WHOLE GENOME TRANSCRIPTOME PROFILING FOR HEAT TOLERANCE AND GENOME-WIDE ASSOCIATION FOR EXOTIC TRAITS IN RICE (*ORYZA SATIVA* L.)

A Thesis

by

WARDAH KHURSHIDA MUSTAHSAN

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

MASTER OF SCIENCE

Chair of Committee,	Michael J. Thomson
Committee Members,	Lee Tarpley
	Endang M. Septiningsih
Head of Department,	David D. Baltensperger

May 2018

Major Subject: Plant Breeding

ABSTRACT

High night temperature (HNT) has strong negative effects on rice plant growth and development. HNT also impacts many physiological characteristics of rice which affect the grain quality of rice grown around the world. One potential mechanism of HNT damage is from the induction of ethylene-triggered reactive oxygen species that can lead to increased membrane damage and negatively impact yield and grain quality. In this study, the changes in physiological behavior due to the interaction between HNT and the ethylene-inhibitor 1-MCP was investigated. Furthermore, genome-wide expression analysis under HNT was performed using RNA-Seq to gain insights into the gene functions underlying tolerance to HNT. Plants were grown under ambient night temperature (ANT) (25 °C) or HNT (30 °C) with or without 1-MCP treatment. RNA extraction was performed on two phenotype-contrasting rice cultivars (Antonio and Colorado) from which in-depth RNA-Seq analysis was used to identify differentially expressed genes involved in heat tolerance in these varieties. Results from this experiment showed a total of 25 transcripts derived from analyzing the effects of various comparisons of treatments on the genotypes used in this study. From these findings we conclude that notable transcripts in this subset played a role in molecular mechanisms pertaining to ethylene interaction and HNT.

High temperature environments are fairly innocuous for some exotic rice varieties; however, these genetic donors for heat tolerance often have various undesirable traits, including red pericarp, black hulls, and awns. To improve the efficiency of using these exotic accessions in

modern rice breeding programs, it would be benficial to develop improved genetic donors by eliminating these exotic traits and thus preventing negative linkage drag when using these donors as parents in a crossing program. Recent advances in CRISPR/Cas9 genome editing can now enable the rapid knock-out of genes underlying negative traits in rice. To gain further insight in the genetic loci controlling these traits, a genome-wide association study (GWAS) was performed on a diversity panel consisting of approximately 300 rice accessions. Traits of interest in this GWAS study included pericarp color, hull color, awn color, and awn length. The accessions were genotyped with an Illumina 7K rice SNP chip to identify genetic loci that control these traits. Results from this GWAS study showed various significant SNPs for exotic rice traits, awn color (chr: 9 & 10), awn length (chr: 4,6,7, 10, and 12), hull color (chr: 12), endosperm color (chr: 1 & 3). These findings may lead to the conclusion that may potentially be novel QTLs. When combined with data on the chromosomal location of known major genes affecting exotic traits, these results can guide the development of improved HNT-tolerant genetic donors for future stress-tolerance breeding programs.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Michael Thomson, and my committee members, Dr. Lee Tarpley, and Dr. Endang Septiningsih. In addition, I would like to extend my sincerest gratitude to Dr. Charles Johnson, Dr. Shichen Wang, Dr. Richard Metz, and Dr. Upendra Devisetty for their guidance and support throughout the course of this research.

I would also like to thank several key members of our lab Dr. Backki Kim, Yuya Liang, Ranjita Thapa, and Dr. Nithya Subramanian for helping overcome various challenges present in this project. Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement, patience, and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Dr. Michael Thomson (chair), Dr. Lee Tarpley, Dr. Endang Septiningsih and Dr. David D. Baltensperger (department head) of the Department of Soil and Crop Sciences.

The data analyzed for Chapter II (HNT experiment) was provided by Dr. Lee Tarpley and Dr. Abdul Razack Mohammed (Texas A&M AgriLife Research Center at Beaumont). The analyses depicted in Chapter II (RNA-Seq) were conducted in part by Dr. Charles Johnson, Director of Texas A&M AgriLife Genomics and Bioinformatics Services, and Dr. Shichen Wang of the Center of Bioinformatics and Genomics Systems Engineering.

All other work for the dissertation was completed independently by the student, Wardah K. Mustahsan.

Funding Sources

Funding for this project was provided by Texas A&M University, Texas A&M AgriLife Research, and the Texas A&M Genomics Seed Grant "Whole genome transcript profiling for heat tolerance in rice" to M. Thomson and L. Tarpley.

NOMENCLATURE

1-MCP	1-Methylcyclopropene
An-1/An-2	Gene for Awn
ANT	Ambient Nighttime Temperature
BAM	Binary Alignment Map
Bh4	Gene for Black Hull
Chr	Chromosome
CRISPR	Clustered Regulatory Interspaced Short Palindromic Repeats
CSSL	Chromosome Segment Substitution Lines
DEG	Differentially Expressed Gene
DGE	Differential Gene Expression
DNA	Deoxyribonucleic Acid
ETC	Electron Transport Chain
FC	Fold-Change
FDR	False Discovery Rate
GBS	Genotyping by Sequencing
GLM	General Linearized Model
GO	Gene Ontology
GTF	Gene Transfer Format
GWAS	Genome-wide Association Study
HNT	High Nighttime Temperature

HT	High Temperature
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MLM	Mixed Linear Model
mRNA	Messenger Ribonucleic Acid
NT	Night Temperature
nTAR	Novel Transcriptional Active Regions
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Loci
RAE	Regulator of Awn Elongation
Rc	Gene for Red Pericarp
RCBD	Randomize Complete Block Design
RNA	Ribonucleic Acid
SAM	Sequence Alignment Map
sgRNA	Single Guide Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
STAR	Spliced Transcripts Aligned to a Reference
TASSEL	Trait Analysis by Association Evolutionary Linkage

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Rice (*Oryza sativa*) is one of the most essential grains produced around the world after wheat. It serves as a principal source of food for more than 3 billion people in different countries around the world. In addition, it is also a major source of calories and carbohydrate (35-75%) intake (Krishan et al., 2011). Since rice is the vital staple crop for an enormous portion of the world population, an increase in rice production is essential to support the rapidly increasing population. Rice is categorized as a semiaquatic annual, self-pollinating grass, belonging to the genus of *Oryza* (Family: Poaece). As a crop, rice is grown in many different countries around the world, mainly Asia, under various ecosystems and cultivation techniques (Global Rice Science Partnership, 2013). There are two major cultivated rice species grown around the world: *O. sativa* (worldwide) and *O. glabberima* (Africa). Geographically there are two significant subspecies groups within the cultivated rice species *Oryza sativa*: *indica* (grown in the tropics) and *japonica* (grown in temperate and tropical upland environments). Human domestication has produced thousands of diverse rice varieties with immense degrees genetic variation that can be used to develop improved rice cultivars (Global Rice Science Partnership, 2013).

Rice goes through three stages of growth: vegetative, reproductive, and ripening (maturity). It takes 3–6 months for a rice plant to grow to maturity depending on both the variety (genotype) and environment. Rice completes the vegetative and reproductive growth phases during this time frame. The vegetative stage is subdivided into germination, early seedling growth, and tillering (Global Rice Science Partnership, 2013).

I.1: Heat Stress: High Nighttime Temperature

High nighttime temperature (HNT) has become a detrimental factor that limits the yield and grain quality of rice globally. According to Peng et al., 2004, "World rice production must increase by 1% annually to meet the growing demand for food due to population growth and economic development. But, due to the drastic changes in the climate today the challenge of improving agronomical traits in crops has risen. Grain yields decline by 10% for each 1°C increase in growing-season minimum temperature in the dry season, whereas the effect of maximum temperature on crop yield was insignificant." Different characteristics that determine rice yield include: the number of productive tillers, spikelet fertility, and grain dimensions. HNT affects various types of plant physiolotical characteristics such as chlorophyll concentration, chlorophyll fluorescence, and photosynthesis. It also has a negative impact on respiration and membrane stability which can cause a decrease in yield and spikelet fertility which can reduce overall rice production (Mohammed et al., 2015). High nighttime temperature also dramatically affects various grain quality traits, especially chalkiness, which negatively affects consumers and the farming community who are dependent on good quaily grain for consumption purposes, as well as the economic markets dependent on rice production.

I.2: Heat Stress: Grain Quality

Traits such as rice grain yield, dimensions, and quality tend to decrease under high heat stress conditions (Counce et al. 2005, Ambardekar et al. 2011). In communities where rice is a major commodity, grain quality is a key element that defines market value and initiation into new cultivar development. Grain quality traits are defined by: physical characteristics, cooking value, and nutritional value (Fitzgerald et al., 2009). Gaining a deeper understanding of these

factors with not only improve overall grain quality of rice, but will also establish a starting point for future breeding programs focusing on yield, climate conditions (heat stress), and economic value (Fitzgerald et al., 2009). The advancement in genetic and genomic techniques can provide information which can be useful in establishing improved cultivars that can withstand heat stress conditions and achieve the goal of producing new high-quality HNT-tolerant varieties.

I.3: Transcriptomics

Transcriptomics is an area of study where whole transcriptome changes can be analyzed under various biological conditions or environments. It is a technology that has been defined by the development of new techniques and tools over time making previous techniques outdated (Lowe et al., 2017). The first attempt was published in 1991,where a partial human transcriptome was analyzed and reported 609 mRNA sequences from the human brain (Adams et al., 1991). The transcriptomes of organisms are constantly evolving to the microcosmic levels for various cells, tissues, and genetic material (Mele et. al 2015, Sandberg et. al 2014, Kolodziejczyk et. al 2015). In the 1980s, low-throughput Sanger sequencing was the most popular technology being used at the time to sequence random individual transcripts, called expressed sequence tags (ESTs) (Pan et al 2008, Sutcliffe et al 1982, Putney et al 1983, Marra et al, 1998). High-throughput techniques eventually replaced the Sanger method when they became openly available for use. Transcriptome analysis is vital for understanding the functional and molecular components of the genome that control the functionality mechanisms of cells and tissues across various organisms. It can be very useful in agriculture for understanding various traits such as yield, abiotic stress, and biotic stress tolerance (Lowe et al., 2017).

There have been studies using transcriptomics to analyzed various abiotic stresses and anatomical mechanisms that affect rice production. Walia et. al (2005) published a

transcriptomics study focusing on salinity stress in rice. In that study, two *indica* rice genotypes, FL478 (a salt tolerant RIL), and IR29 (salt susceptible parent) were used. Transcriptome analysis was performed using an Affymetrix rice genome DNA microarray containing 55,515 probes investigating the contrasting genotypes transcriptomes under various salt stress and unstressed treatments in the vegetative stage. Salinity stress conditions introduced a number of genes involved in the flavonoid biosynthesis pathway present in the IR29 variety. Findings produced in this study showed that cell wall-related genes were highly expressed in both genotypes, indicating that the transcripts related to this trait formed an adaptive defense mechanism under salt stress conditions. Additionally, genes expressed were also mapped to the Saltol gene (salt-tolerance: chromosome 1) present among both genotypes. The findings from this study presented a genome-wide transcriptomics analysis of two genotypically related rice cultivars differing in the degree of salinity tolerance during various salt stress treatments in greenhouse conditions. In contrast to salinity tolerance, understanding heat stress through transcriptomics is a topic that is not well-studied. The application of transcriptomics into heat stress will expand our understanding on the capacity of how *O. sativa* species can handle heat stress in various environmental conditions.

I.4: Whole Genome Transcript Profiling with RNA-Seq

RNA sequencing (RNA-Seq) is an innovative next-generation sequencing technology that can provide a broader understanding of differential gene expression under various environmental conditions, chemical treatments, and genotypes under heat stress conditions (Chougule, 2017). RNA-Seq isolates mRNA from the target tissues under different treatments and used for library preparation. RNA-Seq can provide deeper insights into various genetic

functions such as, alternative gene splicing, post-transcriptional modifications, and differentially expressed genes. Therefore, this tool can offer more in-depth knowledge on the gene structure, expression patterns of genes across various genotypes, and treatments (Chougule, 2017). To better analyze and visualize the parameters in this experiment, different bioinformatics pipelines (Tuxedo, Salmon, and DESeq) provided from the CyVerse platform will be used in this study.

Understanding the transcriptomes is essential for deciphering the functional and molecular entities of the genome in the gene expression process in different tissues of an organism. Wang et al., 2009 defines that, "The principal aims of transcriptomics are:

1. To catalog all species of transcripts, including mRNAs, non-coding RNAs, and small RNAs.
2. To determine the transcriptional structure of genes, regarding their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications.
3. To quantify the changing expression levels of each transcript during development and under different conditions.

Following sequencing, the raw reads obtained from the sequencing results are aligned to a reference genome (*Oryza sativa* IRGSP 1.0), to produce a transcription map which provides information on both the transcriptional structure and variation in gene expression." Various bioinformatic tools can be used to interpret the vast amount of data produced from the RNA-Seq library construction. Dobin et al., 2013 states, "In transcript analysis, two primary tasks that will are accomplished:

1. An accurate alignment of reads that contain mismatches and indels caused by genomic/sequencing variations or errors.

2. Mapping sequences derived from genomic regions consisting of spliced sequence modules that are joined together to make spliced RNAs.

The second task is specific and crucial to RNA-Seq analysis, as it provides the connectivity information needed to reconstruct the full extent of spliced RNA molecules."

Lu et al. (2010) utilized RNA-Seq to analyze the rice transcriptome at a single nucleotide. RNA-Seq was used to look at transcripts from cultivated rice Indica (93-11 & Gla4) and Japonica subspecies (Nipponbare) in order to establish whole-genome transcription profiles for the respective genotypes. Findings from this study produced 15,708 novel transcriptional active regions (nTARs), with >63% samples having putative single-exon transcripts. At the time of this study the available rice gene models showed, 83.1% (46,472 genes) were validated with this model by RNA-Seq. Transcriptome analysis among the two genotypes presented 3464 genes with variation in differential expression. Through this analysis various levels of single-nucleotide polymorphisms (SNPs) were also detected. Among the two-rice subspecies, 67,011 SNPs existed between 93-11 and Nipponbare, and 64,481 SNPs between Gla4 and Nipponbare were identified (http://www. ncgr.ac.cn/english/ .edatabase.htm). The findings showed only half of these SNPs were located in 16,597 annotated gene model.

Zhang et. al (2010), did an in-depth RNA-Sequencing study for a single base pair with the purpose of depicting the intricacy of the rice transcriptome. In this study transcriptomes derived from Illumina sequencing technology are designed for eight organs of cultivated rice (*Oryza sativa L. ssp. indica* cv. 9311) such as, callus, flowering panicle, and filling panicle. stage. RNA-sequencing results, presented a total of >410 million paired-end reads of 35–75 bp in length. Furthermore, >5-million single- end reads from each of the eight organs aligned

to the reference genome of *Oryza sativa subsp. indicia* (The Beijing Gene Finder, http://bgf.genomics.org.cn/). In this portion of the study 73% of the reads were uniquely mapped to the reference genome. Overall deep sequencing of the rice transcriptome covered approximately 99.7% (32,959) of the rice DNA data (Kikuchi et al., 2003). Alternative splicing analysis showed 33% cis-splicing being present in all genes and 234 putative chimeric transcripts identified during trans-splicing. The findings from this study showed the overall complexity of transcriptional regulation in rice.

I.5: Genome-wide Association Study: Exotic Traits in Rice

In 2007, a study on the complexity of human disease is analyzed through the use of a SNP was published by WTCCC and recognized as the first authentic genome-wide association study (GWAS). Hartl et al., 1997 states, "Genome-wide association studies are based on the principle of linkage disequilibrium (LD) at the population level. LD is the non-random association between alleles at different loci. It is created by evolutionary forces such as mutation, drift, and selection and is broken down by recombination." Genome-wide association studies focusing on organisms has promoted further discoveries about genetic mechanisms involved in complex traits and has further encouraged basic research in genetics and genomics (Visscher et al., 2012).

Zhao et al., 2011 states, "In rice GWAS gives an advantage as most rice varieties are homozygous in nature, which makes it possible to employ a 'genotype or sequence once and phenotype many times over ' strategy, whereby once the lines are gnomically characterized, the genetic data can be reused many times over across different phenotypes and environments."

