant and Cell Physiology 58: 185-197.

Toojinda, T., M. Siangliw, S. Tragoonrung and A. Vanavichit. 2003. Molecular genetics of submergence tolerance in rice: QTL analysis of key traits. Annals of Botany 91: 243-253.

Yano, K., E. Yamamoto, K. Aya, H. Takeuchi, P.-c. Lo, L. Hu, et al. 2016. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nature genetics 48: 927.

Yu, J., G. Pressoir, W.H. Briggs, I.V. Bi, M. Yamasaki, J.F. Doebley, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics 38: 203.

Yu, S.-M., S.-F. Lo and T.-H.D. Ho. 2015. Source–sink communication: regulated by hormone, nutrient, and stress cross-signaling. Trends in plant science 20: 844-857.

Zhang, M., Q. Lu, W. Wu, X. Niu, C. Wang, Y. Feng, et al. 2017. Association mapping reveals novel genetic loci contributing to flooding tolerance during germination in Indica rice. Frontiers in plant science 8: 678.

Zhang, X.-Q., P. Hou, H.-T. Zhu, G.-D. Li, X.-G. Liu and X.-M. Xie. 2013. Knockout of the VPS22 component of the ESCRT-II complex in rice (Oryza sativa L.) causes chalky endosperm and early seedling lethality. Molecular biology reports 40: 3475-3481.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature genetics 42: 35

CHAPTER III

GENOME-WIDE ASSOCIATION MAPPING TO IDENTIFY COLD TOLERANCE GENES DURING GERMINATION

Synopsis

Low temperature significantly affects rice growth and yield. Temperatures lower than 30°C are detrimental for germination and uniform seedling stand. To investigate the genetic architecture underlying cold tolerance during germination in rice, we conducted a genome wide association study of an enriched rice panel consisting of 257 rice accessions from Central and South Asia along with rice collections from around the world, using a 7K SNP marker array. The experiment was conducted in controlled growth chambers in dark conditions at 13°C, and the rice accessions were phenotyped for low temperature germinability (LTG), germination index (GI), coleoptile growth under cold condition (CLC), plumule length at 4 d recovery (PLR), and plumule growth rate after cold germination (PLRR). GWAS analysis was performed using the mixed linear model of the GAPIT package. A total of 41 independent QTLs were identified at a p-value < 0.001 and a total of 9 QTLs were identified using a FDR corrected p-value < 0.05 across the different chilling indices with the full set of accessions. At the same threshold of p < 0.001, a total of 6 and 10 independent QTLs were identified in the subset of *japonica* and indica subpopulations, respectively. We identified 4 and 3 common GWAS sites between indica species and ALL and japonica species and ALL accessions respectively. Among the 41 QTLs in all accessions, 13 QTLs were identified in common across at least two of the low temperature indices. Out of 41 QTLs identified in the whole panel, 11 QTLs identified in *japonica* only and 10 QTLs in indica only, 14, 6 and 7 QTLs were colocalized with known genes/QTLs, respectively. From this study, we

identified 27 novel QTLs from all accessions, 5 novel QTLs from *japonica* group and 3 novel QTLs from indica group. The results of our study provided useful information on genetic architecture underlying cold tolerance during germination in rice, which in turn can be used for further molecular study and crop improvement for low-temperature stressed environments.

Introduction

Rice is more susceptible to cold stress than other cereal crops due to its origin in the tropical and subtropical regions (Zhao, Zhang, et al., 2017). Low temperature causes major stress for rice growing in 25 countries (Cruz, Sperotto, et al., 2013) and to more than 15 million ha^2 of rice grown worldwide (Bai, Zhao, et al., 2016). Low temperature sensitivity at the germination stage is a challenge for rice cultivation especially in many temperate Asian countries where direct seeding method is practiced (Iwata, Shinada, et al., 2010). Low temperature stress causes detrimental effects on seed germination, tillering, flowering and yield (Andaya and Mackill, 2003). The optimum temperature for rice growth is 30°C and the temperature below 15°C causes low germination rate, reduced tillering, yellowing, withering, delayed heading or sterility which ultimately reduces yield (Kaneda, 1974, Mackill and Lei, 1997, Yoshida, Kanno, et al., 1996). Good seed germination and robust plant establishment determines the final yield of the plants. The cold stress during germination causes poor germination and retarded plant growth. Vigorous growth during seedling stage is crucial for good plant establishment. Hence, the development of cold tolerant lines has been one of the main targets of rice breeding. However, breeding of rice lines for low temperature has been challenging due to the following reasons: response of rice plants to cold varies with growth stages, i.e. the mechanism of cold tolerance at one stage is not correlated with other stage (Liu, Lu, et al., 2015); low temperature tolerance is controlled by quantitative loci where many genes across the genome have small effects (Ji, Jiang, et al., 2009); and epistatic

interaction among alleles at unlinked loci (Zhang, Ma, et al., 2014). A wide range of variation to cold tolerance among *Oryza sativa* has been reported where *japonica* sub species have been identified to be more tolerant than *indica* (Baruah, Ishigo-Oka, et al., 2009). A few studies have been done to improve cold tolerance of the indica cultivars using japonica cultivars; however, due to lack of genetic diversity in japonica germplasm, further improvement of japonica cultivars has been a big challenge in breeding program (Zhang, Ma, et al., 2014).

Cold stress can lead to severe osmotic and oxidative stress in plants (Wang, Vinocur, et al., 2003, Yun, Park, et al., 2010). Rice is a chilling sensitive crop due to its subtropical origin but so far only very few genes controlling for chilling tolerance have been identified (Cruz, Sperotto, et al., 2013, Zhang, Chen, et al., 2014, Zhang, Zheng, et al., 2014). The first gene identified to be associated with cold tolerance of young seedling (CTS) was *CTS 12* where the gene codes for stress-induced protein with multifunctional glutathione transferase (GST) activity (Kim, Andaya, et al., 2011, Takesawa, Ito, et al., 2002). The first gene identified for low temperature germinability was *qLTG3-1*, where the gene encodes for a secreted hybrid glycine-rich protein and a single nucleotide substitution differentiates between strong and weak alleles (Fujino, Sekiguchi, et al., 2008). The first gene identified for cold tolerance at booting stage was Ctb-1, where the gene codes for F-box protein suggesting a ubiquitin-proteasome pathway is involved in chilling tolerance at the booting stage (Saito, Hayano-Saito, et al., 2010). The first gene presumably involved in signal transduction, "*Chilling-Tolerance Divergence 1*"(*COLD1*) was recently identified where the gene codes for G-protein signaling regulator leading to CA^{++} influx into the cells of chilling tolerant cultivars (Ma, Dai, et al., 2015). It is highly imperative to identify additional chilling tolerance genes or QTLs in order to better understand the chilling tolerance mechanism in rice and for crop improvement.

QTL mapping and GWAS are two widely used tools to discover genetic control of complex traits. Most of the published data on genetic loci controlling chilling tolerance in rice were obtained by bi-parental mapping populations from *O. sativa* ssp Indica X *O sativa* ssp japonica crosses as japonica sub-species are more cold tolerant than indica (Cruz and Milach, 2004, Ma, Dai, et al., 2015, Mackill and Lei, 1997, Mao and Chen, 2012). The major drawback of biparental mapping is the limitation of genetic background to parental lines, these studies do not account for genetic variation in rice beyond the parental lines and are exclusively limited to the use of few mapping populations derived from East Asian cultivars representing indica and japonica subspecies (Schläppi, Jackson, et al., 2017).

Many QTLs controlling low temperature germination (16, 19-22, (Sales, Viruel, et al., 2017, Shakiba, Edwards, et al., 2017) and cold tolerance during seedling stage have been detected. Recently, Pan et al. (Pan, Zhang, et al., 2015) reported 22 QTLs associated to cold tolerance during germination stage using SSR markers in 174 Chinese accessions. Sales et al., 2017 detected 24 SNPs associated with low temperature germination and growth rate at low temperature while Shakiba et al., 2016 discovered 42 QTLs controlling cold tolerance at the seedling stage (Sales, Viruel, et al., 2017, Shakiba, Edwards, et al., 2017). Fujino et al., 2015 conducted GWAS mapping with 117 markers in Hokkaido Rice core panel, comprising 63 Japanese land races and breeding lines and discovered 6 QTLs associated with cold tolerance at heading stage and 17 QTLs with low temperature germinability (Fujino, Obara, et al., 2015). Lv et. al. (Lv, Guo, et al., 2016) reported 132 loci associated with 16 traits evaluated under natural chilling and cold shock stress using large collection of 529 rice accessions with more than 4 million markers. Haplotype analysis of these accessions for *OsMYB2* revealed the differences between indica and japonica accessions, where japonica sub-species was found to be more tolerant than indica.

Different researchers have used different stress treatments (time, temperature and duration) and different indicators (survival of seeds, germination rate, yellowing etc.) to study cold tolerance which might have resulted in variation in the number and location of QTLs related to low temperature (Zhang, Zheng, et al., 2014). Cruz and Milach (Cruz and Milach, 2004) suggested that variation in coleoptile growth and percentage of seeds superior to 5mm coleoptile length at cold temperature were sufficient to identify cold tolerant genotypes. The differences in seedling vigor among genotypes may cause difficulty in the identification of cold tolerant lines at low temperature. Due to which several researchers have emphasized on the evaluation of lines both at control (ambient) and cold temperature to enable the separation of seedling vigor from cold tolerance [25].

In this study, we performed GWAS on a selected rice panel of 257 accessions. The objectives of this study were to evaluate our germplasm collection for cold tolerance at germination stage and to identify SNPs and the underlying candidate genes associated with tolerance to cold stress during germination and very early seedling growth.

Materials and methods

Rice accessions

The 257 accessions/lines used in this study were obtained from the Genetic Stocks-Oryza (GSOR) collection located at the USDA-ARS Dale Bumpers National Rice Research Center (USDA-ARS DBNRRC; Stuttgart, AR) and the inbred rice breeding program in Beaumont, Texas. The panel that we used represented accessions/lines belonging to aus, aromatic,glaberrima, hybridus, indica, tropical japonica, temperate japonica., Texas breeding elite lines and US released varieties. Seed multiplication was performed in the Beaumont rice research station in summer, 2016 (March – August). The seeds from single panicle were sown for seed multiplication of all

the accessions in 2017. After maturity, the seeds were harvested separately from panicles of single plants. The harvested seeds were dried in a heated air dryer at 37°C for 5 days after harvest and then stored at 4°C. In this experiment to maintain the uniformity, we used seeds collected from single plants. Considering seed age effects on germination and vigor, we only used seeds that were about nine months old. To break the seed dormancy, seeds were incubated at 50°C for 5 days. Germination testing was done by following roll paper method to access the germination percentage of each accession. The detail of the method is found in http://www.knowledgebank.irri.org/step-by-step-production/pre-planting/seed-quality. Dormant seeds were excluded from this experiment.

Indices for evaluating cold tolerance

To screen for the cold tolerance variability in the collected germplasm, different parameters were considered, including low temperature germinability (LTG), germination index (GI), coleoptile growth under cold temperature exposure (CLC), plumule length after 4 days recovery at 30°C from 13°C (PLR) and Plumule Growth Rate after cold germination (PLRR). The experiment was conducted in a growth chamber in a controlled-dark condition following completely randomized design with three replications, and, 30-40 seeds per rep were used. Seeds of all accessions were rinsed with 5% tween-20 for 5 min followed by thorough rinsing with 10% bleach (Sodium hypochlorite) for 10 min and later washed with autoclaved distilled water for 3 times. This process of seed washing extremely helped in avoiding contamination of the germinating seeds.

For control samples, 30 - 40 sterilized seeds were placed on water soaked filter paper placed inside the petri-dishes. The petri-dishes were wrapped in aluminum foil and placed for germination in a growth chamber maintained at 30°C. The experiment was conducted in completely

randomized design and the dark condition was provided to mimic the natural dark condition under soil during rice germination stage. After 7 days of germination, the germination percentage of all three replications of each accessions was counted and averaged to represent the germination percentage of the particular accession.

Low temperature germinability (LTG) and germination Index (GI)

Surface sterilized seeds were incubated in water-soaked filter paper in petri dishes, 30-40 seeds were placed in each petri dishes and the petri dishes were then wrapped with aluminum foil to provide darkness mimicking the seed germination under soil. For each accession, three plates were randomly distributed in the growth chamber set at 13^{o}C temperature. Another set of the same accessions was conducted in growth chamber at 30^{o}C temperature. After 28 days, germinated seeds were counted in each petri dish obtained in both chambers. Germination was defined as visible coleoptile emergence (>5mm) through lemma and palea (hull). The low temperature germinability (LTG) was calculated as percent of seeds germination at 13°C after 28 days. The mean LTG scores were recorded from three petri dishes and normalized with the mean percent germinability of seeds at 30^{o}C (HTG) which was used to calculate GI index. The GI index was calculated as LTG divided by HTG times 100.

Coleoptile length under cold exposure (CLC)

After counting the germinated seeds, the images of all the germinating seeds from each replication were taken with Pentax camera. Later the images were imported to ImageJ software and the coleoptile length of all the germinated seeds were measured and averaged to represent the mean of coleoptile length of each accession after cold exposure. The arithmetic means of the measurement were used for GWAS mapping.

Plumule length after recovery (PLR) and Plumule length recovery rate (PLRR)

After getting the pictures of all germinating seeds, the petri dishes with seedlings were moved to a growth chamber maintained at 30°C and were kept for four days. Plumule lengths were then measured 4 days after recovery at 30°C using ImageJ, averaged and represented as plumule length after recovery (PLR) of the respective accession. The mean plumule growth rate after cold germination was estimated by subtracting the mean coleoptile length after 28 days at 13°C from the mean plumule length at day 4 at 30°C after recovery and dividing the obtained value by 4 to represent plumule growth rate after recovery (PLRR). The PLRR value indicated the growth rate of the plumule over a period of 4 days under normal condition (30°C).

To associate the five chilling tolerance indices with the population structure of our panel, we calculated for each subpopulation, a mean and standard error for the five indices, and compared the distributions to the generally more-chilling tolerant *temperate japonica* population using a Student's *t*-test.

Correlation Coefficient

The LSmeans of all three cold indices were used to calculate Spearman's correlation coefficients r between the chilling indices using R software version 3.5.1 (Lenth, 2016). The correlation analysis of different chilling indices was performed for all the accessions used in our study. In addition, we also performed a separate correlation analysis of the chilling indices of for *japonica* and *indica* group.

Genotyping

The young leaves were collected during the seedling stage from all the field planted germplasm in Beaumont, Texas and used for DNA extraction. Genotyping of all the accessions was done using Illumina custom designed arrays by following the Infinium HD Array Ultra Protocol. The array consisted of 7902 SNPs that were chosen from the Rice Haplotype map project. Genotypes were called using Genome Studio (Illumina, USA). SNPs of call rate <90% and minor allele frequency <5% were removed from the dataset. The quality of each SNPs was confirmed manually by re-clustering. For our study, a subset of 5185 high quality SNP markers obtained after removal of rare allele markers at 5% or less and removing heterozygosity of more than 20% were used to perform the genome wide association analysis.

Population structure and geographic origin

Population structure was estimated using the Bayesian model of the Markov Chain Monte Carlo (MCMC) implemented in the STRUCTURE program (Pritchard, Stephens, et al., 2000). The burn-in length and number of replications of MCMC were set to be 100,000 and 100,000, respectively. Five iterations were performed for each number of populations (k) tested from 2 to 10. The model was set to allow for admixture and correlated allele frequencies. The most probable structure number of (K) was determined by the value of data log likelihood *[LNP (D)*] and an ad hoc statistics delta K (*D(K)*), based on the change rate of [LNP (D)] between successive values of k. The *D(K)* perceives the rate of change in log probability of the data between successive (K) values rather than just the log probability of the data. The Structure Harvester program (Earl, 2012) was used to perform the analysis. The coefficient of ancestry (Q) threshold was defined at 70% to consider an individual with its inferred ancestry from one single group; while the accessions which were unable to be clearly assigned to only one group were determined as mixed

ancestry. We also used the Bayesian clustering program fastStructure (Raj, Stephens, et al., 2014) to calculate the different levels of K (K = 1-10) where the command choose K.py was used to identify the model complexity that maximized the marginal likelihood (K=8).

Model comparison and association mapping

GWAS of all japonica (temperate japonica and tropical japonica,), indica (aus and indica) and the full population (all accessions) were conducted using their corresponding data sets. The Genome Association and Prediction Integrated Tool (GAPIT) package (Lipka, Tian, et al., 2012) with a genotype matrix of 5185 SNPs and a phenotype matrix of 257 accessions was used to perform the GWAS analysis. To predict the genomic regions associated with the traits, we used Mixed Linear model (MLM) of GAPIT (Zhang, Ersoz, et al., 2010). For MLM, we used a kinship (K) matrix as the variance–covariance matrix between the individuals complemented with population structure (Q matrix). The structure data was obtained from the STRUCTURE software (Pritchard, Stephens, et al., 2000) and the relationship matrix (K matrix or kinship) was obtained from the TASSEL 4.0 software (Bradbury, Zhang, et al., 2007). The kinship matrix (K) estimated from SNP genotyping data was used jointly with population structure (Q value) to control the false positives.

The MLM model used is: $Y = \beta X + \gamma P + Zu + \varepsilon$; where Y is the vector of the phenotypic data, X is the vector of genotypic data, β represents the SNP effect, P is the vector of the Q matrix resolving population structure, γ is the effect of population structure, u refers to the random effect from kinship, Z is the Kinship matrix and ε corresponds to random error. The expected *p*-value versus the observed *p*-value test statistics for the SSR markers were plotted (Q-Q plot) to assess the control of type I (false positive) errors under multiple run parameters, The markers were defined to be significantly associated to chilling indices based on $p < 0.001$. The Manhattan plot

distribution chart was obtained by the R software. The percent variance explained by all significant SNPs discovered for each trait was given by GAPIT. The percent variance explained by each individual significant SNP was calculated as the squared correlation between the phenotype and genotype of the SNP (Faraway, 2002).

Identification of Potential Candidate Genes in the QTL Regions

Candidate genes at or near the QTLs reported in this study were identified in the QTARO database (Yonemaru, Yamamoto, et al., 2010) and other previously published literatures on QTL mapping and GWAS study of cold stress tolerance in rice.

Results

To assess the level of cold tolerance of rice during germination, several indices were measured, including low temperature germinability (LTG), coleoptile length at cold stress (CLC), plumule growth after recovery (PLR), and plumule recovery rate (PLR). Most of the rice accessions used in this study have more than 90% germination rate at 30°C. However, we observed a wide variation in coleoptile length (Table 8). In most cases, LTG and CLC was found to be significantly reduced when the rice seeds were germinated at a lower temperature (13°C). In the cold germinated condition, LTG ranged from 0% to 100%. Cold temperature delayed the germination rate of rice seeds and many of the lines started germinating after 7 days of sowing. The range of the coleoptile length was found to be 0 cm to 1.69 cm; while the mean was 0.69 cm. All the rice accessions which had been germinated at 13°C for 28 days were later placed at 30°C for 4 days to evaluate their recovery process. The PLR and the PLRR ranged and from 0 cm to 5.33 cm and 0 cm/d to 1.08 cm/d, with the mean values of 2.96 cm and 0.57 cm/day, respectively.

Based on the population structure, the panel was categorized into nine sub-populations, including the admixtures (Table 9). We observed that among the highest LTGs were the Texas

lines and US released varieties, followed by *Oryza hybridus* and temperate japonica, with the means of 86.2 %, 80.81 %, and 80.3 %, respectively; while the lowest were *O. glaberrima* with the mean 41.12 %. Among the japonicas, aromatic has the lowest LTG (64.17 %), similar rate to the aus group (53.76%). Interestingly, the indica lines used in our study (73.67 %) had comparable germination rate under cold stress with several of the japonica lines. The three groups having highest LTG are almost the same as CLC, with temperate japonica having the highest mean for CLC (0.94 cm), followed by Texas and US lines (0.91 cm) and then hybridus (0.80 cm). Similarly, the smallest length for CLC was also observed in aus, followed by glaberrima and aromatic. CLC of the indica (0.63 cm) was generally shorter compared to the japonica groups, with the exception of the aromatic (0.61 cm). The recovery process from cold stress was also evaluated using the PLR and PLRR parameters. For PLR, the Texas lines and TEJ obviously had the longest plumule growth with the mean values of 3.96 cm and 3.68 cm, respectively; whereas the shortest growth were seen in glaberrima with a mean value of 2.07 cm. Similar trend was observed for PLRR (Table 9).

To evaluate the level of chilling tolerance in different rice groups, the mean and standard error of the five indices of each subpopulation were calculated and compared the distributions to the chilling tolerant groups (temperate japonica) using student's t test (Table 9, Figure 5). We found the values of LTG and PLR were significantly lower for glaberrima, aus, and admixtures. For CLC and PLRR, the values were significantly lower for glaberrima, aus, aromatic, indica and admixtures. We also observed significantly higher PLRR of Texas lines compared to TEJ ($p < 0.05$); whereas no significant difference of CLC was observed between Texas lines and temperate japonica.

We performed the pairwise Pearson's correlation analysis to see the relationships between different chilling indices. We observed significant correlations between all the chilling indices (Table 10-12), albeit with different levels of significance. The results show that rice accessions having good germination under cold stress in general also have higher coleoptile length under cold stress and high recovery rate as well. An assessment of the five chilling tolerance indices indicates the different subpopulations of our collection have significantly different chilling tolerance abilities, potentially reflecting different genetic mechanisms. This suggests GWAS study has the potential to uncover subpopulation specific loci with either positive or negative effects on the described chilling tolerance indices.

Table 8 Phenotypic variation among all traits under hypoxia stress

Trait	Mean	Range	Stdev	SE
LTG (%)	69.21	0 - 100	27.10	1.69
Control ger, CTS (%)	95.32	80.75 - 100	4.41	0.27
GI (%)	72.47	0 - 107.60	28.24	1.76
CLC (cm)	0.69	0 - 1.69	0.33	0.02
PLR (cm)	2.96	0 - 5.33	1.04	0.07
PLRR (cm/d)	0.57	0 - 1.08	0.21	0.01

Table 9 Phenotypic variation of all traits under hypoxia stress among different sub-populations

POP	Sample size	LTG (%)	Control germ. (%)	GI	CLC (cm)	PLR (cm)	PLRR (cm/d)
Admixtures	11	60.73 ± 8.85	94.08 ± 4.60	63.78 ± 9.18*	0.67 ± 0.10*	2.63 ± 0.27**	0.49 ± 0.04*
Aro	20	64.17 ± 5.26*	93.59 ± 1	68.43 ± 5.50**	0.61 ± 0.05***	3.13 ± 0.16*	0.63 ± 0.03
Aus	53	53.76 ± 3.86***	96.74 ± 0.48***	55.25 ± 3.93***	0.48 ± 0.03***	2.46 ± 0.11***	0.5 ± 0.02***
Glaberrima	20	41.12 ± 7.85***	94.74 ± 0.92	42.47 ± 8.07***	0.54 ± 0.09***	2.07 ± 0.27***	0.38 ± 0.05***
Hybrid	9	80.81 ± 4.98	92.48 ± 1.45	87.74 ± 5.85	0.8 ± 0.09	2.65 ± 0.15***	0.46 ± 0.02***
Indica	48	73.67 ± 3.56	96.42 ± 0.64***	76.34 ± 3.65*	0.63 ± 0.05***	2.67 ± 0.15***	0.51 ± 0.03***
TEJ	30	80.3 ± 2.59	92.17 ± 0.94	87.04 ± 2.56	0.94 ± 0.06	3.68 ± 0.14	0.69 ± 0.03
Texas	44	86.2 ± 5.43	95.51 ± 0.67**	90.21 ± 5.60	0.91 ± 0.06	3.96 ± 0.17	0.76 ± 0.04*
TRJ	22	77.16 ± 5.43	97.4 ± 0.67***	79.24 ± 5.60	0.74 ± 0.06*	2.89 ± 0.16***	0.54 ± 0.04*

Table 10 Correlation analysis of different chilling indices among all accessions

	LTG	CT	GI	CLC	PGR	PGRR
LTG	1	0.22***	0.97***	0.64***	0.58***	0.47***
CT	0.22***	1	0.03	0.05	-0.04	-0.06
GI	0.97***	0.03	1	0.65***	0.60***	0.50***
CLC	0.64***	0.05	0.65***	1	0.67***	0.44***
PLR	0.58***	-0.04	0.60***	0.67***	1	0.96***
PLRR	0.47***	-0.06	0.50***	0.44***	0.96***	1

Table 11 Correlation analysis of different chilling indices among indica subpopulation

	LTG	CT	GI	CLC	PLR	PLRR
LTG	1	0.27*	0.95***	0.56***	0.44***	0.34**
CT	0.27*	1	0.02	0.04	-0.16	-0.18
GI	0.95***	0.02	1	0.57***	0.51***	0.42***
CLC	0.56***	0.04	0.57***	1	0.61***	0.34**
PLR	0.44***	-0.16	0.51***	0.61***	1	0.94***
PLRR	0.34**	-0.18	0.42***	0.34**	0.94***	1

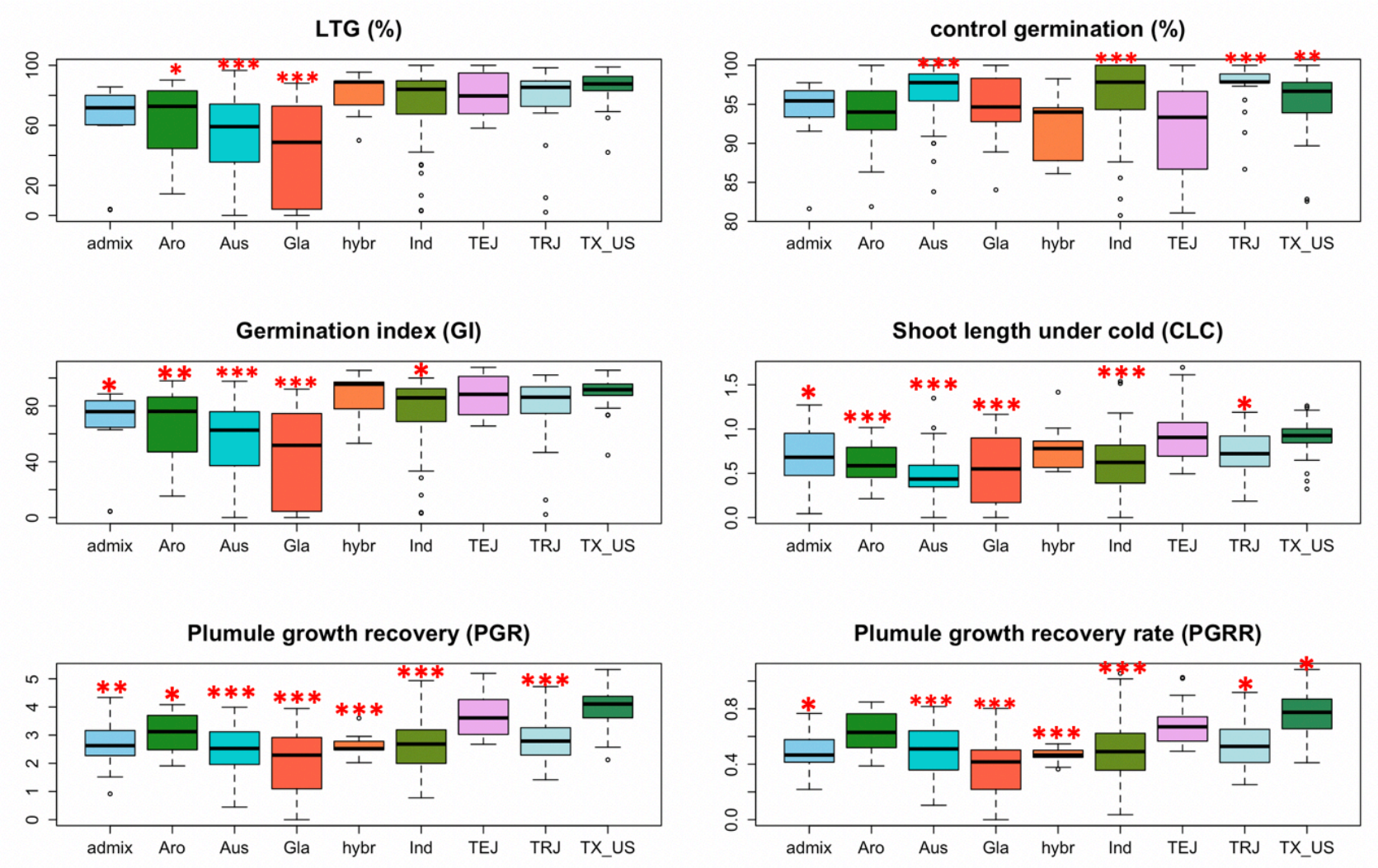
Table 12. Correlation analysis of different chilling indices among japonica subpopulation

	LTG	CT	GI	CLC	PLR	PLRR
LTG	1	0.26**	0.98***	0.68***	0.56***	0.41***

Table 12 (continued)

	LTG	CT	GI	CLC	PLR	PLRR
CT	0.26**	1	0.13	0.16	0.08	0.01
GI	0.98***	0.13	1	0.69***	0.59***	0.44***
CLC	0.68***	0.16	0.69***	1	0.56***	0.30**
PLR	0.56***	0.08	0.59***	0.56***	1	0.94***
PLRR	0.41***	0.01	0.44***	0.30**	0.94***	1

Figure 5 Box plots of chilling indices of different species



GWAS of five chilling tolerance indices for all accessions

A total of 54 significant SNPs were identified at $p < 0.001$, and 9 of them were detected at FDR < 0.05 (Table 13, Figure 6). The SNPs located within an interval of 250 Kb in the same chromosomal region were considered to be a single associated loci. Of the 54 significant associated SNPs, we identified some of the GWAS sites were also shared by multiple chilling indices, considering the reoccurring SNPs in multiple chilling indices, 41 unique SNPs were discovered. At $p < 0.001$, the analysis revealed that 41 unique QTLs within 54 SNPs were identified to be significantly associated with cold tolerance indices within a well-fitted Q-Q plots. Among 54 significant SNPs, 10, 15, 10, 10, and 9 SNPs were found to be significantly associated with LTG, GI, CS, PLR and PLRR, respectively at $p < 0.001$. With FDR < 0.05, 2, 6 and 1 SNPs were found to be associated with CS, PLR and PLRR respectively.

The GWAS site LTG1, LTG2, LTG3, LTG4, LTG5 and LTG7 were found be significantly associated with both LTG and GI. Likewise, the GWAS sites- CS1, CS4 and CS5 were found to be associated with both CLC and PLR. The 4 GWAS sites, PGC1, PGC4, PGC5 and PGC6 were found to be associated with both PLR and PLRR.

The amount of phenotypic variance explained (R^2) ranged from 1.13 % to 3.85 % for LTG, 0.39% to 3.47 % for GI, 1.82 % to 12.87 % for CLC, 0.57 % to 8.59 % for PLR, and 0.58 % to 8.04 % for PLRR.

Figure 6 Manhattan plots and QQ plots of ALL accessions for chilling indices

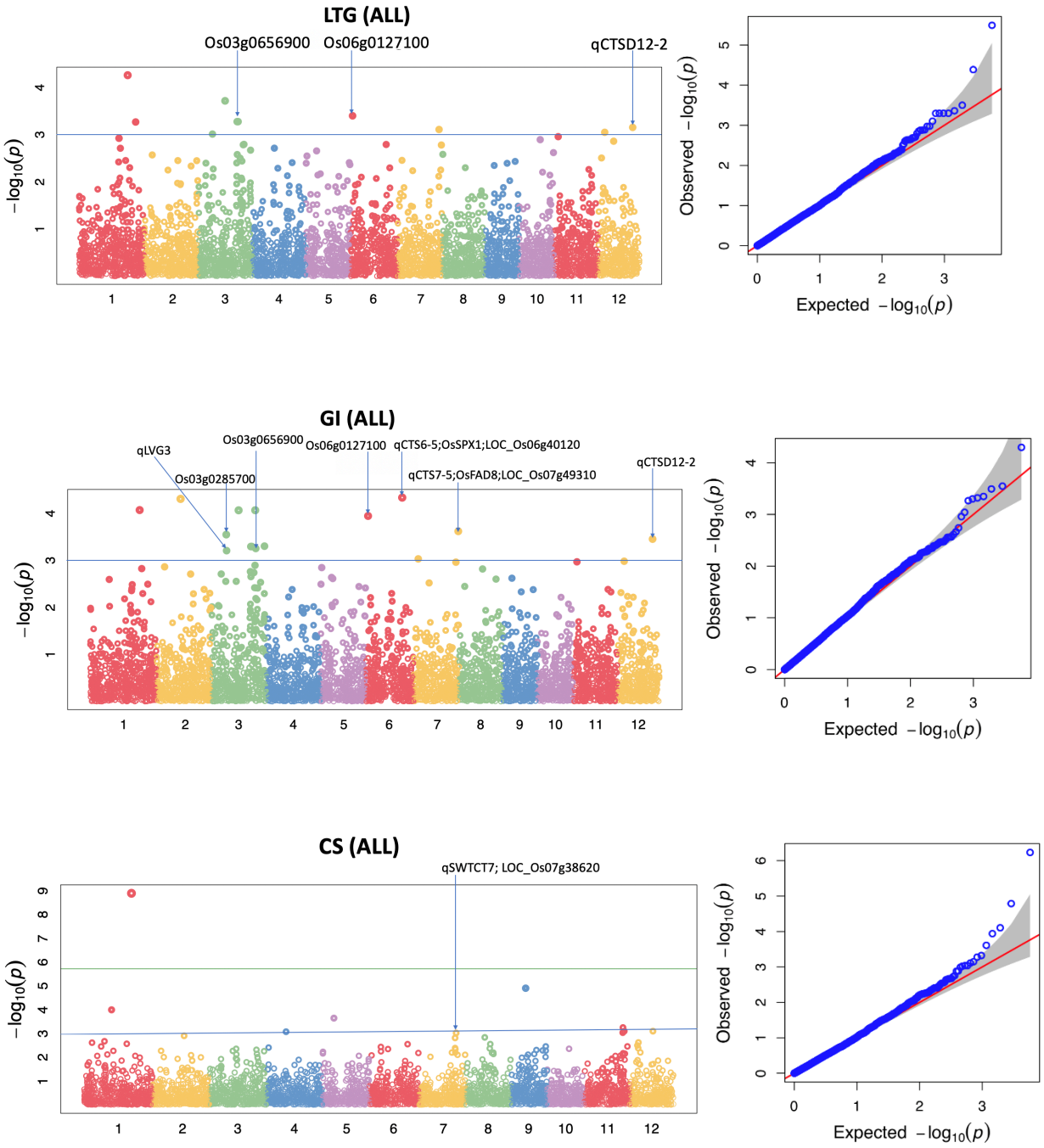


Figure 6 (continued)

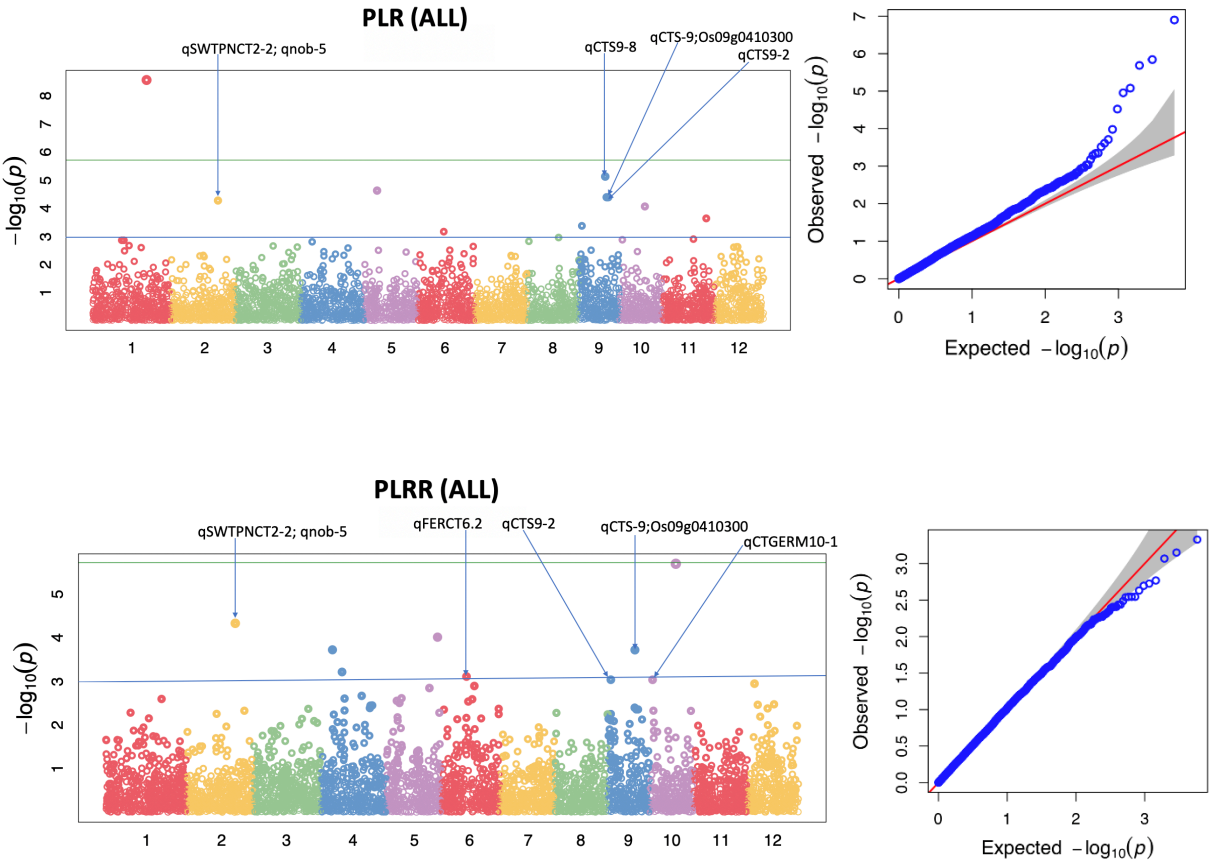


Table 13 SNPs identified from GWAS analysis of all accessions for chilling stress

Chilling indices	SNPs	Chr	Position	P.value	R2 (%)	FDR	Allele effect
LTG	LTG1	1	32323875	5.54E-05	3.856	0.287	-1.0230241
	LTG2	3	17847337	0.00019321	1.72	0.431	6.8428784
	LTG3	6	1444280	0.00040045	2.994	0.431	-0.2889982
	LTG4	3	25833182	0.00053266	1.313	0.431	-0.6990975
	LTG5	3	26456796	0.00053266	1.313	0.431	-2.6570105
	LTG6	1	37615383	0.00054184	1.133	0.431	-6.2436197
	LTG7	12	23074510	0.00070554	2.863	0.431	7.34242857

Table 13 (continued)

Chilling indices	SNPs	Chr	Position	P.value	R2 (%)	FDR	Allele effect
	LTG8	7	27547838	0.00078066	2.398	0.431	-0.114749
	LTG9	12	4552754	0.00089574	2.415	0.431	1.31518816
	LTG10	3	9471911	0.00097599	1.754	0.431	3.44930569
GI	GI1	6	23718296	4.65E-05	2.873	0.09	0.45487244
	GI2	2	15788979	4.95E-05	0.391	0.09	3.37074123
	LTG1	1	32323875	8.56E-05	3.478	0.09	-0.7153672
	GI3	3	28688032	8.63E-05	1.673	0.09	2.18929885
	LTG2	3	17847337	8.65E-05	0.835	0.09	8.68783956
	LTG3	6	1444280	0.00011418	2.644	0.099	-0.5367502
	GI4	7	29079097	0.0002415	2.321	0.179	3.90240881
	GI5	3	9795722	0.00028406	1.252	0.184	11.5694999
	LTG7	12	23074510	0.0003533	2.73	0.204	7.28727296
	GI6	3	34728291	0.00050015	1.906	0.22	2.09818343
	LTG4	3	25833182	0.0005096	1.308	0.22	-0.8587244
	LTG5	3	26456796	0.0005096	1.308	0.22	-1.789968
	GI7	3	29035496	0.00055943	1.738	0.223	-6.797907
	GI8	3	10010232	0.00062319	1.322	0.231	-5.4452627
	GI9	7	2982091	0.00093149	1.082	0.315	3.92516271
CLC	CS1	1	29942776	1.29E-09	12.87	0	0.00231383
	CS2	9	9230514	1.26E-05	5.051	0.033	0.03003859
	CS3	1	17394761	0.00010178	2.357	0.176	0.03868944
	CS4	5	7195992	0.00022752	2.879	0.295	0.02986135
	CS5	11	24782915	0.00056715	2.729	0.496	-0.0411822
	CS6	11	25591959	0.00080267	2.945	0.496	-0.0514784
	CS7	12	14661596	0.00080278	2.863	0.496	-0.036353
	CS8	4	11986644	0.00084796	2.772	0.496	-0.0048604
	CS9	11	24650416	0.00092491	3.186	0.496	0.0523779
	CS10	7	23425700	0.00095739	1.82	0.496	-0.0157005
PLR	CS1	1	29942776	2.74E-09	8.345	0	0.07460478
	PGC1	9	14648157	7.42E-06	3.212	0.019	-0.228834
	CS4	5	7195992	2.35E-05	2.156	0.041	0.11271764

Table 13 (continued)

Chilling indices	SNPs	Chr	Position	P.value	R2 (%)	FDR	Allele effect
	PGC2	9	15399656	4.05E-05	8.596	0.042	-0.0494461
	PGC3	9	16325535	4.05E-05	8.596	0.042	-0.1736818
	PGC4	2	26231409	5.31E-05	1.395	0.046	0.00067455
	PGC5	10	13897640	8.62E-05	1.303	0.064	-0.0858118
	CS5	11	24782915	0.00022899	1.736	0.148	-0.1865644
	PGC6	9	1727512	0.00042242	0.574	0.243	-0.0355241
	PGC7	6	14504992	0.00067857	1.15	0.352	0.04401395
PLRR	PGC5	10	13897640	2.04E-06	2.313	0.011	-0.0223965
	PGC4	2	26231409	4.72E-05	2.043	0.122	0.00792358
	PGCG1	5	27611207	9.79E-05	8.048	0.169	-0.0256048
	PGCG2	4	6193569	0.00019072	1.939	0.199	0.00577457
	PGC1	9	14648157	0.00019193	2.793	0.199	-0.0605513
	PGCG3	4	11331474	0.00061349	1.439	0.53	-0.0126685
	PGCG4	6	13539065	0.00078051	1.407	0.533	-0.0208893
	PGC6	9	1727512	0.00092339	0.589	0.533	-0.0103552
	PGCG5	10	1430027	0.00092595	2.13	0.533	-0.0329187

GWAS of five chilling tolerance indices for japonica species

At cut-off p-value of < 0.001 we identified 13 significant SNPs associated with the chilling tolerance indices in japonica species (Table 14, Figure 7). Among the 13 significant SNPs, we identified 3 SNPs associated with LTG, 3 SNPs associated with GI, 3 SNPs associated with CLC and 7 SNPs associated with PLRR respectively. We also observed a number of SNPs controlling for more than one chilling indices. The 3 SNPs (GI9, JaLTG1 and JaLTG2) were found to be associated with both LTG and GI. Of the 13 significant QTLs, 9 QTLs occupied only one location while 4 QTLs shared 2 sites; one site on chromosome 7 shared by JaPGCG1 and JaPGCG2 and other site on chromosome 7 shared by JaPGCG4 and JaPGCG5. Considering the reoccurring SNPs in multiple chilling indices and linked SNPs, we identified 11 independent GWAS sites. A

significant SNP, GI9 associated with LTG and GI in japonica species was found to be associated with GI in GWAS analysis of ALL accessions. We also identified significant SNPs, CS2 associated with CS and PGCG4, associated with PLRR in japonica species and ALL accessions. The phenotypic variance (R^2) explained by the significant SNPs ranged from 15.98 % to 18.90 % for LTG, 14.68 % to 17.49 % for GI, 15.79% to 20.54% for CLC and 0.64 % to 10.43 % for PLRR.

