Multi-Stress Rice Transcriptome Analysis Using Network-based Approach

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science
In
Computational Natural Sciences by Research

by

Mayank Musaddi
20171115
mayank.musaddi@research.iiit.ac.in

International Institute of Information Technology, Hyderabad
(Deemed to be University)
Hyderabad - 500 032, INDIA
April 2023

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "Multi-stress rice transcriptome analysis using network-based approach" by Mayank Musaddi, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. Nita Parekh

To my family

Acknowledgments

I would like to express my deepest gratitude to all the people around me whose support and guidance led me to complete this thesis. First and foremost, I would like to thank my advisor Dr. Nita Parekh for her guidance and invaluable feedback. The regular discussions and meetings helped me understand the research problem and generated curiosity in this domain. She insisted on high standards which helped instill in me professionalism and diligence towards my work. I am highly indebted to my mentor and friend Dr. Sanchari Sircar without whom this thesis would not have been possible. Working alongside with her gave me the opportunity to have critical discussions about the research topic which helped me learn a great deal. She actively helped me in all aspects of my research work. I would extend my gratitude to Akshay Gupta, Yash Agarwal, Broto Chakrabarty, Suba Suseela, Sri Lakshmi, and others with whom I shared group meetings which led to insightful discussions and feedbacks that were very valuable for my research.

I would like to thank the Centre for Natural Sciences and Bioinformatics (CCNSB) and the respective faculties for providing a conducive environment for research. I am grateful to all my friends Tanuj, Parth, Akshay, Atirek, Shaunak, Aditya and many others who have always held my back and made my years at the institute so memorable. Lastly, I would thank my parents who believed in me and guided me in all aspects of my life.

Abstract

Rice is a fascinating and complex plant. Consumed by more than half of the world's population, it is important to have a comprehensive understanding of the organism to advance crop engineering and breeding strategies. Abiotic stresses like drought, high temperature, salinity and flood have affected its growth and productivity. Furthermore, global climate change has added to the severity of these stresses, suggesting the need for varieties with improved stress tolerance for sustainable crop production. Improving stress tolerance requires an in-depth understanding of the biological processes, transcriptional pathways and hormone signaling involved in stress response. With the surge in omics data, it has paved the way for deciphering the biological information underlying complex traits. However, dealing with such large datasets calls for the development of powerful bioinformatics methods for a thorough transcriptome analysis. A popular approach is the construction and analysis of co-expression networks representing transcriptionally coordinated genes that are often part of the same biological process. Using prior knowledge and data integration further enhances the elucidation of gene regulatory relationships in this network.

With this objective we have developed NetREx, a Network based Rice Expression Analysis Server, that hosts ranked co-expression networks of *Oryza sativa* using publicly available mRNA-seq data across uniform experimental conditions. It provides a range of interactable data viewers and modules for analysing user queried genes across different stress conditions (drought, flood, cold and osmosis) and hormonal treatments (abscisic and jasmonic acid) and tissues (root and shoot). Subnetworks of user-defined genes can be queried in preconstructed tissue-specific networks, allowing users to view the fold-change, module memberships, gene annotations and analysis of their neighborhood genes and associated pathways. The webserver also allows querying of orthologous from Arabidopsis, wheat, maize, barley, and sorghum. Here we demonstrate that NetREx can be used to identify novel candidate genes and tissue-specific interactions under stress conditions and can aid in the analysis and understanding of complex phenotypes linked to stress response in rice. Available at: https://bioinf.iiit.ac.in/netrex/.

In the second part of the thesis, we present a meta-analytic study using co-expressed modules to understand the biological functions associated with different abiotic stresses in the root tissue. The osmotic stress condition is an extremely severe stress condition involving the effects of multiple stresses like drought, salinity and ionic stress and is discussed in detail. The early responsive modules are analyzed and a causal flow of mechanisms and signaling pathways is established.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Rice is the most widely consumed staple food crop across the world and is a model monocot system (Cantrell & Reeves, 2002). It is consumed by more than half of the world's population (Seck et al., 2012) and provides for one-fifth of the calories consumed worldwide. The cultivation of rice is well suited to countries with low labor costs and high rainfall, as it is labor intensive and requires an ample supply of water. There are two main species of rice which are cultivated worldwide, *Oryza sativa* (Asian rice) and *Oryza glaberrima* (African rice). While *Oryza sativa* is native to tropical southern and southeastern Asia, *Oryza glaberrima* is only grown in South Africa. *Oryza sativa* is further classified into three subspecies based on its geographical environment: indica, japonica and javanica. The indica variety refers to the tropical and subtropical varieties native to southern and southeastern Asia. The japonica variety is native to temperate areas of Japan, China and Korea. While the javanica varieties are grown in Indonesia alongside of indica (*Botanical Classification of Rice | Agropedia*, n.d.). About 91% of the rice in the world, accounting for nearly 640 million tons, is grown in Asia. China shares nearly 30% of the total rice worldwide production followed by India (21%), Indonesia (9%) and Bangladesh (6%) (Rosell & Marco, 2008).