Zhao et al. (2011) did a genome-wide association study with the goal of understanding the genetic basis of diverse physiological, developmental, and morphological traits that provide the basis for improving yield, quality and sustainability of rice. In this study 44,100 SNP variants (originally from two sources: the Oryza SNP project and BAC clone Sanger sequencing of wild species) were genotyped across 413 diverse accessions of *O. sativa* collected from 82 countries that were systematically phenotyped for 34 traits. From the 44,100 SNP variants 34,454 (~ 78 %) have a minor allele frequency > 0.05 across the panel. Principal component analysis on this data set showed clear clusters of the five subpopulations *indica*, *aus* , *temperate japonica*, *tropical japonica* and *aromatic* based on the top four PCs. Principal component analysis shows that PC1 separates the samples into *Indica* and *Japonica* (34 % of the genetic variance), PC2 separates the *indica* subgroup from the *aus* subgroup (10 % of the variance), PC3 separates the two *japonica* groups into *temperate* and *tropical* components (~ 6 % of the variance), and PC4 identifies the *aromatic* group as a clear and distinct gene pool (~ 2 % of the variance). The mixed model was used to analyze the associations between 34 phenotypes and 44 K SNP genotypes evaluated in our 413 *O. sativa* rice lines, and identified various associations such as candidate genes and QTLs (Kang et. al, 2008). Cross-population-based mapping identified many common variants that influence many complex traits. Significant levels of heterogeneity were identified in the genetic architecture associated with the subpopulation and its response to the environment (Zhao et al., 2011).

Begum et al., (2015) performed an association mapping study of 19 agronomic traits dealing with yield components derived from a breeding population of elite tropical rice breeding lines. This population was genotyped with 71,710 SNPs using genotyping-by-sequencing (GBS), and GWAS techniques to speed up the process of selection in a breeding program. In this panel

52 QTLs were identified for 11 agronomic traits, including QTLs that effect flowering time and grain length/grain width/grain-length-breadth ratio. Haplotypes identified in this study can be utilized to select for plants with short height (plant height), early flowering time, and high yield. In order to obtain desirable traits by removing undesirable traits indicates the importance of utilizing tools such as association mapping in breeding programs. Following the example of the studies above GWAS can also be very beneficial in removing undesirable traits from certain types of rice varieties such as exotic rice which possess traits such as red pericarp, black hull, and awns. These traits hinder crop production and improvement for rice production around the world. More profound studies in these traits using genome-wide association studies can expand our understanding on the roles these traits play.

I.6: Single Nucleotide Polymorphism (SNPs)

SNP markers have emerged as one of the most potent marker systems for crop genetic studies especially for the study of closely related varieties or species in population genetics and genetic mapping (Bader, 2001). SNPs are single base changes in the genetic codes at a specified location on the chromosome and are the most abundant type of sequence variation in eukaryotic genomes (Batley et. al., 2003; Garg et al., 1999). The most popular genotyping platforms currently used for SNP validation are Illumina's Bead Array technology-based Golden Gate (GG) (Fan et al., 2003) and Infinium assays (Steemers et al., 2007), Life Technologies' TaqMan (Livak et al., 1995) assay coupled with OpenArray platform (TaqMan Open Array Genotyping system, Product bulletin) and KBiosciences' Competitive Allele Specific PCR (KASPar) combined with the SNP Line platform. These genotyping assays and platforms vary in their chemistry, cost and throughput of samples to genotype (Mammadov et al., 2012). Illumina platforms and 454 Roche platforms used second generation technologies to sequence the genomes of crop species (Perez-de-Castro et al., 2012).

SNPs can be applied in various types of molecular studies which can provide deeper information on the genetic basis of different traits in plants. Genome-wide association studies is a method often used to identify genes involved in particular diseases or traits. Typically, in a genome-wide association study hundreds of SNPs are examined in the entire genome to provide a more comprehensive manner which allows SNPs to detective the causing factors for traits.

Konshi et al., 2006 performed a study that demonstrated that a single-nucleotide polymorphism (SNP) in the regulatory region of the qSH1 gene is the cause loss of seed shattering in rice. Haplotype and GWAS analysis across various samples showed that the SNP was significantly associated with shattering among *japonica* varieties. In current rice breeding programs, this seed-shattering habit is still a target, especially in the construction of new *indica* cultivars. A QTL (quantitative trait locus) analysis between a cross consisting of a *indica* cultivar (shattering), Kasalath, and japonica cultivar (non-shattering), Nipponbare identified five QTLs on five chromosomes from this cross. There are three QTLs located on various Nipponbare alleles on chromosomes 1, 2, and 5, and Kasalath has two QTLs on chromosomes 11 and 12. All of the QTLs identified potentially contributed to shattering reduction, suggesting that loss of seed shattering may occur independently in *japonica* and *indica*. Results from a linkage analysis of 10,388 plants segregating at the qSH1 region and fine mapping of the qSH1, revealed that the SNPs are highly associated with the degree of seed shattering among *temperate japonica* rice cultivars. All tested *indica* cultivars exhibited strong seed shattering features based on the fact that they possess functional SNPs. The identified SNP may have been a mutation that occurred in early domestication of *japonica* subspecies (Konshi et al., 2006).

I.7: Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)

Currently there are three types of programmable genome-editing nucleases; these include zinc finger nucleases (ZFNs), transcription activator–like effector nucleases (TALENs), and most recently, RNA-guided nucleases (RGNs) from the clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR- associated proteins (Cas) system (Boch et al., 2009; Moscou and Bogdanove, 2009; Pabo et al., 2001; van der Oost, 2013). Among these programmable nucleases, the Cas9 nuclease in conjunction with the CRISPR subunit, has become the most simple and powerful tool for gene editing. CRISPR/Cas9 technology can be utilized to address biological questions at a molecular level and can promote innovative development in molecular genetics (Pennisi, 2013; Doudna and Charpentier, 2014; Hsu et al., 2014). The early application of this technology was used on the genomes of animals and bacteria (Cong et al., 2013; Hwang et al., 2013; Jiang et al., 2013a; Mali et al., 2013). In future studies it versatility and efficacy was demonstrated on various model crop systems which include *Arabidopsis*, rice, sorghum, and tobacco (Feng et al., 2013; Jiang et al., 2013; Li et al., 2013; Mao et al., 2013; Miao et al., 2013; Nekrasov et al., 2013; Shan et al., 2013; Xie and Yang, 2013). Currently, this modern gene editing tool is largely used in different plant and animal species, with many different of CRISPR/Cas9 vectors available to the public in a public plasmid repository of Addgene (http://www.addgene.org/crispr/plant/). CRISPR has transformed into an efficient and robust platform for gene editing which can significantly promote crop improvement across various crop systems and can in the long term increase global food production (Ding et al., 2016). For this study this tool can potentially be used to knock-out all undesirable exotic rice traits and improve the capacity of heat tolerance in rice plants in the near future.

CHAPTER II

RNA-SEQ OF DIFFERENTIAL GENE EXPRESSION ANALYSIS FOR HEAT TOLERANCE

II.1: Introduction

RNA-Seq refers to the sequencing of transcripts, where data produced indicates the number of counts from each transcript. The technique has been heavily influenced by the development of high-throughput sequencing technologies (Wang et al. 2009, Morozova et al. 2009). The nucleotide sequences generated from the use of this tool are typically around 100 bp in length but can range from 30 bp to over 10,000 bp in length, depending on the sequencing method used. RNA-Seq allows statistical reconstruction of the original RNA transcript by aligning reads from a sample set to a reference genome (Wang, 2009). As RNA-Seq is a quantitative method of analysis, it can determine RNA expression levels more accurately than older technology. One major advantage of RNA-Seq is that it can define the overall transcriptome parameters and across different sample sets.

There have been many successful applications of RNA-Seq in its' ability to accurately monitoring gene expression examples include: yeast meiosis, and mouse embryonic stem-cell differentiation. In these cases, RNA-Seq can keep track of any gene expression that may have occurred during the developmental stages of the organism and different tissues within the organism. (Wilhelm, B. T. et al, 2008, Mortazavi et al., 2008, Nagalakshmi et. al, 2008). These advances in RNA-Seq will undoubtedly be able to expand our understanding of transcriptomic dynamics during various physiological, developmental, molecular changes within an organism or various tissue samples. Data collected from these samples will provide in-depth robust comparison between various types of samples (Wang et. al, 2009).

Liao et al. (2015) performed a study utilizing Illumina sequencing to compare the transcriptome differences between heat tolerant and heat sensitive rice lines responding to high night temperatures during the endosperm development (early milky stage). Previous studies have reported that high nighttime temperatures are more harmful to grain weight characteristics in rice and other crops in comparison to high daytime temperatures (Li et al. 2011, Peng et al. 2004). Plant symptoms caused by high-temperature stress during the rice grain filling stage include: increased rice grain-filling rate, low amylose content, an increased chalkiness degree, and poor milling quality (Singh et al. 2012, Lu et al. 2013, Zhang et al. 2014). Key findings in Liao's study from the sequenced data showed 35 transcripts with different expression levels between heat-tolerant and heat-sensitive rice. Functional analysis through gene ontology of the DGEs indicated 21 genes whose functions are mainly involved in oxidation-reduction, metabolic processes, transcript regulation, and photosynthetic processes. Deeper analysis showed that high night temperature stress disrupts electron transport in the mitochondria, as well as other enzymatic and biochemical pathways in various plant cells (Liao et al., 2015).

RNA-Seq analysis is usually with done with a Tuxedo pipeline. Previous Tuxedo pipelines consist generally of Tophat, Bowtie, Cufflinks, Cuffmerge, and CummRbund. Current Tuxedo pipelines have improved in their efficiency to process data, faster core processors, and higher efficiency in sequencing large volumes of data. In this study, a more efficient and faster Tuxedo pipeline was used from the CyVerse platform. The pipeline consists of STAR Alignment, String-Tie, HT-Seq, Salmon, and edge-R/DESeq2.

II.2: RNA-Seq: Data Analysis Pipeline

The RNA-Seq data analysis pipeline has a number of components to process the raw data into gene expression information. Spliced Transcripts Alignment to a Reference (STAR), is a sequence alignment tool which is designed to address the current challenges in transcriptome analysis such as: high mapping error rate, alignment biases, and low mapping throughput. This tool shows a greater improvement in its ability of alignment precision and sensitivity than other types RNA-Seq aligners for various forms of data (Dobin, 2013).

Petra et al., 2015 defines: "StringTie is a transcriptome assembler, it assembles the genes for each dataset separately, estimating the expression levels of each gene and each isoform as it joins them. StringTie first groups the reads into clusters then creates a splice graph for each cluster from which it identifies transcripts, and then for each transcript, it creates a separate flow network to estimate its expression level using a maximum flow algorithm. As a tool, it has an improved accuracy of reproducing more complete transcriptomes by correctly identify 36–60% more transcripts than the next best assembler (Cufflinks) on multiple real and simulated data sets."

Another component of the bioinformatics pipeline used in this study is HTSeq-count. HTSeq-count is an application within HTSeq that processes RNA-Seq alignments for differential expression calling (Anders et al., 2014). The main application used in this study is Htseq-count. It reads counts for each gene keeping track on how many aligned reads overlap its exons using BAM (Binary Alignment Map) and GTF (Gene Transfer Format) files. The counts produced from HTSeq-count can be further utilized for differential gene expression (DGE) analysis using different visualization tools such as DESeq2 or edgeR through R-Studio, in this study the primary tool used is edgeR.

Accurate quantification of transcript-level abundance is an essential component of high-throughput RNA-Seq data analysis. Salmon is a transcript quantification software tool that possess high speed and accuracy (Patro et al., 2015). EdgeR is the last component of the pipeline which is available in R-Studio. This package performs differential gene analysis of count data (genes or transcripts) using false discovery rate, and fold changes to improve interpretability of the data produced for DGE. Like DESeq, edgeR will provide results on differential gene expression through various visual diagrams, such as scatter plots, histograms, and box plots, which can indicate information about the expression levels of differential genes.

II.3: High Nighttime Temperature & 1-Methylcyclopropene (1-MCP)

Current and future climate change conditions have introduced higher temperature conditions that have detrimental effects on various agronomical traits such as yield which can decrease rice production. One major type of heat stress that negatively impacts rice production is high nighttime temperature. High temperature (HT) is known to reduce rice grain yield and quality. Exposure to HT can increase the rate of grain filling but can reduce grain weight and yield. HT can also affect the quality of rice grain by increases chalky appearance resulting in a reduction in economic value. (Yoshida 1977, Tashiro 1991, Huang, 2000).

High night temperature (HNT) is a type of high-temperature condition that can induce the production of ethylene-triggered reactive oxygen species (ROS). This condition can lead to an increase in membrane damage, which can affect various variables such as consumption, production, and yield. A chemical agent that can inhibit the production of ROS is 1-methylcyclopropene. 1-MCP tricks the plant into “thinking” that it is not under stress by

competitively binding to the ethylene receptor, which causes a decrease in the chemical effects of ethylene and leads to an improvement in the output of agronomical traits (Mohammed et al., 2015). High temperatures tend to increase rate of grain growth but this reduces the time frame for grain filling, thus grain dimensions. High night temperature (HNT) is known to decrease rice yield (Cheng et al. 2009, 2010, Mohammed et al., 2009(a,b), 2010 Zakaria et al. 2002, Yamakawa et al. 2007, Fitzgerald and Resurreccion 2009) through effects on spikelet fertility and seed size.

II.4: Materials & Methods

II.4.1: Plant Materials

A collection of nine rice accessions was utilized in the high nighttime greenhouse experiment. In the GWAS study, approximately 330 rice germplasm accessions were characterized in this genetic diversity panel study for weedy rice. The rice seeds were received from the USDA-ARS National Small Grains Collection, Aberdeen, Idaho, USA.

II.5: High Nighttime Temperature

II.5.1: Greenhouse Experiment

The experimental design for this study is Random Complete Block Design (RCBD) with a total of 9 cultivars, of which 4 cultivars were used for the data analysis. For each variety, eight reps are exposed to ambient conditions (25°C) and a second set of 8 reps is exposed to HNT conditions (30°C). For the nighttime regime within each set of 8 reps, 4 plants receive 1-MCP treatment, and the other 4 plants were controls (without 1-MCP treatment). In total, there were 8 plants per night temperature (NT) x chemical treatment. All leaf samples in this study were analyzed for various types of physiological parameters.

Rice plants were grown in 3-liter size pots filled with a clay soil. Each pot had five seeds placed at a depth of 2.5cm. Once seedlings emerged, the plants were thinned down to one plant per pot.. At the 20th day after emergence, the boxes were filled with water to approximately 3 cm above the top of the soil in each pot. All the plants were maintained at ambient (25°C) conditions until the boot stage of the rice plants. At boot stage of the rice plant, half of the plants were selected randomly and moved under heat lamps (HNT: 30°C). The remaining half of the plants remained at ambient conditions. Within each temperature regime, randomly selected plants (half of the plants) were treated with 1-MCP at the boot stage of the rice plants. A reflective insulation with foil cover (ASTRO-E; Heartland Insulation Supply, Wichita, KS, USA) was placed over the water surface in the boxes to prevent direct infrared heating of the water (Mohammed et al., 2015).

The assignment of heat treatment to greenhouse location was random. The greenhouse was maintained at 25 °C NT and, within this, plants of the HNT treatment were subjected to an elevated NT (30 °C) through the use of nearly continuously controlled (sub-second response) infrared heaters (1100 W; Chromalox, Ogden, UT, USA), as described by Mohammed and Tarpley (2009c), starting at the boot stage of rice plants. The infrared heaters were positioned 1.0 m above the topmost part of the plants and provided infrared radiation enrichment. In ANT treatments, dummy heaters were provided to account for shading. The NT was imposed from 2000 h until 0600 h, starting from the boot stage and maintained until harvest. Temperature treatments in the greenhouse were monitored independently of the temperature-control system through the use of standalone sensor/loggers (HOBO, H08-003-02; Onset Computer Corporation, Bourne, MA, USA), which were placed a few centimeters into the canopy in both portions of the study. (Mohammed et al., 2015). Data collected from this study showed that

Antonio was a heat susceptible variety and Colorado was a heat tolerant variety (Table 1 & Table 2).

Table 1: Effects of night temperature and 1-MCP on rice yield.

Cultivars	ANT	HNT	% difference	ANT-1MCP	HNT-1MCP	% difference at ANT	% difference at HNT
Antonio	9.66 ± 1.09	4.36 ± 0.74	-55	10.31 ± 0.79	11.16 ± 0.72	NS	156
Cheniere	9.87 ± 0.72	3.60 ± 1.03	-64	10.29 ± 1.13	2.47 ± 0.26	NS	NS
CL-151	15.17 ± 0.83	8.63 ± 0.64	-43	15.98 ± 1.18	11.73 ± 1.16	NS	36
Colorado	9.86 ± 1.02	10.74 ± 1.83	NS	9.07 ± 1.06	9.57 ± 0.77	NS	NS

Table 2: Effects of night temperature and 1-MCP on rice spikelet fertility (%).