Figure 7. Manhattan plots and QQ plots of Japonica group for chilling indices

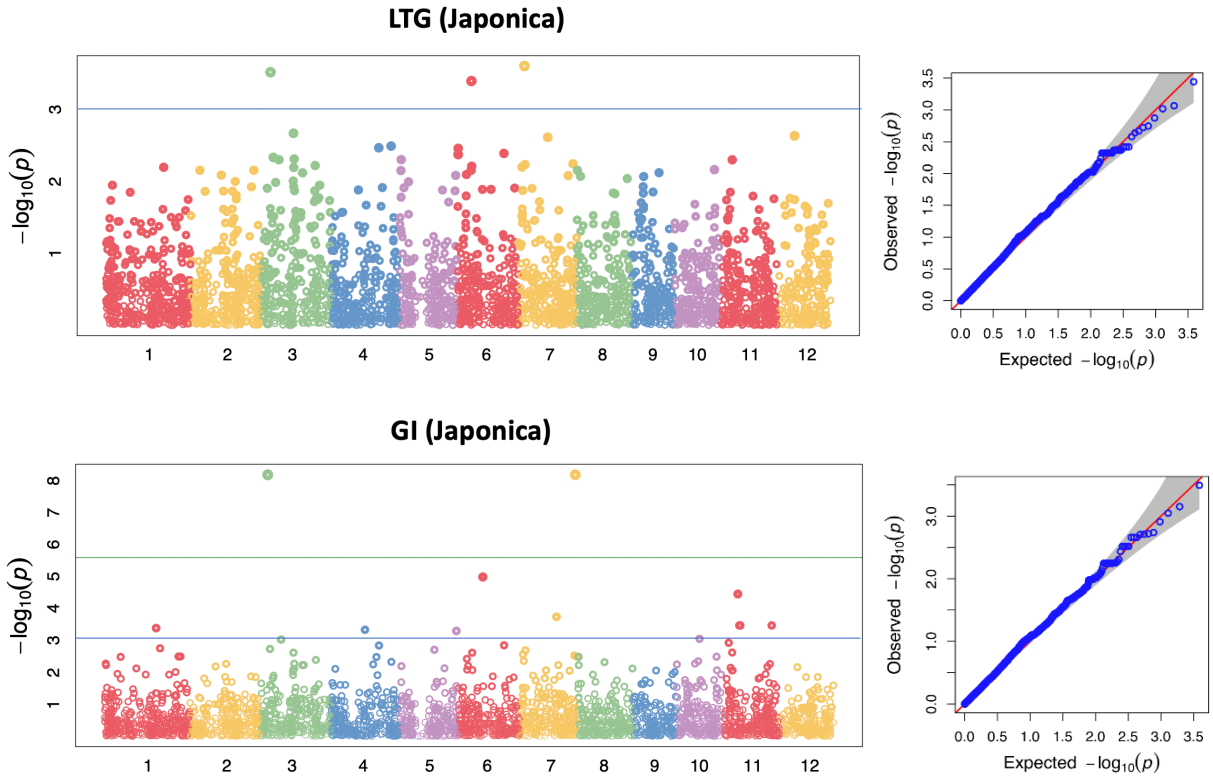


Figure 7 (continued)

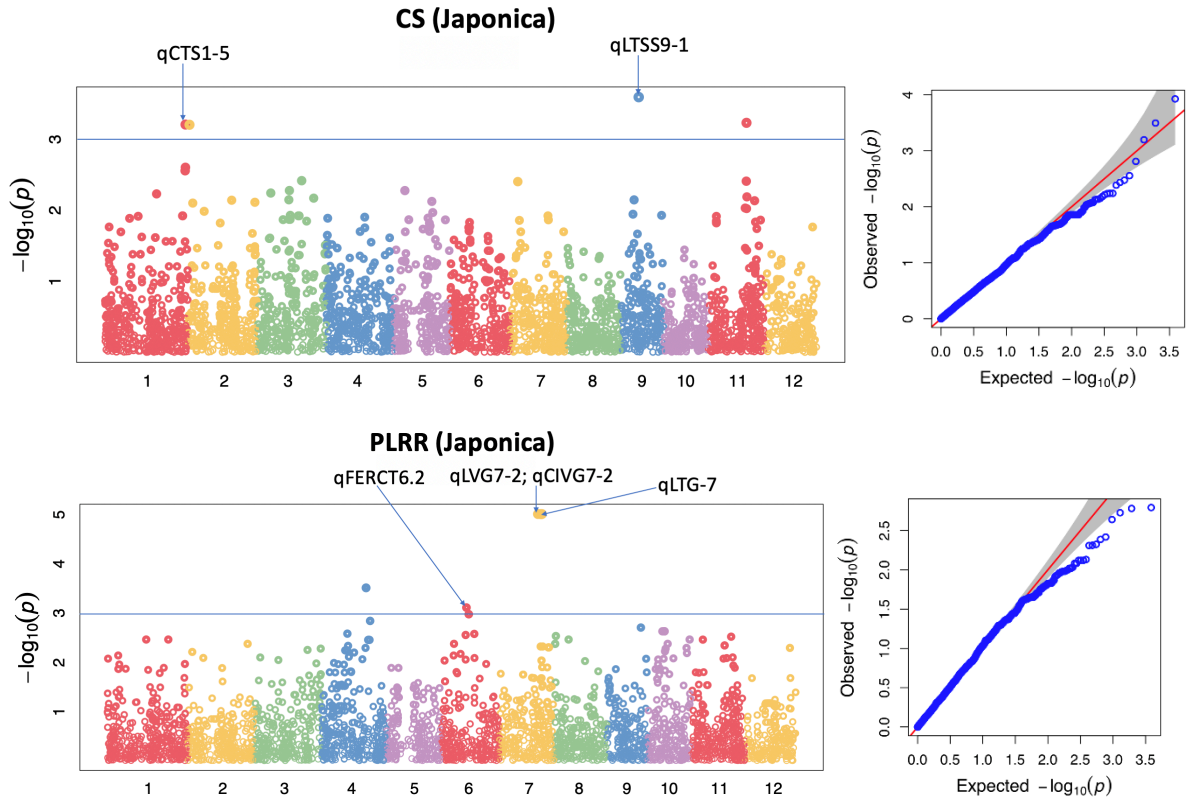


Table 14 SNPs identified from GWAS analysis of japonica species for chilling stress

Chilling indices	SNPs	Chro.	Position	P.value	R2(%)	FDR	Allele effect
LTG	JaLTG1	6	7133939	0.00036083	18.908	0.706	-2.4119481
	GI9	7	2982091	0.00086075	16.297	0.706	-9.5373451
	JaLTG2	3	5124568	0.00095689	15.985	0.706	-2.7723966
GI	JaLTG1	6	7133939	0.00032147	17.492	0.754	-2.653762
	JaLTG2	3	5124568	0.0007048	15.337	0.754	-2.1187457

Table 14 (continued)

Chilling indices	SNPs	Chro.	Position	P.value	R2(%)	FDR	Allele effect
	GI9	7	2982091	0.00089725	14.688	0.754	-9.6650846
CLC	CS2	9	9230514	0.00011812	20.549	0.455	-0.0007151
	JaCS1	1	41999235	0.000319	17.702	0.615	-0.1391191
	JaCS2	2	855183	0.00063464	15.79	0.816	0.12549706
PLRR	JaPGCG1	7	20762320	9.74E-06	0.646	0.00750741	0.05084762
	JaPGCG2	7	22126446	9.74E-06	0.646	0.00750741	0.04359316
	JaPGCG3	7	22277144	9.74E-06	0.646	0.00750741	0.00181973
	JaPGCG4	7	22755972	9.74E-06	0.646	0.00750741	0.00411964
	JaPGCG5	7	22878275	9.74E-06	0.646	0.00750741	0.02188472
	JaPGCG6	4	23859231	0.00030726	9.373	0.19741333	0.03747913
	PGCG4	6	13539065	0.0007749	10.153	0.42675076	0.02204616

GWAS of five chilling tolerance indices for indica species

At cut-off p-value of < 0.001 we identified 14 significant SNPs associated with the chilling tolerance indices in indica species (Table 15, Figure 8). Among the 14 SNPs, 3 SNPs were found to be associated with LTG, 3 SNPs were found to be associated with GI, 4 SNPs were found to be associated with CLC, 3 SNPs were found to be associated with PLR and 1 SNP was found to be associated with PLRR. The 3 SNPs – InLTG1, InLTG2 and InLTG3 were found to be associated with both LTG and GI. A significant SNP, PGC1 was found to be associated with both PGC and PGCG. A significant SNP, LTG3 associated with both LTG and GI was found to be associated

with LTG and GI of ALL accessions too. Similarly, the significant SNPs – CS6 associated with CLC and CS5 associated with PGC in japonica species were identified to be associated with CLC of ALL accessions analysis. We also observed significant SNPs, InLTG1 and InLTG2 associated with LTG were located in the same chromosome region at a distance of 84.56 Kb away. Considering the reoccurring SNPs and linked SNPs, we identified 10 unique SNPs in indica species. The phenotypic variance (R^2) explained by the significant SNPs ranged from 9.96 % to 11.38 % for LTG, 9.78 % to 12.03 % for GI and 10.93 % to 13.40 % for CLC, 8.85 % to 10.75 % for PLR and 10.95 % for PLRR respectively.

A mixed linear model (MLM) was used to control the confounding effect of the subpopulation structure and relatedness in the population of ALL accessions analysis and separate analysis of japonica and indica sub-species. The results for the *japonica* and *indica* group-specific GWAS indicated an overcorrection for the population structure when both subpopulation structure and kinship matrix were considered in the mixed model (*japonica* MLM and *indica* MLM). To address this overcorrection and avoid false negative results, a GAPIT model considering the principal components were used to individually analyze the indica and japonica varietal groups. Only two main subpopulations were observed in indica group and three subpopulations were observed in japonica group as depicted by PCA plot results from GAPIT output.

Figure 8 Manhattan and QQ plots of indica group for traits under chilling indices

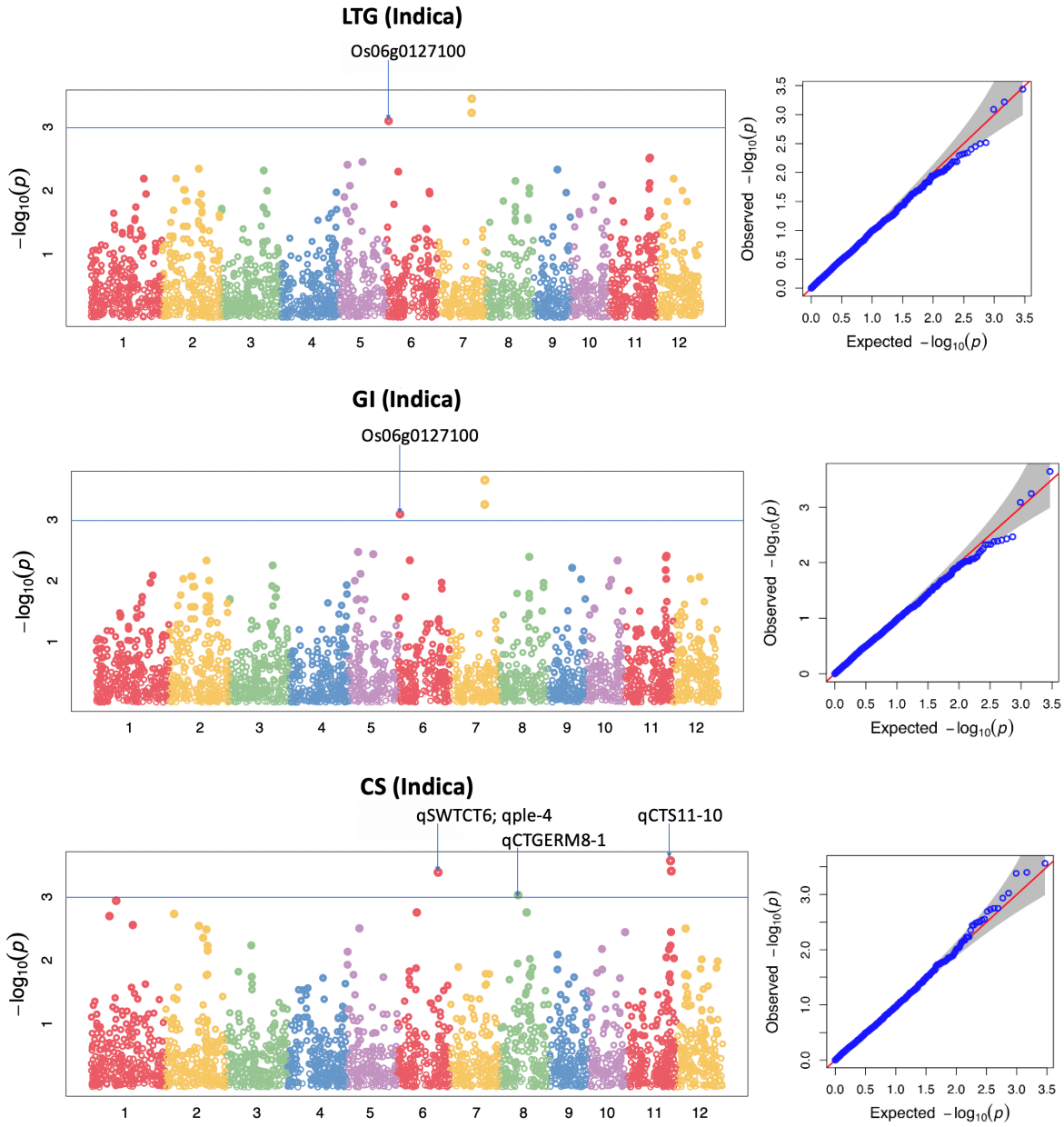


Figure 8 (continued)

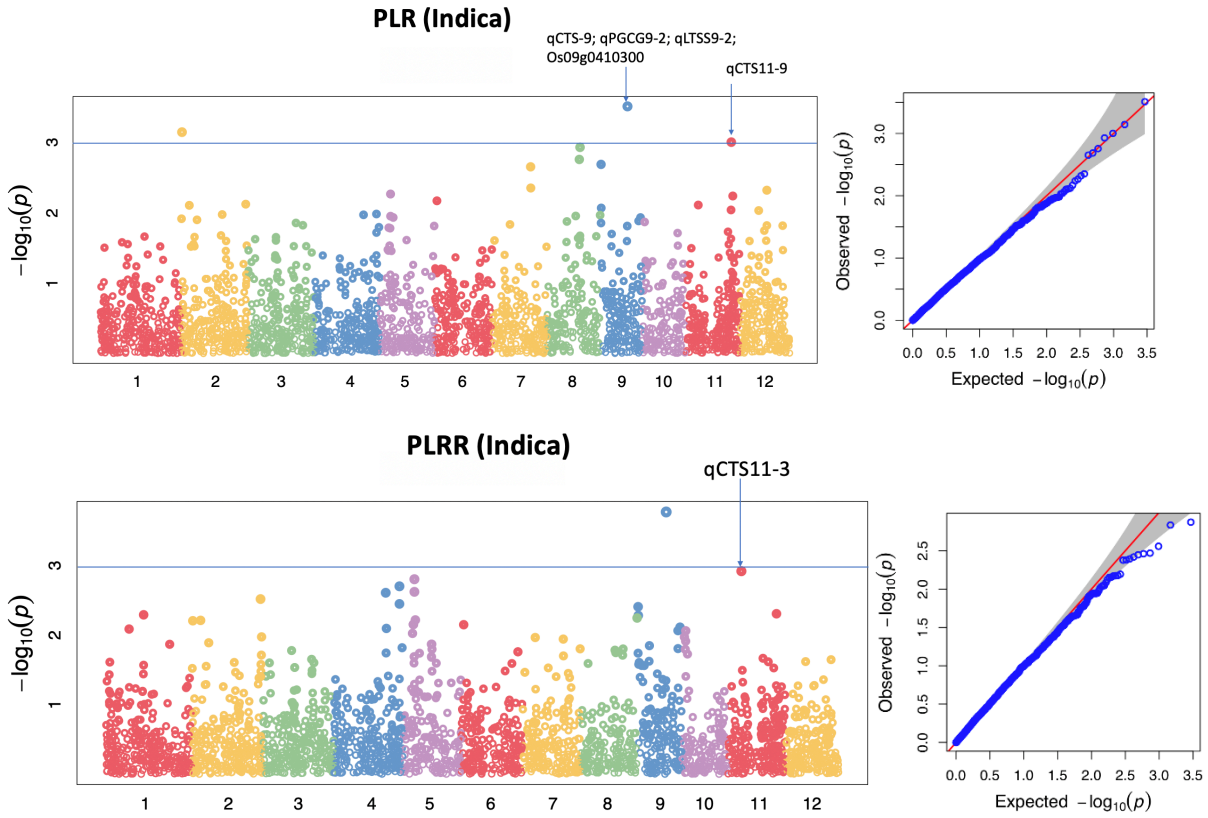


Table 15 SNPs identified from GWAS analysis of indica species for chilling stress

Chilling indices	SNP	Chro	Position	P.value	R2(%)	FDR	Allele effect
LTG	InLTG1	7	20875408	0.00036325	11.382	0.793	-8.0145008
	InLTG2	7	20790840	0.00060439	10.479	0.793	-1.8448415
	LTG3	6	1444280	0.00081147	9.962	0.793	-2.7437131
GI	InLTG1	7	20875408	0.00022682	12.030	0.665	-8.1725897
	InLTG2	7	20790840	0.0005677	10.417	0.799	-1.2611196

Table 15 (continued)

Chilling indices	SNP	Chro	Position	P.value	R2(%)	FDR	Allele effect
	LTG3	6	1444280	0.00081782	9.787	0.799	-2.9723631
CLC	CS6	11	25591959	0.00027289	13.409	0.404	-0.054952
	InCS1	11	26028278	0.00039728	12.651	0.404	0.00767988
	InCS2	6	24039384	0.00041357	12.570	0.404	0.06315769
	InCS3	8	10358484	0.00094445	10.932	0.659	-0.0308898
PLR	PGC1	9	14648157	0.00030923	10.754	0.864	0.29822227
	InPGC1	2	568585	0.00071817	9.377	0.864	-0.060457
	CS5	11	24782915	0.00099619	8.851	0.864	0.17285861
PLRR	PGC1	9	14648157	0.00017225	10.978	0.505	0.04231438

Co-localization of GWAS sites with known genes/QTLs/SNPs

Among the 41 significant QTLs ($p < 0.001$) associated with chilling indices of whole accessions analysis, we found 14 loci co-localized with the previously identified genes/QTLs controlling for cold stress tolerance in rice. We identified 28 novel GWAS sites in GWAS analysis of ALL accessions. A significantly associated SNP, LTG3 associated with both LTG and GI was found to be located at a distance of 10.5 Kb away from a loci, Os06g0127100 which was previously identified to be coding for dehydration-responsive element binding protein responsible for cold, drought and salinity stress tolerance in rice (Ito, Katsura, et al., 2006). Other significant SNP, LTG5 associated with both LTG and GI was found to be 69.39 Kb away from a loci, Os03g0656900 which was identified to be coding for RNA binding protein and controlling for chloroplast development under cold condition (Table 16). A significant SNP, LTG7 associated with LTG and GI was identified to be located at a distance of 11.76 Kb away from a QTL, qCTSD12-2 which was previously identified to be associated with cold tolerance at bud burst stage of rice (Zhang, Ye, et al., 2018). We also identified 4 significant SNPs associated with GI colocalized with previously identified QTLs/genes - GI1 was found to be located at a distance of 157.81 Kb away from QTL, qCTS6-5 and loci, LOC_Os06g40120 which were previously identified to be controlling for cold tolerance at seedling stage in rice (Wang, Liu, et al., 2016); GI4 was found to be 141.58 Kb away from a QTL, qCTS7-5 and loci, LOC_Os07g49310 which were previously discovered to be controlling for cold tolerance at seedling stage; GI5 was found to be located at 98.27 Kb away from a loci, Os03g0285700 previously identified to be responsible for cold stress tolerance at booting stage of rice (Saito, Hayano-Saito, et al., 2010, Wang, Liu, et al., 2016); GI8 was found to be located at 9.713 Kb away from a QTL, qLVG3 which was previously identified to be underlying seedling vigor and germination under low temperature

(Long-Zhi, Zhang, et al., 2006). One significant SNP, CS10 associated with CLC was found to be located at a distance of 231.7 Kb away from QTL, qSWTCT7 and loci, LOC_Os07g38620 which were previously reported to be responsible for cold tolerance at reproductive stage in rice (Shakiba, Edwards, et al., 2017). One significant SNP, PGC1 was found to be located at 116.96 Kb away from *OsWRKY76* which was previously shown to increase cold tolerance when overexpressed (Yokotani, Sato, et al., 2013). The two significant SNPs, PGC3 and PGC6 were found to be located at a distance of 143.79 Kb and 21.39 Kb away from qCTS9-8 and qCTS9-2 respectively which were identified to be cold stress tolerance at seedling stage (Wang, Liu, et al., 2016). A significant SNP, PGC4 associated with both PLR and PLRR was identified to be located at 109.65 Kb away from QTLs, qSWTPNCT2-2 and qnob-5, which were identified to be responsible for cold stress tolerance at reproductive stage in rice (Shakiba, Edwards, et al., 2017). We also identified a SNP, PGCG5 associated with PLRR to be located at a distance of 27.02 Kb away from qCTGERM10-1 which was previously reported to be controlling for cold tolerance at germination stage (Shakiba, Edwards, et al., 2017). Other significant SNP, PGCG4 associated with PLRR was found to be colocalized with a QTL, qFERCT6.2, previously identified to be responsible for cold tolerance during reproductive stage in rice (Shakiba, Edwards, et al., 2017).

Among the significant 11 independent GWAS sites associated with japonica species, we found 6 GWAS sites colocalized with previously identified SNPs. We identified 5 novel GWAS sites in GWAS analysis of japonica species. The 2 SNPs associated with CLC, CS2 and JaCS1 were found to be located at a distance of 169.48 Kb and 140 Kb away from QTLs, qLTSS9-1 and qCTS1-5 which were previously reported to be controlling for cold stress tolerance in seedling stage in rice (Schläppi, Jackson, et al., 2017, Wang, Liu, et al., 2016). A significant SNP, JaPGCG3 ($P = 0.0018$) was found to be located at a distance of 122.11 Kb away from a QTL, qCTS11-2

which was reported to be responsible for cold tolerance at seedling stage in rice (Wang, Liu, et al., 2016). Other significant SNP, PGCG4 associated with PLRR was found to be colocalized with a QTL, qFERCT6.2, previously identified to be responsible for cold tolerance during reproductive stage in rice (Shakiba, Edwards, et al., 2017).

Among the 10 significant GWAS sites discovered in indica species, 7 GWAS sites were found to be colocalized with previously identified genes/QTLs. We identified 2 novel QTLs in indica spcies. A significant SNP, LTG3 was found to be located at a distance of 10.5 Kb away from a loci, Os06g0127100 which was previously identified to be coding for dehydration-responsive element binding protein responsible for cold, drought and salinity stress tolerance in rice (Ito, Katsura, et al., 2006). Three significant SNPs associated with CLC - CS6, InCS2 and InCS3 were found to be colocalized with previously identified genes/QTLs. The significant SNP, CS6 was found to be located at 159.94 Kb away from QTL, qCTS11-10 which was identified to be responsible for cold tolerance at seedling stage (Wang, Liu, et al., 2016). Likewise the SNP, InCS2 was found to be located at 2.8 Kb distance of previously identified QTLs, qSWTCT6 and qple-4 which were reported to be responsible for cold tolerance at reproductive stage (Shakiba, Edwards, et al., 2017), A significant loci, InCS3 associated with CLC was found to be located at 79.516 Kb away from a QTL, qCTGERM8-1 previously identified to be controlling for cold tolerance at germination stage (Shakiba, Edwards, et al., 2017)(ref). Two significant SNPs associated with PLR and PLRR, PGC1 and InPGCG1 were found to be at a distance of 248 Kb and 143.48 Kb away from QTLs, qPGCG9-2; qLTSS9-2 and qCTS11-3 respectively which were reported to be responsible for cold tolerance at seedling stage in rice (Schläppi, Jackson, et al., 2017, Wang, Liu, et al., 2016).

Table 16 Co-localized genes/QTLs in ALL for chilling stress

SNPs id	Chr.	Position	Traits associated	Colocalized genes/QTLs	Function	Distance away (Kb)	References
LTG5	3	26456796	LTG; GI	Os03g0656900	Chloroplast development under cold condition	69.39	Kusumi et al., 2011
LTG3	6	1444280	LTG; GI	Os06g0127100	Cold, drought and salinity tolerance	10.509	Ito et al., 2006
LTG7	12	23074510	LTG; GI	qCTSD12-2	Cold tolerance at bud burst stage	11.76	Zhang et al., 2018
GI1	6	23718296	GI	qCTS6-5; OsSPX1; LOC_Os06g40120	Cold tolerance at seedling stage	157.81	Wang et al., 2016
GI4	7	29079097	GI	qCTS7-5; OsFAD8; LOC_Os07g49310	Cold tolerance at seedling stage	141.58	Wang et al., 2016
GI5	3	9795722	GI	Os03g0285700	Cold tolerance at booting stage.	98.27	Sato et al., 2011
GI8	3	10010232	GI	qLVG3	Low temperature vigor germination	9.713	Han et al, 2006
CS10	7	23425700	CS	qSWTCT7 ; LOC_Os07g38620	Cold tolerance at reproductive stage	231.7	Shakiba et al., 2017
PGC3	9	16325535	PGC	qCTS9-8	Cold tolerance at seedling stage	143.79	Wang et al., 2016
PGC4	2	26231409	PGC; PGCG	qSWTPNCT2-2, qnob-5	Cold tolerance at reproductive stage	109.65	Shakiba et al., 2017
PGC6	9	1727512	PGC	qCTS9-2	Cold tolerance at seedling stage	21.39	Wang et al., 2016
PGCG5	10	1430027	PGCG	qCTGERM10-1	Cold tolerance at germination stage	27.027	Shakiba et al., 2017
PGC1	9	14648157	PGC; PGCG	OsWRKY76; qCTS-9; Os09g0410300	Cold tolerance.	116.96	Peng et al., 2010

Table 16 (continued)

SNPs id	Chr.	Position	Traits associated	Colocalized genes/QTLs	Function	Distance away (Kb)	References
PGCG4	6	13539065	PGCG	qFERCT6.2	Cold tolerance at reproductive stage	17.93	Shakiba et al., 2017

Table 17 Co-localized genes in japonica group for chilling stress

SNPs id	Chr	position	traits associated	colocalized genes/QTLs	function	distance away from the colocalized genes (Kb)	references
JaCS1	1	41999235	CS	qCTS1-5	Cold tolerance at seedling stage	140	Wang et al., 2016
CS2	9	9230514	CS	qLTSS9-1	Cold tolerance at seedling stage	169.48	Schläppi et al., 2017
JaPGCG7	11	6600908	PGCG	qCTS11-2	Cold tolerance at seedling stage	122.11	Wang et al., 2016
PGCG4	6	13539065	PGCG	qFERCT6.2	Cold tolerance at reproductive stage	17.93	Shakiba et al., 2017
JaPGCG4	7	22755972	PGCG	qLVG7-2; qCIVG7-2	Cold tolerance at germination stage	223.62	Han et al., 2006
JaPGCG5	7	22878275	PGCG	qLTG-7	Cold tolerance at germination stage	34.71	Hou et al., 2004
PGCG4	6	13539065	PGCG	qFERCT6.2	Cold tolerance at reproductive stage	17.93	Shakiba et al., 2017

Table 18 Co-localized genes in indica group for chilling stress

SNPs id	Chr	position	traits associated	colocalized genes/QTLs	function	distance away from the colocalized genes (Kb)	references
LTG3	6	1444280	LTG	Os06g0127100	Cold, drought and salinity tolerance	10.509	Ito et al., 2006
CS5	11	24782915	PGC	qCTS11-9	Cold tolerance at seedling stage	175.92	Wang et al., 2016
CS6	11	25591959	CS	qCTS11-10	Cold tolerance at seedling stage	159.94	Wang et al., 2016
InCS2	6	24039384	CS	qSWTCT6; qple-4	Cold tolerance at reproductive stage	2.8	Shakiba et al., 2017
InCS3	8	10358484	CS	qCTGERM8-1	Cold tolerance at germination stage	79.516	Shakiba et al., 2017
PGC1	9	14648157	PGC; PGCG	qCTS-9; qPGCG9-2; qLTSS9-2; Os09g0410300	Cold tolerance at seedling stage	248	Schläppi et al., 2017; Wang et al., 2016
InPGCG1	11	7105792	PGCG	qCTS11-3	Cold tolerance at seedling stage	143.48	Wang et al., 2016

Discussion

It has been a challenge to map loci controlling quantitative abiotic stress tolerance traits like cold tolerance due to the polygenic nature of the loci and they are correlated with the population structure (Shakiba, Edwards, et al., 2017). In our study, we used MLM method incorporating both population structure and kinship relatedness of our phenotypically and genotypically diverse rice accessions to discover the genetic control of cold tolerance. To reflect the response of germinating rice under cold stress scenarios, we assessed five different phenotypic assays, e.g. LTG, GI, CLC, PLR, and PLRR. The growth chamber LTG assay was performed to mimic the cold stress during rice seed germinating stage. GI was measured by dividing the germination rate of each accessions under cold stress by the germination rate under control condition. The separate GWAS analysis of LTG and GI helped us to discover if the chilling tolerance was due to the inherent cold tolerance ability or due to high seedling vigor. The PLR and PLRR assays helped us to determine if there is quantitative effect on subsequent growth and development of seedlings at warmer temperature due to extended exposure to cold temperature during seed germination stage because we noticed that some accessions with good LTG indices did not grow well after a temperature shift to 30°C and *vice versa*. For the direct seeding method, this situation might address a realistic scenario in which rice seeds may experience an extended period of cold during germination followed by warm, growth promoting temperatures. Interestingly, while many accessions developed a robust plumule, others showed only minor elongation of the coleoptile and stopped growth altogether during the 4-day assay period. PLRR is a useful index to select accessions with a superior ability to recover from cold temperature germination by adjusting their genetic pathway and metabolism to a growth promoting temperature.

To identify the chilling indices with different groups of our collection, we calculated mean and standard error for the five indices of each group. We found the highest values of all indices were in Texas elite breeding lines and US released varieties followed by TEJ. This shows that most of the inbred lines developed at the Texas A&M AgriLife Research Center-Beaumont generally have some good level of tolerance under cold stress during germination and very early seedling growth conditions. The lowest value of CLC was observed in aus subpopulation of rice. The lowest values of GI, PLR and PLRR were observed in *O. glaberrima* species depicting that the glaberrima species which are usually grown in Africa are not the good source of cold tolerant genes. However, our sample size representing glaberrima might be too small; so future research focusing on this species with a greater number of samples is needed to have more conclusive results. The other lines belonging to *O. hybridus* are found to have good GI, CLC, PLR and PLRR. The aromatic and aus species were found to have low tolerance to cold stress indicated by the low values of different chilling indices which were similar to the findings of other researchers (Schläppi, Jackson, et al., 2017, Shakiba, Edwards, et al., 2017). Interestingly, we didn't find significant difference of LTG between highly tolerant TEJ and indica species. Because LTG values were relatively similar between different subpopulations, both *INDICA* and *JAPONICA* subspecies are expected to have alleles contributing to superior LTG abilities. This also shows that there are many accesions of indica species which have good germination under cold stress. This is in agreement of the recent findings of Shakiba et al. (2017) where they had identified indica specific LTG QTL and have reported that both indica and japonica sub-species are expected to have alleles contributing to superior LTG abilities. But we observed significantly lower values of CS, PLR and PLRR of indica species than TEJ. These findings depict that although the indica species have good germination ability under cold stress condition, their growth rapidly gets retarded under cold

condition. The results conclude that Texas lines, TEJ, and TRJ of our study are highly tolerant to cold stress whereas aus, aromatic and indica lines are susceptible to cold condition. The phenotypic measurement of different chilling indices revealed that japonica sub-species were generally more tolerant than indica. This finding is consistent with previous findings (Cui, Peng, et al., 2002, Lv, Guo, et al., 2016, Morsy, Almutairi, et al., 2005). It could be because indica accessions are more adapted to low latitude regions while japonica accessions are adapted to low temperature in high latitude and higher elevations as a result of adaptation of sub-species to particular environment. The presence of differences in genetic architecture of cold tolerance among different sub-species and sub-populations analyzed in this study provides the opportunities for enhancing cold tolerance through molecular breeding.

To determine how the means of different chilling tolerance indices for each accession of our study were compared to each other, Spearman's correlation analysis was conducted. The results showed that all the indices were highly correlated with each other. Schlappi et al. (2017) had reported that low temperature germination (LTG) and plumule growth recovery rate (PLRR) were not or weakly correlated with other indices while PLR was highly correlated with other indices unlike our results where we saw high correlation between LTG with all other measured indices. This could be because of the difference in tolerance ability of different accessions used in our study. In this study, the indica species were found to have good germination under cold stress but had retarded coleoptile growth, poor plumule growth and slow plumule recovery. All other species had similar responses to all the measured indices, i.e. the lines having poor germination capacity had retarded growth and the ones with good germination ability had good coleoptile growth and plumule recovery under cold stress condition. An assessment of the chilling tolerance indices among different species indicates that different species have significantly different chilling

tolerance abilities. This reveals that different genetic mechanisms could be responsible for cold tolerance ability in different rice species. This suggests that GWAS mapping could be an efficient tool to uncover the species specific loci with either positive or negative effects on different cold tolerance indices.

The separate analysis of the chilling indices of indica species and japonica species shows similar results. In both japonica and indica species, we observed strong positive correlation of GI with CLC, PLR and PLRR. This shows that the rice lines with good germination ability under cold stress possess ability to grow well under prolonged cold stress. We observed strong correlation between PLR and PLRR ($r = 0.93$, $p < 0.0001$) indicating that there might be common genetic mechanisms underlying PLR and PLRR. The correlation was observed to be highly significant between CLC and PLR, and CLC and PLRR indicating that some common genetic mechanisms might be partly responsible for CLC, PLR and PLRR of japonica and indica sub-species.

The total of 10, 15, 10, 10, and 9 QTLs were found to be associated with LTG, GI, CLC, PLR and PLRR in the GWAS analysis of all accessions, respectively. We observed 6 SNPs (LTG1, LTG2, LTG3, LTG4, LTG5 and LTG7) associated with LTG appearing in GWAS results of GI as well. This is in agreement with the correlation analysis where highly significant correlation was observed between LTG and GI. This result also shows that these two traits may be partly controlled by some common genetic or molecular mechanisms. This result also shows that the association signals discovered from GI are in fact due to the tolerance of the accessions to cold germination and not due to the seedling vigor. We observed 3 SNPs (CS1, CS4 and CS5) associated with both CLC and PLR, this indicates that there may be similar genetic mechanism underlying coleoptile length growth at low temperature and plumule recovery after cold stress exposure. Four of the significant SNPs associated with PLR (PGC1, PGC4, PGC5 and PGC6) were found to be

associated with PLRR. This is in agreement of the correlation analysis of chilling indices PLR and PLRR (Table 7). This further suggests that LTG and GI, PLR and PLRR may share some common genetic mechanisms. Fine mapping and ultimately cloning of the responsible genes could be perform to confirm whether the overlapping QTLs associate with one or more genetic factors.

In a separate GWAS analysis of japonica sub-species, we identified 3, 3 and 3 QTLs associated with LTG, GI and CLC respectively. We observed two SNPs associated with LTG (JaLTG1 and JaLTG2) appearing in GWAS result of GI too. This is in agreement with its correlation analysis where highly significant correlation was observed between LTG and GI. This result also shows that there is similar genetic control of both LTG and GI. We also observed a significant loci, GI9 associated with LTG and GI detected in GWAS analysis of ALL accessions too. This indicate that this SNP might be coming from japonica species.

In a separate GWAS analysis of indica sub-species, we identified 3, 3, 4, 3 and 2 QTLs associated with LTG, GI, CLC, PLR and PLRR respectively. All the 3 SNPs (InLTG1, InLTG2 and LTG3) identified to be associated with LTG were found to be associated with GI too. This indicates that the strong phenotypic correlation between these two indices is due to the underlying similar genetic mechanisms. A significant SNP, PGC1 was found to be associated with both PLR and PLRR indicating the underlying similar genetic mechanisms controlling both indices. The SNP, LTG3 associated with LTG; GI and CS6 associated with CLC and PGC1 associated with both PLR and PLRR were detected in GWAS analysis of ALL accessions too. This shows that these SNPs are from indica subspecies.

Some of the QTLs identified from our GWAS study were located within or near the cold tolerance QTLs reported in previously studied bi-parental populations. We found several GWAS signals located within or near previously reported QTLs including 12 in all accessions (LTG3,

LTG5, LTG7, GI1, GI4, GI5, GI8, CS10, PGC3, PGC4, PGC6 and PGCG5) , 3 in japonica (JaCS1, CS2 and JaPGCG3) and 7 in indica (LTG3, CS5, CS6, InCS2, InCS3, PGC1 and InPGCG1) panel GWAS analyses. The known loci or genes that were identified to be collocated within the significant SNPs in our study were qCTGERM10-1, qCTGERM8-1, qCTS6-5, qCTS7-5, qLVG3, qSWTCT7, qCTS9-8, qSWTPNCT2-2, qCTS9-2, qCTS11-9, qCTS11-10, qSWTCT6, qCTS11-3, qPGCG9-2, qLTSS9-2, qCTS1-5, qLTSS9-1, qCTS11-2, Os06g0127100, Os03g0656900, Os03g0285700 and LOC_Os07g38620 (Fujino, Sekiguchi, et al., 2008, Li, Liu, et al., 2013, Satoh, Tezuka, et al., 2016) (Table 8). One of the interesting QTL is PGC1 (associated with PLR in ALL accessions and indica species), which is near *OsWRKY76*, previously shown to increase cold tolerance when overexpressed (Yokotani, Sato, et al., 2013), and 116.96 Kb away from Os09g24440, a putative subunit of TFIID recently shown to be the functional gene underlying *qCTS-9* and to enhance chilling cold tolerance when overexpressed (Zhao, Zhang, et al., 2017). Allelic differences at some of those genes may thus contribute to different PLRR indices. The other interesting QTL identified in our study is LTG3, associated with LTG and GI in ALL accessions and LTG in indica species which was identified to be located at 10.5 Kb away from OsDREB1C (Os06g0127100), previously reported to be involved in cold, salinity and drought stress tolerance when overexpressed (Ito, Katsura, et al., 2006).

The four major QTLs LTG3, CS5, CS6 and PGC1 were observed in both ALL and indica sub-species. The four significant SNPs (GI1, GI4, PGC3 and PGC6) discovered in all accessions, four SNPs in indica species (CS5, CS6, PGC1, In PGC1) and 3 SNPs (qCTS1-5, qLTS9-1, qCTS11-2) in japonica species were found to be in close LD with the QTL controlling for cold tolerance during seedling stage, qCTS6-5, qCTS7-5, qCTS9-8, qCTS9-2, qCTS11-9, qCTS11-10, qPGCG9-2, qCTS11-3, qCTS1-5, qLTSS9-1 and qCTS11-2 respectively (Wang, Liu, et al., 2016).

Several QTLs related to LTG (qLTG-3-1, qLTG-3-2, qLTG-7, qLTG-9, qLTG-11 and qLTG-12) have been previously mapped in the rice genome (Fujino, Sekiguchi, et al., 2008, Li, Liu, et al., 2013, Satoh, Tezuka, et al., 2016). In addition to validating our GWAS results, many of the identified QTLs near the previously mapped chilling tolerance related genes in rice help us to narrow down the QTL region and provide further support of location of the underlying genes. The most interesting region discovered are near LTG3, LTG5, and PGC1 which were found to be located very close to genes involved in cold stress tolerance (OsDREB1C, v1 and OSWRKY76 respectively.)

Conclusion

We evaluated 257 rice accessions in our rice GWAS panel to map QTLs associated with germination cold stress tolerance. The evaluation indicated that both temperate and tropical japonica rice cultivars are more tolerant in cold stress than indica, aromatic and aus rice. Our GWAS analysis of ALL accessions revealed that 41 QTLs are associated with cold tolerance traits, among them 28 are novel loci, GWAS analysis of japonica species unraveled 13 QTLs of which 7 were found to be novel loci and GWAS analysis of indica species identified 10 GWAS sites of which 3 were discovered to be unique loci. Our study has identified a set of highly tolerant rice cultivars that potentially can be used as donors, provided additional information on genetic structure of cold tolerance in rice, and identified some cold tolerance associated SNP markers for potential rice improvement and molecular studies.

References

Andaya, V. and D. Mackill. 2003. QTLs conferring cold tolerance at the booting stage of rice using recombinant inbred lines from a japonica× indica cross. Theoretical and Applied Genetics 106: 1084-1090.

Bai, X., H. Zhao, Y. Huang, W. Xie, Z. Han, B. Zhang, et al. 2016. Genome-wide association analysis reveals different genetic control in panicle architecture between indica and japonica rice. The plant genome 9.

Baruah, A.R., N. Ishigo-Oka, M. Adachi, Y. Oguma, Y. Tokizono, K. Onishi, et al. 2009. Cold tolerance at the early growth stage in wild and cultivated rice. Euphytica 165: 459-470.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Cruz, R.P.d. and S.C.K. Milach. 2004. Cold tolerance at the germination stage of rice: methods of evaluation and characterization of genotypes. Scientia Agricola 61: 1-8.

Cruz, R.P.d., R.A. Sperotto, D. Cargnelutti, J.M. Adamski, T. de FreitasTerra and J.P. Fett. 2013. Avoiding damage and achieving cold tolerance in rice plants. Food and Energy Security 2: 96-119.

Cui, K., S. Peng, Y. Xing, C. Xu, S. Yu and Q. Zhang. 2002. Molecular dissection of seedling-vigor and associated physiological traits in rice. Theoretical and Applied genetics 105: 745-753.

Earl, D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources 4: 359-361.

Faraway, J. 2002. Practical Regression and Anova using R. r-project.

Fujino, K., M. Obara, T. Shimizu, K.O. Koyanagi and T. Ikegaya. 2015. Genome-wide association mapping focusing on a rice population derived from rice breeding programs in a region. Breed Sci 65: 403-410. doi:10.1270/jsbbs.65.403.

Fujino, K., H. Sekiguchi, Y. Matsuda, K. Sugimoto, K. Ono and M. Yano. 2008. Molecular identification of a major quantitative trait locus, qLTG3–1, controlling low-temperature germinability in rice. Proceedings of the National Academy of Sciences 105: 12623-12628.

Ito, Y., K. Katsura, K. Maruyama, T. Taji, M. Kobayashi, M. Seki, et al. 2006. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. Plant and Cell Physiology 47: 141-153.

Iwata, N., H. Shinada, H. Kiuchi, T. Sato and K. Fujino. 2010. Mapping of QTLs controlling seedling establishment using a direct seeding method in rice. Breeding Science 60: 353-360.

Ji, S., L. Jiang, Y. Wang, W. Zhang, X. Liu, S. Liu, et al. 2009. Quantitative trait loci mapping and stability for low temperature germination ability of rice. Plant Breeding 128: 387-392.

Kaneda, C. 1974. Response of indica japonica rice hybrids to low temperature. SABRAO J. 6: 17-32.

Kim, S.-I., V.C. Andaya and T.H. Tai. 2011. Cold sensitivity in rice (Oryza sativa L.) is strongly correlated with a naturally occurring I99V mutation in the multifunctional glutathione transferase isoenzyme GSTZ2. Biochemical Journal 435: 373-380.

Lenth, R.V. 2016. Least-squares means: the R package lsmeans. Journal of statistical software 69: 1-33.

Li, L., X. Liu, K. Xie, Y. Wang, F. Liu, Q. Lin, et al. 2013. qLTG-9, a stable quantitative trait locus for low-temperature germination in rice (Oryza sativa L.). Theoretical and applied genetics 126: 2313-2322.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, et al. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics 28: 2397-2399.

Liu, W., T. Lu, Y. Li, X. Pan, Y. Duan, J. Min, et al. 2015. Mapping of quantitative trait loci for cold tolerance at the early seedling stage in landrace rice Xiang 743. Euphytica 201: 401-409.

Long-Zhi, H., Y.-Y. ZHANG, Q. Yong-Li, C. Gui-Lan, S.-Y. ZHANG, K. Jong-Hwan, et al. 2006. Genetic and QTL analysis for low-temperature vigor of germination in rice. Acta Genetica Sinica 33: 998-1006.

Lv, Y., Z. Guo, X. Li, H. Ye, X. Li and L. Xiong. 2016. New insights into the genetic basis of natural chilling and cold shock tolerance in rice by genome-wide association analysis. Plant, cell & environment 39: 556-570.

Ma, Y., X. Dai, Y. Xu, W. Luo, X. Zheng, D. Zeng, et al. 2015. COLD1 confers chilling tolerance in rice. Cell 160: 1209-1221.

Mackill, D.J. and X. Lei. 1997. Genetic variation for traits related to temperate adaptation of rice cultivars. Crop Science 37: 1340-1346.

Mao, D. and C. Chen. 2012. Colinearity and similar expression pattern of rice DREB1s reveal their functional conservation in the cold-responsive pathway. PloS one 7: e47275.

Morsy, M.R., A.M. Almutairi, J. Gibbons, S.J. Yun and G. Benildo. 2005. The OsLti6 genes encoding low-molecular-weight membrane proteins are differentially expressed in rice cultivars with contrasting sensitivity to low temperature. Gene 344: 171-180.

Pan, Y., H. Zhang, D. Zhang, J. Li, H. Xiong, J. Yu, et al. 2015. Genetic analysis of cold tolerance at the germination and booting stages in rice by association mapping. PLoS One 10: e0120590.

Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Raj, A., M. Stephens and J.K. Pritchard. 2014. Variational inference of population structure in large SNP datasets. Genetics: genetics. 114.164350.

Saito, K., Y. Hayano-Saito, M. Kuroki and Y. Sato. 2010. Map-based cloning of the rice cold tolerance gene Ctb1. Plant science 179: 97-102.

Sales, E., J. Viruel, C. Domingo and L. Marqués. 2017. Genome wide association analysis of cold tolerance at germination in temperate japonica rice (Oryza sativa L.) varieties. PloS one 12: e0183416.

Satoh, T., K. Tezuka, T. Kawamoto, S. Matsumoto, N. Satoh-Nagasawa, K. Ueda, et al. 2016. Identification of QTLs controlling low-temperature germination of the East European rice (Oryza sativa L.) variety Maratteli. Euphytica 207: 245-254.

Schläppi, M.R., A.K. Jackson, G.C. Eizenga, A. Wang, C. Chu, Y. Shi, et al. 2017. Assessment of five chilling tolerance traits and GWAS mapping in rice using the USDA Mini-Core collection. Frontiers in plant science 8: 957.

Shakiba, E., J.D. Edwards, F. Jodari, S.E. Duke, A.M. Baldo, P. Korniliev, et al. 2017. Genetic architecture of cold tolerance in rice (Oryza sativa) determined through high resolution genome-wide analysis. PloS one 12: e0172133.

Takesawa, T., M. Ito, H. Kanzaki, N. Kameya and I. Nakamura. 2002. Over-expression of ζ glutathione S-transferase in transgenic rice enhances germination and growth at low temperature. Molecular Breeding 9: 93-101.

Wang, D., J. Liu, C. Li, H. Kang, Y. Wang, X. Tan, et al. 2016. Genome-wide association mapping of cold tolerance genes at the seedling stage in rice. Rice 9: 61.

Wang, W., B. Vinocur and A. Altman. 2003. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta 218: 1-14.

Yokotani, N., Y. Sato, S. Tanabe, T. Chujo, T. Shimizu, K. Okada, et al. 2013. WRKY76 is a rice transcriptional repressor playing opposite roles in blast disease resistance and cold stress tolerance. Journal of experimental botany 64: 5085-5097.

Yonemaru, J.-i., T. Yamamoto, S. Fukuoka, Y. Uga, K. Hori and M. Yano. 2010. Q-TARO: QTL annotation rice online database. Rice 3: 194-203.

Yoshida, R., A. Kanno, T. Sato and T. Kameya. 1996. Cool-temperature-induced chlorosis in rice plants (I. Relationship between the induction and a disturbance of etioplast development). Plant physiology 110: 997-1005.

Yun, K.-Y., M.R. Park, B. Mohanty, V. Herath, F. Xu, R. Mauleon, et al. 2010. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. BMC plant biology 10: 16.

Zhang, F., X.-F. Ma, Y.-M. Gao, X.-B. Hao and Z.-K. Li. 2014. Genome-wide response to selection and genetic basis of cold tolerance in rice (Oryza sativa L.). BMC genetics 15: 55.

Zhang, M., J. Ye, Q. Xu, Y. Feng, X. Yuan, H. Yu, et al. 2018. Genome-wide association study of cold tolerance of Chinese indica rice varieties at the bud burst stage. Plant cell reports 37: 529-539.

Zhang, Q., Q. Chen, S. Wang, Y. Hong and Z. Wang. 2014. Rice and cold stress: methods for its evaluation and summary of cold tolerance-related quantitative trait loci. Rice 7: 24.

Zhang, S., J. Zheng, B. Liu, S. Peng, H. Leung, J. Zhao, et al. 2014. Identification of QTLs for cold tolerance at seedling stage in rice (Oryza sativa L.) using two distinct methods of cold treatment. Euphytica 195: 95-104.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature genetics 42: 355.

Zhao, J., S. Zhang, J. Dong, T. Yang, X. Mao, Q. Liu, et al. 2017. A novel functional gene associated with cold tolerance at the seedling stage in rice. Plant biotechnology journal 15: 1141-1148.

CHAPTER IV

TRANSCRIPTOMIC ANALYSIS OF TWO RICE GENOTYPES WITH CONTRASTING PHENOTYPES UNDER COLD STRESS DURING GERMINATION

Synopsis

Rice is one of the most important cereals consumed worldwide. Cold stress is one of the major abiotic factors affecting rice in difference stages of rice growth, including cold stress during germination. Fortunately, a wide array of phenotypic differences for chilling stress tolerance exist within the rice gene pool which may indicate the presence of difference molecular mechanisms underlying tolerance to this stress. Understanding these differences may assist in developing improved rice cultivars that have higher tolerance to cold stress. One of the most recent technologies, RNA-Seq, a high-throughput-sequencing based technology, allows for the identification of differential expressed genes among different cultivars under chilling stress. In our study, we conducted a comparative global gene expression analysis of two rice genotypes with contrasting germination phenotypes under cold stress. Differential gene expression analysis revealed 12,124 and 16,178 differentially regulated genes in cold tolerant and cold susceptible lines, respectively, with 7,954 common genes expressed between them. Gene ontology analysis showed major changes in the carbohydrate metabolic process, DNA metabolic process, photosynthesis and lipid metabolic process due to exposure to cold stress. Genes involved in signal transduction, phytohormones, anti-oxidant system and abiotic stress were identified. The genes identified in our study may be used for future gene validation and marker assisted selection for developing cultivars with increased chilling stress tolerance during germination.

Introduction

Rice is one of the most important cereal crops feeding more than half of the world's population. However rice yield is increasingly affected by different abiotic stresses, such as drought, salinity, heat, flooding, and cold stress. Chilling stress (nonfreezing temperature) is one of the major environmental stresses that affects rice growth, development and yield. Rice originated from tropical and sub-tropical regions. As a thermophilic crop, rice plant is highly sensitive to cold stress (Li, Dai, et al., 2006). Temperatures lower than 15C at the germination stage reduce germination rates, retard seedling vigor, delay seedling emergence and cause high seedling mortality (Fujino, Sekiguchi, et al., 2004). Low temperature at the reproductive stage during the beginning of pollen production affects meiotic and mitotic cell divisions in microspore development, ultimately resulting in pollen sterility (Shinada, Iwata, et al., 2013).

Several molecular mechanisms and metabolic processes are involved in the induction of plant acclimatization and tolerance to cold stress. Significant changes of molecular, physiological, and biochemical processes and changes in appearance may occur during the acclimatization process, including upregulation of anti-oxidant enzymes such as superoxide dismutase (SOD), catalase (CAT), ascorbate peroxidase (APX), synthesis and accumulation of osmo-protectants (polyamines and proline), sugars (sucrose, maltose, glucose, fructose) and proteins (Lissarre, Ohta, et al., 2010). Chilling stress causes changes in membrane fluidity and lipid composition which may be involved in sensing cold and inducing initial signal transduction events. The changes in the membrane and rearrangements in the cytoskeleton caused by cold stress are believed to modulate calcium channels and increases calcium content in the cytosol which is followed by induction of cold-responsive genes (COR) (Heidarvand and Amiri, 2010). Changes in the calcium content modulate phospholipase C (PLC) and D (PLD), which in turn regulate inositol 1,4,5-trisphosphate

(IP3) and phosphatidic acid respectively. Calcium influx is sensed by many protein families, including the calcium dependent protein kinases (CDPKs), calmodulins (CaMs) and Salt overly sensitive 3-like proteins (SOS3-like) thought to function in signaling cold. Besides calcium influx, the accumulation of reactive oxygen species (ROS) in the cytosol will be detected by other protein groups such as the mitogen activated protein kinases (MAPK, MAPKK and MAPKKK), C-repeat binding factor/ dehydration response element binding (CBF/DREB), inducer of CBF expression 1 (ICE1), MYB transcription factors and Zinc transporter proteins (ZAT) (Chinnusamy, Zhu, et al., 2010).

A comprehensive understanding of the molecular mechanisms of cold adaptation or resistance in rice is an important area of research. A number of studies have been performed to better understand cold stress responses in rice (Rabbani, Maruyama, et al., 2003, Shinozaki and Yamaguchi-Shinozaki, 2000, Takasaki, Maruyama, et al., 2010, Thomashow, 1999, Thomashow, 2001, Yun, Park, et al., 2010), and several transcriptomic studies have investigated the underlying mechanisms of cold tolerance in rice using microarray technologies (Lv, Bai, et al., 2010, Maruyama, Todaka, et al., 2011, Rabbani, Maruyama, et al., 2003, Takasaki, Maruyama, et al., 2010, Yun, Park, et al., 2010). These studies revealed that a large number of cold responsive genes can be categorized into three groups. The first group are the signaling components that regulate gene expressions in cold stress responses, which include protein kinases and transcription factors. The second group includes the functional components that directly protect plant cells against cold stress. These include enzymes in metabolic pathways, aquaporins and proteins involved in photosynthesis. The last group includes small noncoding RNAs such as microRNAs (miRNAs). Significant progress has been made using microarray technologies; however, this tool suffers from major limitations including high background levels due to cross-hybridization and the limited

dynamic range of detection due to both background and signal saturation (Okoniewski and Miller, 2006, Royce, Rozowsky, et al., 2007). A more recent technology, the large scale sequencing of RNA (RNA-Seq) provides a more efficient and relatively low-cost way of generating high throughput transcriptomics data such as tissue specific mRNA, new genes, SNPs and alternative splicing (Trapnell, Williams, et al., 2010).

In this study, we used the Illumina sequencing platform to sequence the transcriptomes of one cold- tolerant and one cold-sensitive rice genotypes to compare the transcriptome differences between the two contrasting lines to better understand the molecular mechanisms contributing to the phenotypic differences. The goals of this study were to evaluate the differentially expressed genes in the two contrasting rice cultivars in response to cold during germination and to identify the most promising candidate genes involved in cold stress tolerance in rice. The findings of our study may contribute to our understanding of cold response mechanisms in rice and assist in improving cold tolerance in rice in breeding programs.

Materials and methods

Plant materials and cold treatment

We previously screened 257 accessions in our genome-wide association study (GWAS) for identification of QTLs underlying cold stress tolerance in rice. Several phenotypic traits such as germination percentage and shoot length after 28 days of cold exposure (13°C), and plumule recovery rate after bringing the stressed germinating seeds back to normal temperature (30°C) for 4 days were measured. Two contrasting cultivars, a highly tolerant (Darij) and a very sensitive line (4610), were selected for the transcriptomic study.

For phenotypic evaluation, seeds were dehulled, surface sterilized with 70% ethanol for 3 minutes and rinsed five times with autoclaved distilled water to remove any surface contaminants.

Seeds were then germinated in moist filter paper placed inside petri dishes, with 30 seeds per replication. The experiment was conducted as a completely randomized design with 10 replications in a controlled environment. Only seeds presenting a coleoptile length at least half of the seed size were considered germinated (Cruz and Milach, 2004). A set of phenotypic traits including shoot length under normal growth condition (30°C), shoot length after 28 days of cold exposure (13°C), and plumule recovery rate after bringing the stressed seedlings back to normal temperature (30°C) were measured.

For preparing samples for the RNA extraction, the surface sterilized seeds were germinated in four replications in a growth chamber maintained at 13°C for 7 days in dark condition. For the control samples, both the contrasting lines were germinated for 2 days in a growth chamber maintained at 30°C in the dark. Germinating embryos from four biological replicates of each line grown under cold stress and normal conditions were collected and immediately frozen in liquid nitrogen. Samples were then maintained at -80°C until RNA extraction.

RNA extraction and library preparation

Total RNA was extracted using TRI Reagent (Invitrogen, MA, USA) and Plant RNA extraction kit (Qiagen). RNA quality and concentration were checked using a spectrophotometer (NanoDrop ND-1000) and its integrity was checked using 1% agarose gel electrophoresis. Poly-A RNA containing mRNA was purified using poly-T oligo-attached magnetic beads and then fragmented, and complementary DNA (cDNA) was synthesized using random hexamer primers, followed by purification, end-repairing, poly-A tailing, and adapter ligation. The libraries were prepared using the TruSEQ RNA Sample preparation v2 kit (Illumina), according to manufacturer's instructions. The quality of libraries was evaluated with an Agilent 2100 BioAnalyzer (Agilent technologies) using the Agilent DNA 1000 kit (Agilent). Library sequencing

was performed as pair-end 2*75 bp on an Hiseq2000 platform (Illumina) for the treatment samples and Novaseq (Illumina) was used to generate pair-end 2*150 bp for the control samples.

RNA-Seq analysis and detection of differentially expressed genes (DEGs)

The software FastQC was used for the analysis of read quality and its visualization. The low quality bases and library adapters were removed from each library using Trimmomatic ver. 0.36. Reads were mapped against the reference genome of *Oryza sativa* cv. Nipponbare (IRGSP build 1.0 RAP-DB). The mapping and alignment of the reads on the rice genome was done using HISAT2. Expression levels of each gene were quantified by normalizing total gene counts with the effective library size. The DESEQ2 package of Bioconductor was used to test for pairwise differential expression analysis. We used multi-factorial linear modeling and tested three null hypothesis of effects on gene expression: (1) whether cold treatment has significant effect on the expression of each gene; (2) whether genotype has significant effects on gene expression, and (3) whether gene expression was affected by cold treatment in a genotype-dependent manner. We fitted for models with our experimental data: (1) FM_{trt}: $Y = \tau + \varepsilon$; (2) FM_{geno}: $Y = \gamma + \varepsilon$; (3) FM_{add}: $Y = \tau + \gamma + \varepsilon$, and; (4) FM_{full}: $Y = \tau + \gamma + \tau{:}\gamma + \varepsilon$. In each model, Y is the expression value of each gene, τ is the effect of cold treatment, γ is the genotype effect, and ε is the random error. Comparing FM_{geno} and FM_{trt} to FM_{add} separately, we then tested whether the expression of each gene was regulated by cold, and whether there was a significant genotypic effect. Comparisons of FM_{full} and FM_{add} allowed us to test whether gene expression was affected by cold in a genotype-dependent manner. Genes with FDR p-values less than 0.05 were considered to be differentially expressed.

Gene ontology and enrichment analysis

We used the web based tool AGRIGO to perform GO enrichment analysis on differentially expressed genes (Du, Zhou, et al., 2010), while the Benjamini- Hochberg's false discovery rate (FDR) adjustment were used to control for multiple comparisons (Benjamini and Hochberg, 1995).

Results

Phenotypic variation

The average phenotypic values for the tolerant (Darij) and susceptible (4610) genotypes under control and cold conditions are shown in Table 19. All variables were statistically different by ANOVA ($P \leq 0.05$) and Tukey's test. The growth of the susceptible line was seen to be very vigorous under control conditions (30°C) but the growth was significantly reduced by cold stress.

Table 19 Phenotypic variation among two contrasting lines for cold stress

Treatment	**Tolerant line shoot length (cm)**	**Susceptible line shoot length (cm)**
Control (30C, 2DAS)		
Range	1.28 - 3	2.53 - 3.60
Average	2.48	3.00
CV	0.32	0.18

Table 19 (continued)

Treatment	Tolerant line shoot length (cm)	Susceptible line shoot length (cm)
Cold (13C, 21DAS)		
Range	0.35 - 0.60	0 - 0.1
Average	0.41	0.025
CV	0.16	1.81
Cold (13C, 28 DAS)		
Range	0.62 - 1.16	0.1 -0.2
Average	0.86	0.14
CV	0.19	0.29
Plumule recovery		
Range	2.05 - 3.57	0 - 1.12
Average	2.63	0.63
CV	0.17	0.47

Overview of cold-tolerant and cold-sensitive cDNA library sequencing

The two contrasting rice genotypes were used to identify differentially expressed genes in germinating seeds under cold treatment using the Illumina Platform-based RNA-seq. Four replications from two genotypes under both control and cold stress conditions were used. A total of sixteen RNA samples from control and cold treated materials were prepared to construct cDNA libraries with fragments of 75-150 bp in length. The cDNA libraries were then sequenced with the Illumina HiSeq2000 platform and the Novaseq platform. A total of 336.28 million clean reads

(91.9 % to 96.4 %) were generated and mapped to the Nipponbare reference genome with Hisat2. The detailed sequencing statistics for these cDNA libraries are presented in Table 20 and Table 21. The total gene counts were normalized using the DEseq2 package and the expression signals of each gene were calculated.

Differentially expressed genes for cold stress

A gene count matrix was used for extracting DEGs. The total number of DEGs in response to cold stress in tolerant and susceptible lines was found to be 12,124 and 16,178, respectively (Figure 9). The number of upregulated and downregulated DEGs under cold stress was found to be higher in the susceptible line than in the tolerant one. The number of upregulated genes in response to cold stress in tolerant and susceptible lines was 5579 and 7948, respectively (Figure 11), while the number of downregulated genes was 6542 and 8230, respectively (Figure 10). The Venn diagram highlights the overlapping DEGs expressed in both tolerant and susceptible lines under cold stress. Among the upregulated genes, 3,352 DEGs were commonly expressed in both, while among the downregulated ones, 4,134 DEGs were common.

Table 20 Summary of RNA seq data and sequence assembly control samples in NovaSEQ under cold stress

Parameters	T-spl #1	T-spl #2	T-spl #3	T-spl #4	S-spl #1	S-spl #2	S-spl#3	S-spl#4
Total reads	20362711	18083685	16736485	16170585	20888662	1.8E+07	23694743	18139624
aligned concordantly 0 times	2644232	2159764	2125647	1738368	2646276	2479763	3462627	2620915
aligned concordantly exactly 1 time	13977184	13028337	11866460	11886196	12950032	1.1E+07	14699545	10525109
aligned concordantly > 1 times	3741295	2895584	2744378	2546021	5292354	4344160	5532571	4993600
overall alignment	93.40%	94.37%	93.60%	94.79%	92.94%	92.32%	91.94%	91.97%

T-spl = Tolerant sample, S-spl = Susceptible sample

Table 21 Summary of RNA seq data and sequence assembly control samples in Hiseq4000 under cold stress

Parameters	T-spl #1	T-spl #2	T-spl #3	T-spl #4	S-spl #1	S-spl #2	S-spl#3	S-spl#4
	26697720	26341887	21596650	21507424	19906704	2.3E+07	22724476	21906681
aligned concordantly 0 times	1722596	1684399	1614216	1226206	2357162	2602197	2355298	2376885
aligned concordantly exactly 1 time	21767697	21093367	17114865	17875194	14477202	1.7E+07	16885005	16047429
aligned concordantly > 1 times	3207427	3564121	2867569	2406024	3072340	3246403	3484173	3482367
overall alignment	96.13%	95.80%	95.68%	96.35%	92.04%	93.19%	93.29%	93%

T-spl = Tolerant sample, S-spl = Susceptible sample

Figure 9 Venn diagram (Total DEGs) under cold stress

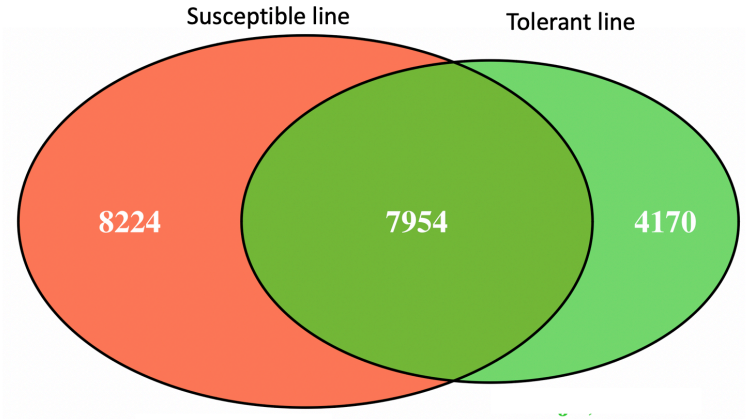
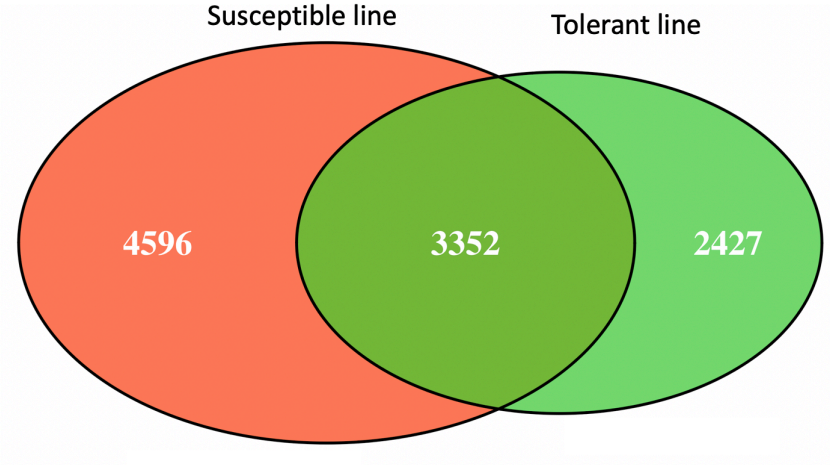


Figure 10 Venn diagram (Downregulated DEGs) under cold stress

Figure 11 Venn diagram (Upregulated DEGs) under cold stress



GO annotation and functional classification

GO enrichment analysis was performed to obtain functional information for the DEGs. A separate GO enrichment analysis was conducted for downregulated and upregulated DEGs. The DEGs were then assigned to three main categories in the GO database: Biological process, Molecular functions and Cellular components. A significantly enriched GO term was identified if the FDR corrected p value < 0.05.

GO of DEGs in the tolerant line compared to the susceptible line in cold stress vs normal condition

The GO enrichment analysis of upregulated genes in tolerant line due to interaction effect (genotype * environment) showed 15, 8 and 33 functional terms were identified for biological process, molecular functions and cellular components, respectively (Figure 12). For the biological process, the highest % of annotated genes were metabolic process (GO: 0008152) and cellular process (GO: 0009987); while those of for the cellular components were cell (GO: 0005623) and cell part (GO: 0044464).

The GO enrichment analysis of downregulated genes due to interaction effects revealed 13, 6 and 2 enriched functional terms for biological process, molecular functions and cellular components, respectively (Figure 13). The highest % of annotated genes in biological process were response to stimulus (GO: 0050896) and response to stress (GO: 0006950).

Figure 12 GO of upregulated DEGs in the tolerant line compared to the susceptible line in cold stress vs normal condition

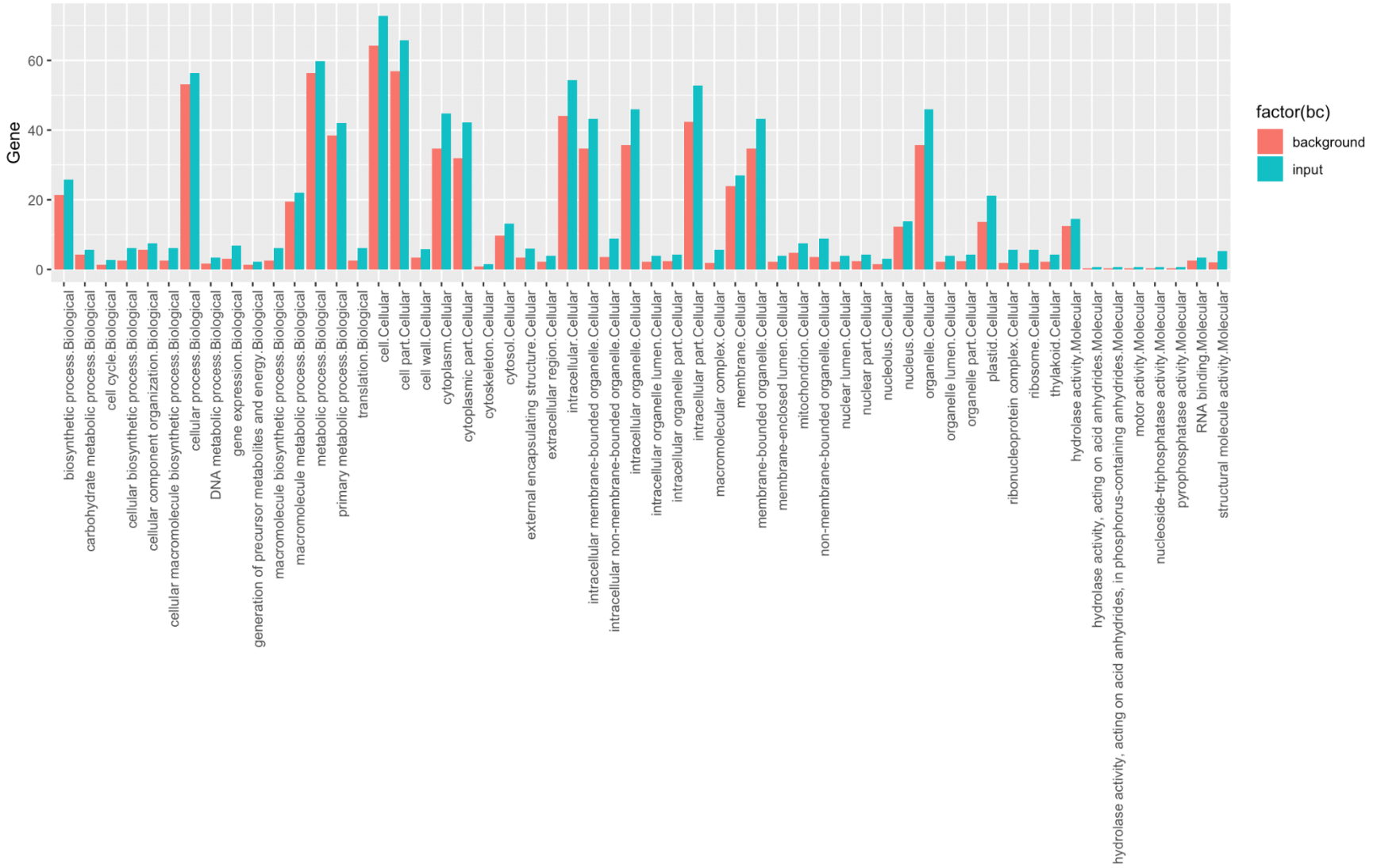
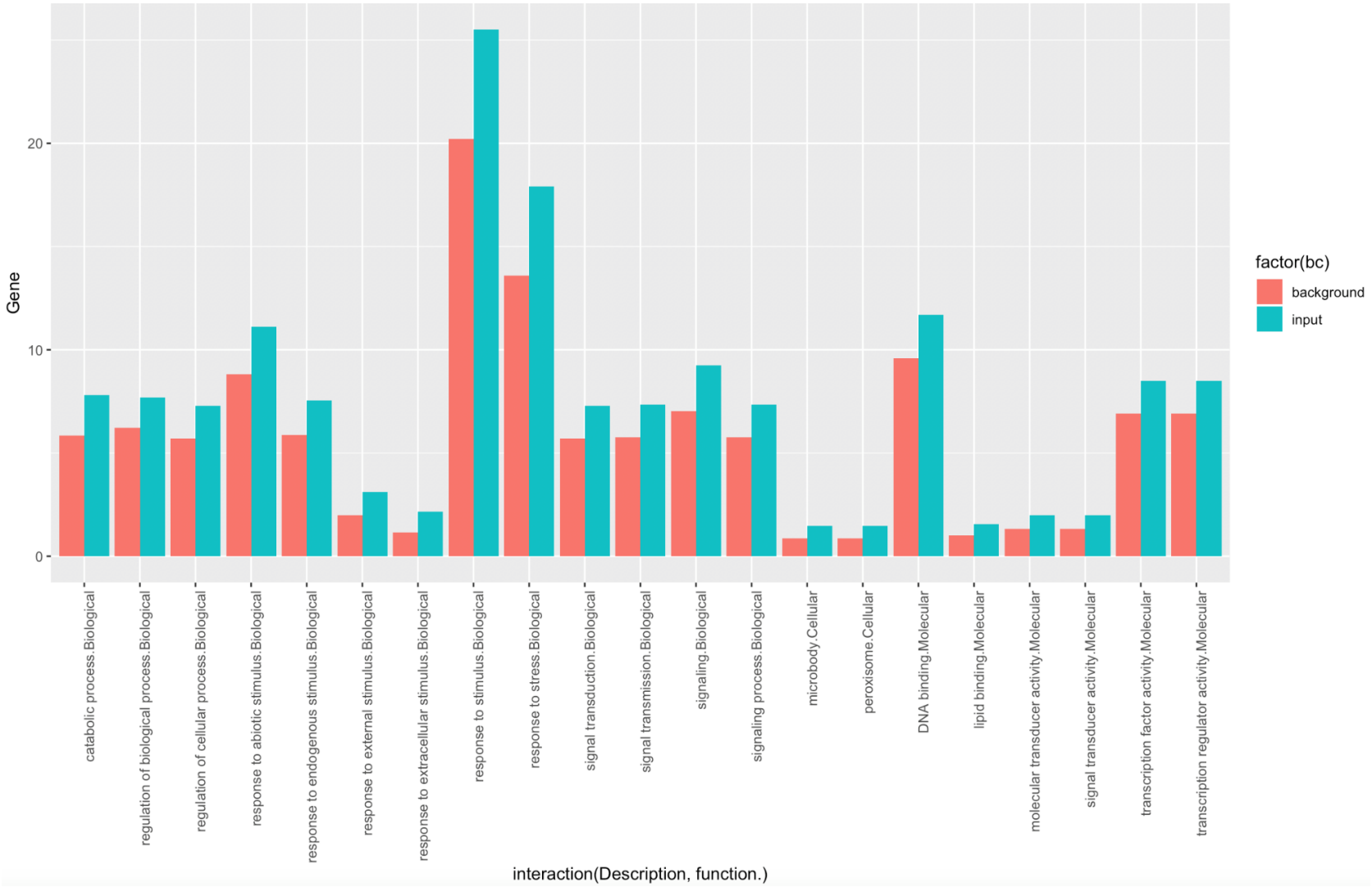


Figure 13 GO of downregulated DEGs in the tolerant line compared to the susceptible line in cold stress vs normal condition



GO of unique DEGs in tolerant genotype (cold stress vs control condition)

The GO enrichment analysis of tolerant genotype in cold stress condition compared to control condition showed 14, 3 and 29 functional terms for biological process, molecular functions and cellular components respectively (Figure 14). The highest % of annotated genes in biological functions were primary metabolic process (GO:0044238) and cellular metabolic process (GO:0044237); those in the cellular components were cell (GO:0005023) and cell part (GO:0044464); while in the molecular function, nucleic acid binding protein (GO:0003676) had the highest percentage of annotated genes.