Cultivars	ANT	HNT	% difference	ANT-1MCP	HNT-1MCP	% difference at ANT	% difference at HNT
Antonio	79.7 ± 0.88	36.86 ± 11.33	-54	70.8 ± 5.57	71.93 ± 6.63	-11	95
Cheniere	66.94 ± 5.55	37.72 ± 8.94	-44	75.12 ± 6.83	14.47 ± 4.93	NS	-62
CL-151	91.76 ± 1.91	63.80 ± 9.42	-30	93.17 ± 2.13	73.6 ± 6.42	NS	NS
Colorado	81.94 ± 3.46	72.86 ± 1.92	-11	78.19 ± 8.71	62.72 ± 2.58	NS	-14

II.6: Materials & Methods

II.6.1: RNA-Sequencing

From the HNT greenhouse experiment, 32 leaf tissue samples were collected from Antonio and Colorado. Each individual accession was moved to the heat treatment based on their developmental progression. These samples are then collected at 5 days post exposure to high night temperatures, at booting stage (~65 days), and flash frozen in liquid nitrogen. Physiological data collected from the HNT greenhouse experiment depicted Antonio (heat susceptible) and Colorado (heat tolerant) as the two-contrasting heat tolerance rice varieties. Antonio and Colorado were selected not only for their contrasting HNT responses but also because they

represent Southern US germplasm which is understudied concerning heat stress. The RNA extraction was done using the Qiagen RNeasy Plant Mini kit. The layout for the treatment and samples for analysis were as follows:

- 2 x contrasting varieties (Colorado and Antonio)
- 2 x night temperatures (25˚C and 30˚C)
- 2 x with 1- MCP pre-treatment and control (without 1-MCP pre-treatment)
- 4 x biological replications

These RNA samples were submitted to Texas A&M AgriLife Genomics and Bioinformatics Service lab for RNA-Seq library preps and sequencing. The genome-wide transcriptome data was used to identify differentially-expressed genes (DEGs) between Colorado and Antonio as biological replicates, temperatures, and 1-MCP treatment.

Many pipelines are available to for performing differential gene expression analysis. For this study, an updated version of the Tuxedo pipeline was used which consists of STAR, String-Tie, HT-Seq Count, Salmon, and DESeq/Edge-R (Figure 1 & 2).

The Tuxedo pipeline is as follows:

1. Align the samples fastq reads to reference (Oryza sativa IRGSP 1.0) using STAR aligner.

2. Assemble transcripts from RNA-Seq reads aligned to the reference (Oryza sativa IRGSP 1.0) with StringTie.

3. Merge all StringTie transcripts into a single transcriptome annotation file using StringTie Merge.

4. Produce a feature count table that counts how many aligned sequencing reads map to the reference genome by utilizing ht-seq count package.

5. Use Salmon to perform transcriptome quantification which will require a FASTA file of the reference transcripts and FASTQ files containing reads of the samples.

6. Visualization analysis of differential gene expression using R package edgeR or DESeq2 (Chougule, 2017).

Data produced from this pipeline was used for transcriptomics analysis. In the transcriptomics analysis, various interactions among different parameters were analyzed. The first major subset of comparisons of interest is among contrasting rice varieties (Antonio/Colorado) and contrasting temperatures. In this setup, Antonio at 25˚C with no 1-MCP treatment is the reference, and Antonio at 30˚C with no 1-MCP was the manipulated treatment (same setup for Colorado). In this comparison, transcripts present in Antonio were identified that change only due to heat stress. The second major subset of comparisons of interest is among contrasting rice varieties (Antonio/Colorado) focusing on both heat stress and 1-MCP application parameters. These subsets of comparisons will contribute to this study for differential gene expression analysis. The results from this transcriptomics analysis were produced through the edgeR package in R-Studio.From this analysis, potential genes that are involved in the ethylene and non-ethylene response to heat stress can be identified. The results from this study can potentially help identify HNT transcripts expressed in Colorado due to different temperature and chemical

treatments in comparison to Antonio; these transcripts would be either indicators of major physiological responses to HNT, or potentially involved in/related to ethylene-signaling in response to HNT.

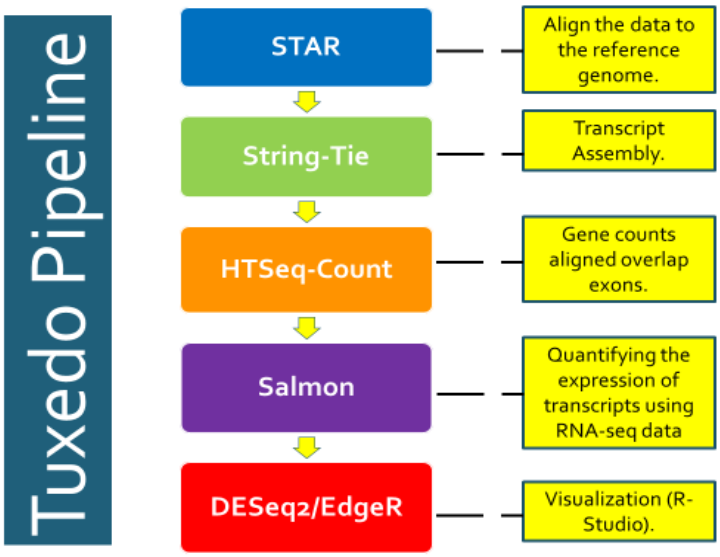

Figure 1: RNA-Seq: Tuxedo Pipeline RNA-Seq processing and analysis workflow for all 32 samples.

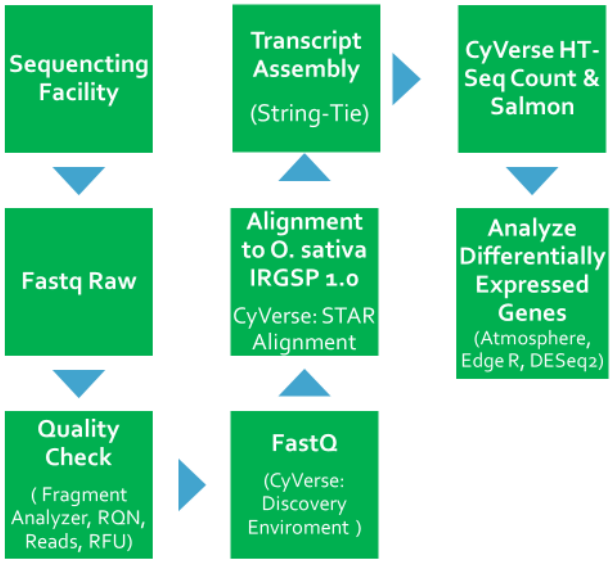

Figure 2: RNA-Seq processing and analysis workflow for all 32 samples.

Table 3: Individual sample raw reads through STAR alignment from the Tuxedo pipeline.

SAMPLE #	NUMBER OF INPUT READS
1	40059105
2	35210489
3	33493174
4	34378819
5	32466506
6	31837903
7	31711274
8	35264501
9	36084031
10	30380552
11	29284544
12	29284544
13	27259016
14	31648633
15	29763203
16	31263477
17	33417734
18	29228126
19	27764425
20	27513175
21	27185517
22	28564361
23	28087483
24	31413266
25	27437717
26	28818771
27	28782612
28	26788085
29	26937232
30	26937232
31	29101369
32	31413266

Table 4: Summary of Illumina sequencing of transcriptome analysis.

Category	Average
Number of input reads	30,562,688.56
Uniquely mapped reads number	27,476,963.78
Uniquely mapped reads %	89.92%
Average mapped length	49.81
Number of splices: Total	3,716,558.78
Number of splices: Annotated (sjdb)	3,525,651.22
Number of splices: GT/AG	3,673,847.81
Number of splices: GC/AG	12,286.31
Number of splices: AT/AC	901.53125
Number of splices: Non-canonical	29,523.13
Number of reads mapped to multiple loci	2,343,239.41
% of reads mapped to multiple loci	7.64%
Number of reads mapped to too many loci	8,863.75
% of reads mapped to too many loci	0.03%
% of reads unmapped: too many mismatches	0.00%
% of reads unmapped: too short	2.39%
% of reads unmapped: other	0.03%

II.7: Results

II.7.1: RNA-Seq: Differential Gene Expression Analysis using DESeq 2 and edgeR

Data derived from the RNA-Seq pipeline (Figure 1 & 2) produced a summary of the Illumina sequencing reads (Table 13) which shows that there is an average of 30,562,688 reads in all 32 samples of which 27,476,963 are uniquely mapped reads (89.92%) concerning the IRGSP 1.0 reference genome. Looking at individual samples sample 1 had the highest number of raw reads with 40,059,105 million reads and sample 27 had the lowest number of reads with 26,788,085 million reads (Table 3& 4).

Focusing on two different parameters: temperature and chemical treatment can show variation in differential gene expression (DGE) among samples. The parameters include:

1. Antonio (25 ˚C) v. Antonio (30 ˚C)
2. Antonio.MCP(30 ˚C) v. Antonio (30 ˚C)
3. Colorado (25 ˚C) v. Colorado (30 ˚C)
4. Colorado.MCP(30 ˚C) v. Colorado (30 ˚C)

Transcripts identified from the Salmon pipeline are used for differential gene expression (DGE) analysis through R-studio (Figure 3-6 & Table 5-7). All samples under temperature treatment parameters and chemical-temperature treatment parameters identified various differentially expressed genes (DEG)s that were filtered at various p-values (0.25-0.80) and log-fold-changes (lfc= 0 to 0.5).

Comparison 1 (C1) looked at the DEGs in Antonio (25 ˚C) vs. Antonio (30 ˚C) filtered at a p-value= 0.50, lfc= 0, FDR<0.25. This produced 2,778 up-regulated transcripts and 1,713 down-regulated transcripts out of a total of 19,681 transcripts (Figure 3). Comparison 2 (C2) looked at the DEGs in Antonio.MCP (30 ˚C) vs. Antonio (30 ˚C) filtered at a p-value= 0.80, lfc=0, and FDR< 1.5. This produced 372 up-regulated transcripts and 328 down-regulated transcripts (Figure 4). Within these comparisons the down-regulated DEGs from comparison 1 were scanned against the up-regulated transcripts from comparison 2 to identify a pattern of transcripts that are repeating themselves in the opposite regulation; similarly, the up-regulated DEGs from comparison 1 were scanned against the down-regulated transcripts from comparison 2. The purpose of this match screening process for these comparisons is to identify transcripts involved in ethylene-response from Antonio (heat sensitive) as well as any transcripts involved in essential physiological behavior (i.e- cellular respiration)). By combining oppositely regulated transcripts into one list (up-regulated (C1-matched transcripts) + down-regulated (C2-matched transcripts)+ down-regulated (C1-matched transcritps)+ up-regulated (C2-matched transcripts)), produce a total of 74 transcripts in total that had matching transcript IDs among the comparisons (Figure 7, Table 5). This subset of 74 transcripts were used for matching among the remaining comparisons. Comparison 3 (C3) looked at Colorado (25˚C) vs. Colorado (30˚C) filtered at a p-value= 0.25, lfc= 0.5, and FDR< 1.5. This produced 884 up-regulated and 93

down-regulated transcripts (Figure 5). The transcripts in comparison 3 that were not significantly up- or down-regulated were then matched against the 74 transcripts to produce a total of 26 transcripts overlapping with the comparisons using the Antonio genotype (Figure 7, Table 6). Comparison 4 (C4) investigated the DGE relationship in Colorado.MCP (30˚C) vs. Colorado (30˚C) filtered at p-value= 0.5, lfc= 0, FDR< 1.5. This filtering produced 490 up-regulated DGEs and 2,000 down-regulated DGEs (Figure 6). The up-regulated and down-regulated transcripts were additionally manually filtered for removing overly expressed up/down-regulated transcripts leaving a total of 596 (up-regulated) and 233 (down-regulated) neutral transcripts. These neutral transcripts are then screened against the 74 transcripts ("C1 & 2 Opposites") to produce final set of 25 matched transcripts (Figure 9, Table 7). This final subset of 25 transcripts may possess some useful functions related to ethylene response and HNT. Potential transcripts of interest include: SAUR family gene (OS09T0545300), the WRKY transcription factor (OS12T0116700), and nitrate reductase (OS08T0468700); these may all potentially affect, or be strongly affected by, the interaction between HNT and ethylene response, Other transcripts may be involved in mechanisms of cellular respiration or other novel mechanisms. This final comparison shows the transcripts that matched through all four comparisons and can used for further gene ontology analysis.

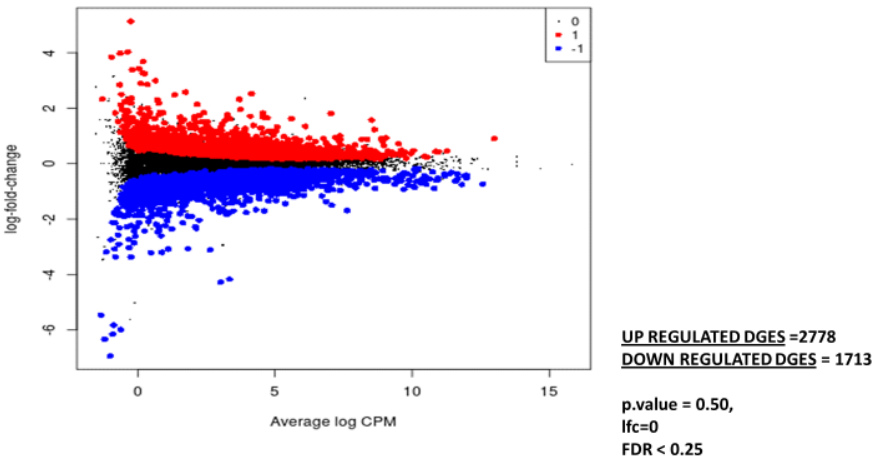

Figure 3: Comparison 1- Antonio (25˚C) vs. Antonio (30˚C) pairwise comparison plot (DGEs).

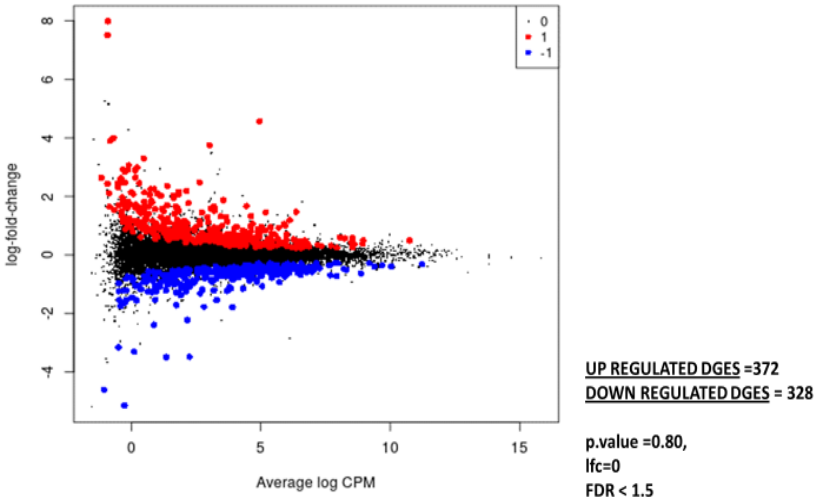

Figure 4: Comparison 2-Antonio.MCP(30˚C) v Antonio (30˚C) pairwise comparison plot (DGEs).

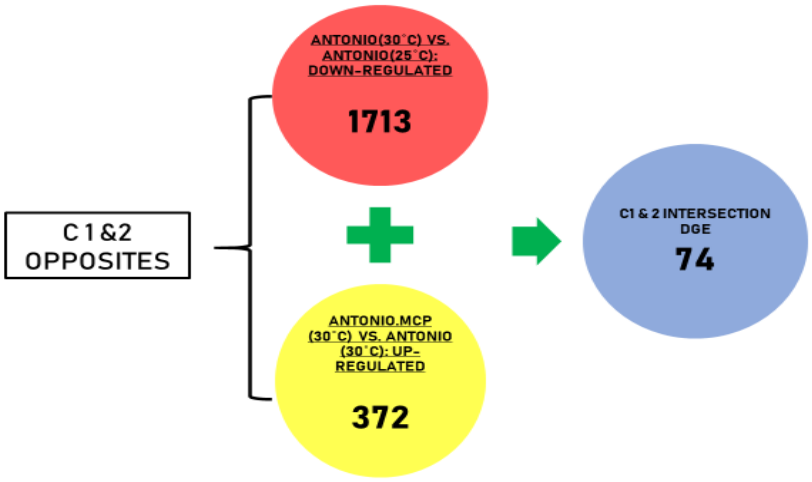

Figure 5: Comparison 1 & Comparison 2 intersecting DGEs. The other set of oppositely regulated transcripts (between comparisons 1 and 2) were also identified.

Table 5: Function of 74 DGEs displaying opposite regulation in "Comparison 1 & 2 (RAP-DB).