Among all the significant GO found in tolerant line (cold vs control), 14, 8 and 2 GO of biological process, molecular function and cellular components respectively were observed in GO enrichment analysis of upregulated genes of tolerant line due to interaction effects. This result shows that the genes related to these common functional terms might be contributing genes in cold tolerance in tolerant lines.

The GO enrichment of downregulated DEGs revealed 5 and 4 functional terms for biological process and molecular function (Figure 15). No enriched GO was found for the cellular components. The enriched GO in biological process were response to stress (GO:0006950), response to stimulus (GO:0050896), response to extracellular stimulus (GO:0009991), secondary metabolic process (GO:0019748), response to external stimulus (GO:0009605). The 4 enriched GO in molecular functions were transferase activity (GO:0016740), catalytic activity (GO:0003824), nucleotide binding (GO:0000166), and binding (GO:0005488). The highest percentage of annotated genes was response to stimulus (GO: 0006950). In molecular function, the highest percentage of annotated genes were binding (GO:0005488) and catalytic activity (GO:0003824).

Figure 14 GO analysis of differentially expressed upregulated genes unique to cold-tolerant line (cold stress vs control condition)

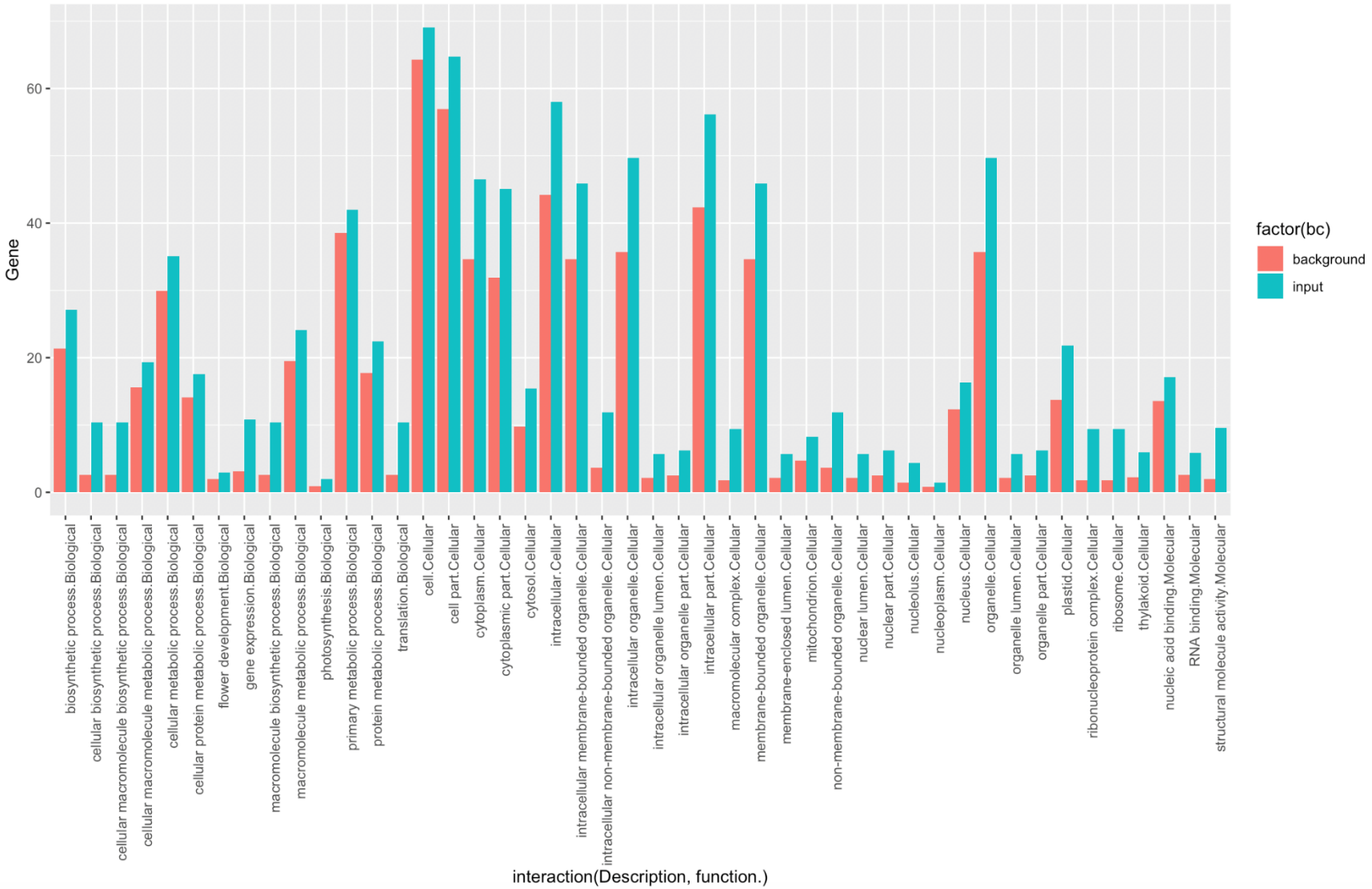
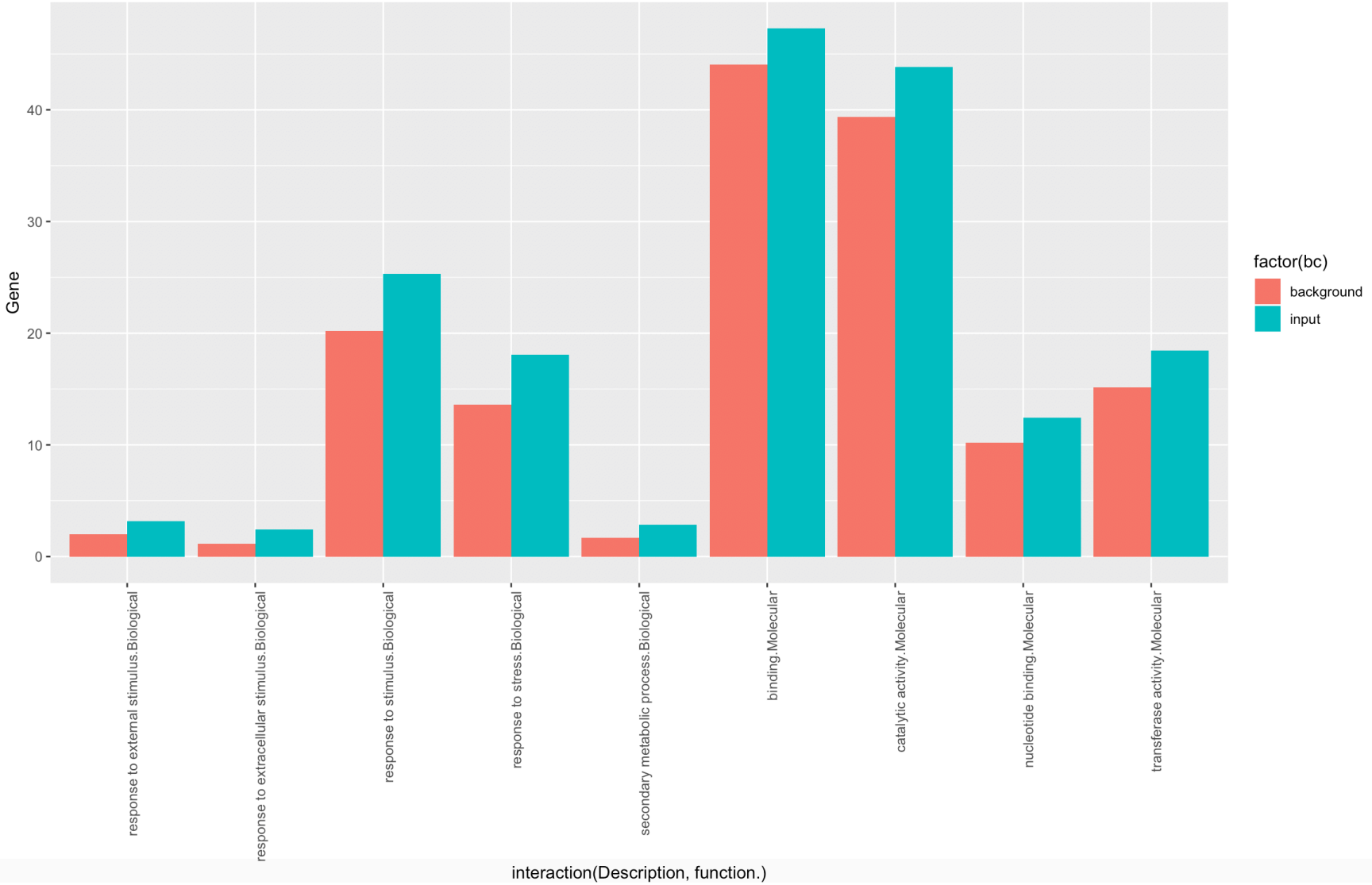


Figure 15 GO analysis of differentially expressed downregulated genes unique to cold-tolerant line (cold stress vs control condition)



GO of unique DEGs in susceptible genotype (cold stress vs control condition)

The number of DEGs upregulated in susceptible genotype was higher than in tolerant genotype. The GO enrichment showed only few functional terms associated with biological process, molecular functions and cellular components. We found 11 functional terms for biological process, 8 for molecular function and 2 for cellular components from upregulated genes (Figure 16). The functional terms enriched in susceptible line were found to be more related to signaling processes. The highest percentage of annotated genes were macromolecule modification (GO:0043412) and protein modification process (GO:0006464)

The GO enrichment analysis of downregulated DEGs showed 11 and 31 enriched functional terms of biological function and cellular components respectively (Figure 17). The highest % of annotated genes was metabolic process (GO: 0008152) among the biological process. The cell (GO:0005623) and cell part (GO:0044464) had the highest percentage of annotated genes in cellular components.

We detected a number of genes involved in carbohydrate metabolism downregulated by chilling stress in the susceptible line. These downregulated genes belong to glucan-endo -1,3-beta-glucosidae-related protein family, hexokinase, cellulose synthase, glycosyl hydrolase, beta galactosidase, malate dehydrogenase, UTP-glucose-1-phosphate, uridyl transferase, glucosyl hydrolase, malate synthase, callose synthase, galactosyl transferase etc. A number of genes related to DNA metabolic process were also downregulated in the chilling sensitive plant. Genes encoding RNA polymerase II transcription factor, DNA binding protein, chromosome condensation protein, and core histone H2A/H2B/H3/H4 domain containing protein were also downregulated in the sensitive cultivar. Histones are the basic proteins packaging DNA into nucleosomes, and histone gene expression is correlated with cell cycle and cell proliferation (de la Paz Sanchez, Caro, et al.,

2008). Histone genes have been reported to be repressed by abiotic stresses, including cold (Steward, Kusano, et al., 2000), drought, and salinity (Kang, Han, et al., 2011, Zhu, Dong, et al., 2012). The downregulation of these genes in the susceptible line under chilling stress suggests that core histone genes are involved in chilling tolerance. Among the histone genes, H2A.Z, encodes a particular histone variant has been previously identified to play an important role in temperature perception (Kumar and Wigge, 2010). Repression of H2A.Z under chilling stress implies that the regulation of this gene is temperature dependent (Zhang, Zhao, et al., 2012).

GO of common DEGs observed in both tolerant and susceptible genotypes (cold vs control)

The GO enrichment analysis of common upregulated DEGs showed that there were 3, 5 and 18 statistically enriched functional terms associated with biological process, molecular functions and cellular components, respectively.

From the GO enrichment analysis of common DEGs downregulated in both tolerant and susceptible lines, we found 6, 10 and 11 functional terms of biological process, molecular function and cellular components.

Figure 16 GO analysis of differentially expressed upregulated genes unique to cold sensitive line (cold stress vs control condition)

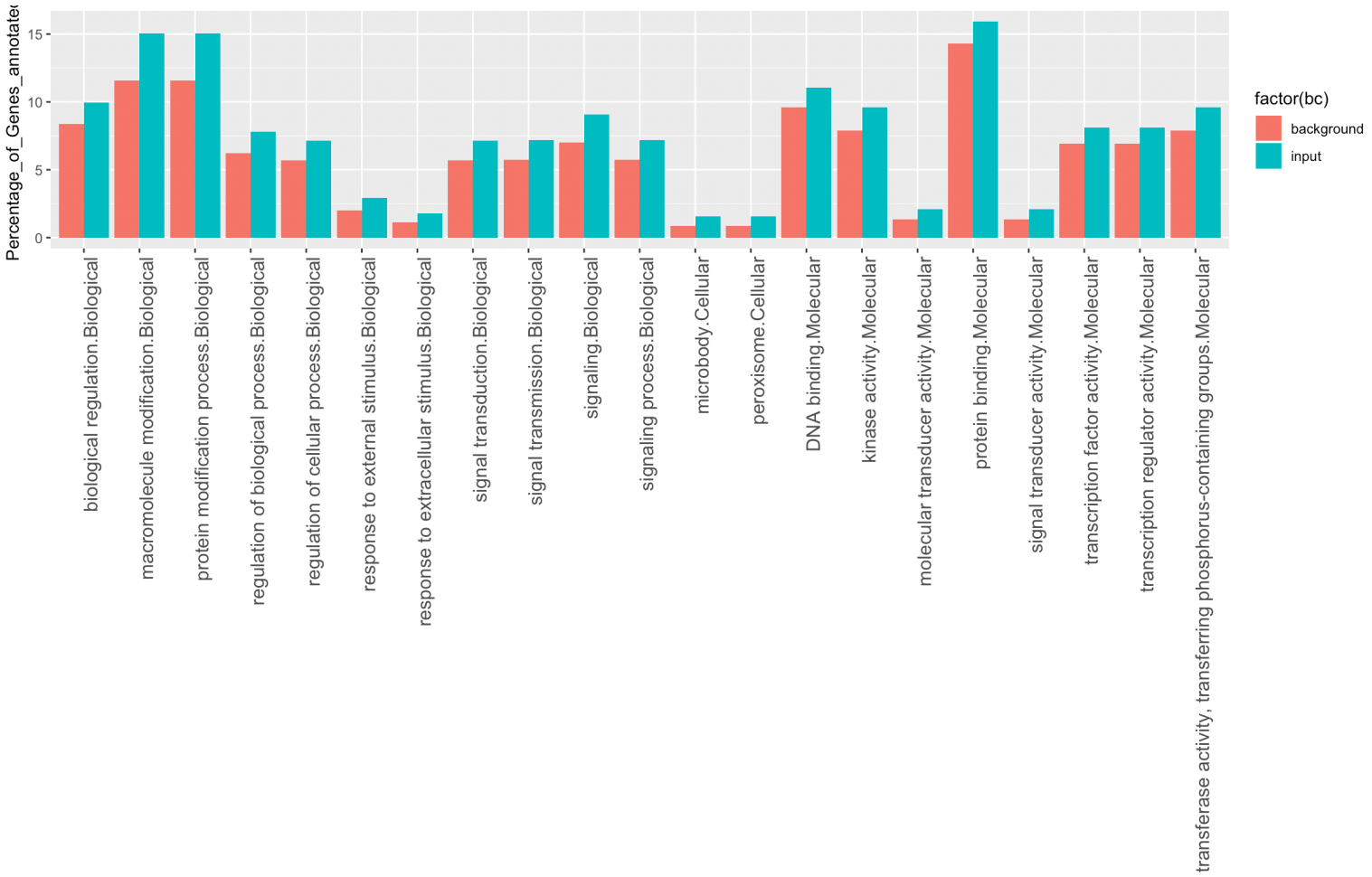
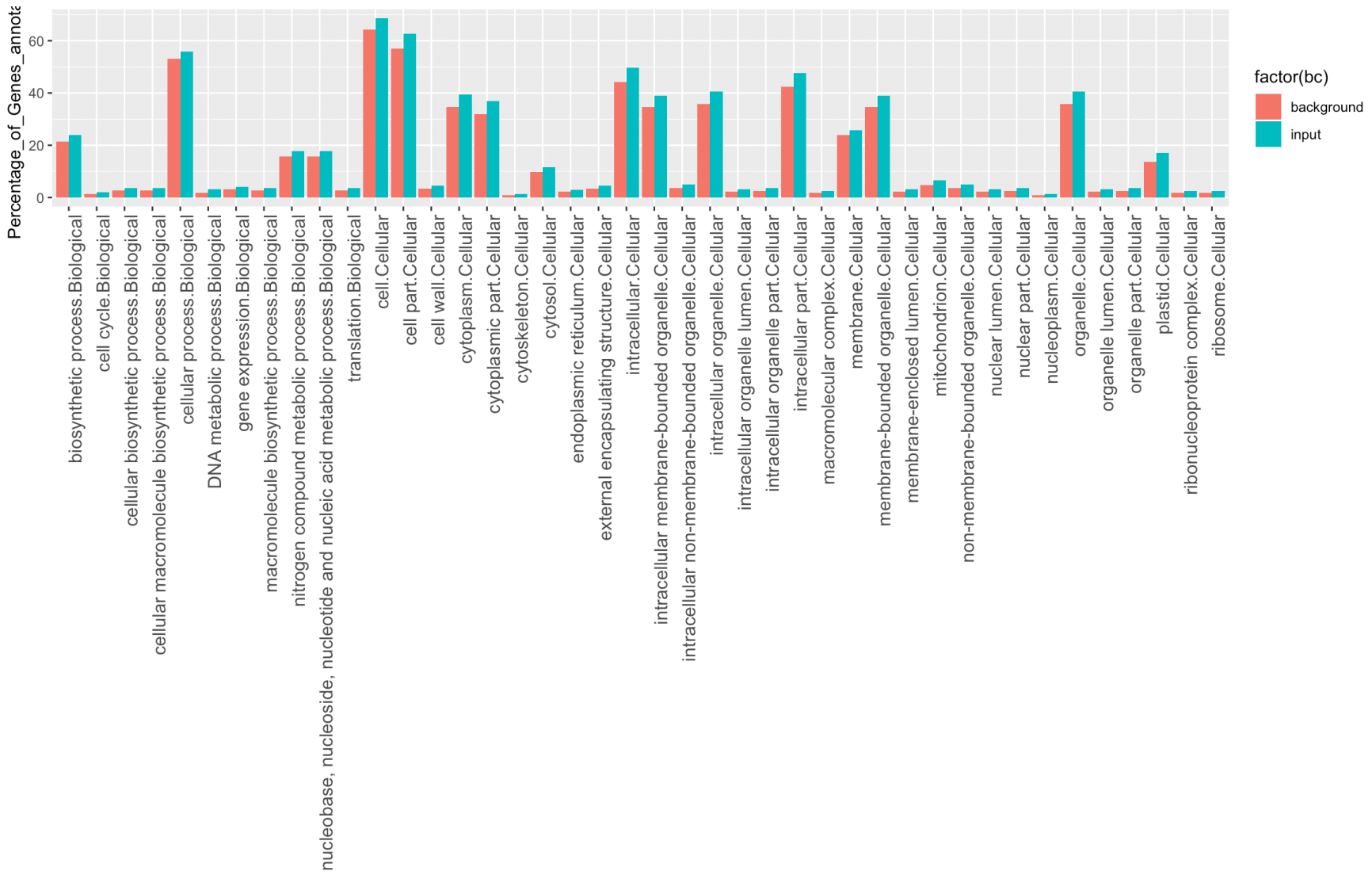


Figure 17 GO analysis of differentially expressed downregulated genes unique to cold sensitive line (cold stress vs control condition)



Cellular components play distinct roles in chilling stress response

It has been previously reported that certain cellular components, such as membrane and mitochondria are highly affected by chilling stress (Kaniuga, 2008, Prasad, Anderson, et al., 1994, Shu, Wang, et al., 2011). In our study, cellular localization annotations in combination with functional annotation were derived from the upregulated genes in both tolerant and susceptible genotypes. This was done using the singular enrichment analysis (SEA) tool available in agriGO (Du, Zhou, et al., 2010). GO enrichment analysis showed 10 GO and 26 GO were significantly enriched for cellular components in susceptible and tolerant lines, respectively. The common classes of GO among both tolerant and susceptible lines were intracellular (GO:0005622), cytoplasm (GO:0005737), organelle (GO:0043226), membrane-bounded organelle (GO:0043227), intracellular organelle (GO:0043229), intracellular membrane-bounded organelle (GO:0043231), intracellular part (GO:0044424), cytoplasmic part (GO:0044444), and cell part (GO:0044464). The percentage of annotated genes for these common functional terms among the tolerant and susceptible line were similar. The only GO that was found to be unique to the susceptible line was vacuole (GO:0005773) with a percentage of gene annotation of 6.09%. The GO terms that were unique to the tolerant lines were ribosome (GO:0005840), ribonucleoprotein complex (GO:0030529), macromolecular complex (GO:0032991), intracellular non-membrane-bounded organelle (GO:0043232), non-membrane-bounded organelle (GO:0043228), cytosol (GO:0005829), nuclear part (GO:0044428), intracellular organelle part (GO:0044446), organelle part (GO:0044422), nucleolus (GO:0005730), membrane-enclosed lumen (GO:0031974), nuclear lumen (GO:0031981), organelle lumen (GO:0043233), intracellular organelle lumen (GO:0070013), mitochondrion (GO:0005739), plastid (GO:0009536), and thylakoid (GO:0009579). The percentage of annotated genes in functional terms related to plastid,

mitochondria and cytosol were 15.55 %, 6.175 % and 12.69% respectively. The genes involved in these unique GO terms could be contributing agents for cold stress tolerance in the tolerant line.

Genes related to biological process are more differentially upregulated in the tolerant line compared to the susceptible line

The biological processes that were found to be significantly overrepresented in the tolerant line compared to susceptible (cold stress compared to control condition) by Panther GO-Slim biological process were lipoprotein biosynthetic process, fatty acid metabolic process and microtubule based process, with a fold enrichment of 14.26, 10.35 and 17.17, respectively. In lipoprotein biosynthetic process, four genes were upregulated in the tolerant line, S- acyl transferase (Os12g0263100, Os03g0275400), mannosyltransferase family protein (Os03g0670200), and GPI mannosyltransferase 2 (Os12g0498700) (Figure 18). The locus Os03g0670200 belonging to mannosyltransferase was found to be more upregulated in the tolerant line by 0.54 log fold than in the susceptible line. The loci Os12g0263100, Os12g0498700 and Os03g0275400 were found to be more upregulated in the tolerant line by 2.10, 0.73 and 0.99 log fold than the susceptible line.

In fatty acid metabolic process, 5 genes were found to be significantly upregulated in the tolerant line (Figure 19). The upregulated genes were acyl carrier protein (Os03g0352800), GNS1/SUR4 family (Os03g0701500), CBS domain-containing protein putative, expressed, (Os12g0169400), Chalcone-flavanone isomerase family protein (Os12g0115700), and AMP-binding enzyme family protein, expressed (Os11g0558300). These genes belong to transfer/carrier protein, acyltransferase, kinase modulator and ligase protein classes. The loci of Os03g0352800, Os03g0701500, Os12g0169400, Os12g0115700, Os11g0558300 were observed to be upregulated by 0.68, 1.56, 0.57, 2.56 and 1 log fold higher than in the susceptible lines, respectively.

In microtubule based process, 5 genes were found to be significantly overexpressed in the tolerant line compared to the susceptible line (Figure 20). The statistically significant upregulated genes were microtubule-associated EB1 family protein, putative, expressed (Os10g0498900), Kinesin-like protein KIN-7L (Os11g0552600), Kinesin-like protein KIN-14O (Os11g0648100), Kinesin-like protein KIN-5B (Os03g0279816) and Kinesin-like protein KIN-14E (Os03g0114000). These genes belong to the protein classes of non-motor microtubule binding proteins and microtubule binding motor proteins. The genes Os10g0498900, Os11g0552600, Os11g0648100, Os03g0279816, and Os03g0114000 were upregulated by 1.9, 1.13, 1.15, 1.59 and 1.36 log folds, respectively compared to the susceptible line.

The heatmap in Figure 18, Figure 19 and Figure 20 represents differentially expressed genes belonging to lipoprotein biosynthetic process, fatty acid metabolic process and microtubule based process in tolerant line compared to susceptible line respectively.

Figure 18 DEGs related to lipoprotein biosynthetic process

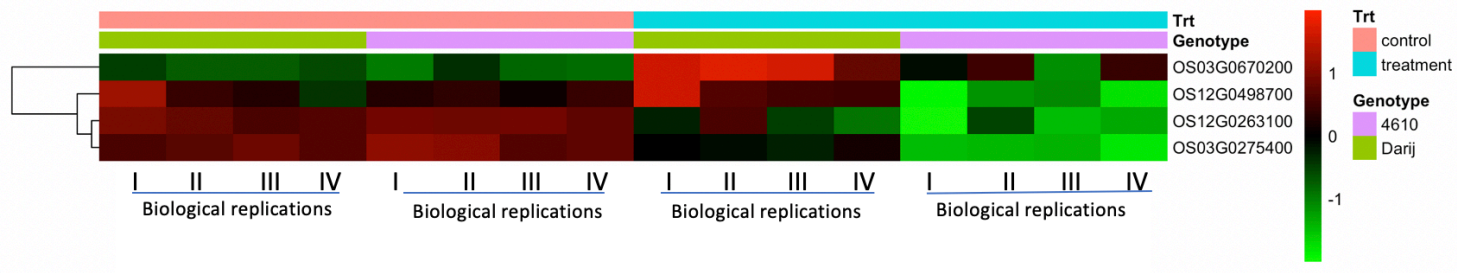


Figure 19 DEGs related to fatty acid metabolic process

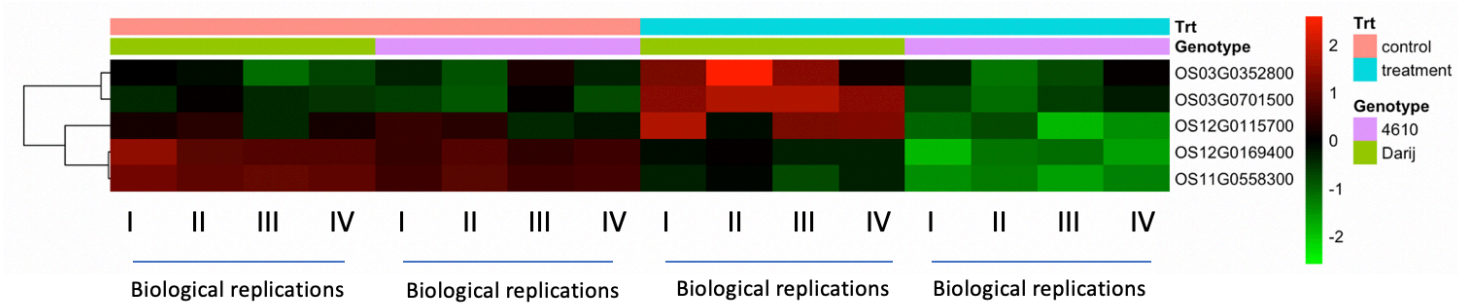
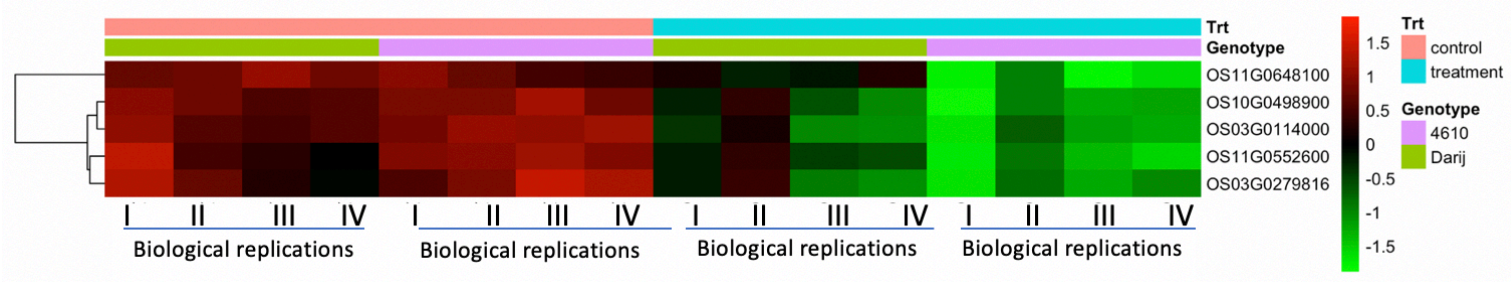


Figure 20 DEGs related to microtubule based process



The GO analysis of upregulated DEGs of tolerant and susceptible lines revealed four significant GOs terms related to molecular functions present in tolerant and susceptible lines. The GO terms of the tolerant line were structural molecule activity (GO:0005198), RNA binding (GO:0003723), nucleic acid binding (GO:0003676), translation factor activity, and nucleic acid binding (GO:0008135); whereas the GO terms of susceptible lines were nucleic acid binding (GO:0003676), transcription regulator activity (GO:0030528), transcription factor activity (GO:0003700), and DNA binding (GO:0003677). The common GO in both tolerant and susceptible lines was found to be nucleic acid binding (GO:0003676). The unique GOs in the tolerant line were structural molecule activity (GO:0005198), RNA binding (GO:0003723), nucleic acid binding (GO:0003676), and translation factor activity, nucleic acid binding (GO:0008135), while the unique GOs in the susceptible line were transcription regulator activity (GO:0030528), transcription factor activity (GO:0003700) and DNA binding (GO:0003677).

We found one heat shock protein 70 kDa protein and heat shock transcription factor upregulated in both genotypes. Previous studies have indicated the involvement of heat shock related genes in plants during abiotic stress response, especially to extreme environmental temperature (Chauhan, Khurana, et al., 2011). Upregulation of these heat shock protein and heat shock transcription factor suggests that they may play similar roles in rice at low temperature as in high temperature. Auxin plays an important role in plant response to cold stress (Rahman, 2013). We observed a number of auxin related genes encoding auxin responsive protein and auxin responsive factor differentially regulated in both genotypes. Additionally, we found two genes (LOC_Os03g04000, LOC_Os05g50140) encoding auxin responsive protein were upregulated in

the tolerant line; while two other loci (LOC_Os03g58350, LOC_Os08g40900) encoding auxin responsive protein were upregulated in the susceptible line. These results may indicate that interactions between auxin signaling and chilling stress response involve different mechanisms in different genotypes. Leucine rich repeat (LRR) proteins are part of a large gene family and play important and functionally diverse roles in plant growth and development (Jones and Jones, 1997). We identified genes belonging to this family, LOC_Os12g38450 and LOC_Os1208180, upregulated in the susceptible and tolerant genotypes, respectively. These results imply that LRR proteins may play an important role in cold stress response.

Genotype-dependent chilling-stress-responsive transcription factors

We found a number of transcription factors (TFs) were differentially regulated in both genotypes under chilling stress. The differentially expressed TF genes were categorized into different TF families, including acetyl transferase, AP2 domain containing protein, auxin family bindng protein, auxin response factor, DNA binding domain containing protein, bZIP transcription factor family, CCT motif family protein, ethylene response element binding protein, dehydration responsive protein, myb-like DNA binding domain, and WRKY DNA binding domain. The vast majority of the differentially expressed transcription factor genes were found to be upregulated. The AP2/EREB family comprises a large group of plant-specific transcription factors and is involved in abiotic stress response (Sharoni, Nuruzzaman, et al., 2010). In our study, we found a number of genes differentially regulated by chilling stress in both genotypes. Previous studies have demonstrated contributing effects of the DREB/CBF family in chilling stress tolerance in plants (Dubouzet, Sakuma, et al., 2003, Mizoi, Shinozaki, et al., 2012, Yun, Park, et al., 2010). More than half of the 103 WRKY TFs have been identified in rice genome and have been shown to be transcriptionally regulated under different abiotic and biotic stresses (Maruyama, Todaka, et al.,

2011, Matsukura, Mizoi, et al., 2010, Ramamoorthy, Jiang, et al., 2008). Yun et al. (2010) reported that the WRKY genes Os05g0343400, Os01g0246700, and Os01g0826400 were chilling induced genes in the japonica rice variety. Our study shows that there are other genes belonging to this family that were differentially regulated under cold stress in rice. Overexpression of OsMYB3R2 significantly improved cold tolerance by mediating the alteration in cell cycle and abnormal expression of stress genes in rice (Ma, Dai, et al., 2009). We identified 25 DEGs from darij and 37 DEGs from 4610 belonging to the WRKY gene family. The number of DEGs from darij and 4610 were 46 and 50 respectively from the MYB family (Figure 22 and Figure 24 respectively). The differentially regulated genes of the MYB family and the WRKY family identified in our study suggest that in depth study of these differentially regulated genes could identify candidate genes responsible for cold stress tolerance in rice.

Figure 21 DEGs of WRKY family in Darij

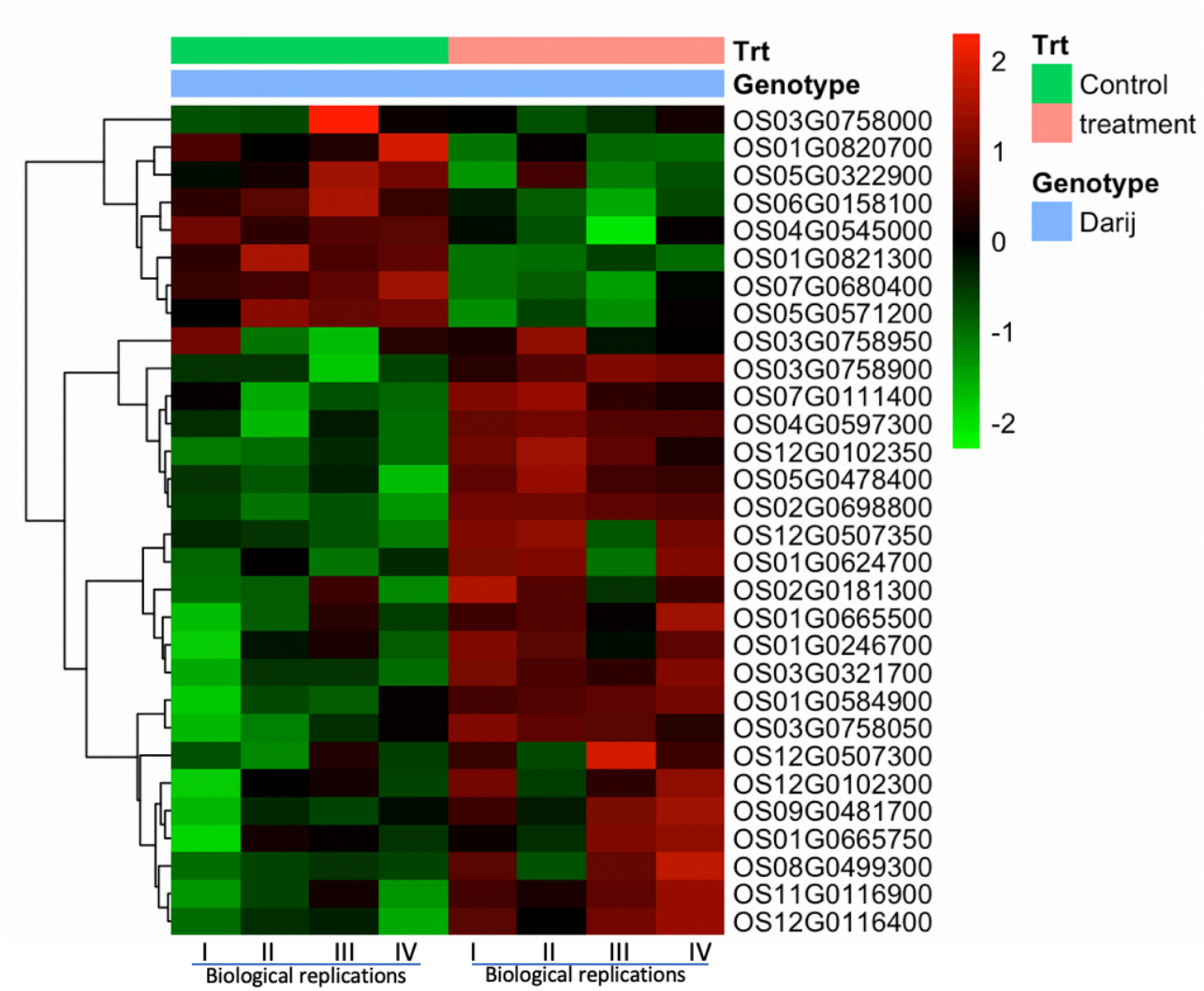


Figure 22 DEGs of MYB family in Darij

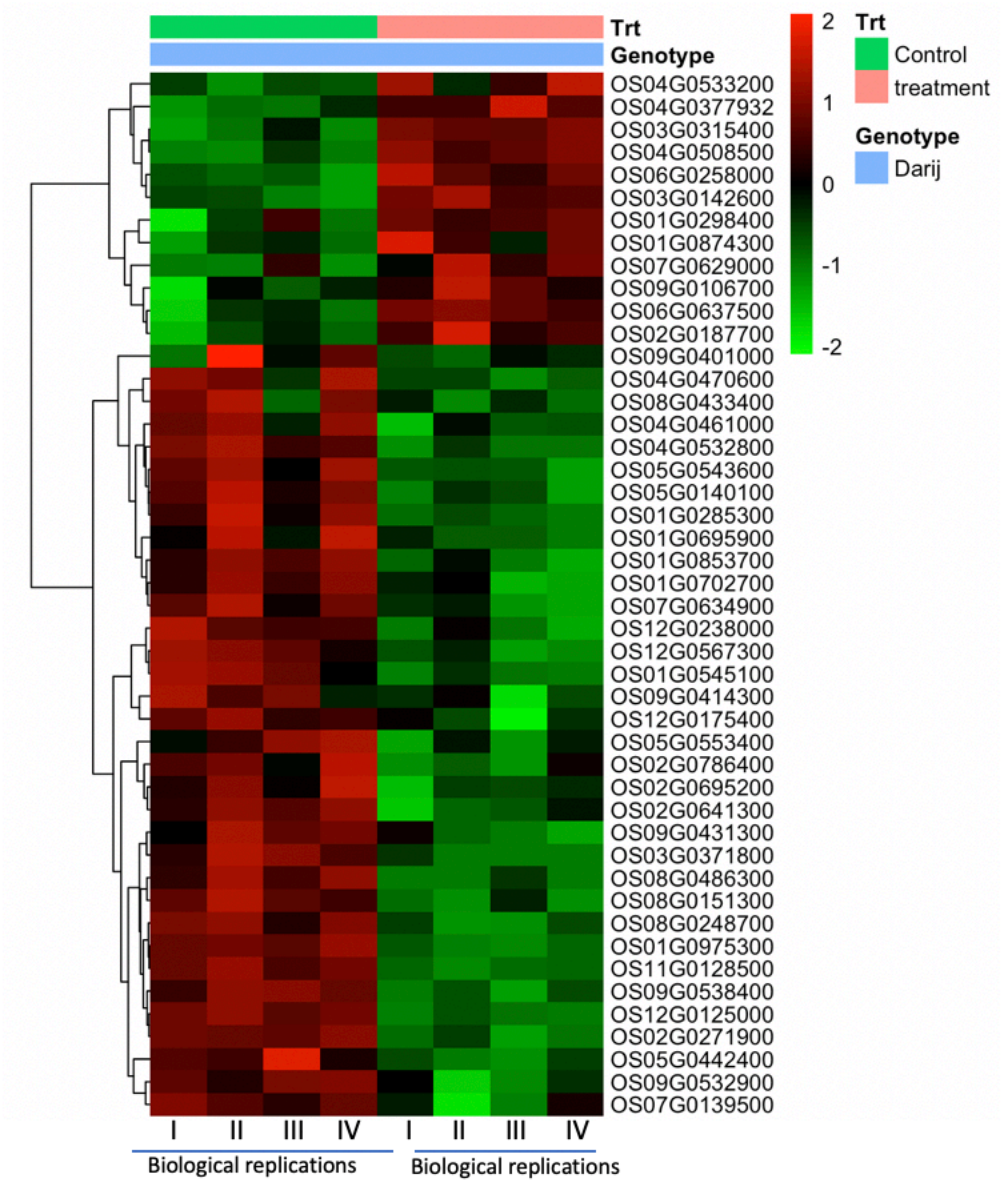


Figure 23 DEGs of WRKY family in 4610

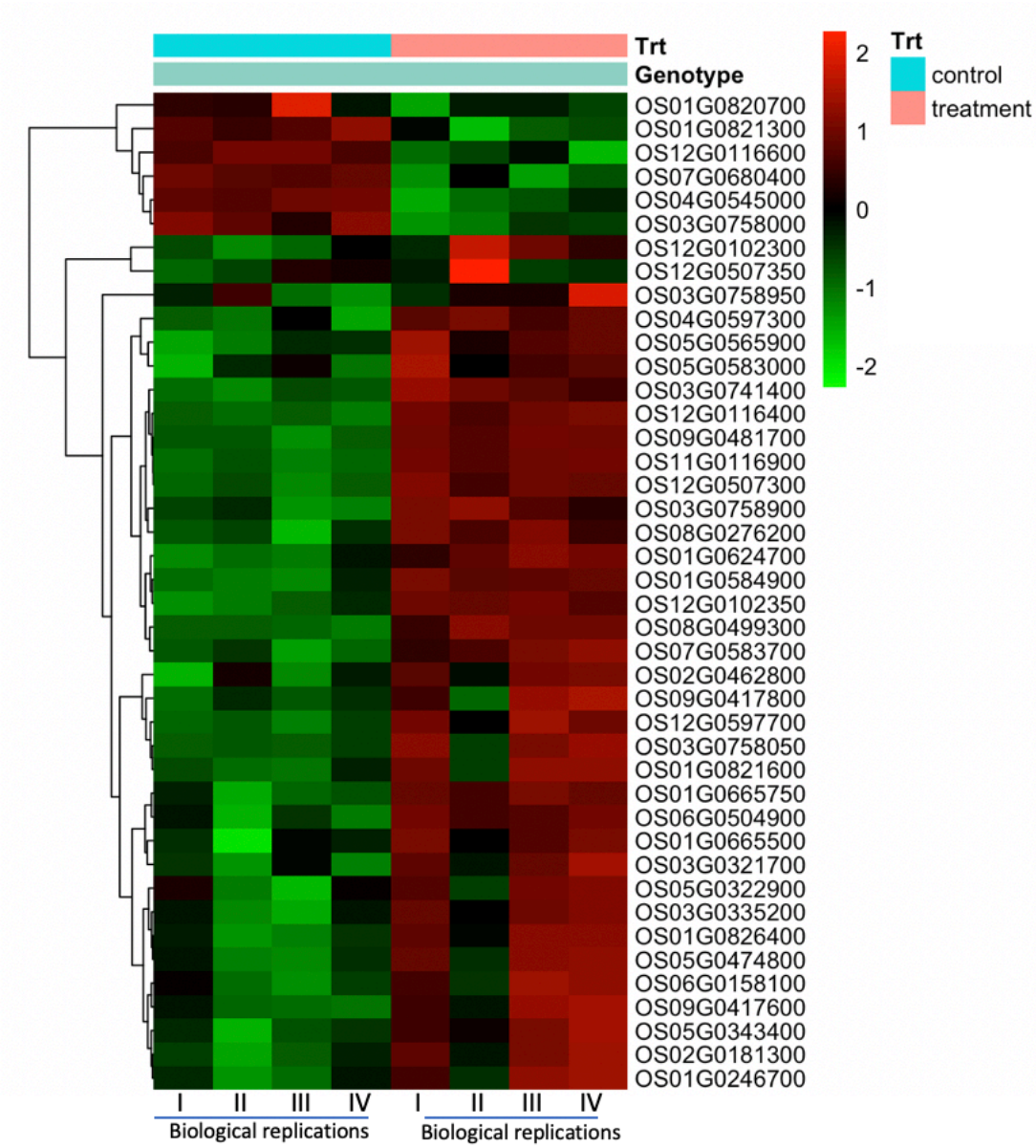
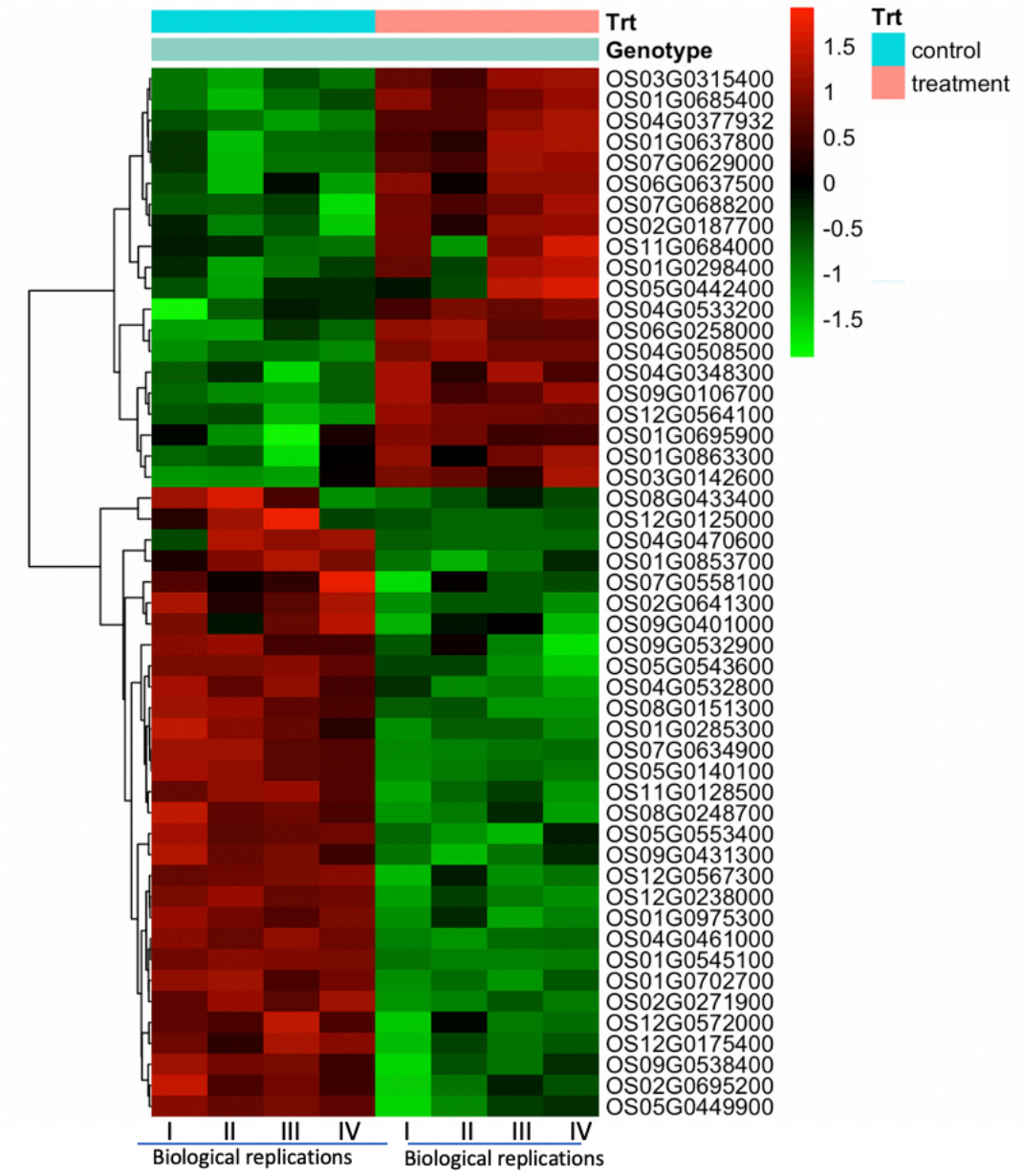


Figure 24 DEGs of MYB family in 4610



Discussion

Rice germination and growth is susceptible to different abiotic stresses that ultimately reduce grain yield. One of the key abiotic stresses detrimental to rice growth and yield is cold stress. A primary goal of rice breeders is to develop rice cultivars with increased chilling stress tolerance. To promote the development of improved cultivars with high chilling stress tolerance, it is crucial to understand the physiological and molecular responses of plants under stress. More importantly, the understanding of dissimilarities in these responses among rice cultivars is essential for genetic manipulation of crops. Previous studies have shown the presence of contrasting difference in molecular responses among the genotypes of common species under chilling stress (Batista-Santos, Lidon, et al., 2011, Ducruet, Roman, et al., 2005, Fortunato, Lidon, et al., 2010, Ikeda, Ohnishi, et al., 2009, Mai, Herbette, et al., 2010, Svensson, Crosatti, et al., 2006). Such differences have been previously identified in rice (Morsy, Almutairi, et al., 2005, Morsy, Jouve, et al., 2007). This study provides insights on different biological process and molecular function responses to chilling stress in the two different genotypes.

Phenotyping – cold effect on seedling germination and growth

All the variables (shoot length under cold exposure and plumule growth recovery) showed superior performance of the tolerant line. The tolerant line was superior for shoot length under cold condition, and seedling growth increased in the tolerant line 21 DAS as well. On the other hand, the seedling growth of the susceptible line was negligible. Likewise, the plumule growth recovery of the tolerant line was also much better than the susceptible one. The reduction of coleoptile growth under cold conditions during the initial stage of germination in rice is a limiting factor for good crop establishment (Cruz and Milach, 2004). The selection of genotypes with higher coleoptile length is key for cold tolerance during germination (Miura, Kuroki, et al., 2002).

Gene ontology analysis

In the broad GO annotation, we observed some similarities and some differences among the tolerant and susceptible lines. Among the upregulated genes, genes involved in nitrogen compound metabolic process, nucleic acid metabolic process and embryonic development in biological process classes were found to be the same in both tolerant and susceptible lines. In the tolerant line, the classes of GO related to cellular biosynthetic process, macromolecule biosynthetic process, gene expression, protein metabolic process, and photosynthesis were significantly enriched whereas in the susceptible line, the GO class of response to abiotic stimulus was significantly enriched. This shows that in the susceptible line, genes related to signal reception and transduction due to abiotic stress stimuli are highly expressed whereas in the tolerant line, genes related to the biological function affecting protein metabolism, cellular and macromolecule biosynthetic process are significantly overexpressed, possibly helping the plants to recuperate and adapt to the chilling stress.

In molecular function class, GO terms related to nucleic acid binding, RNA binding, transcription factor activity, translation factor activity, DNA binding were commonly enriched in both genotypes. However, genes contributing to structural molecular activity were significantly overrepresented in the tolerant line. GO terms belonging to the cellular components class like intracellular membrane-bound organelle, plastid and thylakoid were common in both genotypes, whereas GO terms related to mitochondria, cytosol, ribosome, ribonucleoprotein complex were significantly enriched in the tolerant line. This shows that genes related to mitochondria, cytosol and ribonucleoprotein complex may have some contributing effects on enhancing stress tolerance in the tolerant line. Chilling stress manifests in a decrease of photosynthesis, increase of reactive oxygen species (ROS), over-reduction of the chloroplast electron transport chain and damage to

membrane integrity which causes cellular dehydration, osmotic imbalance and other deleterious effects (Ruelland, Vaultier, et al., 2009, Thomashow, 1999). Our study shows that the tolerant line has more genes expressed for photosynthesis and maintenance of structural membrane integrity. Previous studies have shown that circadian rhythms influence photosynthetic processes and play a vital role in coordinating photosynthetic activity with the diurnal changes in light, and showed that chilling temperatures disrupt photosynthetic processes in warm-climate plants (Allen and Ort, 2001, Hennessey and Field, 1991, Hennessey, Freeden, et al., 1993, Mansfield, Carabasi, et al., 1973). Our results supports previous findings that some tolerant lines have developed responses to cope with the changes in membrane compositions to sustain photosynthesis under cold stress conditions.

The GO analysis of downregulated genes showed that the GO terms of biological class related to carbohydrate metabolic process, lipid metabolic process, transport, and localization were significantly enriched in both genotypes. In the tolerant line, GO classes of response to stress, response to stimulus and secondary metabolic process were significantly enriched, whereas in the susceptible line genes related to DNA metabolic process were significantly downregulated. This suggests that chilling stress causes detrimental effects on carbohydrate metabolism, lipid metabolism, ion transport and DNA metabolism. The internal disturbance in biological process might be the cause of physiological symptoms of wilting, yellowing and stunted growth of chilling induced crops. Genes related to hydrolase activity, catalytic activity, pyrophosphatase activity, nucleoside-tri-phosphatase activity, lipid binding were found to be downregulated in both genotypes. Among the GO terms of cellular components, genes related to cell wall, cell part, peroxisome, vacuole, external encapsulating structure, cytoskeleton, plasma membrane and extracellular structure were found to be significantly downregulated in both lines. Additionally,

genes related to mitochondria, endoplasmic reticulum, membrane bound organelle, and Golgi apparatus were found to be significantly downregulated in the susceptible line.

Hormone signals under cold stress

Many hormones including auxin, ABA, GA and ethylene are known to play an important role in plants during adaptive responses to abiotic stresses (Jain and Khurana, 2009, Peleg and Blumwald, 2011). In our study, the transcriptome data showed the expression of genes related to several plant hormones were altered by chilling stress. The role of auxin following the cold stress is still not yet clear but recent studies have shown that auxin response is involved in cold-stress. Several DEGs related to auxin response factor and auxin carrier, including LOC_Os12g04000, LOC_Os05g50140, LOC_Os06g48950, LOC_Os08g40900 and LOC_Os11g32110 were detected in our study. Hegedus et al. (2013) have reported the response of auxin related genes to various biotic and abiotic stresses such as pathogen infection, physical wounding, low temperature, and drought (Hegedus, Yu, et al., 2003). The phytohormone GA has also been associated with the response to abiotic stress (Colebrook, Thomas, et al., 2014). In our study, we identified genes LOC_Os10g38700 and LOC_Os03g41060 related to Gibberellin regulated protein, upregulated in tolerant and susceptible genotype, respectively.

ROS signals in cold response

Suzuki and Mittler, 2006 have reported the production of ROS like superoxide, hydroxyl radicals and hydrogen peroxide during cold stress (Suzuki and Mittler, 2006). ROS accumulation following abiotic stress has been identified to be toxic to cellular functions of plants (Jaspers and Kangasjärvi, 2010). The major cause of ROS accumulation is due to repressed photosynthesis (Noctor, Veljovic-Jovanovic, et al., 2002). Plants have developed various antioxidant systems to prevent lethal damage caused by ROS. Several enzymes like catalase, glutathione transferase,

superoxide dismutase have been characterized for their role in ROS scavenging. In our study we found several genes (LOC_Os10g38700, LOC_Os10g38720, LOC_Os03042240, LOC_Os11g37730) related to glutathione transferase upregulated in both genotypes. We found a gene, LOC_Os03g25300 related to peroxidase upregulated only in the tolerant genotype.

Transcription factor (TF)

The TFs related to cold stress and the genes regulated by them are vital for cold stress tolerance in plants. The TFs and genes which regulate enzymes, regulatory proteins and metabolites comprise a complex regulatory network in plants (Cushman and Bohnert, 2000). In our study, we identified DEGs from GNAT, AP2-EREBP, AUX/IAA, ARF, ABI3VP1, bZIP, CCAAT, C2C2-CO-like, HB, G2-like, MYB, NAC, Alfin-like, WRKY and Zinc finger families. Many genes regulated by different transcription factors were previously reported to be induced under cold stress. Kanneganti and Gupta (2008) reported the expression of 0siSAP8 (Os06g0612800) under cold stress (Kanneganti and Gupta, 2008). Kong et al. (2006) have identified the expression of OsDOS (Os01g0192000) under drought stress. In our study, we found 0s01g0192000 which encodes a zinc finger protein upregulated in both genotypes under cold stress. This shows that this gene may be involved not only in drought stress, but also in cold stress.

A few genes belonging to the MYB family have been previously reported to be induced under cold stress (Agarwal, Hao, et al., 2006, Guan, Xu, et al., 2019). OsMYBS3 represses the DREB1/CBF-dependent cold signaling pathway in rice (Guan, Xu, et al., 2019). These MYBs might be involved in a additional pathways that regulate cold adaptation in rice (Suh, Jeung, et al., 2010). In our study, we found several differentially regulated genes of the MYB family induced under cold stress in both genotypes. This suggests that MYB family members may play an important role in the regulatory pathway of cold tolerance in rice. The P2/EREB family contains

a large number of plant specific TFs and participates in abiotic stress responses in plants (Sharoni, Nuruzzaman, et al., 2010). We were able to identify many genes belonging to the AP2/EREB family differentially regulated under cold stress.

Many WRKY genes have been found to be transcriptionally regulated in response to abiotic stress treatments in rice (Berri, Abbruscato, et al., 2009, Ramamoorthy, Jiang, et al., 2008). Yun et al. (2010) reported three DEGs of WRKY genes (*Os03g0758900*, *Os11g0490900* and *Os01g0826400)* regulated by cold or H_2O_2 stress in rice (Yun, Park, et al., 2010). In our study, we found some genes of the WRKY family upregulated and downregulated in both genotypes. Additionally, heat shock proteins have been reported to be involved in heat shock responsiveness in plants (Kotak, Larkindale, et al., 2007). In our study, we found a few genes related to heat shock factor and heat shock protein differentially regulated. But the role of heat shock-related proteins in cold stress response of plants still remains obscure.

Many bZIP TFs have been discovered to function in plant stress signaling. In our study, we also found many genes of the bZIP family differentially regulated in both tolerant and susceptible lines.

NAC TF genes have been previously reported to encode a set of plant specific TFs involved in responses to biotic and abiotic stress (Lin, Zhao, et al., 2007, Nakashima, Tran, et al., 2007, Zheng, Chen, et al., 2009). The expression of NAC genes was found to be upregulated during cold stress (Zhang, Zhao, et al., 2012), but in our study we found both upregulated and downregulated NAC genes in both genotypes. This shows NAC genes have diverse roles in cold stress response in plants. Many genes related to bHLH protein and bHLH transcription factor have also been discovered to be differentially regulated in our study. ICE1, a bHLH protein in Arabidiopsis, has been reported to specifically bind to the CBF3 promoter region and promote the expression of

CBF3 which leads to improved tolerance to cold stress (Chinnusamy, Ohta, et al., 2003). Yun et al. (2010) have reported differential regulation of Os030741100, a bHLH gene induced by cold stress.

In this study, we identified more DEGs related to abiotic stress tolerance mechanism in the cold stress tolerant line than in the susceptible line. Functional analysis of the most promising candidate genes identified for cold tolerance in our study may be performed by both gain and loss of function analysis in future studies.

Conclusion

The differential gene expression analysis of our study revealed a genetic difference between two contrasting rice lines under chilling stress. The results indicated that many genes involved in signal transduction, phytohormone (auxin, GA), abiotic stress, anti-oxidant system and unknown function genes are differentially regulated under cold stress. Our findings have increased the understanding of molecular and physiological basis of cold response in rice. The results of our study may also be useful for identification of candidate genes underlying QTLs contributing to cold tolerance. Together with the GWAS study and transcriptomics study, development of cold tolerance-associated markers for MAS would be done in the near future.

References

Agarwal, M., Y. Hao, A. Kapoor, C.-H. Dong, H. Fujii, X. Zheng, et al. 2006. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. Journal of Biological Chemistry 281: 37636-37645.

Allen, D.J. and D.R. Ort. 2001. Impacts of chilling temperatures on photosynthesis in warm-climate plants. Trends in plant science 6: 36-42.

Batista-Santos, P., F. Lidon, A. Fortunato, A. Leitão, E. Lopes, F. Partelli, et al. 2011. The impact of cold on photosynthesis in genotypes of Coffea spp.—photosystem sensitivity, photoprotective mechanisms and gene expression. Journal of plant physiology 168: 792-806.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57: 289-300.

Berri, S., P. Abbruscato, O. Faivre-Rampant, A.C. Brasileiro, I. Fumasoni, K. Satoh, et al. 2009. Characterization of WRKY co-regulatory networks in rice and Arabidopsis. BMC plant biology 9: 120.

Chauhan, H., N. Khurana, P. Agarwal and P. Khurana. 2011. Heat shock factors in rice (Oryza sativa L.): genome-wide expression analysis during reproductive development and abiotic stress. Molecular Genetics and Genomics 286: 171.

Chinnusamy, V., M. Ohta, S. Kanrar, B.-h. Lee, X. Hong, M. Agarwal, et al. 2003. ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. Genes & development 17: 1043-1054.

Chinnusamy, V., J.-K. Zhu and R. Sunkar. 2010. Gene regulation during cold stress acclimation in plants. Plant Stress Tolerance. Springer. p. 39-55.

Colebrook, E.H., S.G. Thomas, A.L. Phillips and P. Hedden. 2014. The role of gibberellin signalling in plant responses to abiotic stress. Journal of experimental biology 217: 67-75.

Cruz, R.P.d. and S.C.K. Milach. 2004. Cold tolerance at the germination stage of rice: methods of evaluation and characterization of genotypes. Scientia Agricola 61: 1-8.

Cushman, J.C. and H.J. Bohnert. 2000. Genomic approaches to plant stress tolerance. Current opinion in plant biology 3: 117-124.

2008. Chromatin dynamics during the plant cell cycle. Seminars in cell & developmental biology, Elsevier.

Du, Z., X. Zhou, Y. Ling, Z. Zhang and Z. Su. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research 38: W64-W70.

Dubouzet, J.G., Y. Sakuma, Y. Ito, M. Kasuga, E.G. Dubouzet, S. Miura, et al. 2003. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt-and cold-responsive gene expression. The Plant Journal 33: 751-763.

Ducruet, J.-M., M. Roman, M. Havaux, T. Janda and A. Gallais. 2005. Cyclic electron flow around PSI monitored by afterglow luminescence in leaves of maize inbred lines (Zea mays L.): correlation with chilling tolerance. Planta 221: 567-579.

Fortunato, A.S., F.C. Lidon, P. Batista-Santos, A.E. Leitão, I.P. Pais, A.I. Ribeiro, et al. 2010. Biochemical and molecular characterization of the antioxidative system of Coffea sp. under cold conditions in genotypes with contrasting tolerance. Journal of Plant Physiology 167: 333-342.

Fujino, K., H. Sekiguchi, T. Sato, H. Kiuchi, Y. Nonoue, Y. Takeuchi, et al. 2004. Mapping of quantitative trait loci controlling low-temperature germinability in rice (Oryza sativa L.). Theoretical and applied genetics 108: 794-799.

Guan, S., Q. Xu, D. Ma, W. Zhang, Z. Xu, M. Zhao, et al. 2019. Transcriptomics profiling in response to cold stress in cultivated rice and weedy rice. Gene 685: 96-105.

Hegedus, D., M. Yu, D. Baldwin, M. Gruber, A. Sharpe, I. Parkin, et al. 2003. Molecular characterization of Brassicanapus NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant molecular biology 53: 383-397.

Heidarvand, L. and R.M. Amiri. 2010. What happens in plant molecular responses to cold stress? Acta Physiologiae Plantarum 32: 419-431.

Hennessey, T.L. and C.B. Field. 1991. Circadian rhythms in photosynthesis: oscillations in carbon assimilation and stomatal conductance under constant conditions. Plant Physiology 96: 831-836.

Hennessey, T.L., A.L. Freeden and C.B. Field. 1993. Environmental effects on circadian rhythms in photosynthesis and stomatal opening. Planta 189: 369-376.

Ikeda, T., S. Ohnishi, M. Senda, T. Miyoshi, M. Ishimoto, K. Kitamura, et al. 2009. A novel major quantitative trait locus controlling seed development at low temperature in soybean (Glycine max). Theoretical and applied genetics 118: 1477-1488.

Jain, M. and J.P. Khurana. 2009. Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. The FEBS journal 276: 3148-3162.

Jaspers, P. and J. Kangasjärvi. 2010. Reactive oxygen species in abiotic stress signaling. Physiologia Plantarum 138: 405-413.

Jones, D.A. and J.D. Jones. 1997. The role of leucine-rich repeat proteins in plant defences. Advances in botanical research. Elsevier. p. 89-167.

Kang, Y., Y. Han, I. Torres-Jerez, M. Wang, Y. Tang, M. Monteros, et al. 2011. System responses to long-term drought and re-watering of two contrasting alfalfa varieties. The Plant Journal 68: 871-889.

Kaniuga, Z. 2008. Chilling response of plants: importance of galactolipase, free fatty acids and free radicals. Plant Biology 10: 171-184.

Kanneganti, V. and A.K. Gupta. 2008. Overexpression of OsiSAP8, a member of stress associated protein (SAP) gene family of rice confers tolerance to salt, drought and cold stress in transgenic tobacco and rice. Plant molecular biology 66: 445-462.

Kotak, S., J. Larkindale, U. Lee, P. von Koskull-Döring, E. Vierling and K.-D. Scharf. 2007. Complexity of the heat stress response in plants. Current opinion in plant biology 10: 310-316.

Kumar, S.V. and P.A. Wigge. 2010. H2A. Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell 140: 136-147.

Li, X., C. Dai, R. Cheng, T. Chen and D. Jiao. 2006. Identification for cold tolerance at different growth stages in rice ({\sl Oryza sativa} L.) and physiological mechanism of differential cold tolerance. Zuo wu xue bao 32: 76-83.

Lin, R., W. Zhao, X. Meng, M. Wang and Y. Peng. 2007. Rice gene OsNAC19 encodes a novel NAC-domain transcription factor and responds to infection by Magnaporthe grisea. Plant science 172: 120-130.

Lissarre, M., M. Ohta, A. Sato and K. Miura. 2010. Cold-responsive gene regulation during cold acclimation in plants. Plant signaling & behavior 5: 948-952.

Lv, D.-K., X. Bai, Y. Li, X.-D. Ding, Y. Ge, H. Cai, et al. 2010. Profiling of cold-stress-responsive miRNAs in rice by microarrays. Gene 459: 39-47.

Ma, Q., X. Dai, Y. Xu, J. Guo, Y. Liu, N. Chen, et al. 2009. Enhanced tolerance to chilling stress in OsMYB3R-2 transgenic rice is mediated by alteration in cell cycle and ectopic expression of stress genes. Plant physiology 150: 244-256.

Mai, J., S. Herbette, M. Vandame, E. Cavaloc, J.-L. Julien, T. Ameglio, et al. 2010. Contrasting strategies to cope with chilling stress among clones of a tropical tree, Hevea brasiliensis. Tree physiology 30: 1391-1402.

Mansfield, C., R. Carabasi, W. Wells and K. Borman. 1973. Circadian rhythm in the skin temperature of normal and cancerous breasts. International journal of chronobiology 1: 235.

Maruyama, K., D. Todaka, J. Mizoi, T. Yoshida, S. Kidokoro, S. Matsukura, et al. 2011. Identification of cis-acting promoter elements in cold-and dehydration-induced transcriptional pathways in Arabidopsis, rice, and soybean. DNA research 19: 37-49.

Matsukura, S., J. Mizoi, T. Yoshida, D. Todaka, Y. Ito, K. Maruyama, et al. 2010. Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Molecular Genetics and Genomics 283: 185-196.

Miura, K., M. Kuroki, H. Shimizu and I. Ando. 2002. Introduction of the long-coleoptile trait to improve the establishment of direct-seeded rice in submerged fields in cool climates. Plant production science 5: 219-223.

Mizoi, J., K. Shinozaki and K. Yamaguchi-Shinozaki. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1819: 86-96.

Morsy, M.R., A.M. Almutairi, J. Gibbons, S.J. Yun and G. Benildo. 2005. The OsLti6 genes encoding low-molecular-weight membrane proteins are differentially expressed in rice cultivars with contrasting sensitivity to low temperature. Gene 344: 171-180.

Morsy, M.R., L. Jouve, J.-F. Hausman, L. Hoffmann and J.M. Stewart. 2007. Alteration of oxidative and carbohydrate metabolism under abiotic stress in two rice (Oryza sativa L.) genotypes contrasting in chilling tolerance. Journal of plant physiology 164: 157-167.

Nakashima, K., L.S.P. Tran, D. Van Nguyen, M. Fujita, K. Maruyama, D. Todaka, et al. 2007. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. The Plant Journal 51: 617-630.

Noctor, G., S. Veljovic-Jovanovic, S. Driscoll, L. Novitskaya and C.H. Foyer. 2002. Drought and oxidative load in the leaves of C3 plants: a predominant role for photorespiration? Annals of Botany 89: 841-850.

Okoniewski, M.J. and C.J. Miller. 2006. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC bioinformatics 7: 276.

Peleg, Z. and E. Blumwald. 2011. Hormone balance and abiotic stress tolerance in crop plants. Current opinion in plant biology 14: 290-295.

Prasad, T.K., M.D. Anderson and C.R. Stewart. 1994. Acclimation, hydrogen peroxide, and abscisic acid protect mitochondria against irreversible chilling injury in maize seedlings. Plant Physiology 105: 619-627.

Rabbani, M.A., K. Maruyama, H. Abe, M.A. Khan, K. Katsura, Y. Ito, et al. 2003. Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant physiology 133: 1755-1767.

Rahman, A. 2013. Auxin: a regulator of cold stress response. Physiologia plantarum 147: 28-35.

Ramamoorthy, R., S.-Y. Jiang, N. Kumar, P.N. Venkatesh and S. Ramachandran. 2008. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. Plant and cell physiology 49: 865-879.

Royce, T.E., J.S. Rozowsky and M.B. Gerstein. 2007. Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic acids research 35: e99.

Ruelland, E., M.-N. Vaultier, A. Zachowski and V. Hurry. 2009. Cold signalling and cold acclimation in plants. Advances in botanical research 49: 35-150.

Sharoni, A.M., M. Nuruzzaman, K. Satoh, T. Shimizu, H. Kondoh, T. Sasaya, et al. 2010. Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. Plant and cell physiology 52: 344-360.

Shinada, H., N. Iwata, T. Sato and K. Fujino. 2013. Genetical and morphological characterization of cold tolerance at fertilization stage in rice. Breeding science 63: 197-204.

Shinozaki, K. and K. Yamaguchi-Shinozaki. 2000. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Current opinion in plant biology 3: 217-223.

Shu, D.-F., L.-Y. Wang, M. Duan, Y.-S. Deng and Q.-W. Meng. 2011. Antisense-mediated depletion of tomato chloroplast glutathione reductase enhances susceptibility to chilling stress. Plant Physiology and Biochemistry 49: 1228-1237.

Steward, N., T. Kusano and H. Sano. 2000. Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic acids research 28: 3250-3259.

Suh, J., J. Jeung, J. Lee, Y. Choi, J. Yea, P. Virk, et al. 2010. Identification and analysis of QTLs controlling cold tolerance at the reproductive stage and validation of effective QTLs in cold-tolerant genotypes of rice (Oryza sativa L.). Theoretical and Applied Genetics 120: 985-995.

Suzuki, N. and R. Mittler. 2006. Reactive oxygen species and temperature stresses: a delicate balance between signaling and destruction. Physiologia plantarum 126: 45-51.

Svensson, J.T., C. Crosatti, C. Campoli, R. Bassi, A.M. Stanca, T.J. Close, et al. 2006. Transcriptome analysis of cold acclimation in barley Albina and Xantha mutants. Plant physiology 141: 257-270.

Takasaki, H., K. Maruyama, S. Kidokoro, Y. Ito, Y. Fujita, K. Shinozaki, et al. 2010. The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. Molecular Genetics and Genomics 284: 173-183.

Thomashow, M.F. 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annual review of plant biology 50: 571-599.

Thomashow, M.F. 2001. So what's new in the field of plant cold acclimation? Lots! Plant physiology 125: 89-93.

Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. Van Baren, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28: 511.

Yun, K.-Y., M.R. Park, B. Mohanty, V. Herath, F. Xu, R. Mauleon, et al. 2010. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. BMC plant biology 10: 16.

Zhang, T., X. Zhao, W. Wang, Y. Pan, L. Huang, X. Liu, et al. 2012. Comparative transcriptome profiling of chilling stress responsiveness in two contrasting rice genotypes. PloS one 7: e43274.

Zheng, X., B. Chen, G. Lu and B. Han. 2009. Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. Biochemical and biophysical research communications 379: 985-989.

Zhu, Y., A. Dong and W.-H. Shen. 2012. Histone variants and chromatin assembly in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1819: 343-348.

CHAPTER V

TRANSCRIPTOMIC ANALYSIS OF TWO RICE GENOTYPES WITH CONTRASTING PHENOTYPE UNDER HYPOXIA STRESS DURING GERMINATION

Synopsis

Among all the cultivated cereal crops, rice possess unique ability to thrive under flooding conditions at different stages of development. Rice can germinate under water, but the rate of coleoptile elongation varies among different rice accessions. The tolerant rice cultivars exhibit faster coleoptile elongation rate which enable them to timely reach the water surface and to survive under the stress conditions. However, the basic understanding of underlying molecular mechanisms, cellular components and biological functions enhancing flooding tolerance during germination in rice still remains unclear. In this study, we conducted RNA-Seq analysis from 4 days old coleoptile tissue of germinating seeds of two contrasting lines germinated under control and hypoxia stress conditions. We identified that genes involved in photosynthesis, lipid metabolism, glycolysis, anaerobic fermentation, nitrogen metabolism and nucleotide binding were differentially regulated within and between genotypes. A few genes belonging to ABC transporters, AP2 domain containing proteins, AP2-like ethylene-responsive transcription factors, basic helix-loop-helix, bZIP transcription factor domain containing proteins, helix-loop-helix DNA-binding domain containing proteins, HSF-type DNA-binding domain containing proteins, MYB family transcription factors, no apical meristem proteins, pentatricopeptides, phospholipid-transporting ATPase 1, RNA recognition motif containing proteins, WRKY and C2H2 zinc finger proteins were uniquely expressed only in the tolerant line. The findings of our study will help in identification of promising candidate genes underlying flooding tolerance during germination in

rice and would help in rice breeding program to develop improve tolerant rice cultivars under direct-seeded ecosystems.

Introduction

Rice is one of the major food crop of the world and consumed by more than half of the world population. It is a staple crop grown in South and South-east Asia where majority of the rice growing are low-land and flood prone (Ismail, 2013). Flash flood due to unpredicted rainfall is a common phenomenon in such areas which severely affects rice crops in different stages of their development. Rice is adapted to aquatic ecology, therefore it possesses a unique mechanism of germinating and developing coleoptile under water (Magneschi and Perata, 2008). In general, rice coleoptile under water exhibits growth rate of $1mmhr^{-1}$ (Vijayan, Senapati, et al., 2018). The coleoptile elongation is achieved by rapid elongation of basal cells (up to 200 um in 12 h) immediately after emerging from embryo (Narsai, Edwards, et al., 2015). However, not all rice accessions possess vigorous growth under water. The germination potential varies greatly among different rice genotypes under flooded condition, and a few genotypes perform better than others under hypoxic stress conditions.

Germination in rice takes place into three phases: rapid water uptake phase (I), plateau of water uptake phase (II), and a further increase of water uptake phase or the initiation of growth phase (He and Yang, 2013, Narsai, Wang, et al., 2013). During phase I and phase II, major metabolic reprogramming occurs, including the initiation of mRNA biosynthesis, anaerobic degradation of starch at phase I, and degradation of storage proteins at phase II. Embryo germination occurs at the end of phase II. TCA cycle starts operating at the end of phase II and coleoptile growth takes place with the commencement of photosynthetic starch biosynthesis at phase III (He and Yang, 2013). A few works on QTL mapping to identify the genes underlying

anaerobic germination tolerance in rice have been reported (Angaji, Septiningsih, et al., 2010, Baltazar, Ignacio, et al., 2014) and one of the QTLs, *qAG-9-2*, has been fine mapped to *trehalose-6-phosphate-phosphatase* (*OsTPP7*) gene which is involved in starch mobilization during germination (Kretzschmar, Pelayo, et al., 2015). Previous biochemical and enzymatic experiments have demonstrated the involvement of starch breakdown mechanisms and the induction of amylase during low oxygen germination in rice (Guglielminetti, Perata, et al., 1995, Ismail, Ella, et al., 2008, Perata, Guglielminetti, et al., 1997). A positive correlation between coleoptile length and total amylolytic activities have been reported (Pompeiano, Fanucchi, et al., 2013). The energy for the coleoptile growth originates from the germinating seed. In anoxic condition, the level of ethanol is elevated in the fast growing coleoptiles which suggest that the energy is generated from the fermentative metabolism pathway (Gibbs, Morrell, et al., 2000, Magneschi, Kudahettige, et al., 2009, Setter, Ella, et al., 1994). Recent studies have also shown that maintaining a high rate of energy production as well as fluxes between glycolytic and fermentative pathways is crucial for survival during anaerobic conditions.