TRANSCRIPT ID	FUNCTION	POSITION
OS01T0165200	Kelch-type beta propeller domain containing protein. (Os01t0165200-01);Similar to KEAP1.	chr01:3362001..3367658
OS01T0566800	Hypothetical conserved gene	chr01:21687799..21689199
OS01T0750900	Conserved hypothetical protein.	chr01:31492428..31492979
OS01T0764850	Hypothetical protein.	chr01:32223295..32224590
OS01T0803300	Protein of unknown function DUF6, transmembrane domain containing protein. (Os01t0803300-01)	chr01:34043179..34044975
OS02T0119250	Hypothetical gene	chr02:1014535..1016006
OS02T0455400	Hypothetical conserved gene	chr02:15106328..15107381
OS02T0743400	Auxin transport protein REH1	chr02:31158979..31161998
OS02T0771666	Hypothetical gene	chr02:32567076..32569028
OS02T0771700	Glycoside hydrolase, family 17 protein.	chr02:32567078..32570042
OS02T0774400	Hypothetical protein.	chr02:32688877..32691426
OS02T0777400	Similar to ER (ERECTA); transmembrane receptor protein kinase.	chr02:32886861..32889740
OS03T0103200	Similar to Physical impedance induced protein.	chr03:211082..211817
OS03T0140200	Similar to Cytochrome P450 86A1 (EC 1.14.-.-)(CYPLXXXVI) (P450-dependent fatty acid omega-hydroxylase).	chr03:2211807..2213793
OS03T0782200	Hypothetical protein.	chr03:32423241..32424487
OS04T0401000	Proline-rich protein, Blast resistance	chr04:19835206..19836892
OS04T0415000	Similar to H0622F05.5 protein.	chr04:20508625..20509566
OS04T0447166	Hypothetical protein.	chr04:22308320..22309009
OS05T0307100	Similar to tetracycline transporter protein.	chr05:13927674..13929113
OS05T0426100	Similar to predicted protein	chr05:20877789..20879887
OS05T0479475	Hypothetical gene	chr05:23579212..23579928
OS06T0552900	Similar to SP3D.	chr06:20971660..20974252
OS06T0592500	Similar to Ethylene-responsive transcriptional coactivator.	chr06:23293383..23294073
OS07T0108900	Similar to MADS-box transcription factor 15. (Os07t0108900-01);APETALA1 (AP1)/FRUITFULL (FUL)-like MADS box transcription factor, Specification of inflorescence meristem identity, sexual reproduction	chr07:471234..477655
OS07T0491800	Similar to taxane 13-alpha-hydroxylase.	chr07:18328958..18330796
OS07T0603300	TON1 RECRUIT MOTIF (TRM)-containing protein, Regulation of grain size and shape	chr07:24664328..24669321
OS08T0425500	Endonuclease/exonuclease/phosphatase domain containing protein.	chr08:20450032..20452270
OS08T0425600	Hypothetical protein.	chr08:20450507..20451061
OS08T0468700	Similar to Nitrate reductase [NADH] 1 (EC 1.7.1.1) (NR1).	chr08:23051707..23055631
OS08T0520600	Conserved hypothetical protein.	chr08:25888333..25889040
OS08T0546800	Similar to Heat stress transcription factor B-2b. (Os08t0546800-01)	chr08:27382872..27384470
OS08T0560000	Similar to gibberellin 20 oxidase 2.	chr08:28043209..28045010
OS10T0555900	Similar to Beta-expansin. (Os10t0555900-01);Beta-expansin precursor.	chr10:21847828..21849722
OS11T0170900	Similar to BURP domain-containing protein 17.	chr11:3463917..3467995
OS11T0439600	Similar to Nod factor binding lectin-nucleotide phosphohydrolase.	chr11:14390405..14398122
OS11T0440200	Similar to GDA1/CD39 family protein, expressed.	chr11:14428631..14430579
OS11T0678200	Conserved hypothetical protein.	chr11:27408007..27409934

TRANSCRIPT ID	FUNCTION	POSITION
OS01T0165200	Kelch-type beta propeller domain containing protein. (Os01t0165200-01);Similar to KEAP1.	chr01:3362001..3367658
OS01T0566800	Hypothetical conserved gene	chr01:21687799..21689199
OS01T0750900	Conserved hypothetical protein.	chr01:31492428..31492979
OS01T0764850	Hypothetical protein.	chr01:32223295..32224590
OS01T0803300	Protein of unknown function DUF6, transmembrane domain containing protein. (Os01t0803300-01)	chr01:34043179..34044975
OS02T0119250	Hypothetical gene	chr02:1014535..1016006
OS02T0455400	Hypothetical conserved gene	chr02:15106328..15107381
OS02T0743400	Auxin transport protein REH1	chr02:31158979..31161998
OS02T0771666	Hypothetical gene	chr02:32567076..32569028
OS02T0771700	Glycoside hydrolase, family 17 protein.	chr02:32567078..32570042
OS02T0774400	Hypothetical protein.	chr02:32688877..32691426
OS02T0777400	Similar to ER (ERECTA); transmembrane receptor protein kinase.	chr02:32886861..32889740
OS03T0103200	Similar to Physical impedance induced protein.	chr03:211082..211817
OS03T0140200	Similar to Cytochrome P450 86A1 (EC 1.14.-.-)(CYPLXXXVI) (P450-dependent fatty acid omega-hydroxylase).	chr03:2211807..2213793
OS03T0782200	Hypothetical protein.	chr03:32423241..32424487
OS04T0401000	Proline-rich protein, Blast resistance	chr04:19835206..19836892
OS04T0415000	Similar to H0622F05.5 protein.	chr04:20508625..20509566
OS04T0447166	Hypothetical protein.	chr04:22308320..22309009
OS05T0307100	Similar to tetracycline transporter protein.	chr05:13927674..13929113
OS05T0426100	Similar to predicted protein	chr05:20877789..20879887
OS05T0479475	Hypothetical gene	chr05:23579212..23579928
OS06T0552900	Similar to SP3D.	chr06:20971660..20974252
OS06T0592500	Similar to Ethylene-responsive transcriptional coactivator.	chr06:23293383..23294073
OS07T0108900	Similar to MADS-box transcription factor 15. (Os07t0108900-01);APETALA1 (AP1)/FRUITFULL (FUL)-like MADS box transcription factor, Specification of inflorescence meristem identity, sexual reproduction	chr07:471234..477655
OS07T0491800	Similar to taxane 13-alpha-hydroxylase.	chr07:18328958..18330796
OS07T0603300	TON1 RECRUIT MOTIF (TRM)-containing protein, Regulation of grain size and shape	chr07:24664328..24669321
OS08T0425500	Endonuclease/exonuclease/phosphatase domain containing protein.	chr08:20450032..20452270
OS08T0425600	Hypothetical protein.	chr08:20450507..20451061
OS08T0468700	Similar to Nitrate reductase [NADH] 1 (EC 1.7.1.1) (NR1).	chr08:23051707..23055631
OS08T0520600	Conserved hypothetical protein.	chr08:25888333..25889040
OS08T0546800	Similar to Heat stress transcription factor B-2b. (Os08t0546800-01)	chr08:27382872..27384470
OS08T0560000	Similar to gibberellin 20 oxidase 2.	chr08:28043209..28045010
OS10T0555900	Similar to Beta-expansin. (Os10t0555900-01);Beta-expansin precursor.	chr10:21847828..21849722
OS11T0170900	Similar to BURP domain-containing protein 17.	chr11:3463917..3467995
OS11T0439600	Similar to Nod factor binding lectin-nucleotide phosphohydrolase.	chr11:14390405..14398122
OS11T0440200	Similar to GDA1/CD39 family protein, expressed.	chr11:14428631..14430579
OS11T0678200	Conserved hypothetical protein.	chr11:27408007..27409934

Table 5: Continued

Table 5: Continued

TRANSCRIPT ID	FUNCTION	POSITION
OS07T0603300	TON1 RECRUIT MOTIF (TRM)-containing protein, Regulation of grain size and shape (Os07t0603300-01).	chr07:24664328..24669321
OS08T0425500	Endonuclease/exonuclease/phosphatase domain containing protein.	chr08:20450032..20452270
OS08T0425600	Hypothetical protein.	chr08:20450507..20451061
OS08T0468700	Similar to Nitrate reductase [NADH] 1 (EC1.7.1.1) (NR1).	chr08:23051707..23055631
OS08T0520600	Conserved hypothetical protein.	chr08:25888333..25889040
OS08T0546800	Similar to Heat stress transcription factor B-2b.	chr08:27382872..27384470
OS08T0560000	Similar to gibberellin 20 oxidase 2.	chr08:28043209..28045010
OS10T0550900	Proline oxidase domain containing protein	chr10:21618211..21620554
OS10T0555900	Similar to Beta-expansin. (Os10t0555900-01);Beta-expansin precursor.	chr10:21847828..21849722
OS11T0170900	Similar to BURP domain-containing protein 17.	chr11:3463917..3467995
OS11T0439600	Similar to Nod factor binding lectin-nucleotide phosphohydrolase.	chr11:14390405..14398122
OS11T0440200	Similar to GDA1/CD39 family protein, expressed.	chr11:14428631..14430579

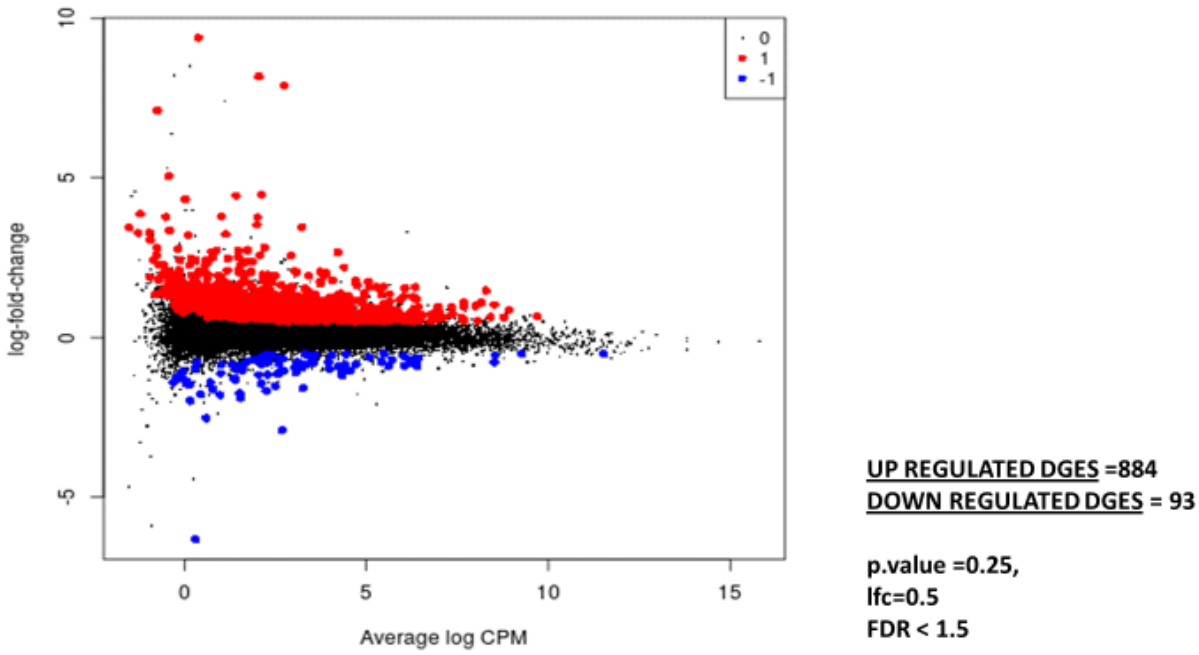

Figure 6: Comparison 3-Colorado (30˚C) v Colorado (25˚C) pairwise comparison plot (DGEs).

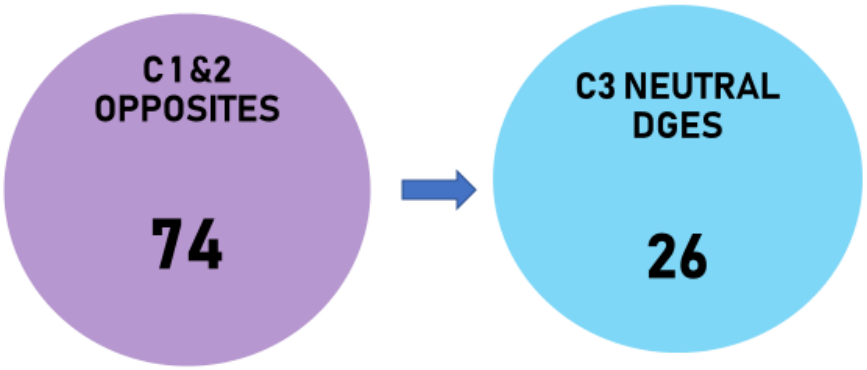

Figure 7: Intersecting DGEs present in "Comparison 3" derived from "Comparison 1 & Comparison 2 "oppositely expressed" DGEs.

Table 6: Functions of 26 intersecting DGEs present among "Comparison 3" "not significantly up-or down-regulated " and "Comparison 1 &2 Opposites".

TRANSCRIPT ID	FUNCTION	POSITION
OS01T0642000	Carboxylesterase, type B family protein.	chr01:25792699..25797614
OS01T0740300	Hypothetical protein.	chr01:30909430..30916476
OS01T0764850	Hypothetical protein	chr01:32223295..32224590
OS01T0901600	Similar to 4-coumarate--CoA ligase-like 6.	chr01:39257838..39261362
OS02T0455400	Hypothetical conserved gene.	chr02:15106328..15107381
OS02T0743400	Auxin transport protein REH1.	chr02:31158979..31161998
OS02T0774400	Hypothetical protein	chr02:32688877..32691426
OS05T0124600	Ankyrin repeat domain containing protein.	chr05:1397305..1404503
OS07T0491800	Similar to taxane 13-alpha-hydroxylase.	chr07:18328958..18330796
OS10T0569900	Conserved hypothetical protein.	
OS11T0170900	Similar to BURP domain-containing protein 17.	chr11:3463917..3467995
OS12T0116700	Similar to WRKY transcription factor 64.	chr12:824221..825793
OS12T0610600	Similar to NAM / CUC2-like protein. (Os12t0610600-01)	chr12:25800944..25806504
OS01T0165200	Kelch-type beta propeller domain containing protein. (Os01t0165200-01);Similar to KEAP1.	chr01:3362001..3367658
OS01T0566800	Hypothetical conserved gene	chr01:21687799..21689199
OS02T0119250	Hypothetical gene.	chr02:1014535..1016006
OS01T0332150	Hypothetical protein.	chr01:12870691..12873687
OS01T0532300	Conserved hypothetical protein.	chr01:19200090..19201308
OS01T0647200	Hypothetical conserved gene. (Os01t0647200-01);Non-protein coding transcript. (Os01t0647200-04)	chr01:26086062..26088337
OS03T0782200	Hypothetical protein.	chr03:32423241..32424487
OS04T0401000	Proline-rich protein, Blast resistance (Os04t0401000-01);Similar to Pi21 protein.	chr04:19835206..19836892
OS10T0555900	Similar to Beta-expansin. (Os10t0555900-01);Beta-expansin precursor. (Os10t0555900-02)	chr10:21847828..21849722
OS01T0803300	Protein of unknown function DUF6, transmembrane domain containing protein.	chr01:34043179..34044975
OS02T0119250	Protein of unknown function DUF6, transmembrane domain containing protein.	chr01:34043179..34044975
OS02T0777400	Similar to ER (ERECTA); transmembrane receptor protein kinase.	chr02:32886861..32889740
OS04T0447166	Hypothetical protein.	chr04:22308320..22309009

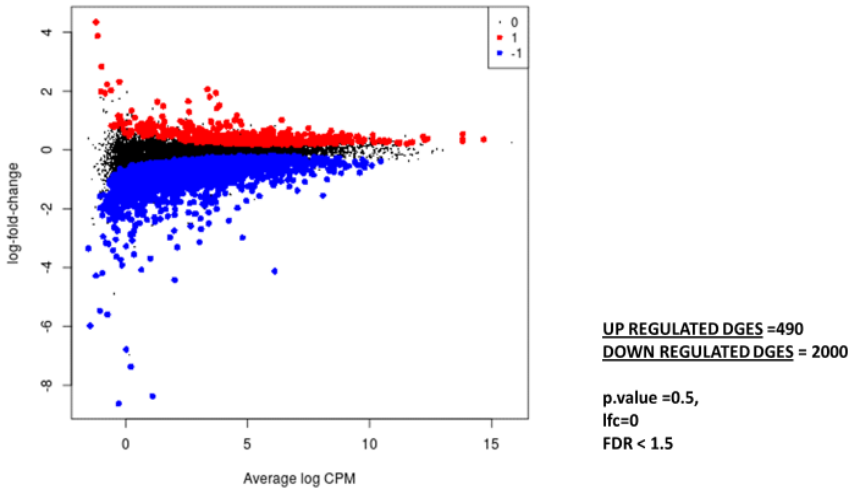

Figure 8: Comparison 4- Colorado.MCP (30˚C) v Colorado (30˚C) pairwise comparison plot (DGEs).

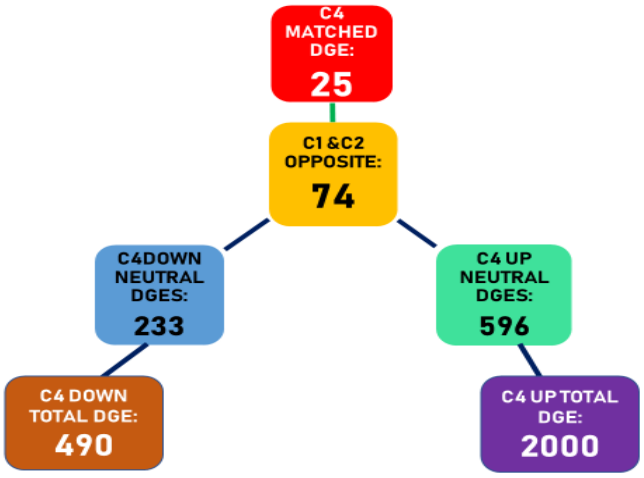

Figure 9: Intersecting DGEs among "Comparison 4" and "Comparison 1&2 Opposite" DGEs.

Table 7: Functions of 25 matched DGEs derived from "Comparison 4" DGEs and "Comparison 1&2" DGEs.

TRANSCRIPT ID	FUNCTION	POSITION
OS01T0233900	Similar to cDNA clone:J013003E06, full insert sequence. (Os01t0233900-01)	chr01:7404911..7409638
OS01T0332100	Similar to Neutral invertase-like protein (Fragment). (Os01t0332100-01);Similar to Invertase. (Os01t0332100-02)	chr01:12870240..12874218
OS01T0532300	Conserved hypothetical protein.	chr01:19200090..19201308
OS01T0543600	Cytochrome P450 domain containing protein.	chr01:20124194..20126848
OS01T0543800	Hypothetical protein.	chr01:20124383..20126778
OS01T0740300	Hypothetical protein.	chr01:30909430..30916476
OS01T0750900	Conserved hypothetical protein.	chr01:31492428..31492979
OS03T0782200	Hypothetical protein.	chr03:32423241..32424487
OS05T0479475	Hypothetical gene.	chr05:23579212..23579928
OS08T0520000	Mitochondrial substrate carrier family protein.	chr08:25848686..25852963
OS09T0545300	SAUR family protein, Negative regulator of auxin synthesis and transport	chr09:21565770..21566461
OS01T0566800	Hypothetical conserved gene.	chr01:21687799..21689199
OS01T0706800	Similar to agglutinin.	chr01:29323176..29323982
OS02T0119250	Hypothetical gene.	chr02:1014535..1016006
OS02T0771700	Glycoside hydrolase, family 17 protein.	chr02:32567078..32570042
OS03T0327800	NAC Family transcriptional activator, Abiotic stress response, Positive regulator of leaf senescence (Os03t0327800-01)	chr03:11973989..11976381
OS03T0431600	Hypothetical conserved gene.	chr03:18119653..18121181
OS04T0398700	Similar to Pollen specific protein C13 precursor.	chr04:19693167..19694275
OS04T0415000	Similar to H0622F05.5 protein.	chr04:20508625..20509566
OS07T0192000	ATPase, AAA-type, core domain containing protein.	chr07:4969408..4971145
OS08T0425500	Endonuclease/exonuclease/phosphatase domain containing protein.	chr08:20450032..20452270
OS08T0468700	Similar to Nitrate reductase [NADH] 1 (EC 1.7.1.1) (NR1). (Os08t0468700-00)	chr08:23051707..23055631
OS09T0545300	SAUR family protein, Negative regulator of auxin synthesis and transport (Os09t0545300-01)	chr09:21565770..21566461
OS10T0529600	Hypothetical protein.	chr10:20597016..20599106
OS12T0116700	Similar to WRKY transcription factor 64.	chr12:824221..825793

II.8: RNA-Seq: Gene Ontology

To identify the molecular and biological components related to the highly expressed DEGs in the given parameters a gene ontology analysis was done using WEGO 2.0 (http://wego.genomics.org.cn/). The 25 transcripts from comparison 4 was input into the WEGO tool for gene annotation analysis. There were no significant GO terms discovered with this subset of 25 transcript terms (Figure 8 & 9). But, with respect to the reference these transcripts may have potentially had a strong cellular component (cell part, marcomolecular complex, and organelle part: ~75%). Further analysis would need to be done to determine the ontology of these transcripts.

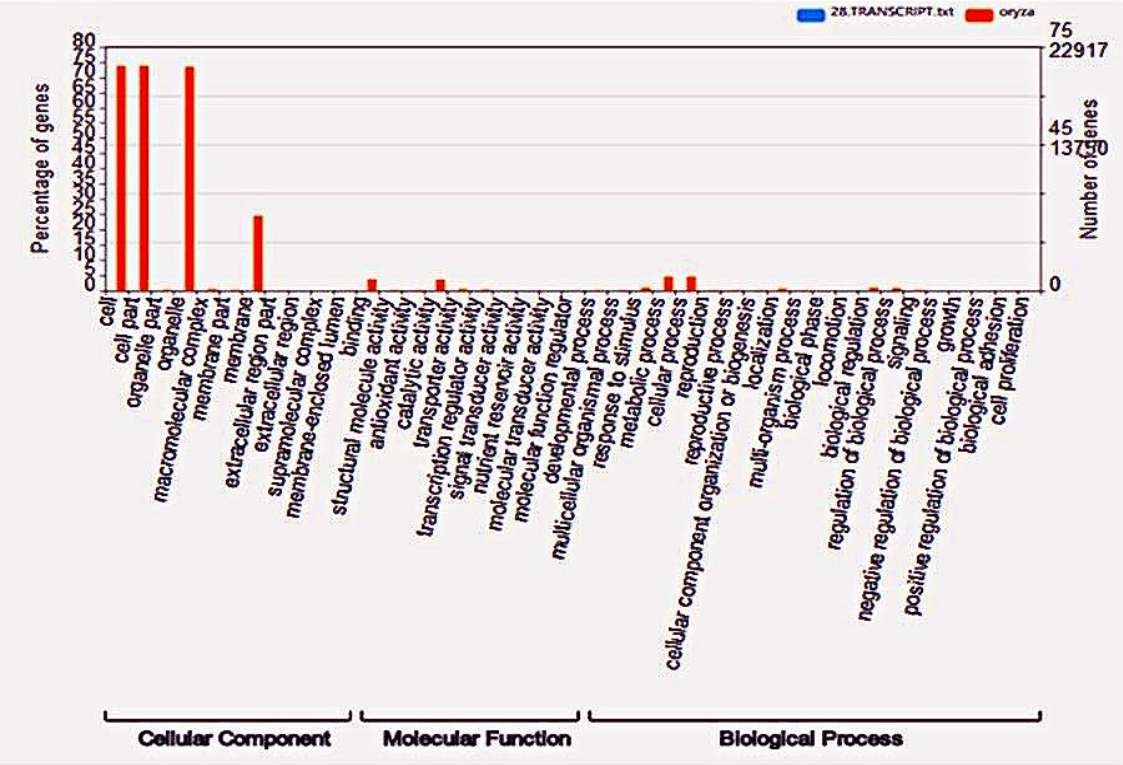

Figure 10: Gene Annotation graph representing all potential cellular, molecular, and biological components among the final 25 transcripts with respect to the reference genome (*Oryza sativa-Japonica*).

Summary

		25.TRANSCRIPT.txt	oryza	Total
Gene		100	30,557	30,657
Annotated Genes		0	30,557	30,557
GO Terms	Biological	0	1,906	1,906
	Cellular	0	29,131	29,131
	Function	0	2,027	2,027
	Total	0	33,064	33,064

Figure 11: WEGO (Web Gene Ontology Annotation Plotting) summary of final subset of 25 transcripts present across all comparisons.

II.9: Discussion

RNA-Seq is a powerful and efficient tool for mining genes that are related to specific functions. In this study, transcriptome analysis for heat tolerance DGE in contrasting rice varieties (Antonio and Colorado) was performed and obtained a total 37.4 Gb of transcriptome data. Over 27 million clean reads were obtained from the samples of TruSeq RNA (total RNA) library and found many upregulated and downregulated DEGs among the library. The filters (p-value, FDR, and lfc) for each comparison were very loose (p-value > 0.05, FDR > 0.05, and lfc = 0) due to the extremely low numbers of differential genes being expressed, which would hinder complete comparison analysis in this study. The DEGs data provided comprehensive gene expression information for heat tolerance/1-MCP treatments on identical genotypes, facilitating our understanding of the molecular mechanisms of rice under these treatments in response to high nighttime temperature and 1-MCP.

Findings from this study showed that the 29 transcripts did not have a unique GO but that the reference genome associated with these transcripts did show >75% cellular component behavior. Further analysis needs to be done on the potential transcripts that have functions which impact ethylene response, cellular respiration, and HNT behavior (Refer to Results). The specific transcripts that responded oppositely (with respect to up- or down-regulation) to HNT and MCP

(blocks ethylene response) in the sensitive variety, yet showed no significant differential response to either HNT or MCP in the tolerant variety, provide clues on the nature of HNT tolerance in Colorado involving regulation of the ethylene response.

CHAPTER III
GENOME-WIDE ASSOCIATION STUDY (GWAS) OF EXOTIC RICE TRAITS

III.1: Literature Review

III.1.1: Exotic Traits in Rice

Many genetic donors for stress-tolerance traits are exotic landrace varieties that share a number of negative traits with wild and weedy rice species, such as red pericarp color, black hulls, and the presence of awns. Weedy red rice (*Oryza sativa spontonea*) is a problematic weed present in many major rice-producing countries around the world. It can be speculated that weedy rice may have been produced due to the hybridization between cultivated rice (*O. sativa*) and wild rice (*O. rufipogon*) (Londo et al., 2007). In the United States, red rice is grouped into two subclasses: straw-hulled and black-hulled. Straw-hulled is the most common type of rice openly available and consumed in the U.S. Within these subclasses, their varying traits can classify red rice in seed color and awning (Vaughan et al., 2001). Weedy rice tends to have some undesirable agronomic traits that can pose a significant threat to sustainable global rice production (Nadir et al., 2017). In this study the exotic rice traits of focus will be red pericarp, awns, and black hull.

III.1.2: Red Pericarp (Rc)

Rice, *Oryza sativa*, was domesticated in Asia and is now grown in every country around the world with exception of extremely cold areas (Antartica). Rice consists of two subspecies, *indica* and *japonica*, whose unique genetic identities are identified since ancient times. The genetic makeup of these subgroups are maintained by inbreeding and sterility barriers encompassed in *O. sativa* (Oka et al. 1988, Glaszmann et al. 1987, Garris et al. 2005). The major

subspecies consist of five distinct subpopulations, indica, temperate japonica, tropical japonica, aus, and aromatic (Sweeney et al., 2007). Genetic analysis and calculations can derive the most recent time of divergence for Indica and Japonica concerning their most recent common ancestor to be more than 100,000 years ago (Vitte et al. 2004, Vaughan et al. 2003, Ma et al. 2004). Most rice cultivars grown and consumed today have a white or light brown pericarp, unpolished rice grains ‘‘brown rice.’’ The most important trademark of rice domestication is the attribute of changing pericarp color from red to white. The Rc gene is a domestication gene vital for red pericarp appearance in rice. Red grain color is common among the wild ancestors of rice and is linked with seed shattering and dormancy traits (Sweeney et al., 2007).

Sweeney et al. (2006), studied the red pericarp traits with the goal of gaining a better understanding of the genetics, association of the red pericarp with other weedy traits, and the molecular biology of the red pericarp trait. This information will potentially lead to better management practices associated with red rice. In this experiment a BC2F2 population was constructed using *Oryza sativa Jefferson* and *Oryza rufipogon.* QTL analysis was used to identify the locus for red grain on chromosome 7. A combination of various analysis tools such as, fine-mapping and sequence comparisons showed that a bHLH protein on LOC_Os07g11020.1 was responsible for mutant alleles *Rc* and *Rc-s* related to the red pericarp. Seed color change from red to white is defined by a SNP that contains a 14-bp deletion that knocked out the gene function of color change (Sweeney et al., 2006). The 14-bp deletion tends to create a white pericarp instead of a red pericarp in rice (Sweeney *et al*. 2006; Furukawa *et al* . 2007). The *rc* allele which contains the 14-bp deletion is present in more than 97% of white pigmented cultivars located on exon 7.

Gross et al. (2010), performed a study looking at the origin of red rice in the United States. This study consisted of a diverse panel containing 156 weedy, domesticated, and wild rices. This experiment aims to evaluate DNA sequence patterns related to the *Rc* locus which can help in determining the origin of the red pericarp rice in the United States. Analyzing this diversity panel for variation at the *Rc* locus showed weed strains with the most considerable degree of differentiation from *japonica* varieties. The *japonica* varieties contain a derived *rc* allele that dominates over the other varieties. The least amount of variation is present in *aus* varieties and *O. rufipogon.* Genome-wide variation studies on US weeds tend to show that blackhull awn (BHA) and straw hull awnless (SH) traits are closely related to domesticated varieties of rice (Londo & Schaal 2007; Gealy *et al*. 2010). This evidence leads one to believe that no US weeds have the 14-bp deletion or any accessions that have with reversion attributes from nonfunctional alleles (Gross et al., 2010).

III.1.3: Awns

Information gathered from archeological and genetic studies have shown that the Asian cultivated rice was domesticated from its ancestor of the wild rice species *Oryza rufipogon* 8000 years ago (Zong et al., 2007; Fuller et al., 2009; Izawa et al., 2009; Huang et al., 2012). Wild rice tends to express a various number of number of traits, such as seed shattering, long awns, black hulls, and few grains per panicle. These unique traits are critical for the survival of wild rice survival under harsh environmental conditions. Cultivated rice on the other hand, has reduced seed shattering and dormancy, awn length, and less likelihood of pericarp and hull color changes (Kovach et al., 2007; Sweeney and McCouch, 2007). The awn is one of the most visible morphological trait of rice seeds and is also found in other cereal crops as well, such as wheat

(*Triticum aestivum*), barley (*Hordeum vulgare*), oats (*Avena sativa*), and sorghum (*Sorghum bicolor*). The awn is defined as an extension-like structure located on the apex of the lemma of a spikelet. Awns can be of varying length, color, texture and shape. Awns tend to have essential roles within the rice ecosystem specifically, long awns are reported to aid in seed dispersal and protect cereal grains from animal predation (Elbaum et al., 2007). However, long awns are not favorable during harvest and storage purposes; hence, they were artificially selected against during domestication. Although this trait is considered undesirable in rice, it is still useful in other cereal crops such as wheat and barley by contributing to aid in the process of photosynthesis and yield (Abebe et al., 2010). Most cultivated rice have no awns or very short awns and does not aid much in the photosynthesis process (Toriba et al., 2010). Much is still unknown about the various mechanisms that control the behavior of awns so, more studies need to be implemented to gain a deeper understanding of awns.

In addition to the known *An-1* locus (Os04g0350700) which regulates the long-awn trait in wild rice (Luo et al., 2013), Gu et al. (2015), performed a study analyzing the gene *An-2* (Os04g0518800) with respect to its role in increasing awn length and grain production in rice. This experiment showed that *An-2* has genetic variation that may reduce awn length and increase grain numbers. Nucleotide diversity analysis of the *An-2* locus in cultivated rice was found to be significantly reduced compared with that of wild rice. This suggests that the *An-2* locus has undergone artificial selection due to reduced awn length and increased grain yield in cultivated rice.

Furata et al. (2015), looked at the convergent loss of the awn trait among two rice varieties (*O. sativa and O. glaberrima*) due to mutations at two different loci. Three sets of chromosome segment substitution lines (CSSLs) are evaluated in a conventional *O. sativa*

genetic background (cv. *Koshihikari*) that harbor genomic fragments from *Oryza nivara, Oryza rufipogon*, and *Oryza glaberrima* donors. Phenotypic analyses revealed the existence of three awn-related genes, Regulator of Awn Elongation 1 (RAE1), RAE2, and RAE3, who are involved in the process of losing the long awns in cultivated rice. Donor segments at RAE1 and RAE2, induced long awn formation in the CSSLs in *O. sativa*, but RAE3 induced long awn formation in *O. glaberrima*. These findings showed that *O. sativa* and *O. glaberrima* take independent pathways to become awnless.

III.1.4: Black Hull

The hulls of cereals are considered to play a role in the protection of seeds from physical damages and also oxidative damages (Ramarathnam et al., 1986). Pigments inside plants have important roles in antioxidant activity, defense against fungi, and protection against UV radiation (Shirley, 1996; Huang et al., 2012), but the pigments inside hulls have an essential role to protect the seeds. Cultivated rice (*Oryza sativa*) has its origin derived from the domestication of *Oryza rufipogon* and *Oryza nivara* (Khush, 1997; Cheng et al., 2003; Kovach et al., 2007; Sang and Ge, 2007). Wild rice species have unique traits such as, seed-shattering and black-colored seed hulls. There are a number of genes known to control seed-shattering habit, red grain pericarp, and grain discoloration (Konishi et al. 2006; Li et al. 2006, Sweeney et al. 2006, Yu et al., 2008). The color of the black hull is controlled by a few corresponding genes, Maekawa (1984) reports that there are three complementary genes, Bh-a, Bh-b, and Bh-c, which control the black hull trait. But the basis for the color change from straw hull to black hull is still unknown.