Several studies on genome-wide transcriptome analysis to investigate the gene expression analysis of rice coleoptiles under hypoxic or anoxic conditions have been reported. These experiments were performed at varying oxygen levels using microarrays. The results of the studies showed that a number of common molecular mechanisms are involved in coleoptile growth, including carbohydrate metabolism, fermentation, hormone induction, and cell division and expression (Huang, Taylor, et al., 2009, Lasanthi-Kudahettige, Magneschi, et al., 2007, Shingaki-Wells, Huang, et al., 2011). A study conducted by Narsai et al. (2015) on tips and basal segments of oxygen-deprived coleoptiles of rice showed that the gene expression is region specific and the metabolic activities is different along the length of the coleoptile under normoxic (air), hypoxic

(3%), and anoxic conditions. Tolerant plants have developed the ability to generate ATP in absence of oxygen through fermentative metabolism and to develop specific morphologies (air channels, enhanced shoot elongation) that improve the entrance of oxygen (Armstrong, Brändle, et al., 1994, Crawford, 1992, Drew, He, et al., 2000, Gibbs and Greenway, 2003, Gibbs, Morrell, et al., 2000, Jackson, 1985, Perata and Alpi, 1993, Sauter, 2000, Voesenek, Colmer, et al., 2006). Further, gene expression studies on plants exposed to low oxygen conditions revealed the up-regulation of genes coding for transcription factors (Hoeren, Dolferus, et al., 1998, Liu, VanToai, et al., 2005), signal transduction components (Baxter-Burrell, Yang, et al., 2002), non-symbiotic haemoglobin (Dordas, Hasinoff, et al., 2004), ethylene biosynthesis (Vriezen, Hulzink, et al., 1999), nitrogen metabolism (Mattana, Coraggio, et al., 1994), and cell wall loosening (Saab and Sachs, 1996).

Despite quite extensive knowledge of adaptive mechanisms and regulation at gene and protein levels, our understanding of global gene networks related to anaerobiosis as a whole is limited. In this study, we conducted comparative RNA-seq analysis using 4-d-old aboveground shoot tissue (coleoptile) from two rice genotypes exhibiting contrasting coleoptile growth under submerged conditions. It is expected that the results of this study will enable us to have a better understanding on the overall schemes of genotype-specific gene expressions and the specific submergence responsive gene expressions under submerged conditions in the two genotypes. The main objectives of this study were to identify the differentially expressed genes (DEGs) between the two rice genotypes and to better understand the underlying molecular mechanisms that contribute to coleoptile growth.

Materials and methods

Plant materials and hypoxia stress treatment

The oxygen concentration of the autoclaved distilled water used in the experiment was measured using Extech dissolved oxygen meter (Extech instruments, Nashua, NH 03063, U.S.A.). Two contrasting lines were selected out of 293 accessions which were previously used for GWAS analysis of hypoxia stress tolerance in rice. All of the seeds used in this study were freshly harvested and air dried at 37°C for 5 days and then stored at 4°C prior to experiment. The dormancy of seeds was broken by oven drying the seeds at 50°C for 5 days. The seeds were dehulled and surface sterilized with 70% ethanol for 5 min followed by rinsing with autoclaved distilled water for three times. The sterilized seeds were then germinated for 4 days on moist filter paper placed inside petri dishes for phenotyping in control condition. For submergence treatments, sterilized seeds were germinated in 8 cm deep water in glass beakers for 4 days. The experiment was conducted in completely randomized condition in five replications. The photos of germinating seeds were captured by a Pentax camera and imported in ImageJ software for coleoptile length measurement.

Extraction of RNA and construction of library

Coleoptiles from four biological replicates of each line grown under hypoxia stress and normal condition were collected and immediately frozen in liquid nitrogen. Samples were then maintained at -80°C until RNA extraction. Total RNA was extracted using TRI Reagent (Invitrogen, MA, USA) and Plant RNA extraction kit (Qiagen). RNA quality and concentration were checked using spectrophotometer NanoDrop ND-1000 and its integrity was checked using 1% agarose gel electrophoresis. Poly-A RNA containing mRNA was purified using poly-T oligo-attached magnetic beads and fragmented, and complementary DNA (cDNA) was synthesized

using random hexamer primers, followed by purification, end-repairing, poly-A tailing, and adapter ligation. The libraries were prepared using the kit TruSEQ RNA Sample preparation v2 (Illumina), according to manufacturer's instructions. The quality of libraries was evaluated using Agilent 2100 BioAnalyser (Agilent technologies) using the kit Agilent DNA 1000 (Agilent). The library sequencing was performed as pair-end 2*75 bp on Hiseq4000 platform (Illumina) for the treatment samples and Novaseq (Illumina) was used to generate pair-end 2*150 bp for the control samples.

RNA-seq reads analysis and differential expressed genes (DEGs)

The software FastQC was used for the analysis of read quality and its visualization. The low quality bases and library adapters were removed from each library using Trimmomatic version. Reads were mapped against the reference genome of *Oryza sativa* cv. Nipponbare (IRGSP build 1.0 RAP-DB). The mapping and alignment of the reads on the rice genome was done using HISAT2. Expression levels of each gene were quantified by normalizing total gene counts with the effective library size. DESEQ2 package Bioconductor was used to test for pairwise differential expression analysis. Genes with p-value less than 0.05 were considered to be differentially expressed. We used the multi-factorial linear modeling and tested three null hypothesis of effects on gene expression: (1) whether hypoxia stress treatment has significant effect on the expression of each gene; (2) whether genotype has significant effects on gene expression, and (3) whether gene expression was affected by cold treatment in genotype-dependent manner. We fitted for models with our experimental data: (1) $\mathbf{FM_{trt}}$**:** $\boldsymbol{Y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}$; (2) $\mathbf{FM_{geno}}$**:** $\boldsymbol{Y} = \boldsymbol{\tau} + \boldsymbol{\varepsilon}$; (3) $\mathbf{FM_{add}}$**:** $\boldsymbol{Y} = \boldsymbol{\tau} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, and; (4) $\mathbf{FM_{full}}$**:** $\boldsymbol{Y} = \boldsymbol{\tau} + \boldsymbol{\gamma} + \boldsymbol{\tau}\boldsymbol{:}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$. In each model, $\boldsymbol{Y}$ is the expression value of each gene, $\boldsymbol{\tau}$ is the effect of cold treatment, $\boldsymbol{\gamma}$ is the effect of different genotype, and $\boldsymbol{\varepsilon}$ is the random error. Comparing $\mathbf{FM_{geno}}$ and $\mathbf{FM_{trt}}$ to $\mathbf{FM_{add}}$ separately, we then tested whether the expression of each

gene was regulated by cold, and whether there was a significant genotypic effect. Comparisons of $\mathbf{FM_{full}}$ and $\mathbf{FM_{add}}$ allowed us to test whether gene expression was affected by cold in a genotype-dependent manner. In all cases, expression values for genes were standardized using the expression $z=(x-\bar{x})/sx$ for cross-genotype comparisons.

Gene ontology and enrichment analysis

We used the web based tool AGRIGO to perform GO enrichment analysis on DEGs (Du, Zhou, et al., 2010), while the Fisher's exact test and Benjamini-Hochberg's false discovery rate (FDR) adjustment were used to control for multiple comparisons (Benjamini and Hochberg, 1995).

Results

Oxygen concentration and Phenotyping variation

The oxygen concentration of the autoclaved distilled water used in the experiment was measured to be 5.2 mg/lit. Average of the phenotypic values for the tolerant (Darij) and suceptible (4610) genotypes under control and hypoxic stress conditions are shown in Table 22. Significant differences was observed in the coleoptile length of the two genotypes under flooding stress condition ($P \leq 0.05$), whereas there was no significant differences in the coleoptile lengths of the two genotypes under control condition ($P \leq 0.05$).

Table 22 Phenotypic variation among two contrasting lines under control and flooded conditions

Treatment	Darij (cm)	4610 (cm)
Control (30C, 4DAS)		
Range	1.42 - 1.65	1.09 - 1.3
Average	1.54	1.2
Stdev	0.09	0.08
Hypoxia stress (30C, 4DAS)		
Range	1.93 - 2.55	0.17 - 0.33
Average	2.2	0.25
Stdev	0.26	0.07

Overview of hypoxia stress tolerant and hypoxia stress sensitive cDNA library sequencing

Illumina Platform-based RNA-seq was used to identify the differentially expressed genes in germinating seeds under hypoxia stress treatment. A total of 16 RNA samples from control and cold treated materials were prepared to construct cDNA libraries with fragments of 75-150 bp in length. The cDNA libraries were then sequenced with the Illumina HiSeq4000 platform and the Novaseq platform. A total of 451,402,131 clean reads were generated and mapped to the Nipponbare reference genome, and about 91.93 % to 96.78 % clean reads were successfully mapped to Nipponbare. The detailed sequencing statistics for these cDNA libraries are presented in Table 23 and Table 24. The total gene counts were normalized using DESeq2 package and the expression signals of each gene were calculated.

Table 23 Summary of RNA seq data and sequence assembly control samples in NovaSEQ under hypoxia stress

Parameters	T-spl #1	T-spl #2	T-spl #3	T-spl #4	S-spl #1	S-spl #2	S-spl#3	S-spl#4
Total reads	35203779	45779199	36065291	39596552	38422888	3.2E+07	40002161	43580521
aligned concordantly 0 times	2055103	2655340	2115356	2082729	4343032	3232098	4069481	4766254
aligned concordantly exactly 1 time	28611944	37389333	29586704	32573082	27358263	2.3E+07	28835485	31657075
aligned concordantly > 1 times	4536732	5734526	4363231	4940741	6721593	5132735	7097195	7157192
overall alignment	96.59%	96.65%	96.64%	96.78%	92.75%	93.66%	93.55%	93.30%

T-spl = Tolerant sample, S-spl = Susceptible sample

Table 24 Summary of RNA seq data and sequence assembly Treatment samples in HiSEQ4000 under hypoxia stress

Parameters	T-spl #1	T-spl #2	T-spl #3	T-spl #4	S-spl #1	S-spl #2	S-spl#3	S-spl#4
Total reads	20967074	20010205	17693506	19412278	15253253	1.3E+07	19310286	14916738
aligned concordantly 0 times	2700064	2429236	2096323	2363606	2337948	2001792	3045511	2251266
aligned concordantly exactly 1 time	14793571	14351151	12786643	13602715	9156346	8037908	11160744	8636842
aligned concordantly > 1 times	3473439	3229818	2810540	3445957	3758959	3318512	5104031	4028630
overall alignment	94.01%	94.54%	94.80%	94.16%	92.04%	91.96%	91.62%	91.93%

T-spl = Tolerant sample, S-spl = Susceptible sample

DEGs for hypoxic stress

To further examine whether the gene expression was affected by (1) genotype only (2) submergence treatment only or (3) both genotype and submergence, we performed multi-factorial linear model testing on our whole transcriptome dataset using DESeq2. Applying an FDR corrected p-value of < 0.05, we identified 7,286 DEGs expressed in Darij when compared to 4610 under control condition and 13,828 DEGs in Darij when compared to 4610 under submerged condition, and 5,571 DEGs due to interaction effect. The separate analysis of Darij and 4610 (hypoxic stress vs control) showed that the total number of DEGs in Darij and 4610 were 16,867 and 15,429, respectively, with 9,968 common genes among them (Figure 25). In this analysis, the total number of upregulated genes in Darij and 4610 were 8,091 and 7,579, respectively, with 4,491 common genes among them (Figure 26). The total number of downregulated genes were 8,776 and 7,850, respectively in Darij and 4610 with 5,064 common genes among them (Figure 27). We found statistically significant interaction effects indicating that the gene expression under submergence is regulated differently depending on the genetic background, i.e. variation of coleoptile elongation in diverse rice accessions could be potentially determined by genotype specific gene expression. In order to understand the underlying changes in biological process, molecular functions and cellular components, we conducted GO analysis of upregulated and downregulated genes identified from interaction effect and genotype effect separately.

Figure 25 Venn diagram of total DEGs under hypoxia stress

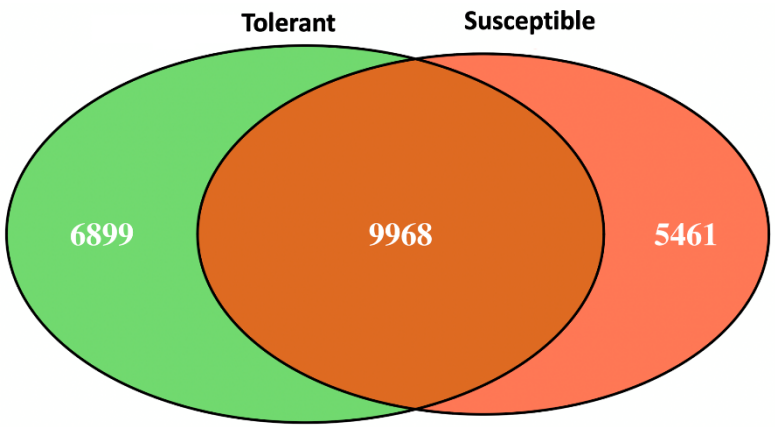


Figure 26 Venn diagram of upregulated DEGs under hypoxia stress

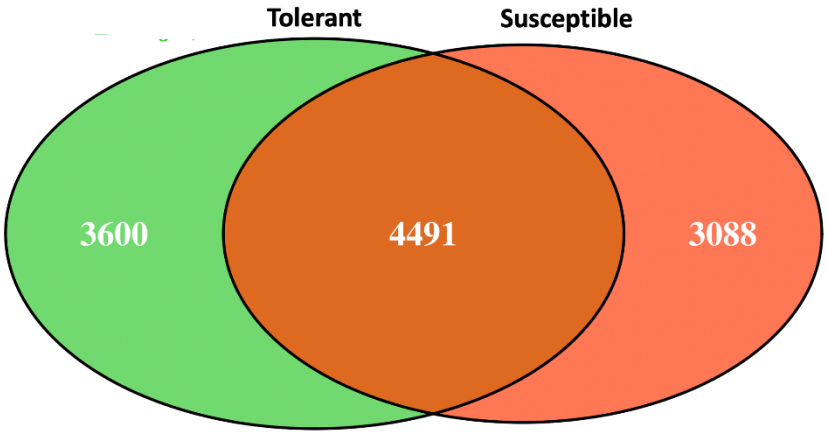
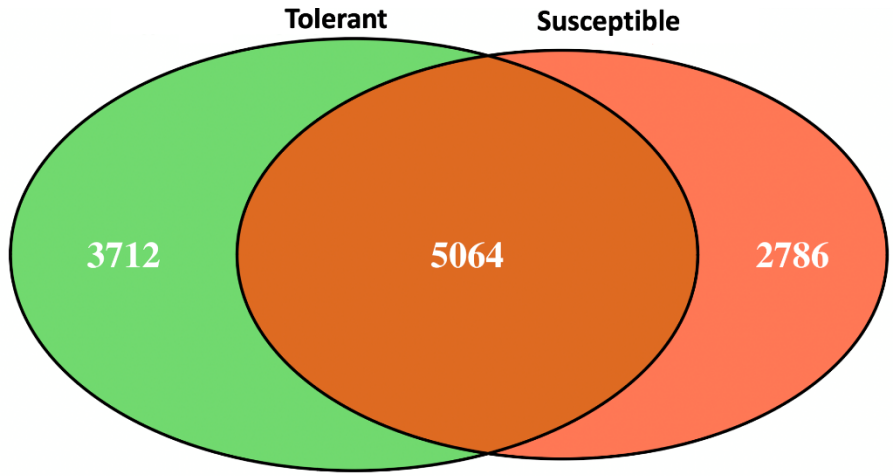


Figure 27 Venn diagram of downregulated DEGs under hypoxia stress



GO of DEGs in tolerant line compared to susceptible line in hypoxia stress vs normal condition

The GO enrichment analysis of upregulated genes in Darij due to interaction effect (genotype*treatment) showed 22 enriched GO terms (Figure 28). There were 12, 8, and 2 function terms found to be enriched in biological process, molecular function and cellular components respectively. Among the GO terms related to biological process, the highest percentage of annotated genes were found in GO terms - nitrogen compound metabolic process (GO:0006807), nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139) with 19.06 % and 19.06% annotated genes respectively. In molecular function, genes related to response to stimulus (GO:0050896) were found to be highly enriched with 23% of genes annotated. The percentage of genes annotated in GO term - binding (GO:0005488) was 48% among all the GO related to cellular components.

The GO analysis of downregulated DEGs in Darij compared to 4610 due to interaction effect (genotype*treatment) revealed 20 significant GO terms with 6, 3, and 11 GO terms related to biological process, molecular functions and cellular components respectively (Figure 29). Among the enriched GO terms related to cellular components, cell (GO:0005623), cytoplasm (GO:0005737) and cytoplasmic part (GO:0044444) were found to have high percentage of annotated genes with annotated gene percentage of 66.94%, 40.27% and 38.49%, respectively.

Figure 28 GO of upregulated DEGs in tolerant line compared to susceptible line in hypoxia stress vs normal conditions

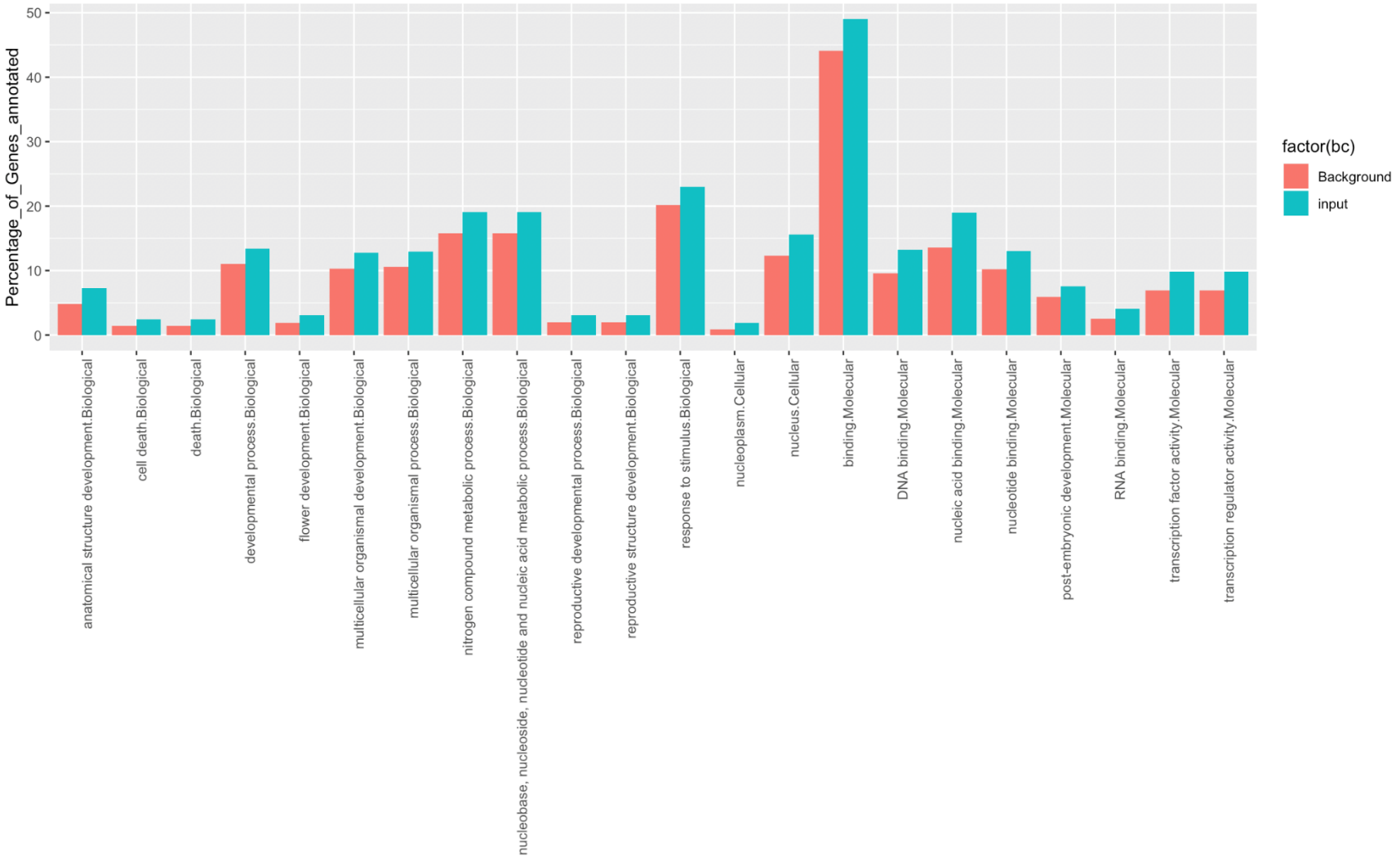
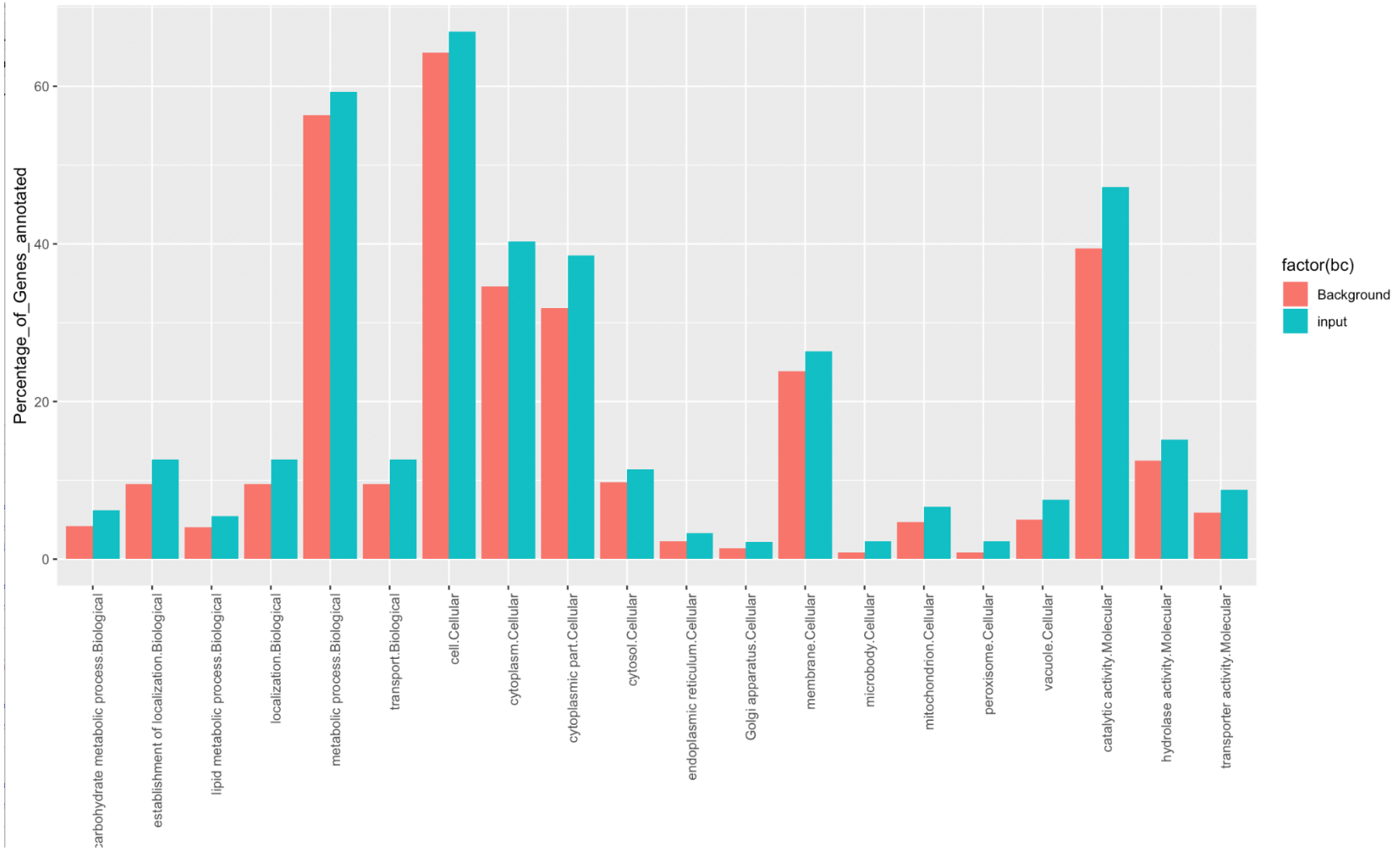


Figure 29 GO downregulated DEGs in tolerant line compared to susceptible line in hypoxia stress vs normal conditions



GO of unique DEGs in Darij (tolerant genotype; hypoxia stress vs control condition)

The GO analysis of the upregulated DEGs in Darij under hypoxia stress compared to control condition resulted in 26 GO terms (Figure 30). Among the 26 GO terms, 12, 5 and 9 functional GO terms were found to be related to biological process, molecular function and cellular components, respectively. The percentage of annotated genes in the biological process was found to be highest in cellular metabolic process (GO:0044237) with 34.47% of the annotated genes. The five significant GO terms related to molecular function found were binding (GO:0005488), protein binding (GO:0005515), nucleotide binding (GO:0000166), nucleic acid binding (GO:0003676) and RNA binding (GO:0003723). The GO term, binding (GO:0005488) was found to have the highest percentage of annotated genes (50.05 %). The highest percentage of annotated gens in cellular components was found in nucleus (GO:0005634) and plasma membrane (GO:0005886) with 14.62 % and 14.59 %, respectively. We found 5 GO terms from biological process and 1 GO term from cellular components discovered in Darij under hypoxia stress compared to normal condition were common due to genotype-treatment interaction effect. The common GO of biological process were nitrogen compound metabolic process (GO:0006807), nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139), death (GO:0016265) and cell death (GO:0008219). And the common GO from cellular components was nucleus (GO:0005634).

We found 27 significant GO terms from GO analysis of downregulated DEGs in Darij when compared Darij under hypoxia stress to control condition (Figure 31). The GO terms related to biological process, molecular function and cellular components were 13, 5 and 9, respectively. Among the biological process, the GO terms related to transport (GO:0006810), establishment of localization (GO:0051234) and localization (GO:0051179) had highest percentage of annotated

genes with 13.09% of annotated genes in all of them. The highest percentage of GO term was found in catalytic activity (GO:0003824) with 48.23% of annotated genes among molecular functions. Among the significant GO terms of cellular components, cytoplasm (GO:0005737) and cytoplasmic part (GO:0044444) had the highest percentage of annotated genes of 39.64% and 37.05%, respectively.

The identified 5, 2 and 8 GO terms in Darij under hypoxia stress compared to control condition were present in GO enrichment analysis of downregulated genes of Darij due to interaction effects. The common GO of biological process were transport (GO:0006810), establishment of localization (GO:0051234), localization (GO:0051179), lipid metabolic process (GO:0006629) and carbohydrate metabolic process (GO:0005975). The common GO of molecular function were catalytic activity (GO:0003824) and transporter activity (GO:0005215). The GO terms of cellular components common in Darij of hypoxia vs control and Darij of interaction effects were cytoplasmic part (GO:0044444), peroxisome (GO:0005777), microbody (GO:0042579), cytoplasm (GO:0005737), vacuole (GO:0005773), mitochondrion (GO:0005739), membrane (GO:0016020), and endoplasmic reticulum (GO:0005783).

Figure 30 GO of unique upregulated DEGs in Darij (tolerant line) genotype (hypoxia stress vs control conditions)

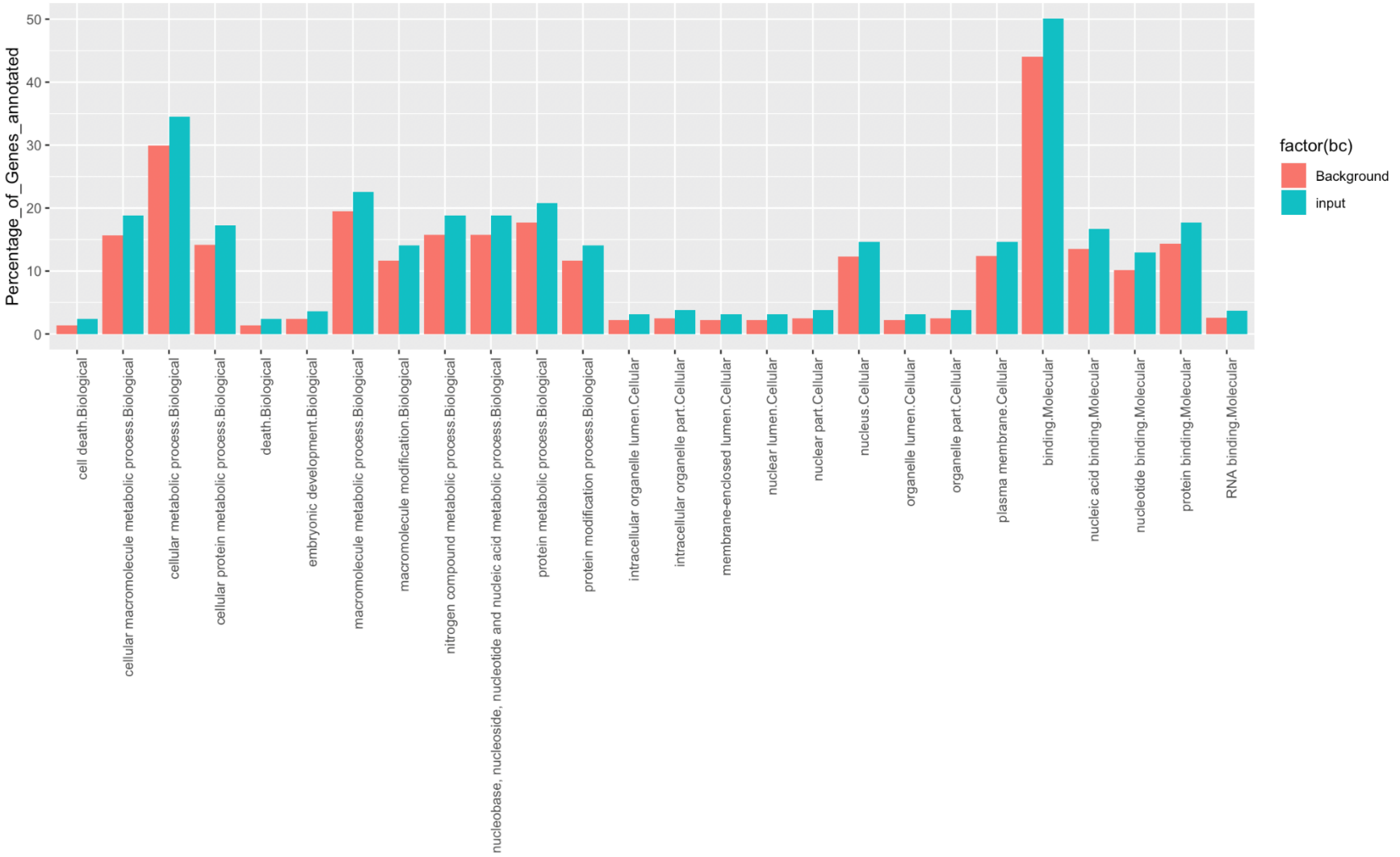
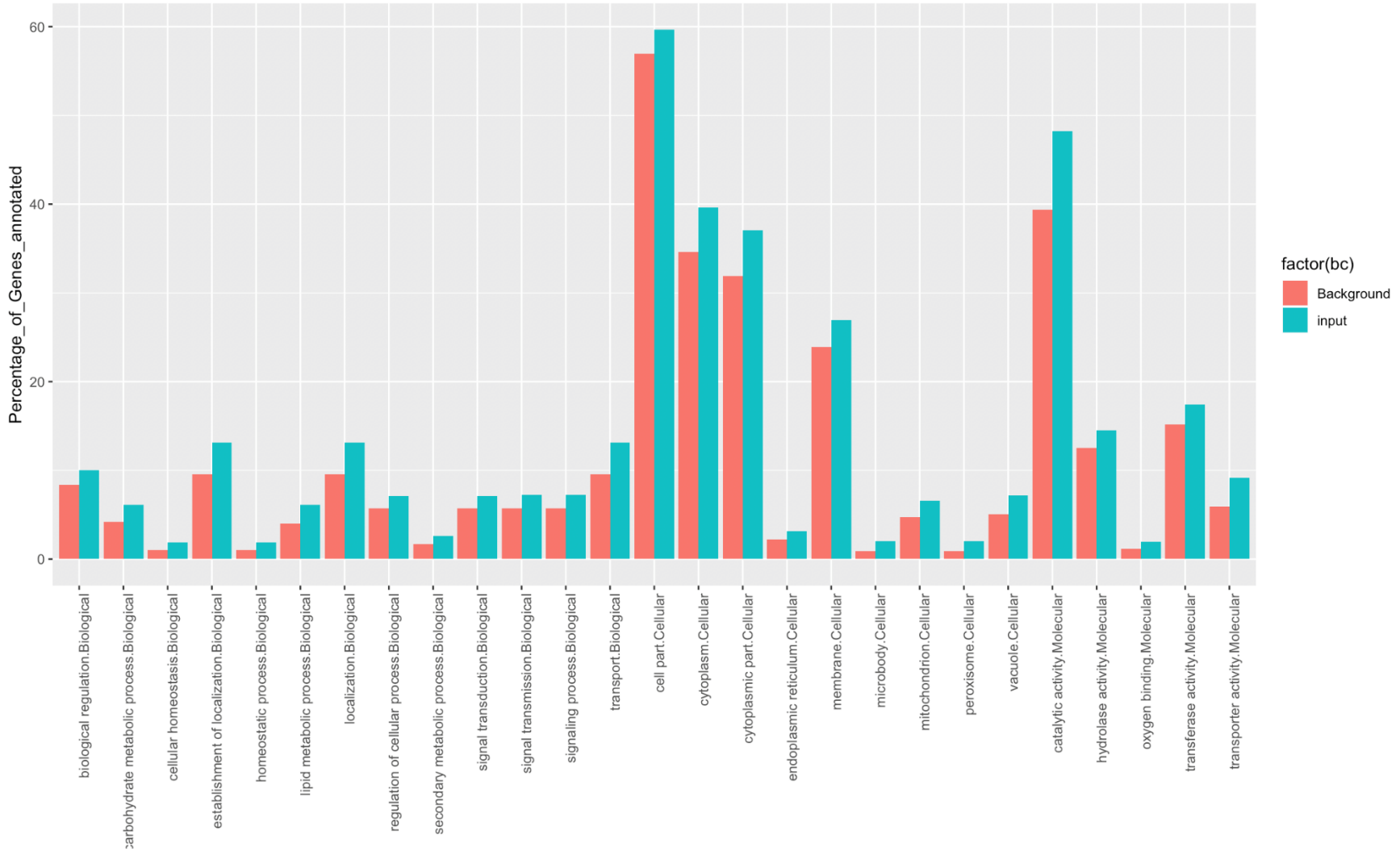


Figure 31 GO of unique downregulated DEGs in Darij (tolerant line) genotype (hypoxia stress vs control conditions)



GO of unique DEGs in 4610 genotype (hypoxia stress vs control condition)

From GO analysis of significant upregulated DEGs in 4610 germinated under hypoxic stress vs control conditions, we found only 4 significant GO terms. The GO terms were related to cellular components (Figure 32). We didn't find any GO related to biological process and cellular components. The significant GO were cytosol (GO:0005829), Golgi apparatus (GO:0005794), cytoplasmic part (GO:0044444) and endosome (GO:0005768). The GO term, cytoplasmic part (GO:0044444) had highest percentage of annotated genes, i.e. 35.03%.

The GO analysis of significantly downregulated genes depicted 31 GO terms. (Figure 33) Among the 31 GO terms, 18, 11 and 2 GO terms were related to biological process, cellular components and molecular functions, respectively. The highest percentage of annotated genes of 21. 32 % and 21.32 % were observed in the GO terms related to biological process of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139) and nitrogen compound metabolic process (GO:0006807), respectively. The highest percentage of annotated genes in molecular function was found in binding (GO:0005488), i.e. 48.32 %. The 2 GO terms related to cellular components were nucleus (GO:0005634) and nucleoplasm (GO:0005654). Among the two significant GO terms of cellular components, highest percentage of annotated genes were identified in nucleus (GO:0005634). Interestingly, we didn't find any common GO terms from the separate GO analysis of unique genes downregulated in the tolerant line and unique genes downregulated in the susceptible line. However, we identified 6 GO terms common in between GO analysis of upregulated genes in the tolerant line and GO analysis of downregulated genes in the susceptible line. The common GO identified from GO analysis of upregulated genes in the tolerant line and downregulated genes in the susceptible line were nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139), nitrogen compound metabolic

process (GO:0006807), nucleic acid binding (GO:0003676), binding (GO:0005488), RNA binding (GO:0003723) and nucleus (GO:0005634).

GO of common DEGs observed in both tolerant and susceptible genotypes (cold vs control)

The GO analysis of upregulated DEGs identified in both Darij and 4610 revealed 18 significant GO terms (Figure 34). There were significant 9, 6 and 3 GO terms discovered in biological process, molecular function and cellular components, respectively. The highest percentage of annotated genes, i.e. 56.20% and 41.52%, respectively were discovered in cellular process (GO:0009987) and primary metabolic process (GO:0044238). The significant GO of molecular function were binding (GO:0005488), kinase activity (GO:0016301), transferase activity, transferring phosphorus-containing groups (GO:0016772), protein binding (GO:0005515), translation factor activity, nucleic acid binding (GO:0008135) and transferase activity (GO:0016740). The GO, binding (GO:0005488) had highest percentage of annotated genes, i.e. 46.87%. The 3 significant GO in cellular components were cytosol GO:0005829), plasma membrane (GO:0005886) and Golgi apparatus (GO:0005794). The highest percentage of annotated genes were identified in plasma membrane (GO:0005886) and cytosol (GO:0005829) with 16.05% and 13.35%, respectively.

We identified a total of 29 significant GO terms from the GO analysis of significant downregulated DEGs (Figure 35). The biological process, molecular function and cellular components had 5, 7 and 17 significant GO terms. The metabolic process (GO:0006629) was identified to have highest percentage of annotated genes, i.e. 60.19 %. Among molecular functions, the GO, catalytic activity (GO:0003824) was identified to have annotated genes of 43.07%. Among the cellular components, the highest percentage of annotated genes were found in cell (GO:0005623) and cell part (GO:0044464) with 68.67% and 60.75%, respectively.