Zhu et al. (2011), performed a study looking at genetic factors that control the hull color change process from black hull to straw-white hull in diverse rice cultivars.

A cross is performed between W1943 (*O. rufipogon*: black hull) and Guangluai 4 (*O. sativa indica:* straw-white hull) is used to look into the hull color change factors. The Black controlled the black hull of O. rufipogon hull4 (*Bh4:* Os04t0460200) gene, which is located on chromosome 4 and functions as an amino acid transporter. The genetic mechanism of the Bh4 gene for hull color is controlled by a 22-bp deletion located in exon 3 which disrupts the hull color mechanism. This gene disruption causes a straw-white hull to appear in cultivated rice. Results obtained from the Kreitman-Aguade test used in this study presented a reduction in nucleotide diversity in rice cultivars used in this experiment which may be caused by artificial selection. This study was able to confirm that *Bh4 gene*, that causes the black hull trait to appear in wild rice species (Zhu et al., 2011).

Advancements in molecular genetics allow the application of gene editing techniques to be used to knock-out the major undesirable characteristics associated with exotic rice accessions and cultivate more versatile genetic donors for future stress-tolerance breeding programs. To reach this long-term goal, a genome-wide association study can be used to identify the genetic basis for exotic traits in rice.

III.1.5: Exotic Rice: Genome-wide Association Study

Genome-wide association studies (GWAS) has the capacity to overcome various limitations present in traditional gene mapping by:

1. Establishing a higher resolution at the geneitic level;
2. Improved analysis by using samples from previously used populations where commonly occurring genetic variations can be associated with phenotypic variation (Brachi et al., 2011).

In this study, three important exotic rice traits are investigated: red pericarp (Rc), black hull (Bh4), and awns (An-1). Most rice that is grown and consumed in the world today has a white pericarp. But, due the range of diversity present in various rice subgroups rice grains have a brown, red, and purple pericarp. Red pericarp is common among the wild ancestors of cultivated rice, and in some regions of the world red rice cultivars are considered a rare delicacy, and are heirloom passed down from their ancestors to current generations. Red rice cultivars are also preferred for their taste, texture, and ceremonial value. *Rc* is a domestication gene that produces a red pericarp in rice that also encodes for a basic helix-loop-helix (bHLH) protein (Sweeney et al., 2006). Red rice typically has traits related to seed shattering, seed dormancy, and red pericarp, which may belong to cultivated rice (*O. sativa*) or wild rice (*O. rufipogon*) which are not native to the United States (Vaughan et al.,2001).

The black seed hull is common among wild rice species. Black hulls seem to always be associated with the seed-shattering trait present in wild rice. The seed-shattering trait is the key functional trait for survival in wild rice and therefore, the black hull is assumed to be the natural color for wild rice hulled grains (Zhu, 2011). The major-effect gene responsible for black hull vs. straw hull phenotype is Black hull 4 (Bh4) on rice chromosome 4, which encodes an amino acid transporter that is only expressed in fully maturing hulls (Zhu et al., 2011). Loss of the black hull will potentially reduce seed shattering and improve yield in rice production.

The awn is one of the morphological characteristics of rice seeds and is also found in other cereal species. Although awns have some usefulness in certain cereal crops such as wheat and barley it is typically undesirable in rice (Abebe et al., 2010). Awn-1 (An-1), which is located on chromosome 4, encodes a basic helix-loop-helix (bHLH) protein and regulates the long-awn

trait in wild rice (Luo et al., 2013). Gene editing for awn loss would increase grain numbers and subsequently improve grain yield in the cultivated rice domestication process over time.

Identification of single nucleotide polymorphisms (SNP) markers using an Illumina 7K SNP (developed by Dr. Susan McCouch at Cornell University) will aid in the genotyping process for exotic rice traits for this GWAS study. Illumina Genome Studio, Tassel, and Gapit are the SNP analysis software tools that will help with the in-depth analysis of SNP variants associated with the exotic rice traits. Once significant QTLs are identified across various chromosomes, candidate genes in these regions can be selected for further validation. The information from these projects will help with the development of sgRNAs (single guide RNA) as part of the CRISPR/Cas9 gene editing system to test if specific candidate genes control the traits of interest. The sgRNA contains a 20-nt target sequence along with PAM motifs which are specific to target DNA sequences (Biolabs, 2018). Designing sgRNAs will be useful in targeting the genes controlling undesirable traits of exotic rice in future CRISPR/Cas9 experiments.

III.2: Materials & Methods

III.2.1: Phenotyping of Exotic Rice Traits

In this project, phenotypic data on exotic rice traits was collected from a seed increase plot consisting of 300 accessions. In this design, each row contained five plants belonging to a single accession with no replications. The main exotic rice characteristics of focus were: awn, black hull, and red pericarp color. Specifically, data on morphological traits relating to awn length and color, hull color, and post-harvest endosperm color.

Awn length was taken from 3 random plants for each accession, and measurements (mm) were made from 1 random grain on each plant with the average calculated for each set. With respect to awn, color panicles were observed overall to determine uniformity in color by using

the International Rice Research Institute SES awn color scale (Chaudhary, 1996). This same procedure was used for analyzing hull color except the scale used was the USDA GRIN hull color scale. In GWAS post-harvest analysis focuses on endosperm color. A total of 20 rice grains were collected from each accession where 10 grains are de-hulled, and data on bran color will be collected using USDA-ARS bran color scale ("Rice bran color samples: USDA ARS", 2018).

III.2.2: Genome-wide Association Study (GWAS) on Exotic Rice

The complied phenotypic data along with Illumina 7K SNP genotypic data was analyzed through Illumina Genome Studio and Tassel software to identify SNPs that control these exotic rice characteristics through a mix linear model (MLM). A mixed model includes both fixed and random effects. Including random effects gives MLM the ability to incorporate information about relationships among individuals (Buckler, 2014). MLM performs an association test for each combination of traits and markers. The visualization diagrams produced from the MLM model will be further put in use to design Manhattan plots which can indicate the most robust associated SNPs on a particular chromosome which may be highly correlated to the level of expression of various traits (awn, black hull, and red pericarp) relating to exotic rice. The GWAS pipeline is as follows:

1. Load raw data in Genome Studio (Sample Sheet + Raw Data + Illumina Map File).
2. Filter out poorly clustered SNPs (use GenTrain/GenCall scores).
3. Export SNP file into Plink format using Report Wizard
4. Import Plink files into TASSEL and format the files into Hapmap (genotype).
5. Import phenotype data into TASSEL and merge genotype/phenotype file.
6. Run GLM to verify validity through Manhattan plot.

7. Run Structure for population structure analysis.
8. Obtain Kinship matrix.
9. Run MLM for final GWAS analysis.

Furthermore, the results obtained in this GWAS study will lead to candidate genes that can be used to initiate gene editing projects (CRISPR/Cas9, sgRNA design) to validate the genes controlling these exotic rice traits.

III.3: Results

III.3.1: Genome-wide Association Study of Exotic Traits

A total of 303 accessions were analyzed for this study. Samples that did not have complete data on awn length, awn color, hull color, due to unexpected field conditions were removed from the final phenotypic data set. The pipeline discussed in material and methods was followed through till the Manhattan plot design. The Illumina 7K SNP was used to detect SNPs expressed in exotic rice. Figure 12 shows the Manhattan plot (Gapit) for awn color among these samples, the most significant SNPs were located on chromosome 9, 10, and 12. Previous studies indicate that the awn trait is located on chromosome 4 but these findings for awn color may be a novel finding. Figure 13 shows the highly expressed SNPs for the awn length trait, the SNPs were located on chromosome 4,6,7, 10, and 12 being the most significant. Figure 14 evaluates SNPs for hull color with the most highly expressed SNP located on chromosome 12. The final trait evaluated in figure 15 is endosperm (pericarp) color, the Manhattan plot indicated a significant SNP on chromosome 1 and 3. Further analysis has to be done with this data by completing population structure analysis, kinship matrix, and a final MLM analysis.

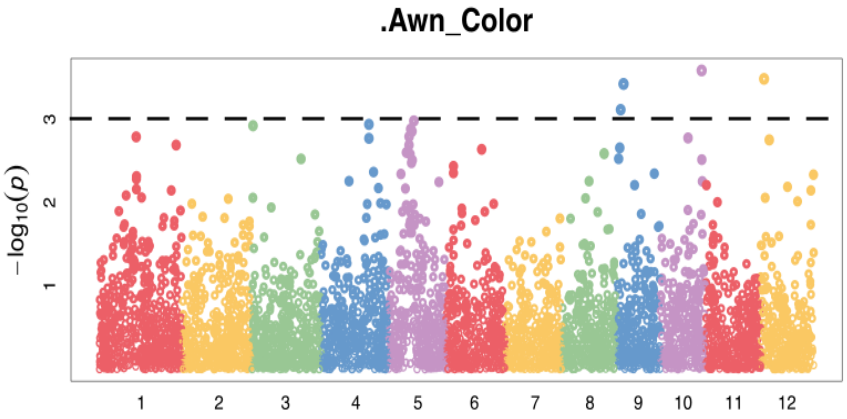

Figure 12: Manhattan plot identifying SNPs of awn color trait.

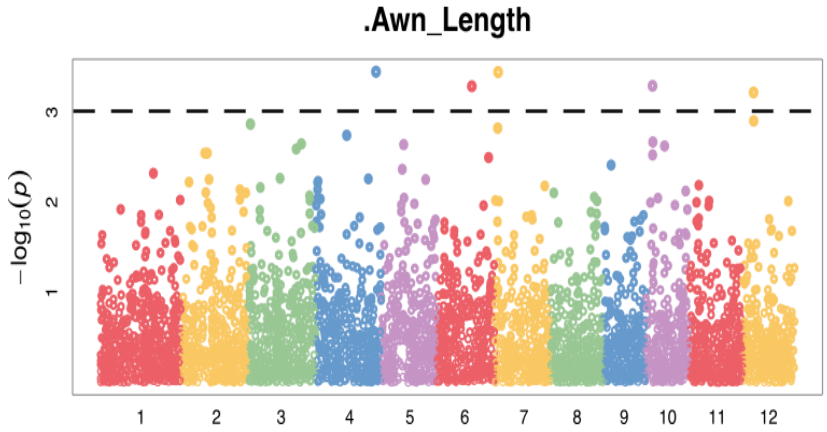

Figure 13: Manhattan plot identifying SNPs of awn length trait.

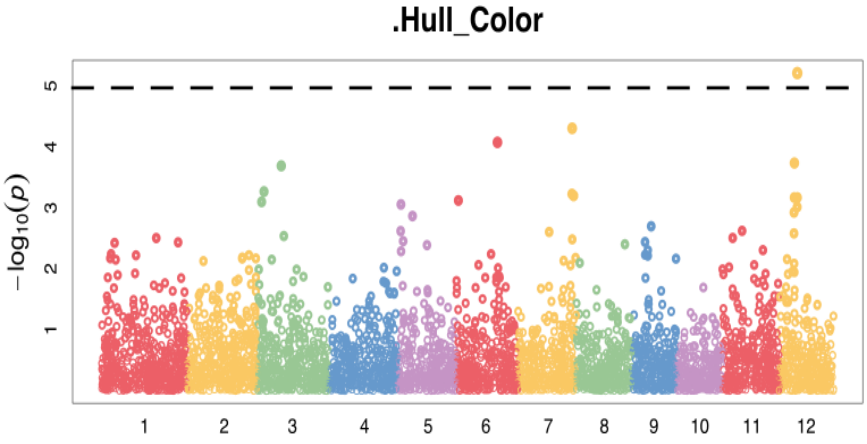

Figure 14: Manhattan plot identifying SNPs of awn length trait.

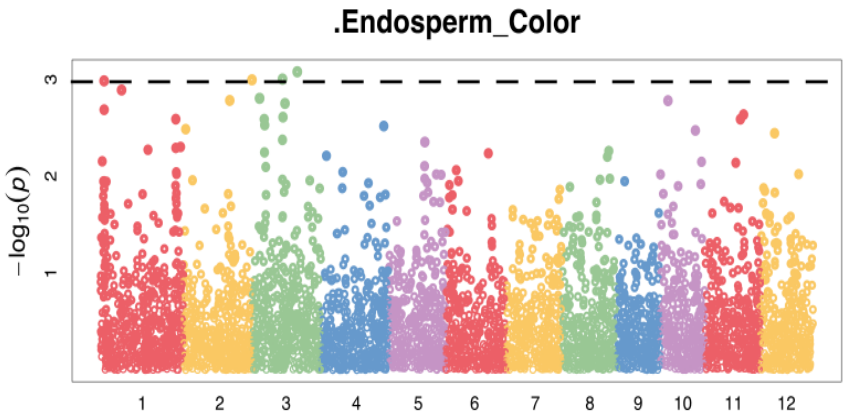

Figure 15: Manhattan plot identifying SNPs of endosperm color (pericarp) color trait.

III.4: Discussion

Weedy rice (*Oryza sativa f. spontanea*) is of the Poaceae is a weed present in rice fields around the world most commonly present in South/Southeast Asia, North and South America, and Europe (Ferrero et al., 1999; Mortimer et al., 2000). Weedy rice is generally defined by its undesirable traits of seed shattering and seed dormancy, which increases the chances of weedy rice to be spread across rice fields. As an infamously detrimental weed occurring in rice fields, it typically causes decreases yield output and negatively impacts grain quality (Hoagland and Paul, 1978). Most weedy rice grains have red pericarps once dehulled which is why it is called red rice (Gealy et al., 2003). Analyzing various exotic rice traits will allow provide an in-depth understanding about the degree and distribution of the genetic diversity.

In this study a diversity panel of 330 accession were to be used for a genome-wide association study focusing on exotic rice traits such as, red pericarp, black hull, and awns. Due to unfavorable environmental conditions (hurricane Harvey) the panel size shrunk down to 303 accessions. Phenotypic data collection for awn color, awn length, hull color, and pericarp color varied from accession to accession. Some accession had complete data for each category but others did not due not flowering, lodging, and seed shattering. GWAS analysis is used to detect

the Illumina 7k SNP markers associated with the genotype data. Illumina Genome Studio after filtering based on GenTrain score had a total 6098 SNPs remaining from the 7K SNP data which is utilized for further analysis in Tassel. The merged data (genotype+ phenotype) produced respective Manhattan plots, using Gapit, pertaining to the exotic rice traits in this GWAS through general linearized model (GLM; Figures: 12-15). Previous studies have indicated that the *Rc* gene is located on chromosome 7 (Gross et al., 2010; Sweeney et al., 2006 & Gu et al., 2011) but our findings based on the Manhattan plot indicate the most significant SNP located on chromosome 3. The black hull is located on chromosome 4 (Zhao et al., 2012 & Zhu et al., 2011) our results of analyzing the diversity panel indicated SNPs located on chromosome 12. Awn traits can be located on chromosome 1, 3,4,5, 6, or 8 (Matsushita et al., 2003; Kinoshita et al., 1984; Takamure et al., 1991; Yoshimura et al., 2004; Li et al., 2017; Huang et al., 2012). With respect to awn color (chromosome 6) the results from this study showed SNPs were located on chromosome 9 and 10. Awn length (previous literature: chromosome 1,3,4,5 and 8) SNPs were identified on chromosome 4,6,7, 10, and 12.

CHAPTER IV

CONCLUSION

IV.1: High Nighttime Temperature

Overall data showed a maintenance in physiological responses of Colorado and Antonio under HNT:1-MCP treatment in comparison to ambient conditions (Tables 1&2). The results clearly identified Colorado as the heat tolerant variety and Antonio as the heat susceptible variety.

IV.2: RNA-Seq

Results identified a subset of 74 unique transcripts of which 29 transcripts matched across all comparisons. Several transcripts may be involved in ethylene and HNT behavior. Potential transcripts include: SAUR family gene (OS09T0545300), the WRKY transcription factor (OS12T0116700), and nitrate reductase (OS08T0468700). Other transcripts may be involved in cellular respiration related mechanisms. Gene ontology (GO) did not identify any unique GO terms with respect to the 25 transcripts. Future studies can analyze these transcripts at the protein-level and potentially be used for future gene editing projects.

IV.3: Genome-wide Association Study

The results from this GWAS study may have discovered mostly novel QTLS on different chromosomes. QTLs belonging to awn color are located on chromosome: 9,10, & 12. Awn length QTLs are on chromosome: 4,6,7,10, &12. QTLs for hull color were located on chromosome 12 and endosperm color were located on chromosome 3. The various SNPs

discovered for exotic rice traits (awns color, awn length, black hull, and red pericarp) can be further tested through cloning the genes from major QTLs and design markers based on the QTLS.