Figure 32 GO of unique upregulated DEGs in 4610 genotype (hypoxia stress vs control conditions)

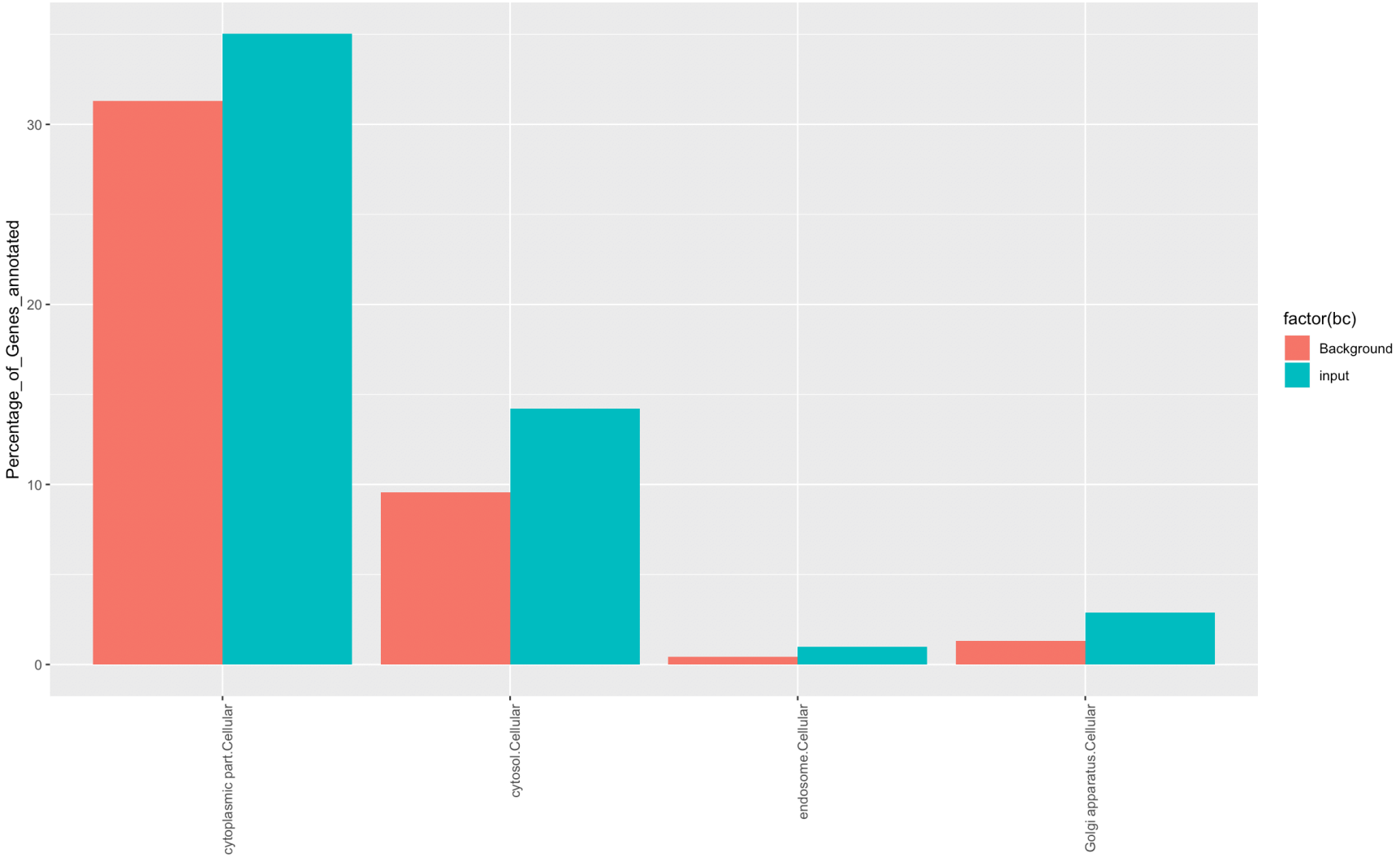


Figure 33 GO of unique downregulated DEGs in 4610 genotype (hypoxia stress vs control conditions)

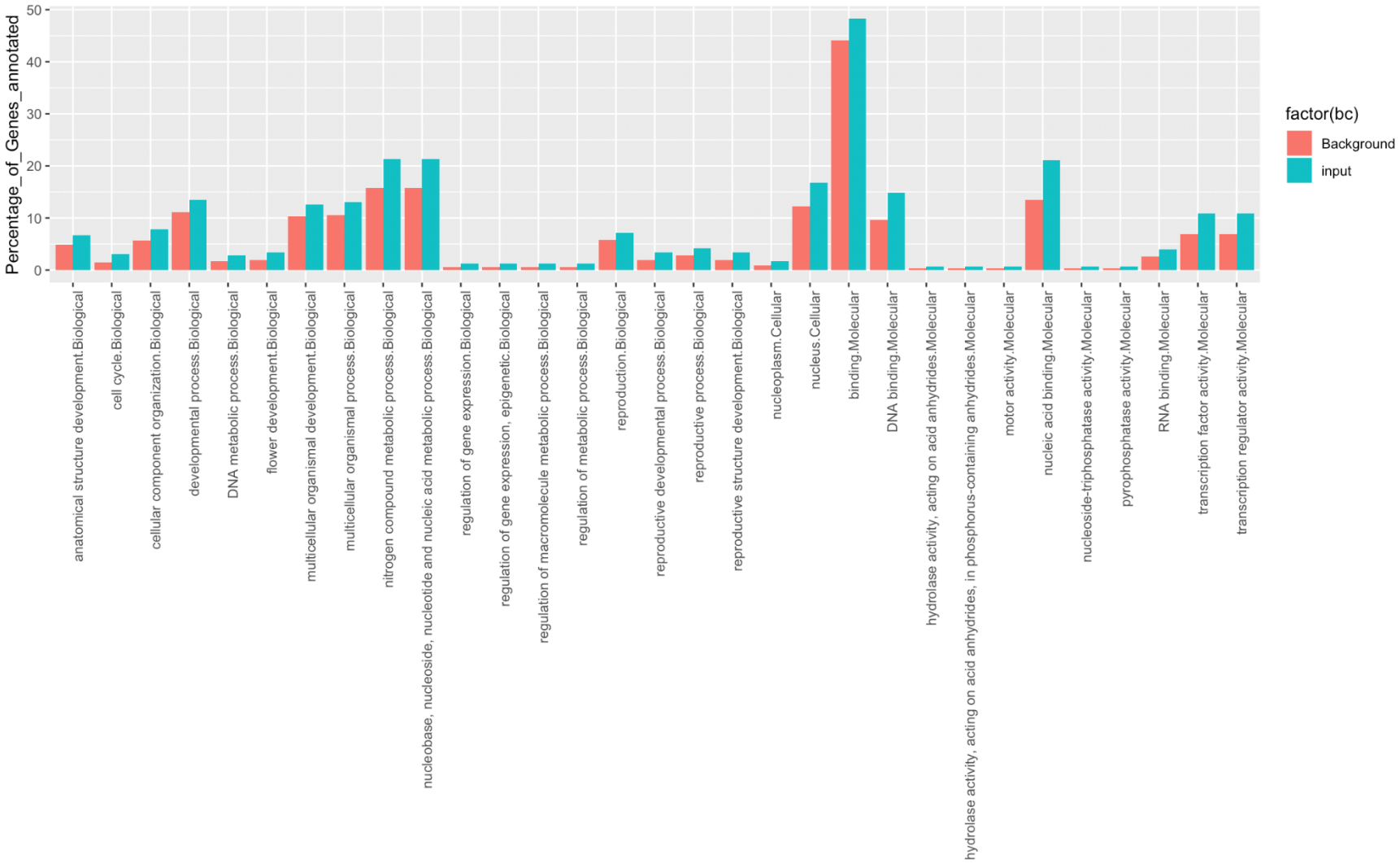


Figure 34 GO of common upregulated DEGs observed in both tolerant and susceptible genotypes (cold vs control)

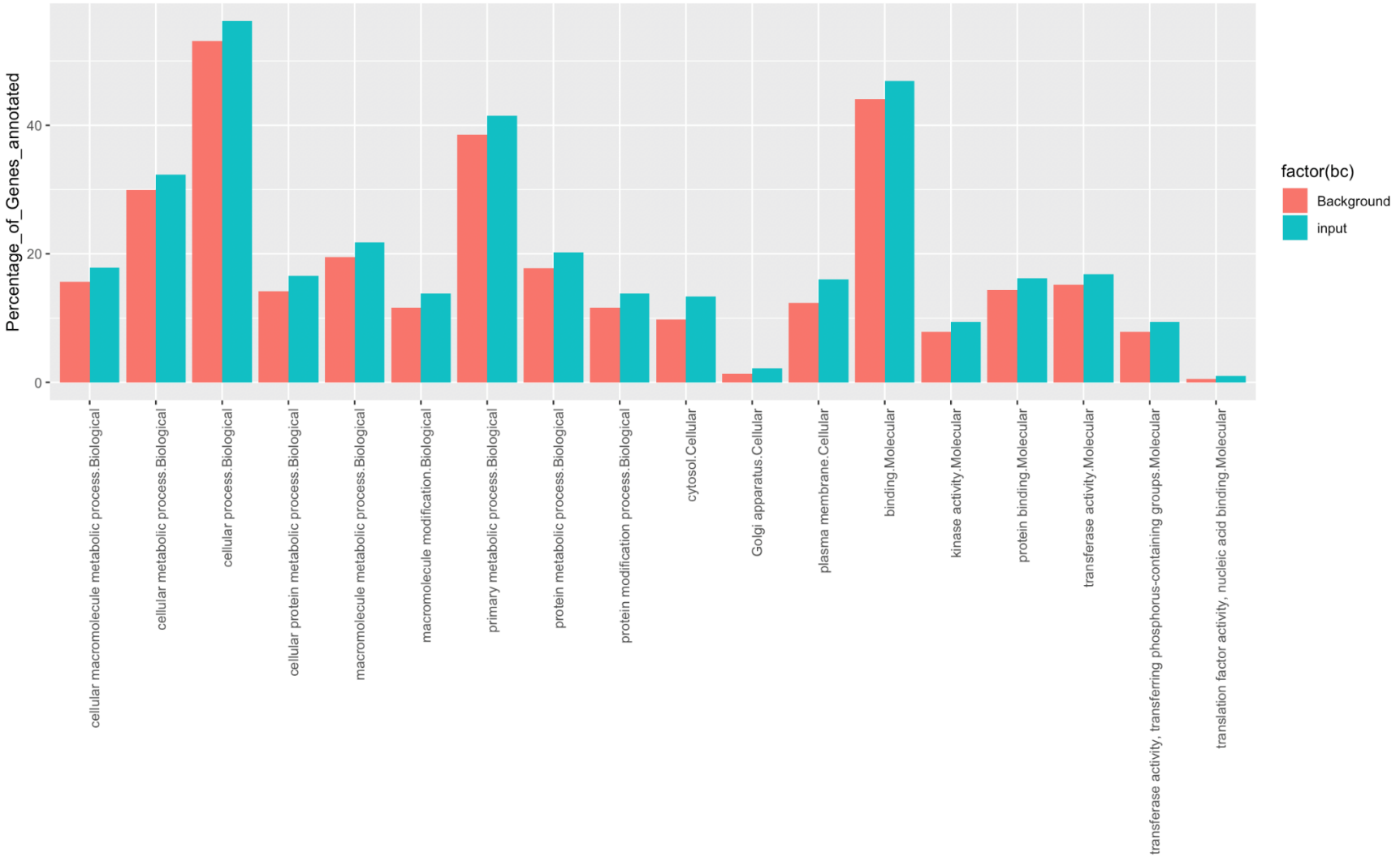
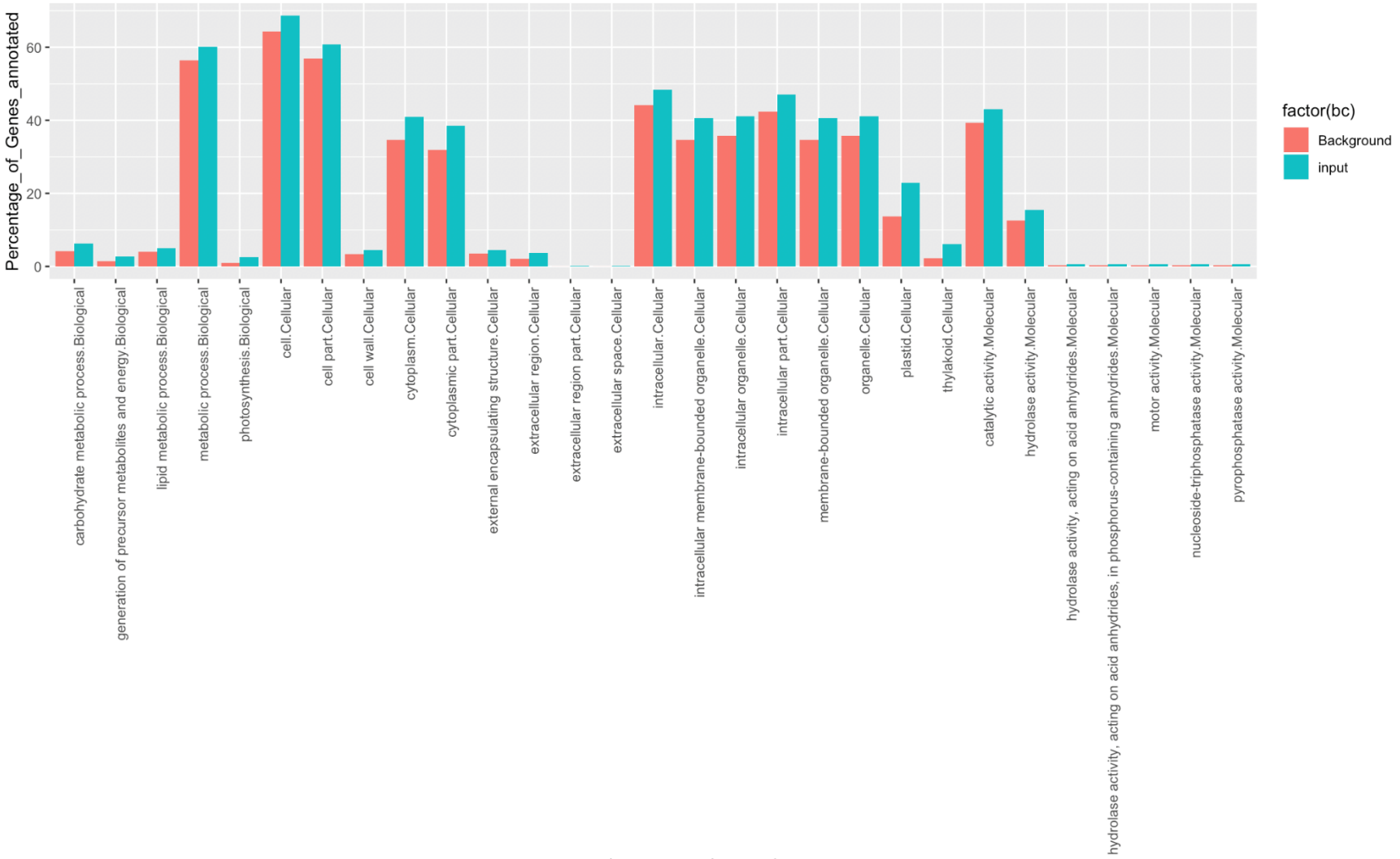


Figure 35 GO of common downregulated DEGs observed in both tolerant and susceptible genotypes (cold vs control)



GO terms and related transcription factors common in upregulated genes in tolerant line and downregulated genes in susceptible lines

The comparison of enriched GO terms derived from GO enrichment analysis of upregulated genes of tolerant line (hypoxia stress vs control) and downregulated genes of susceptible line (hypoxia stress vs control) revealed 6 GO terms. The common GO were nucleobase, nucleoside, nucleotide and nucleic acid metabolic process (GO:0006139), nitrogen compound metabolic process (GO:0006807), nucleic acid binding (GO:0003676), binding (GO:0005488), RNA binding (GO:0003723) and nucleus (GO:0005634). The detailed analysis of the common GO terms showed those genes are related to ABC transporters, AP2 domain containing proteins, AP2-like ethylene-responsive transcription factors, basic helix-loop-helix, bZIP transcription factor domain containing proteins, expressed protein, helix-loop-helix DNA-binding domain containing proteins, HSF-type DNA-binding domain containing proteins, MYB family transcription factors, no apical meristem proteins, putative, pentatricopeptide, phospholipid-transporting ATPase 1, RNA recognition motif containing proteins, and WRKY and C2H2 zinc finger proteins.

Figure 36 DEGs of ABC transporter family

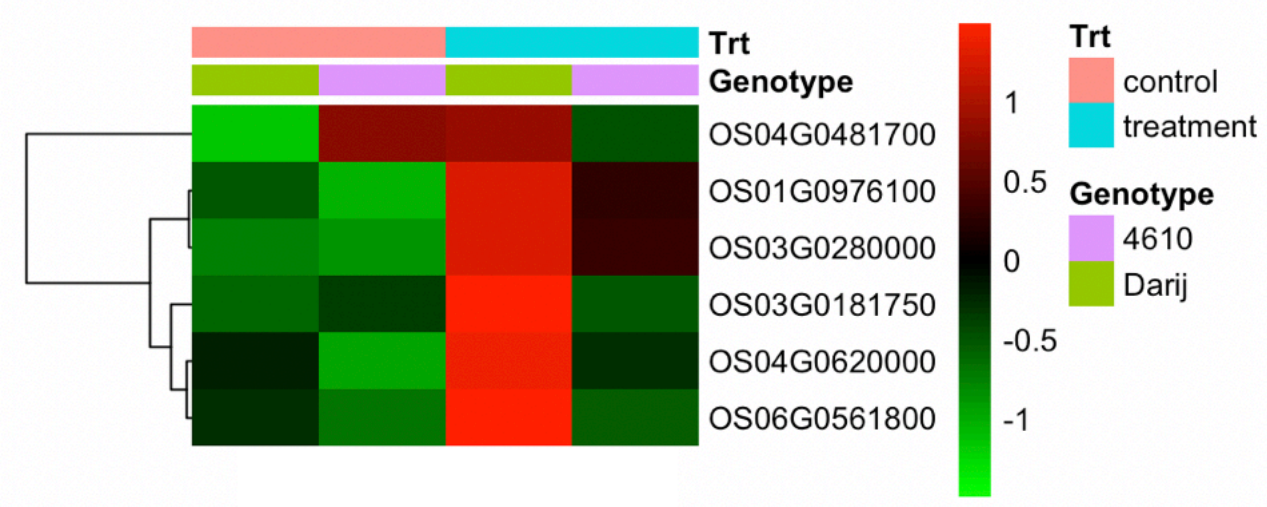


Figure 37 DEGs of AP2 domain containing protein family

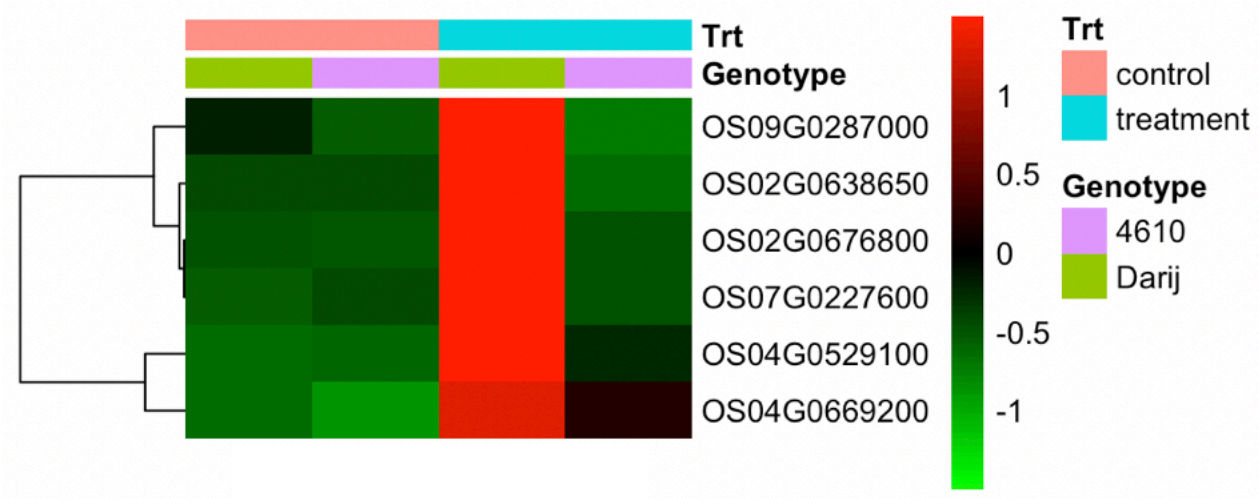


Figure 38 DEGs of helix-loop-helix DNA-binding domain containing protein family

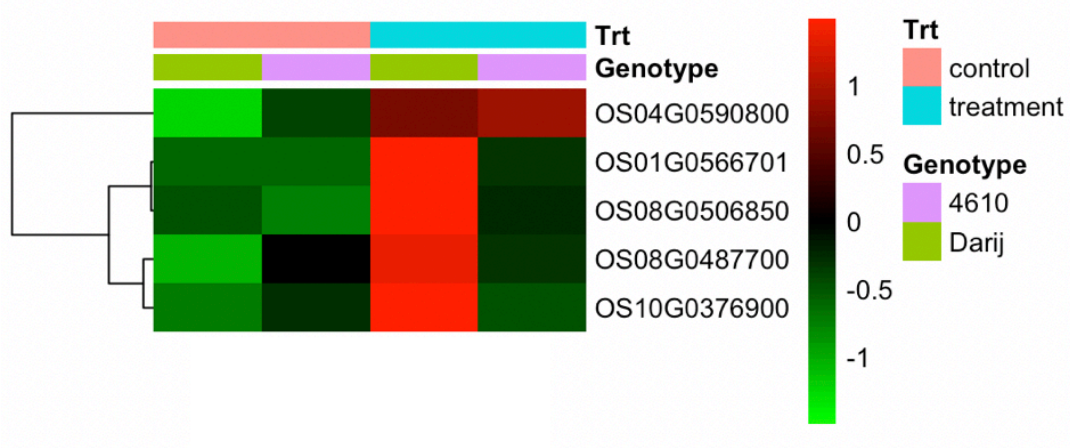


Figure 39 DEGs of pentatricopeptide (PPR) gene family

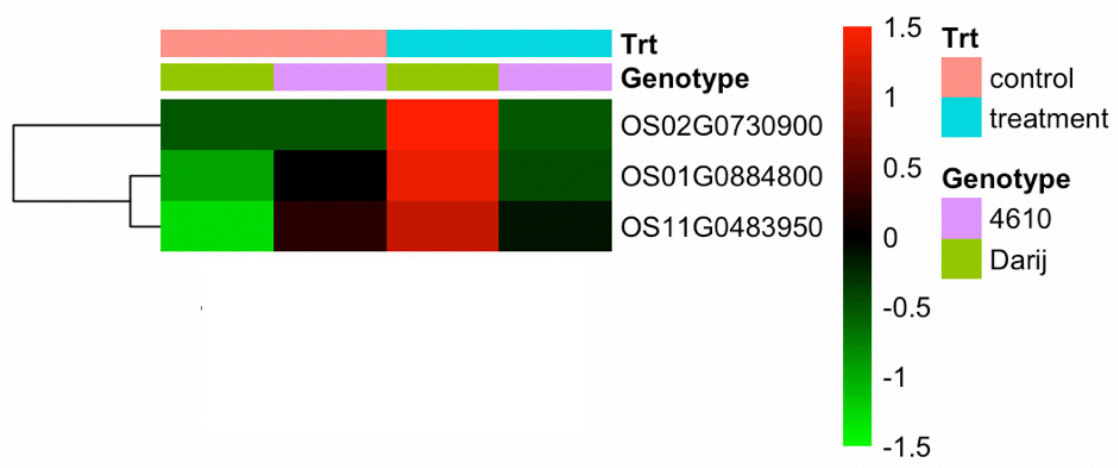


Figure 40 DEGs of No apical meristem (NAM) protein family

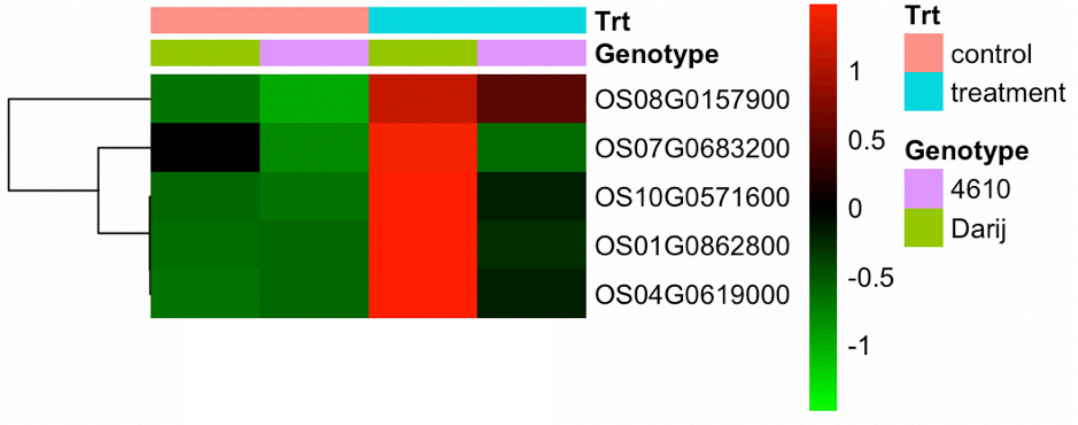


Figure 41 DEGs of RNA binding family

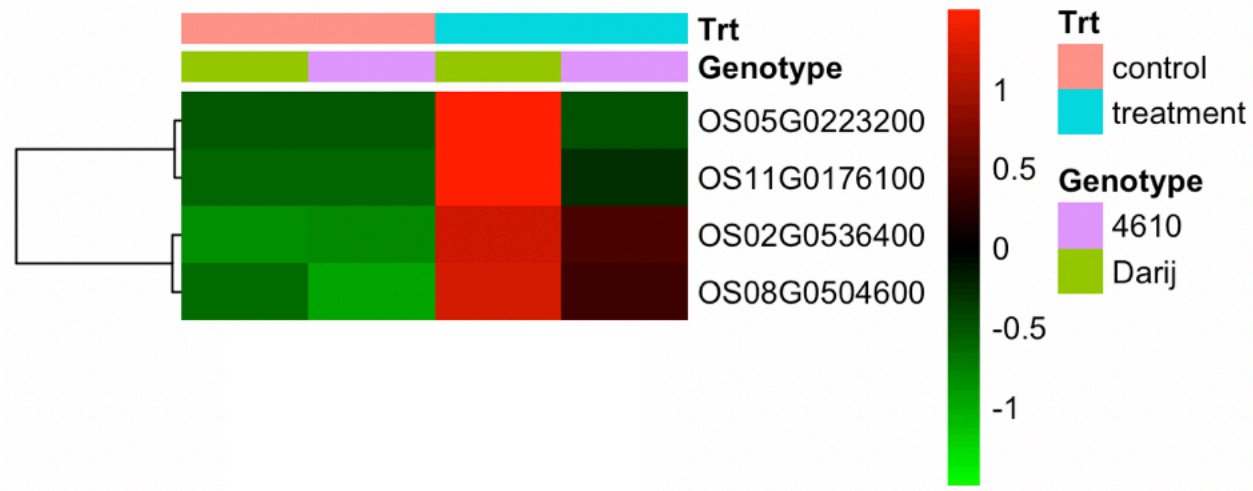


Figure 42 DEGs of MYB transcription factor family

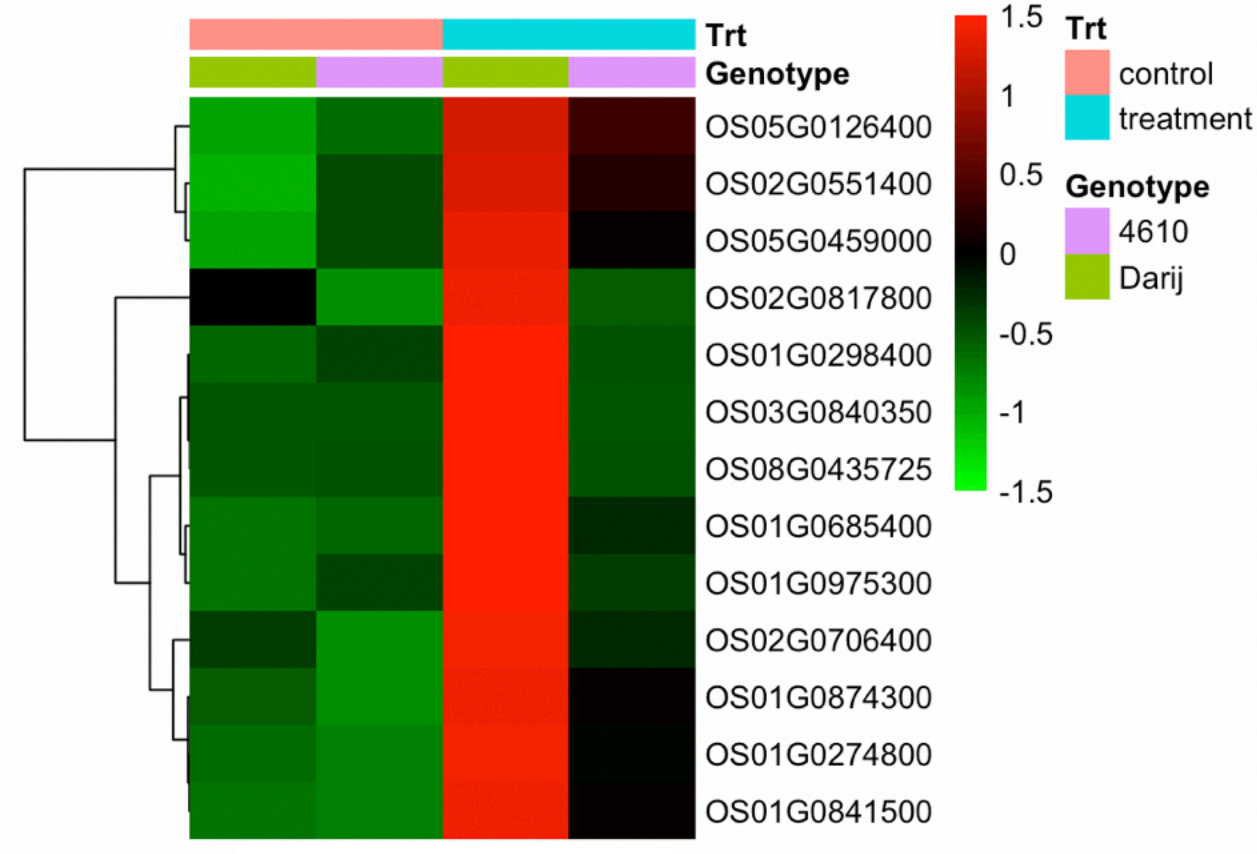


Figure 43 DEGs of WRKY family

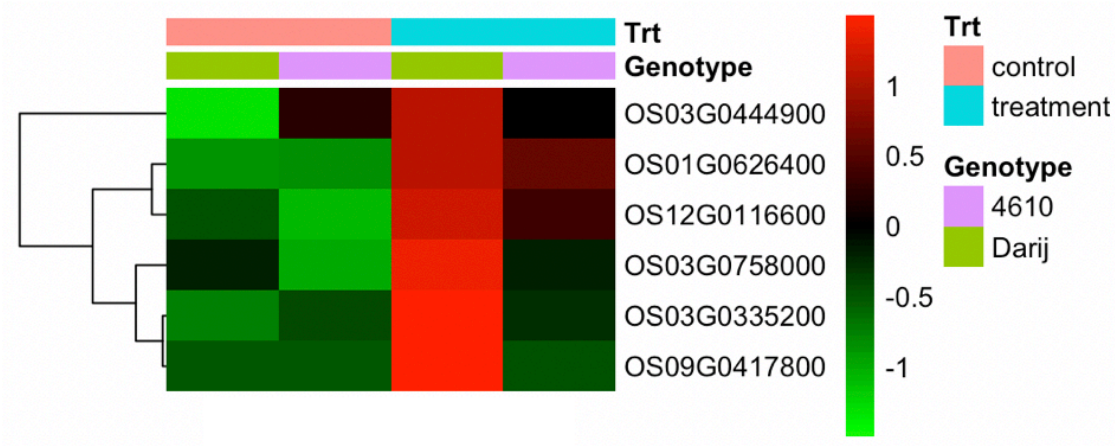
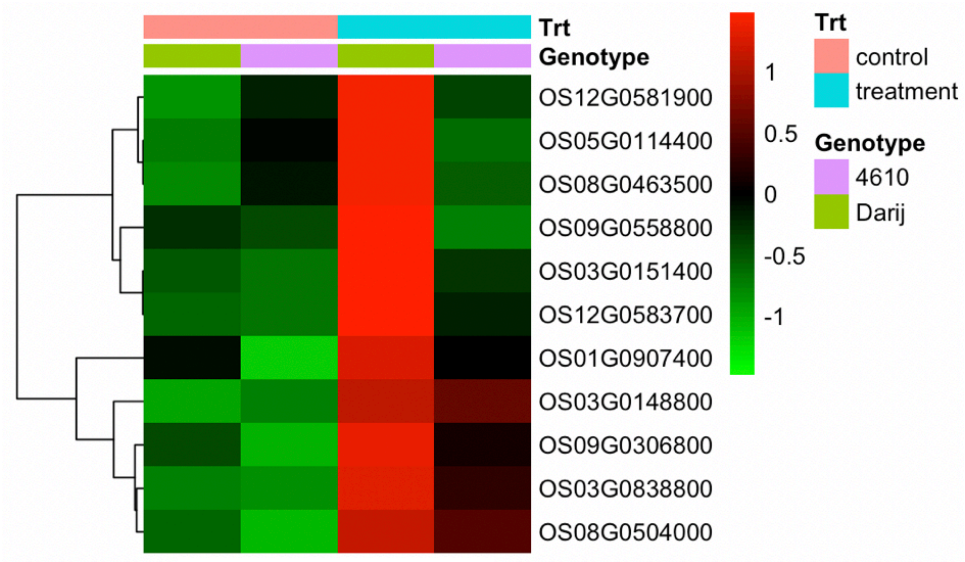


Figure 44 DEGs of C2H2 zinc finger protein family



Genes that exhibit significant differential regulation ($p < 0.05$) in tolerant line but not in susceptible line

The genes differentially regulated in tolerant line that are responsive to submergence are the promising candidate genes for involvement in anaerobic germination in tolerant line. We identified 6,899 genes that were significantly regulated only in tolerant line. Among them, 3466 were upregulated and 3433 were downregulated. The results showed that *trehalose-6-phosphate-synthase-7* (*OsTPS7*; Os08g0414700) and *trehalose-6-phosphate-phosphatase-9* (*OsTPP9*; Os06g0222100) were significantly upregulated in tolerant line. The other genes that were significantly upregulated in tolerant line were *aldehyde dehydrogenase 12B1* (*OsALDH12B1*; Os120596000) and *aldehyde dehydrogenase 2A* (*OsALDH2a*; Os02g0730000). We also found that 6 genes from the ERF family, 2 genes from cytochrome p450, and 4 genes from the heat stress transcription factors were significantly upregulated in tolerant line. The genes upregulated from ERF family and DREB family were ERF20 (Os02g0676800), ERF45 (Os04g0529100), ERF57 (Os07g0227600), ERF76 (Os04g0669200), ERF137 (Os03g0117900), ERF141 (Os02g0638650) and DREB1E (Os04g0572400). The 4 genes significantly upregulated genes from heat stress transcription factor were OsHsfA2a (Os03g0745000), OsHsfA6 (Os06g0565200), OsHsfB4a (Os08g0471000) and OsHsfB4d (Os03g0366800). The significantly upregulated genes from Cytochrome 450 were CYP714B1 (Os07g0681300) and OsCYP71X14 (Os02g0601500). Ethylene biosynthesis relies on the rapid oxidation of 1-aminocyclopropane-1-carboxylic acid (ACC), so we investigated the expression pattern of ACC genes in both tolerant and susceptible line. We identified 8 common ACC genes (LOC_Os01g39860, LOC_Os02g53180, LOC_Os03g48430, LOC_Os03g64280, LOC_Os08g30080, LOC_Os09g27750, LOC_Os09g27820 and LOC_Os09g39720) differentially regulated in both tolerant and

susceptible lines. However, we also identified unique DEGs (LOC_Os01g35230, LOC_Os08g30100 and LOC_Os08g30210) only in tolerant line and unique DEGs (LOC_Os03g63900, LOC_Os04g10350 and LOC_Os06g14400) only in susceptible line.

We identified *trehalose-6-phosphate synthase 3* (*OsTPS3*; Os01g0730300) and *heat stress transcription factor 5* (*HSF05*; Os02g0232000) were differentially downregulated in tolerant line. We also found 2, 4, 7 and 13 differentially downregulated genes from beta-expansin, alpha-expansin, ERF and Cyt 450, respectively. The genes from beta-expansin family were: OsEXPB4 (Os10g0556100) and OsEXPB6 (Os10g0555600); while the genes from apha-expansin family were: OsEXP1 (Os04g0228400), OsEXPA10 (Os04g0583500), OsEXPA5 (Os02g0744200) and OsEXPA7 (Os03g0822000). The significant downregulated genes found from ERF family were: ERF19 (Os11g0242300), ERF32 (Os02g0656600), ERF35 (Os02g0657000), ERF37 (Os04g0549800), ERF44 (Os08g0565200), ERF69 (Os03g0183200) and ERF72 (Os09g0434500). Likewise, those from Cyt 450 family were: OsCYP51H3 (Os07g0464700), OsCYP51H6 (Os05g0415800), OsCYP704A5 (Os10g0525000), CYP714C2 (Os12g0119000), OsCYP71AK1 (Os09g0530275), Oscyp71Z6 (Os02g0570500), Oscyp71Z8 (Os10g0439924), OsCYP72A19 (Os01g0627600), OsCYP72A32 (Os01g0602400), CYP76M8 (Os02g0569400), CYP76M9 (Os06g0501900), CYP93G2 (Os06g0102100) and CYP97A4 (Os02g0817900).

DEGs common in both genotypes

Among all the DEGs in both genotypes, there were 9,968 common DEGs. Among the conservatively regulated genes, genes encoding fermentative enzymes, vacuolar protein phosphatase, genes encoding protein related to cell wall or membrane structure, cell wall modification gene and genes related to ethylene response factor were observed. The two important genes observed in both genotypes were genes encoding the important fermentative enzymes,

pyruvate decarboxylase 1 (PDC1, LOC_Os05g39310), which channels pyruvate into the alcohol fermentative pathway (Gibbs, Morrell, et al., 2000, Hossain, Huq, et al., 1996) and *alcohol dehydrogenase 2* (*ADH2*, LOC_Os11g10510) which converts acetaldehyde to ethanol (Perata and Alpi, 1991). The other gene identified for encoding v*acuolar proton phosphatase* (*H+-PPase*, a pyrophosphate related active proton transporter that maintains cytosolic Ph homeostasis, LOC_Os02g55890) (Liu, Zhang, et al., 2010). The genes encoding proteins related to cell wall or membrane structure were identified in both genotypes were LOC_Os01g67030, LOC_Os10g40520, LOC_Os10g40510, and LOC_Os08g40690. The other commonly regulated genes encoding a cellulase (LOC_Os10g22570) which play an important role in cell wall modification and an AP2-domain containing ethylene response transcription factor (ERF, LOC_Os01g04800). Additional DEGs include genes encoding glycolytic enzymes, *pyrophosphate-dependent phosphofructokinase* (*PPi-PFK*; LOC_Os8225720), *glyceraldehyde 3-phosphate dehydrogenase* (*G3PDH*, LOC_Os0238920), *phosphoenolpyruvate carboxylase* (*PEPC*, LOC_Os08g27840) and cytochrome P450 (LOC_Os03g45619). Although these genes have been previously identified to be involved in fundamental submergence response mechanisms (Lasanthi-Kudahettige, Magneschi, et al., 2007, Narsai, Howell, et al., 2009), they were found to be differentially regulated with various fold change of expressions in both genotypes in the current study. This results suggest that stronger induction or suppression of these fundamentally responsive genes may play a vital role on enhancing the submergence tolerance during germination.