IV.4: Future Direction: Gene Editing and sgRNA Designing

Candidate gene targets of exotic rice traits of red pericarp, awns, and black hull are the gene editing targets. Therefore, they will have sgRNAs designed for the first exon of each gene to knock-out the gene function using CRISPR/Cas9. The weedy rice genes for their respective traits: red pericarp (Rc: Os07g0211500), awns (An-1: Os04G0350700), and black hull (Bh4: OS04G0460200) can be found on Oryzabase. The sequences associated with these genes are found on Gramene. DNASTAR: MegAlign Pro, designs multi-alignment reports for each individual weedy rice target trait. The reports can be used to identify SNPs for each attribute which will aid in primer design.

For red pericarp, comparisons are made with the following sequences: Nipponbare Rc, Nipponbare AP014963, Rc BAC Clone Ap005098.4, and Kasalath. The awn trait can have comparisons made among An-1 Nipponbare, Nipponbare Awn BLAST AP014960.1, Awn 1 BC Clone AC090882.10, An-1 BLAST mRNA_XM-015780181.1, and Kasalath reverse complement. Sequences to be compared for black hull include Bh4 Nipponbare, Nipponbare Bh4 BLAST AP014960.1, Bh4 BAC Clone AL606460.3, Bh4 cDNA FQ 377519.1, and Kasalath.

When designing primers for the candidate genes, all primers are designed for Nipponbare (*Japonica*) from Gramene. Primer design can help in finding overlapping exonic regions

between variants. In the sequences for candidate genes only exon 1 or exon 2 are selected as target sequences for primer design. The chosen exon sequences are put into Primer3, a primer design tool, which produces outputs that indicate the forward and reverse primers for the respective sequences. For each exotic rice trait, approximately two sets of primers (forward and reverse) are designed, where one set serves as backup primers (Table 8).

Table 8: Potential primers for Rc, Bh4, and An-1

Name	Sequence
Rc. EX1-EX2. F1	TCTCGATCATCCACGAGCTA
Rc. EX 1-EX2. R1	TTTAGGGTTTCTGGCTCCAA
Rc EX 1-EX2. F2	TCAATTCTTCCATCCCCAAC
Rc EX 1-EX2. R2	CGTGGATCAACACACCGATT
Bh4 EX 1 F1	CTTGCATGAATGGCCTCAAT
Bh4 EX 1 R1	TGACCTTTGCAAATTGGTTG
Bh4 EX 1 F2	CTTGCATGAATGGCCTCAAT
Bh4 EX 1 R2	TGACCTTTGCAAATTGGTTG
An1 EX2 F1	CGTGCTGTAGCAGGAGTTGT
AN1 EX2 R1	CCATCACTTCTCCGATCTCC

To confirm if the primers are working correctly PCR is run extracted DNA (N22 and Pokkali) to verify the target PCT product size. If the size is correct, then the samples will be sent for Sanger sequencing once purification is complete. Based on the Sanger sequence results gRNAs can be designed and CRISPR/Cas9 construct vectors can be developed using the Golden Helix protocol. The vectors will be used for future studies to knock out undesirable exotic rice traits. Each of the respective findings in these projects can be utilized for further advancement of rice crop improvement for heat tolerance. Essentially the results obtained in this study can be used to design breeding programs for heat tolerance in the near future.

REFERENCES

Abebe, T., Melmaiee, K., Berg, V., & Wise, R. P. (2009). Drought Response in the Spikes of Barley: Gene Expression in the Lemma, Palea, Awn, and Seed. *Functional & Integrative Genomics, 10*(2), 191-205. doi:10.1007/s10142-009-0149-4.

Adams, M., Kelley, J., Dubnick, M., Polymeropoulos, M., Xiao, H., & Merril, C. et al. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, *252*(5013), 1651-1656. http://dx.doi.org/10.1126/science.2047873

Ambardekar, A., T. J. Siebenmorgen, P. A. Counce, S. Lanning, and A. Mauromoustakos, 2011: Impact of Field-Scale Night- Time Air Temperatures During Kernel Development on Rice Milling Quality. *Field Crops* Res. 122, 179–185.

Anders, S., Pyl, P., & Huber, W. (2014). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169. http://dx.doi.org/10.1093/bioinformatics/btu638

Bader, J. (2001). The relative power of SNPs and haplotype as genetic markers for association tests. *Pharmacogenomics*, *2*(1), 11-24. http://dx.doi.org/10.1517/14622416.2.1.11.

Batley, J. (2003). Mining for Single Nucleotide Polymorphisms and Insertions/Deletions in Maize Expressed Sequence Tag Data. *Plant Physiology*, 132(1), pp.84-91.

Begum, H., Spindel, J., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., Virk, P., Collard, B. and McCouch, S. (2015). Genome-Wide Association Mapping for Yield and Other Agronomic Traits in an Elite Breeding Population of Tropical Rice (Oryza sativa). *PLOS ONE*, 10(3), p.e0119873.

Biolabs, N. (2018). *sgRNA Template Construction for Cas9 Gene Editing | NEB*. *Neb.com*. Retrieved 17 April 2018, from https://www.neb.com/applications/genome-editing/sgrna-template-construction-for-cas9-gene-editing.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326(5959), pp.1509-1512.

Brachi, Benjamin, Geoffrey P Morris, and Justin O Borevitz. "Genome-Wide Association Studies in Plants: The Missing Heritability Is in the Field." *Genome Biology* 12.10 (2011): 232. *PMC*. Web. 21 Dec. 2017.

Buckler, Ed, et al. "User Manual for TASSEL." *Buckler Lab at Cornell University*, www.maizegenetics.net/tassel.

Chaudhary, R. (1996). *Standard evaluation system for rice* (4th ed., pp. 44-45). Los Banos (Phillipines): International Rice Research Institute.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H. and Ohtsubo, E. (2003). Polyphyletic Origin of Cultivated Rice: Based on the Interspersion Pattern of SINEs. *Molecular Biology and Evolution*, 20(1), pp.67-75.

Cheng, W., H. Sakai, K. Yagi, and T. Hasegawa, 2009: Interactions of Elevated [CO_2] and Night Temperature on Rice Growth and Yield. *Agric. For. Meteorol*. 149, 51–58.

Cheng, W., H. Sakai, K. Yagi, and T. Hasegawa, 2010: Combined Effects of Elevated [CO_2] and High Night Temperature on Carbon Assimilation, Nitrogen Absorption, and the Allocations of C and N by Rice (*Oryza sativa L.*). *Agric. For. Meteorol*. 150, 1174–1181.

Chougule, K. (2017). *RNA-seq Tutorial- STAR, StringTie and DESeq2*. *CyVerse*. Retrieved 15 April 2018, from https://pods.iplantcollaborative.org/wiki/display/TUT/RNA-seq+Tutorial-+STAR%2C+StringTie+and+DESeq2.

Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P., Wu, X., Jiang, W., Marraffini, L. and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819-823.

Counce, P. A., R. J. Bryant, C. J. Bergman, R. C. Bautista, Y.-J. Wang, T. J. Siebenmorgen, K. A. K. Modenhauer, and J.-F. C. Meullenet, 2005: Rice Milling Quality, Grain Dimensions, and Starch Branching as Affected by High Night Temperatures. *Cereal Chem*. 82, 645–648.

Ding, Y., Li, H., Chen, L. and Xie, K. (2016). Recent Advances in Genome Editing Using CRISPR/Cas9. *Frontiers in Plant Science*, 7.

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., & Jha, S. et al. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. http://dx.doi.org/10.1093/bioinformatics/bts635.

Doudna, J. and Charpentier, E. (2014). The New Frontier of Genome Engineering with CRISPR-Cas9. *Science*, 346(6213), pp.1258096-1258096.

Elbaum, R., Zaltzman, L., Burgert, I., & Fratzl, P. (2007). The Role of Wheat Awns in the Seed Dispersal Unit. *Science*, 316(5826), 884-886. doi:10.1126/science.1140097.

Fan, J., Oliphant et al. (2003). Highly Parallel SNP Genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, 68(0), pp.69-78.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y. and Zhu, J. (2013). Efficient Genome Editing in Plants using a CRISPR/Cas System. Cell Research, 23(10), pp.1229-1232.

Ferrero, A., Vidotto, F., Balsari, P., & Airoldi, G. (1999). Mechanical and Chemical Control of Red Rice (Oryza sativa L. var. sylvatica) in rice (Oryza sativa L.) Pre-Planting. *Crop Protection*, *18*(4), 245-251. http://dx.doi.org/10.1016/s0261-2194(99)00022-8

Fitzgerald, M., and A. P. Resurreccion, 2009: Maintaining the Yield of Edible Rice in a Warming World. *Funct. Plant Biol*. 35, 1037–1045.

Fuller, D., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L., & Sun, G. (2009). The Domestication Process and Domestication Rate in Rice: Spikelet Bases from the Lower Yangtze. *Science*, *323*(5921), 1607-1610. http://dx.doi.org/10.1126/science.1166605

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Kadowaki, K. (2007). The *Rc* and *Rd* Genes are involved in Proanthocyanidin Synthesis in Rice Pericarp. *The Plant Journal*, 49(1), 91-102. doi:10.1111/j.1365-313x.2006.02958.x.

Furuta, T., Komeda, N., Asano, K., Uehara, K., Gamuyao, R., Angeles-Shim, R., Nagai, K., Doi, K., Wang, D., Yasui, H., Yoshimura, A., Wu, J., McCouch, S. and Ashikari, M. (2015). Convergent Loss of Awn in Two Cultivated Rice Species *Oryza sativa* and *Oryza glaberrima* Is Caused by Mutations in Different Loci. *Genes|Genomes|Genetics*, 5(11), pp.2267-2274.

Garg, K. (1999). Identification of Candidate Coding Region Single Nucleotide Polymorphisms in 165 Human Genes Using Assembled Expressed Sequence Tags. *Genome Research*, 9(11), pp.1087-1092.

Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., & Mccouch, S. (2005). Genetic Structure and Diversity in Oryza sativa L. *Genetics, 169*(3), 1631-1638. doi:10.1534/genetics.104.035642.

Gealy, D. R., Agrama, H. A., & Eizenga, G. C. (2009). Exploring Genetic and Spatial Structure of U.S. Weedy Red Rice (*Oryza sativa*) in Relation to Rice Relatives Worldwide. *Weed Science*, 57(06), 627-643. doi:10.1614/ws-09-018.1.

Glaszmann, J. C. (1987). Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics, 74*(1), 21-30. doi:10.1007/bf00290078.

GRiSP (Global Rice Science Partnership. (2013). *Rice almanac* (4th ed., pp. 2-4, 6-10). Los Baños (Philippines): International Rice Research Institute.

Gross, B. L., Reagon, M., Hsu, S., Caicedo, A. L., Jia, Y., & Olsen, K. M. (2010). Seeing Red: The Origin of Grain Pigmentation in US Weedy Rice. *Molecular Ecology, 19*(16), 3380-3393. doi:10.1111/j.1365-294x.2010.04707.x.

Gu, B., Zhou, T., Luo, J., Liu, H., Wang, Y., Shangguan, Y., Han, B. (2015). An-2 Encodes a Cytokinin Synthesis Enzyme that Regulates Awn Length and Grain Production in Rice. *Molecular Plant*, 8(11), 1635-1650. doi:10.1016/j.molp.2015.08.001.

Hartl, D., & Clark, A. (1997). *Principles of population genetics*. Sunderland, Mass.: Sinauer Associates.

Hoagland, R., & Paul, R. (1978). A Comparative SEM Study of Red Rice and Several Commercial Rice (Oryza sativa) Varieties. *Weed Science*, *26*(6), 619-625.

Hsu, P., Lander, E. and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*, 157(6), pp.1262-1278.

Huang, S., Pollack, H. and Shen, P. (2000). Temperature Trends Over the Past Five Centuries Reconstructed from Borehole Temperatures. Nature, 403(6771), pp.756-758.

Huang, S., Weigel, D., Beachy, R. and Li, J. (2016). A Proposed Regulatory Framework for Genome-Edited Crops. *Nature Genetics*, 48(2), pp.109-111.

Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., Han, B. (2012). A Map of Rice Genome Variation Reveals the Origin of Cultivated Rice. *Nature, 490*(7421), 497-501. doi:10.1038/nature11532

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., & Zhao, Q. et al. (2011). Genome-wide Association Study of Flowering Time and Grain Yield Traits in a Worldwide Collection of Rice Germplasm. *Nature Genetics*, *44*(1), 32-39. http://dx.doi.org/10.1038/ng.1018

Hwang, W., Fu, Y., Reyon, D., Maeder, M., Tsai, S., & Sander, J. et al. (2013). Efficient Genome Editing in Zebrafish Using a CRISPR-Cas System. *Nature Biotechnology*, *31*(3), 227-229. http://dx.doi.org/10.1038/nbt.2501

Izawa, T., Konishi, S., Shomura, A., & Yano, M. (2009). DNA Changes tell us About Rice Domestication. *Current Opinion in Plant Biology, 12*(2), 185-192. doi:10.1016/j.pbi.2009.01.004.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L. (2013). RNA-Guided Editing of Bacterial Genomes Using CRISPR-Cas Systems. *Nature Biotechnology*, 31(3), pp.233-239.

Kang, H., Zaitlen, N., Wade, C., Kirby, A., Heckerman, D., Daly, M. and Eskin, E. (2008). Efficient Control of Population Structure in Model Organism Association Mapping. *Genetics*, 178(3), pp.1709-1723.

Khush, G. (1997). Origin, Dispersal, Cultivation and Variation of Rice. *Plant Molecular Biology*, [online] 35(2), pp.25-34. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9291957 [Accessed 21 Feb. 2018].

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., & Kishimoto, N. et al. (2003). Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from japonica Rice. *Science*, *301*(5631), 376-379. http://dx.doi.org/10.1126/science.1081288

Kolodziejczyk, A., Kim, J., Svensson, V., Marioni, J. and Teichmann, S. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4), pp.610-620.

Konishi, S., Izawa, T., Lin, S., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006). An SNP Caused Loss of Seed Shattering During Rice Domestication. *Science*, 312(5778), pp.1392-1396.

Kinoshita,T. (1984) Gene analysis and linkage map. In “Biology of Rice” Tsunoda, S. and N. Takahashi (eds.), Japan Sci. Soc. Press, Tokyo. p. 187–273.

Kovach, M. J., Sweeney, M. T., & Mccouch, S. R. (2007). New Insights into the History of Rice Domestication. *Trends in Genetics*, 23(11), 578-587. doi:10.1016/j.tig.2007.08.012.

Krishnan, P., B. Ramakrishnan, K.R. Reddy and V. Reddy. 2011. Chapter Three-High-Temperature Effects on Rice Growth, Yield, and Grain Quality. *Adv. Agron*., 111: 87-206.

Kuriyama, H. and Kudo, M. (1967). Complementary Genes Ph and Bh Controlling Ripening-Black Coloration of Rice Hulls and their Geographical Distribution. *Ikushugaku zasshi*, 17(1), pp.13-19.

Li, C. (2006). Rice Domestication by Reducing Shattering. *Science*, 311(5769), pp.1936-1939.

Li, H., Chen, Z., Hu, M., Wang, Z., Hua, H., Yin, C. and Zeng, H. (2011). Different Effects of Night Versus Day High Temperature on Rice Quality and Accumulation Profiling of Rice Grain Proteins During Grain Filling. *Plant Cell Reports*, 30(9), pp.1641-1659.

Li, J., Norville, J., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G. and Sheen, J. (2013). Multiplex and Homologous Recombination–Mediated Genome Editing in *Arabidopsis* and *Nicotiana Benthamiana* Using Guide RNA And Cas9. *Nature Biotechnology*, 31(8), pp.688-691.

Li, L., Li, Y., Jia, Y., Caicedo, A., & Olsen, K. (2017). Signatures of Adaptation in the Weedy Rice Genome. *Nature Genetics*, *49*(5), 811-814. http://dx.doi.org/10.1038/ng.3825

Liao, J., Zhou, H., Peng, Q., Zhong, P., Zhang, H., He, C. and Huang, Y. (2015). Transcriptome Changes in Rice (*Oryza Sativa L.*) in Response to High Night Temperature Stress at the Early Milky Stage. BMC Genomics, 16(1), p.18.

Livak, K., Flood, S., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Research*, *4*(6), 357-362. http://dx.doi.org/10.1101/gr.4.6.357

Londo, J., & Schaal, B. (2007). Origins and population genetics of weedy red rice in the USA. *Molecular Ecology*, *16*(21), 4523-4535. http://dx.doi.org/10.1111/j.1365-294x.2007.03489.x

Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. (2017). Transcriptomics Technologies. *PLOS Computational Biology*, 13(5), p.e1005457.

Lu, G., Wu, Y., Bai, W., Ma, B., Wang, C. and Song, J. (2013). Influence of High Temperature Stress on Net Photosynthesis, Dry Matter Partitioning and Rice Grain Yield at Flowering and Grain Filling Stages. *Journal of Integrative Agriculture*, 12(4), pp.603-609.

Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., Feng, Q., Zhao, Y., Guo, Y., Li, W., Huang, X. and Han, B. (2010). Function Annotation of the Rice Transcriptome at Single-Nucleotide Resolution by RNA-Seq. *Genome Research*, 20(9), pp.1238-1249.

Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., & Shangguan, Y. et al. (2013). An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. *The Plant Cell*, *25*(9), 3360-3376. http://dx.doi.org/10.1105/tpc.113.113589

Ma, J., & Bennetzen, J. L. (2004). Rapid Recent Growth and Divergence of Rice Nuclear Genomes. *Proceedings of the National Academy of Sciences, 101*(34), 12404-12410. doi:10.1073/pnas.0403715101.

Maekawa, M. (1984). Geographical Distribution of the Genes for Black Hull Coloration. *Rice Genetics*, 1(1), pp.104-105.

Mali, P., Aach, J., Stranges, P., Esvelt, K., Moosburner, M., Kosuri, S., Yang, L. and Church, G. (2013). CAS9 Transcriptional Activators for Target Specificity Screening and Paired Nickases for Cooperative Genome Engineering. *Nature Biotechnology*, 31(9), pp.833-838.

Mammadov, J., Aggarwal, R., Buyyarapu, R. and Kumpatla, S. (2012). SNP Markers and their Impact on Plant Breeding. *International Journal of Plant Genomics*, 2012, pp.1-11.

Mammadov, J., Chen, W., Mingus, J., Thompson, S. And Kumpatla, S. (2011). Development of Versatile Gene-Based SNP Assays in Maize (*Zea mays L.*). *Molecular Breeding*, 29(3), pp.779-790.

Mao, Y., Zhang, H., Xu, N., Zhang, B., Gou, F. and Zhu, J. (2013). Application of the CRISPR–Cas System for Efficient Genome Engineering in Plants. *Molecular Plant*, 6(6), pp.2008-2011.

Marra, M., Hillier, L. and Waterston, R. (1998). Expressed Sequence Tags — Establishing Bridges Between Genomes. *Trends in Genetics*, 14(1), pp.4-7.

Matsushita,S., T. Kurakazu, Sobrizal, K. Doi and A. Yoshimura (2003) Mapping of genes for awn in rice using *Oryza meridionalis* introgression lines. Rice Genet. Newsl. 20: 17–18.

Mele, M., Ferreira, P., Reverter, F., DeLuca, D., Monlong, J., Sammeth, M., Young, T., Goldmann, J., Pervouchine, D., Sullivan, T., Johnson, R., Segre, A., Djebali, S., Niarchou, A., Consortium, T., Wright, F., Lappalainen, T., Calvo, M., Getz, G., Dermitzakis, E., Ardlie, K. and Guigo, R. (2015). The Human Transcriptome Across Tissues and Individuals. *Science*, 348(6235), pp.660-665.

Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H. and Qu, L. (2013). Targeted Mutagenesis in Rice Using CRISPR-Cas System. *Cell Research*, 23(10), pp.1233-1236.

Mohammed, A. R., and L. Tarpley, 2009a: High Nighttime Temperature Affects Rice Productivity through Altered Pollen Germination and Spikelet Fertility. *Agric. For. Meteorol*. 149, 999– 1008.

Mohammed, A. R., and L. Tarpley, 2009b: Impact of High Night-Time Temperature on Respiration, Membrane Stability, Antioxidant Capacity and Yield of Rice Plants. *Crop Sci*. 49, 313–322.

Mohammed, A. R., and L. Tarpley, 2009c: Instrumentation Enabling Study of Plant Physiological Response to Elevated Nighttime Temperature. *Plant Methods* 5, 7.

Mohammed, A. R., and L. Tarpley, 2010: Effects of High Night Temperature and Spikelet Position on Yield-Related Parameters of Rice (*Oryza Sativa L.*) *Plants. Eur. J.* Agron. 33, 117–123.

Mohammed, A. R., and L. Tarpley, 2015: 1-Methylcyclopropene (1-MCP)-Induced Alteration in Leaf Photosynthetic Rate, Chlorophyll Fluorescence, Respiration and Membrane Damage in Rice (*Oryza sativa L.*) Under High Night Temperature. *J. Agron. Crop Sci*. 201, 105-116.

Morita, S., Yonemaru, J. and Takanashi, J. (2005). Grain Growth and Endosperm Cell Size Under High Night Temperatures in Rice (Oryza sativa L.). *Annals of Botany*, 95(4), pp.695-701.

Morozova, O., Hirst, M. and Marra, M. (2009). Applications of New Sequencing Technologies for Transcriptome Analysis. *Annual Review of Genomics and Human Genetics*, 10(1), pp.135-151.

Mortazavi, A., Williams, B., McCue, K., Schaeffer, L. and Wold, B. (2008). Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq. *Nature Methods*, 5(7), pp.621-628.

Mortimer M, Pandey S, Piggin C. 2000. Weedy rice: Approaches to Ecological Appraisal and Implications for Research Priorities. In: Baki BB, Chin DV, Mortimer M, eds. Proceedings of Wild and Weedy Rice in Rice Ecosystems in Asia. A review. Los Banos, Philippines: International Rice Research Institute, 97–105

Moscou, M. and Bogdanove, A. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, 326(5959), pp.1501-1501.

Nadir, S., Xiong, H., Zhu, Q., Zhang, X., Xu, H., & Li, J. et al. (2017). Weedy Rice in Sustainable Rice Production. A Review. *Agronomy For Sustainable Development*, *37*(5). http://dx.doi.org/10.1007/s13593-017-0456-4

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008). The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. *Science*, 320(5881), pp.1344-1349.

Nagao, S. and Takahashi, M. (1954). Genetical Studies on Rice Plant. XVI. : Some Genes Responsible for Yellow, Brown and Black Color of Glume. *Ikushugaku zasshi*, 4(1), pp.25-30.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. and Kamoun, S. (2013). Targeted Mutagenesis in The Model Plant *Nicotiana Benthamiana* Using Cas9 RNA-guided Endonuclease. *Nature Biotechnology*, 31(8), pp.691-693.

Oka, H. I. (1988). *Origin of cultivated rice*. Elsevier.

Oost, J. (2013). New Tool for Genome Surgery. *Science*, 339(6121), pp.768-770.

Ozsolak, F. and Milos, P. (2010). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2), pp.87-98.

Ozsolak, F. and Milos, P. (2010). RNA Sequencing: Advances, Challenges and Opportunities. *Nature Reviews Genetics*, 12(2), pp.87-98.

Pabo, C., Peisach, E. and Grant, R. (2001). Design and Selection of Novel Cys2His2 Zinc Finger Proteins. *Annual Review of Biochemistry*, 70(1), pp.313-340.

Pan, Q., Shai, O., Lee, L., Frey, B. and Blencowe, B. (2008). Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing. *Nature Genetics*, 40(12), pp.1413-1415.

Pan, Q., Shai, O., Lee, L., Frey, B. and Blencowe, B. (2009). Erratum: Addendum: Deep Surveying of Alternative Splicing Complexity in The Human Transcriptome by High-Throughput Sequencing. *Nature Genetics*, 41(6), pp.762-762.

Patro, R., Duggal, G., Love, M., Irizarry, R., & Kingsford, C. (2015). Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment. http://dx.doi.org/10.1101/021592

Perez-de-Castro, M., Vilanova, A., Canizares, S., Pascual, J., Blanca, L., Diez, J., Prohens, J. and Pico, B. (2012). Application of Genomic Tools in Plant Breeding. *Current Genomics*, 13(3), pp.179-195.

Peng, S., Huang, J., Sheehy, J., Laza, R., Visperas, R., Zhong, X., Centeno, G., Khush, G. and Cassman, K. (2004). Rice Yields Decline with Higher Night Temperature from Global Warming. Proceedings of the National Academy of Sciences, 101(27), pp.9971-9975.

Pennisi, E. (2013). The CRISPR Craze. *Science*, 341(6148), pp.833-836.

Pertea, Mihaela, et al. "StringTie enables improved reconstruction of a transcriptome from RNA-Seq reads." *Nature Biotechnology*, vol. 33, no. 3, 2015, pp. 290–295., doi:10.1038/nbt.3122

Putney, S., Herlihy, W. and Schimmel, P. (1983). A New Troponin T and cDNA Clones for 13 Different Muscle Proteins, found by Shotgun Sequencing. *Nature*, 302(5910), pp.718-721.

Marra, M., Hillier, L. and Waterston, R. (1998). Expressed sequence tags — Establishing Bridges Between Genomes. *Trends in Genetics*, 14(1), pp.4-7.

Ramarathnam, N., Osawa, T., Namiki, M. and Tashiro, T. (1986). Studies on the Relationship Between Antioxidative Activity of Rice Hull and Germination Ability of Rice seeds. *Journal of the Science of Food and Agriculture*, 37(8), pp.719-726.

Rice Bran Color Samples : USDA ARS. (2018). *Ars.usda.gov*. Retrieved 16 April 2018, from https://www.ars.usda.gov/southeast-area/stuttgart-ar/dale-bumpers-national-rice-research-center/docs/rice-bran-color-samples/

Sandberg, R. (2014). Entering the Era of Single-Cell Transcriptomics in Biology and Medicine. *Nature Methods*, 11(1), pp.22-24.

Sang, T. and Ge, S. (2007). The Puzzle of Rice Domestication. *Journal of Integrative Plant Biology*, 49(6), pp.760-768.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J., Qiu, J. and Gao, C. (2013). Targeted Genome Modification of Crop Plants Using a CRISPR-Cas System. *Nature Biotechnology*, 31(8), pp.686-688.

Sheehy, J., Dionora, M. and Mitchell, P. (2001). Spikelet Numbers, Sink Size and Potential Yield in Rice. Field Crops Research, 71(2), pp.77-85.

Shirley, B. (1996). Flavonoid Biosynthesis: 'New' Functions for an 'Old' Pathway. *Trends in Plant Science*, 1(11), pp.377-382.

Singh, R., Singh, A., Sharma, T., Singh, A., & Singh, N. (2012). Fine mapping of grain length QTLs on chromosomes 1 and 7 in Basmati rice (Oryza sativa L.). *Journal Of Plant Biochemistry And Biotechnology*, *21*(2), 157-166. http://dx.doi.org/10.1007/s13562-011-0080-3

Steemers, F. and Gunderson, K. (2007). Whole Genome Genotyping Technologies on the BeadArray™ Platform. *Biotechnology Journal*, 2(1), pp.41-49.

Sutcliffe, J., Milner, R., Bloom, F. and Lerner, R. (1982). Common 82-Nucleotide Sequence Unique to Brain RNA. Proceedings of the National Academy of Sciences, 79(16), pp.4942-4946.

Sweeney, M. T. (2006). Caught Red-Handed: Rc Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. *The Plant Cell Online*, 18(2), 283-294. doi:10.1105/tpc.105.038430.

Sweeney, M., Thomson, M., Cho, Y., Park, Y., Williamson, S., Bustamante, C., & McCouch, S. (2007). Global Dissemination of a Single Mutation Conferring White Pericarp in Rice. Plos Genetics, 3(8), e133. http://dx.doi.org/10.1371/journal.pgen.0030133

Takamure, I. and T. Kinoshita (1991). Linkage analysis in chromosomes 3 and 6. Rice Genet. Newsl. 8: 98-100.

Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., Fu, Y., Cai, H., Wang, X., Xie, D. and Sun, C. (2008). Control of a key Transition from Prostrate to Erect Growth in Rice Domestication. *Nature Genetics*, 40(11), pp.1360-1364.

Tashiro, T. and Wardlaw, I. (1991). The Effect of High Temperature on the Accumulation of Dry Matter, Carbon and Nitrogen in the Kernel of Rice. Australian Journal of Plant Physiology, 18(3), p.259.

Toriba, T., Suzaki, T., Yamaguchi, T., Ohmori, Y., Tsukaya, H., & Hirano, H. (2010). Distinct Regulation of Adaxial-Abaxial Polarity in Anther Patterning in Rice. *The Plant Cell Online, 22*(5), 1452-1462. doi:10.1105/tpc.110.075291.

Vaughan, D. A., Morishima, H., & Kadowaki, K. (2003). Diversity in the *Oryza* Genus. *Current Opinion in Plant Biology, 6*(2), 139-146. doi:10.1016/s1369-5266(03)00009-8.

Visscher, P., Brown, M., McCarthy, M. and Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, 90(1), pp.7-24.

Vitte, C., Ishii, T., Lamy, F., Brar, D. and Panaud, O. (2004). Genomic Paleontology provides Evidence for Two Distinct Origins of Asian Rice (*Oryza sativa L.*). *Molecular Genetics and Genomics*, 272(5), pp.504-511.

Walia, H. (2005). Comparative Transcriptional Profiling of Two Contrasting Rice Genotypes under Salinity Stress during the Vegetative Growth Stage. *Plant Physiology*, 139(2), pp.822-835.

Wang, Zhong, Mark Gerstein, and Michael Snyder. "RNA-Seq: A Revolutionary Tool for Transcriptomics." *Nature Review Genetics* 10.1 (2009): 57-63. Web. 25 July 2017.

Wilhelm, B., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C., Rogers, J. and Bähler, J. (2008). Dynamic Repertoire of a Eukaryotic Transcriptome Surveyed at Single-Nucleotide Resolution. *Nature*, 453(7199), pp.1239-1243.

Xie, K. and Yang, Y. (2013). RNA-Guided Genome Editing in Plants Using a CRISPR–Cas System. *Molecular Plant*, 6(6), pp.1975-1983.

Yamakawa, H., T. Hirose, M. Kuroda, and T. Yamaguichi, 2007: Comprehensive Expression Profiling of Rice Grain Filling-Related Genes Under High Temperature Using DNA Microarray. *Plant Physiol*. 144, 258–277.

Yoshida, S., 1981: Physiological Analysis of Rice Yield. In: S. Yoshida, ed. *Fundamentals of Rice Crop Science*, pp. 231–251. International Rice Research Institute, Los Banos, Philippines.

Yoshida, S., Forno, D., Cock, J. and Gomez, K. (1977). Laboratory Manual for Physical Studies of Rice. 3rd ed. Manila: International Rice Research Institute.

Yoshimura, A., Nagayama, H., Sobrizal, Kurakazu, T., Sanchez, P., & Doi, K. et al. (2010). Introgression lines of rice (Oryza sativa L.) carrying a donor genome from the wild species, O. glumaepatula Steud. and O. meridionalis Ng. *Breeding Science*, *60*(5), 597-603. http://dx.doi.org/10.1270/jsbbs.60.597

Yu, Y., Tang, T., Qian, Q., Wang, Y., Yan, M., Zeng, D., Han, B., Wu, C., Shi, S. and Li, J. (2008). Independent Losses of Function in a Polyphenol Oxidase in Rice: Differentiation in Grain Discoloration Between Subspecies and the Role of Positive Selection Under Domestication. *The Plant Cell Online*, 20(11), pp.2946-2959.

Zakaria, S., T. Matsuda, S. Tajima, and Y. Nitta, 2002: Effect of High Temperature at Ripening Stage on the Reserve Accumulation in Seed in Some Rice Cultivars. *Plant Prod*. *Sci*. 5, 160– 168.

Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X., Chen, L., Tian, W., Tao, Y., Kristiansen, K., Zhang, X., Li, S., Yang, H., Wang, J. and Wang, J. (2010). Deep RNA Sequencing at Single Base-Pair Resolution Reveals High Complexity of the Rice Transcriptome. *Genome Research*, 20(5), pp.646-654.

Zhang, H., Duan, L., Dai, J., Zhang, C., Li, J., Gu, M., Liu, Q. and Zhu, Y. (2014). Major QTLs Reduce the Deleterious Effects of High Temperature on Rice Amylose Content by Increasing Splicing Efficiency of Wx pre-mRNA. *Theoretical and Applied Genetics*, 127(2), pp.273-282.

Zhao, K., Tung, C., Eizenga, G., Wright, M., Ali, M., Price, A., Norton, G., Islam, M., Reynolds, A., Mezey, J., McClung, A., Bustamante, C. and McCouch, S. (2011). Genome-wide Association Mapping Reveals a Rich Genetic Architecture of Complex Traits in *Oryza sativa*. *Nature Communications*, 2, p.467.

Zhou, H. and Steffenson, B. (2013). Genome-wide Association Mapping Reveals Genetic Architecture of Durable Spot Blotch Resistance in US Barley Breeding Germplasm. *Molecular Breeding*, 32(1), pp.139-154.

Zhu, B., Si, L., Wang, Z., Jingjie Zhu, Y., Shangguan, Y., Lu, D., Fan, D., Li, C., Lin, H., Qian, Q., Sang, T., Zhou, B., Minobe, Y. and Han, B. (2011). Genetic Control of a Transition from Black to Straw-White Seed Hull in Rice Domestication. *Plant Physiology*, 155(3), pp.1301-1311.

Zong, Y., Chen, Z., Innes, J. B., Chen, C., Wang, Z., & Wang, H. (2007). Fire and Flood Management of Coastal Swamp Enabled First Rice Paddy Cultivation in East China. *Nature*, 449(7161), 459-462. doi:10.1038/nature06135.