Common TFs

We found 1, 1, 1, 3, 4, 4 and 13 differentially upregulated genes belonging to gene family trehalose-6-phospahte phosphatase (TPP), alpha-expansin, dehydradation-responsive element-

binding protein (DREB), heat shock protein and transcription factor (HSF), cytochrome 450 (Cyt.450), trehalose-6-6phosphate synthase (TPS) and ethylenene response factor (ERF) that were common in both tolerant and susceptible lines. The gene belonging to TPP, alpha-expansin, and DREB were OsTPP3 (Os07g0624600), OsEXPA13 (Os02g0267200) and DREB2A (Os01g0165000), respectively. The genes belonging to HSF protein family were OsHsp70CP2 (Os12g0244100), HSP70 (Os11g0703900) and OsHSP82 (Os09g0482100) . The genes belonging to Cyt 450 were OsCYP450 (Os05g0445100), OsCYP709C5 (Os07g0635500) , CYP86A7-1 (Os04g0560100) and CYP97B4 (Os02g0173100) . The common genes upregulated from TPS were OsTPS11 (Os09g0376800), OsTPS1 (Os05g0518600), OsTPS2 (Os01g0749400) and OsTPS6 (Os05g0517200). The 13 common genes from ERF were ERF9 (Os03g0263000), ERF34 (Os04g0550200), ERF65 (Os07g0617000), ERF66 (Os03g0341000), ERF68 (Os01g0313300), ERF74 (Os05g0497300), ERF94 (Os04g0547600), ERF99 (Os01g0868000), ERF102 (Os09g0457900), ERF107 (Os02g0521100), ERF117 (Os05g0351200), ERF122 (Os06g0553700) and ERF126 (Os02g0767700).

Among the common downregulated genes between both tolerant and susceptible lines, we found 1, 2, 2, 2, 3, 4, 5, 10 and 15 genes from DREB, TPS, TPP, beta expansin, alpha expansin, HSF, Aldehyde dehydrogenase (ADH), ERF and Cyt.450 respectively. The gene identified from DREB family was DREB1D (Os06g0165600). The genes from TPS family were: OsTPS10 (Os09g0397300), OsTPS4 (Os03g0224300): from TPP family were OsTPP4 (Os02g0753000) and OsTPT2 (Os05g0241200); from beta expansin were: OsEXPB17 (Os04g0530100) and OsEXPB3 (Os10g0555900); alpha expansin were: OsEXPA12 (Os03g0155300), OsEXP2 (Os01g0823100), and OsEXPA21 (Os03g0377100). Those from ADH were ALDH1a (Os01g0591000), OsALDH2B1 (Os06g0270900), OsALDH3B1 (Os04g0540600), OS-ALDH (Os02g0646500) and

OsALDH3E2 (Os02g0647900); from ERF were: ERF1 (Os06g0604000), ERF3 (Os02g0202000), ERF33 (Os04g0549700), ERF52 (Os05g0572000), ERF54 (Os01g0657400), ERF62 (Os03g0183000), ERF64 (Os03g0183300), Os06g0194000 (ERF71), ERF129 (Os04g0655700), ERF142 (Os05g0389000); from HSF were OsHsp70CP1 (Os05g0303000), OsHSP70 (Os03g0821100), OsHsfA2b (Os07g0178600) and OsHsfC1b (Os01g0733200). The common genes from Cyt 450 were: OsCYP51H4 (Os02g0323600), OsCYP704A3 (Os04g0573900), CYP714B2 (Os03g0332100), OsCYP71AA2 (Os01g0957800), OsCYP71AA3 (Os01g0957600), OsCYP71E5 (Os12g0512800), OsCYP71K9 (Os06g0641800), OsCYP71Y7 (Os06g0639800), OsCYP78C6 (Os09g0528700), CYP86A7-2 (Os01g0854800), CYP86A7-3 (Os10g0486000), OsCYP87C2 (Os03g0658800), OsCYP96B5 (Os03g0140200), OsCYP96B6 (Os03g0138200) and CYP97C2 (Os10g0546600). This finding suggest that the genes belonging to these family were significantly affected by hypoxia stress.

Pyrophosphate (PPi) dependent energy supply pathway under submergence

Many previous studies have reported that since ATP-consuming processes are not favored by hypoxia, PPi should be considered as an alternative energy donor to potentially maintain cell growth (Atwell, Greenway, et al., 2015, Bailey-Serres, Fukao, et al., 2012, Gibbs and Greenway, 2003, Magneschi, Kudahettige, et al., 2009). To evaluate the contribution of ATP and PPi-dependent enzymes in anaerobic germination, we investigated the expression patterns of genes encoding ATP-dependent phophofructokinase (ATP-PFK), PPi dependent phosphofructokinase (PPi-PFK or PPi-PFP), and a vacuolar proton ATPase (V-ATPase). We found 9, 5 and 4 DEGs related to ATP-PFK, PPi-PFK and V-ATPase, respectively in the tolerant line. Likewise, those were 7, 5 and 3 DEGs in the susceptible line, respectively. All the DEGs related to ATP-PFK, PPi-PFK and V-ATPase observed in susceptible line were identified in tolerant line too. The common

DEGs observed were: ATP-PFK (LOC_Os08g34050, LOC_Os10g26570, LOC_Os09g24910, LOC_Os06g05860, LOC_Os01g09570, LOC_Os05g44922 and LOC_Os05g10650), PPi-PFK (LOC_Os06g22060, LOC_Os02g48360, LOC_Os06g13810, LOC_Os08g25720 and LOC_Os09g12650) and V-ATPase (LOC_Os01g23580, LOC_Os02g55890 and LOC_Os06g08080). The unique DEGs identified only in Darij were: ATP-PFK (LOC_Os09g30240 and LOC_Os04g39420) and V-ATPase (LOC_Os06g43660). This findings shows that although most of the genes are conservatively regulated in both tolerant and susceptible lines, some of them are regulated in genotype specific pattern, which may contribute to higher tolerance of these lines.

Discussion

Due to uneven distribution of rain and periodic changes in frequencies and intensities of flooding caused by extreme weather conditions, the impact of flood is worsening in recent years (Coumou and Rahmstorf, 2012). Flooding and water logging cause significant yield loss in major cereal crops grown worldwide. Rice is one of the unique cereal crops having peculiar ability to germinate and grow under flooded condition. However, vast diversity and plasticity of rice in adaptation to different hydrological conditions have been reported (Ismail, Singh, et al., 2013, Kirk, Greenway, et al., 2014). When flooding occurs after direct seeding, tolerant line germinate faster and their coleoptiles grow faster enabling them to reach the more aerated zone and oxygen which is referred as snorkel effect (Alpi and Beevers, 1983, Kawai and Uchimiya, 2000). The underlying biological process, molecular function and metabolic process have not been fully explored in rice. In this study, we have conducted comparative transcriptome analysis of two contrasting rice lines under submerged conditions to evaluate the differential gene expressions between them.

GO ontology

In broad GO annotation, we observed some differences and some similarities among the tolerant and susceptible lines. Among the upregulated gens, the genes involved in cellular process, cellular metabolic process, macromolecule modification, protein metabolic process, protein modification process, kinase activity, transferase activity and translation factor activity were found to be same in both tolerant and susceptible lines. The GO analysis of downregulated genes observed in both tolerant and susceptible lines showed that the genes related to carbohydrate metabolic process, lipid metabolic process, photosynthesis, thylakoid, plastid, catalytic activity, hydrolase activity and phosphatase activity were downregulated in both genotypes. This finding suggest that biological process related to carbohydrate metabolism, lipid metabolism and photosynthesis, genes related to cellular components of plastid, thylakoids and membrane bound organelles were significantly affected by hypoxia stress. In tolerant line, GO related to biological process like cellular macromolecule metabolic process, cellular protein metabolic process, embryonic development, nitrogen compound metabolic process, nucleotide and nucleic acid metabolic process, protein modification process, and GO related to nucleotide binding, nucleic acid binding, RNA binding and protein binding were found to be significantly enriched. But in the susceptible line, we identified GO terms related to only Golgi apparatus, endosome, cytosol and cytoplasmic part were found to be enriched. On the other hand, we observed genes related to cell cycle, cellular components organization, DNA metabolic process, nitrogen compound metabolic process, nucleotide and nucleic acid metabolic process, regulation of gene expression, regulation of metabolic process, DNA binding, hydrolase activity, motor activity, transcription factor activity were significantly downregulated. This suggests that the genes involved in important biological

process, molecular functions, and cellular components are significantly downregulated in susceptible line.

Significance of Nitrogen metabolism for hypoxia stress tolerance

Flooding and water logging due to heavy rainfall are expected to occur more frequently due to global climate change (http://www.ipcc.ch/). During the flooded condition, rice plants are threatened by a reduction in the availability of oxygen either due to slower diffusion of oxygen through water or due to continuing competition for oxygen with respiring microorganisms (Drew, 1997, Jackson, 1985). Major disorders associated with low oxygen stress are related to energy crisis caused by inhibition of mitochondrial oxidative phosphorylation and subsequent reduction in ATP production. The shortage of ATP causes impaired functioning of the plasma membrane H+ pumping ATPase which lead to hypoxia induced acidification of cytoplasm (Felle, 2006). The adaptability of plants to low oxygen stress is generally considered in terms of anaerobic carbon metabolism (Limami, Diab, et al., 2014). However, few researches have reported the contribution of nitrogen metabolism in cellular acclimation to low oxygen stress in plants (Bailey-Serres, Fukao, et al., 2012, Bailey-Serres and Voesenek, 2008). Low oxygen stress has significant effects on several nitrogen metabolism process like nitrogen absorption, nitrate and nitrite reduction, nitric oxide (NO) production, ammonium assimilation and amino acid metabolism (Limami, Diab, et al., 2014). Several studies have previously shown that nitrate fertilization improves tolerance to hypoxia and limit the hypoxia stress in the roots (Allègre, Silvestre, et al., 2004, Horchani, Aschi-Smiti, et al., 2010, Morard, Silvestre, et al., 2004).

Transcriptional response related to photosynthesis efficiency and chlorophyll under hypoxia stress

The genes involved in photosynthesis were found to be downregulated under flooding stress in both lines. This agrees with the hypothesis that plant's photosynthetic ability is affected under stresses, inlcuding flooding stress. Interestingly, few researches have reported the upregulation of photosynthesis related genes under flooding. Nanjo et al. (2011) reported the upregulation of photosynthesis related genes after imposing flooding at 2 d old seedling stage of soybean (Nanjo, Maruyama, et al., 2011); Kreuzwieser et al. (2009) reported upregulation of photosynthetic related genes in Arabidopsis after imposing low oxygen stress for 3h and in poplar after imposing flooding stress for 5 h (Kreuzwieser, Hauberg, et al., 2009). However, Lee et al. (2014) reported downregulation of photosynthesis related genes in the leaves of rape seedlings after 3 d of water logged condition (Lee, Kim, et al., 2014). This suggests that flood stress stimulates the expression of photosynthesis related genes during initial stress response and repress the expression of these genes at later stages. Reduced photosynthesis during flooding stress could be due to production of ROS (Klok, Wilson, et al., 2002). High levels of ROS can damage cell through peroxidation of lipids, oxidation of proteins and other pathways finally leading to cell death (Clement, Lambert, et al., 2008, Wrzaczek, Vainonen, et al., 2011).

Transcriptional factors and genes upregulated only in tolerant line

In this study, the detailed analysis of the GO terms showed that some genes related to ABC transporter, AP2 domain containing protein, AP2-like ethylene-responsive transcription factor, basic helix-loop-helix, bZIP transcription factor domain containing protein, expressed protein, helix-loop-helix DNA-binding domain containing protein, HSF-type DNA-binding domain containing protein, MYB family transcription factor, no apical meristem protein, putative,

pentatricopeptide, phospholipid-transporting ATPase 1, RNA recognition motif containing protein, WRKY and C2H2 zinc finger protein were upregulated only in tolerant line and downregulated in susceptible line. This shows that the genes of these gene family might be contributing factor of hypoxia stress tolerance in the tolerant line.

We found 6 genes of ATP binding cassette (ABC) upregulated in tolerant line and downregulated in susceptible line. The ATP binding cassette (ABC) protein family consists of membrane proteins is active in ATP-powered transport of structurally unrelated substrate across the membrane (Decottignies and Goffeau, 1997, Martinoia, Klein, et al., 2002, Sánchez-Fernández, Davies, et al., 2001). The role of ABC transporters in plants in response to abiotic stress have been poorly studied. Previously, *Ospdr9*, gene encoding an ABC protein had been identified to be induced by hypoxia stress in rice (Moons, 2003). In recent years, ABC proteins have been identified to be an important elements in the transport of phytohormones, heavy metals, lipids, chlorophyll catabolites and secondary metabolites (Yazaki, Shitan, et al., 2009). ABS proteins in plastid play an important role in biosynthesis of membrane lipids like galactolipid in the thylakoid membrane (Yazaki, Shitan, et al., 2009).

From our study, we identified 6 genes of APETELLA2/ethylene response factor (AP2/ERF) upregulated in tolerant line and downregulated in susceptible line. It has been previously reported that AP2/ERF genes, a large multigene superfamily of transcription factors play an important role in water logging and flooding stress (Fukao, Xu, et al., 2006, Hattori, Nagai, et al., 2009, Hinz, Wilson, et al., 2010). Hattori et al. (2009) reported the deep water response of two genes encoding ethylene response factors, *SNORKEL1* and *SNORKEL2* in rice. Under deep water condition, ethylene accumulates in water and induces the expression of both genes which ultimately lead to internode elongation via gibberellin (Hattori, Nagai, et al., 2009). Hinz et al.

(2010) reported that overexpression of *RAP2.2* gene belonging to AP2/ERF transcription factor improves the plant survivability under hypoxia stress condition (Hinz, Wilson, et al., 2010).

A study conducted in *O. coarcata* transcriptomic changes under salt and submergence stresses (alone or combined, compared to control conditions) and found several transcription factors upregulated in leaves under stress conditions, NAC, WRKY, MYB gene family members indicating extensive transcriptional regulation in stress responses (Garg, Verma, et al., 2013). MYB related genes were found to be significantly upregulated in soybean seeding due to flooding stress (Nanjo, Maruyama, et al., 2011). In this study, we identified 13 genes of MYB family upregulated in tolerant line and downregulated in susceptible line. The MYB transcription factors contain MYB domain that is highly conserved in all eukaryotes and involves in regulation of wide range of functions MYB protein (Butt, Yang, et al., 2017). Several previous researches have shown the involvement of MYB transcription factor in regulatory networks associated with plant growth and development (Huang, Lv, et al., 2017, Tak, Negi, et al., 2017, Xu, Feng, et al., 2017). Many other researches have suggested the significant role of MYB proteins in regulating plant responses to abiotic stress (Lv, Yang, et al., 2017, Wang, Chen, et al., 2017, Wei, Luo, et al., 2017, Zhou, Yang, et al., 2016).

Pentatricopeptide repeat (PPR) protein represents a large family in plants and associated with processing of organelle mRNA molecules (Lurin, Andrés, et al., 2004). PPR proteins are chloroplast binding protein and part of the constitutive expression machinery of the chloroplast. Chloroplast recognition motif proteins respond to various signals and appear to be prime candidates for regulating chloroplast gene expression (Ruwe, Kupsch, et al., 2011). PPR proteins are known to be involved in RNA processing within mitochondria and chloroplasts and highly expressed during germination (Howell, Narsai, et al., 2009, Narsai, Law, et al., 2011). Narsai et

al. (2015) reported the expression of pentatricopeptide repeats (PPR) proteins in coleoptile tips and bases during hypoxia (Narsai, Edwards, et al., 2015). In our study, 3 genes of PPR family were found to be upregulated in tolerant line and downregulated in susceptible line. The unique expression of the PPR gene in tolerant line indicates that these genes might be involved in hypoxia stress in rice.

RNA recognition motif, an RNA binding domain is also important constituent of chloroplast proteins. Several members of this family have been identified to be respond to external and internal stimuli by changing their expression levels and protein modification state (Ruwe, Kupsch, et al., 2011). They have been reported as ideal candidates for regulating chloroplast RNA processing under shifting environmental condition (Ruwe, Kupsch, et al., 2011). RNA binding proteins act directly or indirectly in post-transcriptional regulation (Nakaminami, Matsui, et al., 2012). There are many kinds of RNA binding proteins and they are categorized based on their structure and bind specificity. RNA recognition motif (RRM) is the most conserved domain. Recently, it has been reported that cold shock domain proteins (CSDs) as an important RRM for stress responses and plant development (Chaikam and Karlson, 2010, Karlson and Imai, 2003). RNA binding proteins are found to be involved in transcription, splicing and control of decay and stabilization (Nakaminami, Matsui, et al., 2012). Our findings also shows 4 genes of RRM family upregulated only in tolerant line and downregulated in susceptible line. However, it is important to analyze the function of these RNA binding proteins to understand the mechanism of RNA regulation in hypoxia stress responses.

The Cyst2/His2 (C2H2) type zinc finger proteins (ZFPs) has 176 members in Arabidopsis and 1,889 members in rice, and it constitute one of the largest families of transcriptional regulators in plants (Agarwal, Arora, et al., 2007, Ciftci-Yilmaz and Mittler, 2008). C2H2 type ZFPs are

found to be important components in the regulation of plant growth, development, hormone responses and tolerance to biotic and abiotic stresses (Ciftci-Yilmaz and Mittler, 2008, Kiełbowicz-Matuk, 2012, Miller, Shulaev, et al., 2008). In rice, several members of C2H2 type ZFPs, such as ZFP182, ZFP245, ZFP252, ZFP179 have shown to be involved in the responses of rice to drought, salinity and oxidative stresses (Huang, Sun, et al., 2012, Huang, Sun, et al., 2009, Huang, Yang, et al., 2007, Sun, Guo, et al., 2010, Xu, Huang, et al., 2008). In this study, we observed 11 genes of C2H2 upregulated in tolerant line and downregulated in susceptible line. This findings suggest that some members of C2H2 type ZFPs are important regulators of ROS signaling in rice under abiotic stress.

Conclusion

In this study, we present the findings of whole genome transcriptome profiles of two rice genotypes that exhibit contrasting response to flooding stress. Our study highlights that genes involved in carbohydrate metabolism, pyrophosphate-dependent energy conservation, ethylene signaling pathways, nitrogen metabolism and nucleotide binding are important contributor for submergence tolerance in rice. We found that significant number of DEGs has contributed to significantly phenotypic difference in submergence tolerance between the tolerant and susceptible genotypes. This study has highlighted the importance of studying expression profile in rice across diverse genetic backgrounds and has provided the basis for identification of potential candidate genes involved in hypoxia stress tolerance in rice and possible use of the information in breeding program.

References

Agarwal, P., R. Arora, S. Ray, A.K. Singh, V.P. Singh, H. Takatsuji, et al. 2007. Genome-wide identification of C 2 H 2 zinc-finger gene family in rice and their phylogeny and expression analysis. Plant molecular biology 65: 467-485.

Allègre, A., J. Silvestre, P. Morard, J. Kallerhoff and E. Pinelli. 2004. Nitrate reductase regulation in tomato roots by exogenous nitrate: a possible role in tolerance to long-term root anoxia. Journal of Experimental Botany 55: 2625-2634.

Alpi, A. and H. Beevers. 1983. Effects of O2 concentration on rice seedlings. Plant Physiology 71: 30-34.

Angaji, S.A., E.M. Septiningsih, D. Mackill and A.M. Ismail. 2010. QTLs associated with tolerance of flooding during germination in rice (Oryza sativa L.). Euphytica 172: 159-168.

Armstrong, W., R. Brändle and M.B. Jackson. 1994. Mechanisms of flood tolerance in plants. Acta Botanica Neerlandica 43: 307-358.

Atwell, B.J., H. Greenway and T.D. Colmer. 2015. Efficient use of energy in anoxia-tolerant plants with focus on germinating rice seedlings. New Phytologist 206: 36-56.

Bailey-Serres, J., T. Fukao, D.J. Gibbs, M.J. Holdsworth, S.C. Lee, F. Licausi, et al. 2012. Making sense of low oxygen sensing. Trends in plant science 17: 129-138.

Bailey-Serres, J. and L. Voesenek. 2008. Flooding stress: acclimations and genetic diversity. Annu. Rev. Plant Biol. 59: 313-339.

Baltazar, M.D., J.C.I. Ignacio, M.J. Thomson, A.M. Ismail, M.S. Mendioro and E.M. Septiningsih. 2014. QTL mapping for tolerance of anaerobic germination from IR64 and the aus landrace Nanhi using SNP genotyping. Euphytica 197: 251-260.

Baxter-Burrell, A., Z. Yang, P.S. Springer and J. Bailey-Serres. 2002. RopGAP4-dependent Rop GTPase rheostat control of Arabidopsis oxygen deprivation tolerance. Science 296: 2026-2028.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological) 57: 289-300.

Butt, H.I., Z. Yang, Q. Gong, E. Chen, X. Wang, G. Zhao, et al. 2017. GaMYB85, an R2R3 MYB gene, in transgenic Arabidopsis plays an important role in drought tolerance. BMC plant biology 17: 142.

Chaikam, V. and D.T. Karlson. 2010. Comparison of structure, function and regulation of plant cold shock domain proteins to bacterial and animal cold shock domain proteins. BMB reports 43: 1-8.

Ciftci-Yilmaz, S. and R. Mittler. 2008. The zinc finger network of plants. Cellular and Molecular Life Sciences 65: 1150-1160.

Clement, M., A. Lambert, D. Herouart and E. Boncompagni. 2008. Identification of new up-regulated genes under drought stress in soybean nodules. Gene 426: 15-22.

Coumou, D. and S. Rahmstorf. 2012. A decade of weather extremes. Nature climate change 2: 491.

Crawford, R. 1992. Oxygen availability as an ecological limit to plant distribution. Advances in ecological research. Elsevier. p. 93-185.

Decottignies, A. and A. Goffeau. 1997. Complete inventory of the yeast ABC proteins. Nature genetics 15: 137.

Dordas, C., B.B. Hasinoff, J. Rivoal and R.D. Hill. 2004. Class-1 hemoglobins, nitrate and NO levels in anoxic maize cell-suspension cultures. Planta 219: 66-72.

Drew, M.C. 1997. Oxygen deficiency and root metabolism: injury and acclimation under hypoxia and anoxia. Annual review of plant biology 48: 223-250.

Drew, M.C., C.-J. He and P.W. Morgan. 2000. Programmed cell death and aerenchyma formation in roots. Trends in plant science 5: 123-127.

Du, Z., X. Zhou, Y. Ling, Z. Zhang and Z. Su. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic acids research 38: W64-W70.

Felle, H.H. 2006. Apoplastic pH during low-oxygen stress in barley. Annals of botany 98: 1085-1093.

Fukao, T., K. Xu, P.C. Ronald and J. Bailey-Serres. 2006. A variable cluster of ethylene response factor–like genes regulates metabolic and developmental acclimation responses to submergence in rice. The Plant Cell 18: 2021-2034.

Garg, R., M. Verma, S. Agrawal, R. Shankar, M. Majee and M. Jain. 2013. Deep transcriptome sequencing of wild halophyte rice, Porteresia coarctata, provides novel insights into the salinity and submergence tolerance factors. DNA research 21: 69-84.

Gibbs, J. and H. Greenway. 2003. Mechanisms of anoxia tolerance in plants. I. Growth, survival and anaerobic catabolism. Functional plant biology 30: 1-47.

Gibbs, J., S. Morrell, A. Valdez, T. Setter and H. Greenway. 2000. Regulation of alcoholic fermentation in coleoptiles of two rice cultivars differing in tolerance to anoxia. Journal of Experimental Botany 51: 785-796.

Guglielminetti, L., P. Perata and A. Alpi. 1995. Effect of anoxia on carbohydrate metabolism in rice seedlings. Plant Physiology 108: 735-741.

Hattori, Y., K. Nagai, S. Furukawa, X.-J. Song, R. Kawano, H. Sakakibara, et al. 2009. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature 460: 1026.

He, D. and P. Yang. 2013. Proteomics of rice seed germination. Frontiers in plant science 4: 246.

Hinz, M., I.W. Wilson, J. Yang, K. Buerstenbinder, D. Llewellyn, E.S. Dennis, et al. 2010. Arabidopsis RAP2. 2: an ethylene response transcription factor that is important for hypoxia survival. Plant Physiology 153: 757-772.

Hoeren, F.U., R. Dolferus, Y. Wu, W.J. Peacock and E.S. Dennis. 1998. Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (ADH1) by low oxygen. Genetics 149: 479-490.

Horchani, F., S. Aschi-Smiti and R. Brouquisse. 2010. Involvement of nitrate reduction in the tolerance of tomato (Solanum lycopersicum L.) plants to prolonged root hypoxia. Acta physiologiae plantarum 32: 1113-1123.

Hossain, M.A., E. Huq, A. Grover, E.S. Dennis, W.J. Peacock and T.K. Hodges. 1996. Characterization of pyruvate decarboxylase genes from rice. Plant molecular biology 31: 761-770.

Howell, K.A., R. Narsai, A. Carroll, A. Ivanova, M. Lohse, B. Usadel, et al. 2009. Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. Plant physiology 149: 961-980.

Huang, J., S. Sun, D. Xu, H. Lan, H. Sun, Z. Wang, et al. 2012. A TFIIIA-type zinc finger protein confers multiple abiotic stress tolerances in transgenic rice (Oryza sativa L.). Plant molecular biology 80: 337-350.

Huang, J., S.-J. Sun, D.-Q. Xu, X. Yang, Y.-M. Bao, Z.-F. Wang, et al. 2009. Increased tolerance of rice to cold, drought and oxidative stresses mediated by the overexpression of a gene that encodes the zinc finger protein ZFP245. Biochemical and Biophysical Research Communications 389: 556-561.

Huang, J., X. Yang, M.-M. Wang, H.-J. Tang, L.-Y. Ding, Y. Shen, et al. 2007. A novel rice C2H2-type zinc finger protein lacking DLN-box/EAR-motif plays a role in salt tolerance. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression 1769: 220-227.

Huang, S., N.L. Taylor, R. Narsai, H. Eubel, J. Whelan and A.H. Millar. 2009. Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. Plant physiology 149: 719-734.

Huang, W., H. Lv and Y. Wang. 2017. Functional characterization of a novel R2R3-MYB transcription factor modulating the flavonoid biosynthetic pathway from Epimedium sagittatum. Frontiers in Plant Science 8: 1274.

Ismail, A.M. 2013. Flooding and submergence tolerance. Genomics and breeding for climate-resilient crops. Springer. p. 269-290.

Ismail, A.M., E.S. Ella, G.V. Vergara and D.J. Mackill. 2008. Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (Oryza sativa). Annals of Botany 103: 197-209.

Ismail, A.M., U.S. Singh, S. Singh, M.H. Dar and D.J. Mackill. 2013. The contribution of submergence-tolerant (Sub1) rice varieties to food security in flood-prone rainfed lowland areas in Asia. Field Crops Research 152: 83-93.

Jackson, M.B. 1985. Ethylene and responses of plants to soil waterlogging and submergence. Annual review of plant Physiology 36: 145-174.

Karlson, D. and R. Imai. 2003. Conservation of the cold shock domain protein family in plants. Plant Physiology 131: 12-15.

Kawai, M. and H. Uchimiya. 2000. Coleoptile senescence in rice (Oryza sativa L.). Annals of Botany 86: 405-414.

Kiełbowicz-Matuk, A. 2012. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. Plant Science 185: 78-85.

Kirk, G., H. Greenway, B. Atwell, A. Ismail and T. Colmer. 2014. Adaptation of rice to flooded soils. Progress in botany. Springer. p. 215-253.

Klok, E.J., I.W. Wilson, D. Wilson, S.C. Chapman, R.M. Ewing, S.C. Somerville, et al. 2002. Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. The Plant Cell 14: 2481-2494.

Kretzschmar, T., M.A.F. Pelayo, K.R. Trijatmiko, L.F.M. Gabunada, R. Alam, R. Jimenez, et al. 2015. A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. Nature Plants 1: 15124.

Kreuzwieser, J., J. Hauberg, K.A. Howell, A. Carroll, H. Rennenberg, A.H. Millar, et al. 2009. Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia. Plant physiology 149: 461-473.

Lasanthi-Kudahettige, R., L. Magneschi, E. Loreti, S. Gonzali, F. Licausi, G. Novi, et al. 2007. Transcript profiling of the anoxic rice coleoptile. Plant Physiol 144: 218-231. doi:10.1104/pp.106.093997.

Lee, Y.-H., K.-S. Kim, Y.-S. Jang, J.-H. Hwang, D.-H. Lee and I.-H. Choi. 2014. Global gene expression responses to waterlogging in leaves of rape seedlings. Plant cell reports 33: 289-299.

Limami, A.M., H. Diab and J. Lothier. 2014. Nitrogen metabolism in plants under low oxygen stress. Planta 239: 531-541.

Liu, F., T. VanToai, L.P. Moy, G. Bock, L.D. Linford and J. Quackenbush. 2005. Global transcription profiling reveals comprehensive insights into hypoxic response in Arabidopsis. Plant physiology 137: 1115-1129.

Liu, Q., Q. Zhang, R.A. Burton, N.J. Shirley and B.J. Atwell. 2010. Expression of vacuolar H+-pyrophosphatase (OVP3) is under control of an anoxia-inducible promoter in rice. Plant molecular biology 72: 47.

Lurin, C., C. Andrés, S. Aubourg, M. Bellaoui, F. Bitton, C. Bruyère, et al. 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. The Plant Cell 16: 2089-2103.

Lv, Y., M. Yang, D. Hu, Z. Yang, S. Ma, X. Li, et al. 2017. The OsMYB30 transcription factor suppresses cold tolerance by interacting with a JAZ protein and suppressing β-amylase expression. Plant physiology 173: 1475-1491.

Magneschi, L., R. Kudahettige, A. Alpi and P. Perata. 2009. Comparative analysis of anoxic coleoptile elongation in rice varieties: relationship between coleoptile length and carbohydrate levels, fermentative metabolism and anaerobic gene expression. Plant Biology 11: 561-573.

Magneschi, L. and P. Perata. 2008. Rice germination and seedling growth in the absence of oxygen. Annals of Botany 103: 181-196.

Martinoia, E., M. Klein, M. Geisler, L. Bovet, C. Forestier, Ü. Kolukisaoglu, et al. 2002. Multifunctionality of plant ABC transporters–more than just detoxifiers. Planta 214: 345-355.

Mattana, M., I. Coraggio, A. Bertani and R. Reggiani. 1994. Expression of the enzymes of nitrate reduction during the anaerobic germination of rice. Plant Physiology 106: 1605-1608.

Miller, G., V. Shulaev and R. Mittler. 2008. Reactive oxygen signaling and abiotic stress. Physiologia plantarum 133: 481-489.

Moons, A. 2003. Ospdr9, which encodes a PDR-type ABC transporter, is induced by heavy metals, hypoxic stress and redox perturbations in rice roots 1. Febs Letters 553: 370-376.

Morard, P., J. Silvestre, L. Lacoste, E. Caumes and T. Lamaze. 2004. Nitrate uptake and nitrite release by tomato roots in response to anoxia. Journal of plant physiology 161: 855-865.

Nakaminami, K., A. Matsui, K. Shinozaki and M. Seki. 2012. RNA regulation in plant abiotic stress responses. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1819: 149-153.

Nanjo, Y., K. Maruyama, H. Yasue, K. Yamaguchi-Shinozaki, K. Shinozaki and S. Komatsu. 2011. Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings. Plant molecular biology 77: 129-144.

Narsai, R., J.M. Edwards, T.H. Roberts, J. Whelan, G.H. Joss and B.J. Atwell. 2015. Mechanisms of growth and patterns of gene expression in oxygen-deprived rice coleoptiles. The Plant Journal 82: 25-40.

Narsai, R., K.A. Howell, A. Carroll, A. Ivanova, A.H. Millar and J. Whelan. 2009. Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. Plant Physiology 151: 306-322.

Narsai, R., S.R. Law, C. Carrie, L. Xu and J. Whelan. 2011. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and

organelle metabolism that are essential for germination in Arabidopsis. Plant physiology 157: 1342-1362.

Narsai, R., C. Wang, J. Chen, J. Wu, H. Shou and J. Whelan. 2013. Antagonistic, overlapping and distinct responses to biotic stress in rice (Oryza sativa) and interactions with abiotic stress. BMC genomics 14: 93.

Perata, P. and A. Alpi. 1991. Ethanol-induced injuries to carrot cells: the role of acetaldehyde. Plant Physiology 95: 748-752.

Perata, P. and A. Alpi. 1993. Plant responses to anaerobiosis. Plant science 93: 1-17.

Perata, P., L. Guglielminetti and A. Alpi. 1997. Mobilization of endosperm reserves in cereal seeds under anoxia. Annals of Botany 79: 49-56.

Pompeiano, A., F. Fanucchi and L. Guglielminetti. 2013. Amylolytic activity and carbohydrate levels in relation to coleoptile anoxic elongation in Oryza sativa genotypes. Journal of plant research 126: 787-794.

Ruwe, H., C. Kupsch, M. Teubner and C. Schmitz-Linneweber. 2011. The RNA-recognition motif in chloroplasts. Journal of plant physiology 168: 1361-1371.

Saab, I.N. and M.M. Sachs. 1996. A flooding-induced xyloglucan endo-transglycosylase homolog in maize is responsive to ethylene and associated with aerenchyma. Plant Physiology 112: 385-391.

Sánchez-Fernández, R.o., T.E. Davies, J.O. Coleman and P.A. Rea. 2001. The Arabidopsis thaliana ABC protein superfamily, a complete inventory. Journal of Biological Chemistry 276: 30231-30244.

Sauter, M. 2000. Rice in deep water:" How to take heed against a sea of troubles". Naturwissenschaften 87: 289-303.

Setter, T., E. Ella and A. Valdez. 1994. Relationship between coleoptile elongation and alcoholic fermentation in rice exposed to anoxia. II. Cultivar differences. Annals of Botany 74: 273-279.

Shingaki-Wells, R.N., S. Huang, N.L. Taylor, A.J. Carroll, W. Zhou and A.H. Millar. 2011. Differential molecular responses of rice and wheat coleoptiles to anoxia reveal novel metabolic adaptations in amino acid metabolism for tissue tolerance. Plant physiology 156: 1706-1724.

Sun, S.-J., S.-Q. Guo, X. Yang, Y.-M. Bao, H.-J. Tang, H. Sun, et al. 2010. Functional analysis of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice. Journal of experimental botany 61: 2807-2818.

Tak, H., S. Negi and T. Ganapathi. 2017. Overexpression of MusaMYB31, a R2R3 type MYB transcription factor gene indicate its role as a negative regulator of lignin biosynthesis in banana. PloS one 12: e0172695.

Vijayan, J., S. Senapati, S. Ray, K. Chakraborty, K.A. Molla, N. Basak, et al. 2018. Transcriptomic and physiological studies identify cues for germination stage oxygen deficiency tolerance in rice. Environmental and experimental botany 147: 234-248.

Voesenek, L., T. Colmer, R. Pierik, F. Millenaar and A. Peeters. 2006. How plants cope with complete submergence. New phytologist 170: 213-226.

Vriezen, W.H., R. Hulzink, C. Mariani and L.A. Voesenek. 1999. 1-Aminocyclopropane-1-carboxylate oxidase activity limits ethylene biosynthesis in Rumex palustrisduring submergence. Plant Physiology 121: 189-196.

Wang, F.-Z., M.-X. Chen, L.-J. Yu, L.-J. Xie, L.-B. Yuan, H. Qi, et al. 2017. OsARM1, an R2R3 MYB transcription factor, is involved in regulation of the response to arsenic stress in rice. Frontiers in plant science 8: 1868.

Wei, Q., Q. Luo, R. Wang, F. Zhang, Y. He, Y. Zhang, et al. 2017. A wheat R2R3-type MYB transcription factor TaODORANT1 positively regulates drought and salt stress responses in transgenic tobacco plants. Frontiers in plant science 8: 1374.

Wrzaczek, M., J. Vainonen, A. Gauthier, K. Overmyer and J. Kangasjärvi. 2011. Reactive oxygen in abiotic stress perception-from genes to proteinsInTech.

Xu, D.-Q., J. Huang, S.-Q. Guo, X. Yang, Y.-M. Bao, H.-J. Tang, et al. 2008. Overexpression of a TFIIIA-type zinc finger protein gene ZFP252 enhances drought and salt tolerance in rice (Oryza sativa L.). FEBS letters 582: 1037-1043.

Xu, Z.-S., K. Feng, F. Que, F. Wang and A.-S. Xiong. 2017. A MYB transcription factor, DcMYB6, is involved in regulating anthocyanin biosynthesis in purple carrot taproots. Scientific reports 7: 45324.

Yazaki, K., N. Shitan, A. Sugiyama and K. Takanashi. 2009. Cell and molecular biology of ATP-binding cassette proteins in plants. International review of cell and molecular biology 276: 263-299.

Zhou, Y., P. Yang, F. Cui, F. Zhang, X. Luo and J. Xie. 2016. Transcriptome analysis of salt stress responsiveness in the seedlings of Dongxiang wild rice (Oryza rufipogon Griff.). PLoS One 11: e0146242.

CHAPTER VI

SUMMARY

Flash flooding due to unpredicted rainfall and chilling stress are two major abiotic stresses that are detrimental to rice germination, growth, development and productivity. In this study, we screened more than 250 rice accessions belonging to different subspecies of rice and evaluated different phenotypic traits under chilling stress and hypoxia stress. A wide range of phenotypic variations was observed among the rice panel used in this study. Recently, association mapping has been widely used to identify candidate genes underlying complex quantitative traits. In our study, we used Compressed Mixed Linear Model (CMLM) of GAPIT using both population structure and Kinship relatedness to conduct GWAS study for identification of QTLs controlling chilling stress tolerance and hypoxia stress tolerance during germination stage in rice. From the GWAS study, we identified 27 and 13 novel QTLs associated with traits measured under chilling stress and hypoxia stress respectively. The results of our study provided useful information on genetic architecture underlying cold tolerance and hypoxia stress tolerance during rice germination which in turn can be used for further molecular study and crop improvement. Two contrasting lines based on the phenotypic evaluation under both stresses were selected and used for whole genome transcriptome profiling study. The differential gene expression analysis of this study revealed genetic and molecular differences between the two contrasting rice lines under chilling stress and hypoxia stress. Our transcriptomics study of two contrasting lines under hypoxic stress highlighted the importance of the genes involved in carbohydrate metabolism, pyrophosphate-dependent energy conservation, ethylene signaling pathway, nitrogen metabolism, and nucleotide binding for submergence tolerance during germination in rice. The transcriptomics study under

chilling stress indicated that many genes involved in signal transduction, phytohormone (auxin, GA), abiotic stress, anti-oxidant system and unknown function genes, were differentially regulated under cold stress. Our findings have increased the understanding of molecular and physiological basis of the rice responses under cold stress and hypoxia stress. Both the GWAS study and transcriptomics study have provided the basis for identification of potential candidate genes involved in hypoxia stress and cold stress tolerance in rice. In future, validation for some target genes of interest can be done using CRISPR/Cas9 system or other approaches. The results of our study may also be useful for development of DNA markers for marker-assisted breeding.