The demand for rice is increasing with the exponential growth in world population. Hence its yield needs to be increased by agricultural or biotechnical approaches. Climate changes including variability in temperature and rainfall pattern and other factors severely affect the yield of rice. Rice is subjected to various kinds of abiotic and biotic stresses that affect its yield during different growth stages. Abiotic stresses include drought, salinity, extreme temperatures, flooding, mineral deficiency, heavy metals, pollutants, wind and mechanical injury. Biotic stresses affecting rice yield majorly include bacterial leaf blight (BLB), sheath blight (ShB), blast, brown spot (BS), false smut (FS), brown plant hopper (BPH) and gall midge (GM). There are a variety of ways in which plants respond to stress. Common response towards abiotic stress includes stomatal closure, increased reactive oxygen scavenging activity, reduced photosynthesis, delayed growth, and increased root length (Maiti & Satya, 2014). Response towards biotic stress such as pathogen also causes stomatal closure and reduced photosynthesis (Bilgin et al., 2010; Melotto et al., 2006). Other responses toward pathogen infection include production of toxic compounds like phytoalexins and reactive oxygen species, and induction of localized cell death (Wojtaszek, 1997).

Most of these responses are mediated by phytohormones (Nguyen et al., 2016). Abscisic acid (ABA) and jasmonic acid (JA) are critical regulators of tolerance toward abiotic stress. For pathogen immunity, plants rely on ethylene, salicylic acid (SA) and jasmonic acid signaling. Many transcription factor (TF) families including both ABA-dependent and ABA-independent, play a significant role in abiotic stress response. ABA-dependent TFs include basic leucine zipper (bZIP) TFs which induce stomatal closure, dehydration tolerance gene expression, and other adaptive changes (Banerjee & Roychoudhury, 2017; Kim et al., 2010). However, ABA acts antagonistically to SA thus increasing the plant's susceptibility to biotic interaction (de Torres Zabala et al., 2009). When compared to other crops, rice has an antagonistic characteristic about tolerances and susceptibilities towards abiotic stresses. Since it grows in standing water containing soil, it can tolerate submergence at levels that would kill other crops. However, it is highly susceptible to drought and cold and moderately tolerant towards salinity and soil acidity.

In this thesis, we have made an attempt to get insights into the mechanisms involved in abiotic stress tolerance in rice. Several studies are present exploring broad plant stress response by analyzing microarray data (Hahn et al., 2013). We have attempted to expand on these studies with robust meta-analysis of publicly available rice RNA-seq data sets.

1.2 Abiotic Stress Mechanisms in Rice

Exposure to abiotic stresses is essentially unavoidable. It is the most harmful factor as it affects the growth and productivity of crops worldwide (Gao et al., 2007). A combination of abiotic stressors is more detrimental than when they act in isolation (Mittler, 2006). Drought is the most damaging stress for rice farming, followed by salinity which is determined by high concentration of salts in the soil. Due to global climate change, heat stress has also become detrimental to rice production. At high altitudes, low temperatures hamper rice production, as it has a negative impact throughout the germination, development, and reproductive phases. Multiple stress like salt and drought, or drought followed by flooding, may also act at the same time, resulting in a high decrease in production. Hence combined tolerance to several forms of abiotic stress would significantly increase rice productivity.

Due to the sessile nature of plants, they must confront the stress and potentially develop adaptations to avoid or tolerate the harsh effects of stress in order to survive. Over the years, there have been plenty of cellular, morphological, and physiological adaptations in crops to defend them against stress. The cuticle is the most apparent defense which is a universal outmost shell (Fich et al., 2016; Shepherd & Griffiths, 2006). Recretohalophytes even developed a specialized organ to excrete salt (Yuan et al., 2016). Understanding of the biochemical and molecular processes involved in defense of various model species like *Arabidopsis thaliana* has provided insights into the defense response at the cellular level. Generalized and conserved cellular defense responses involve desaturation of membrane lipids, activation of reactive species, scavengers, and induction of molecular chaperones. Stress defense involves a complex regulatory network of molecules including

hormones, reactive oxygen species (ROS), hydrogen sulphide (H_2S), nitric oxide (NO), polyamines (PAs), and calcium (Ca^{2+}), as well as protein kinases and transcription factors (**Figure 1.1**).

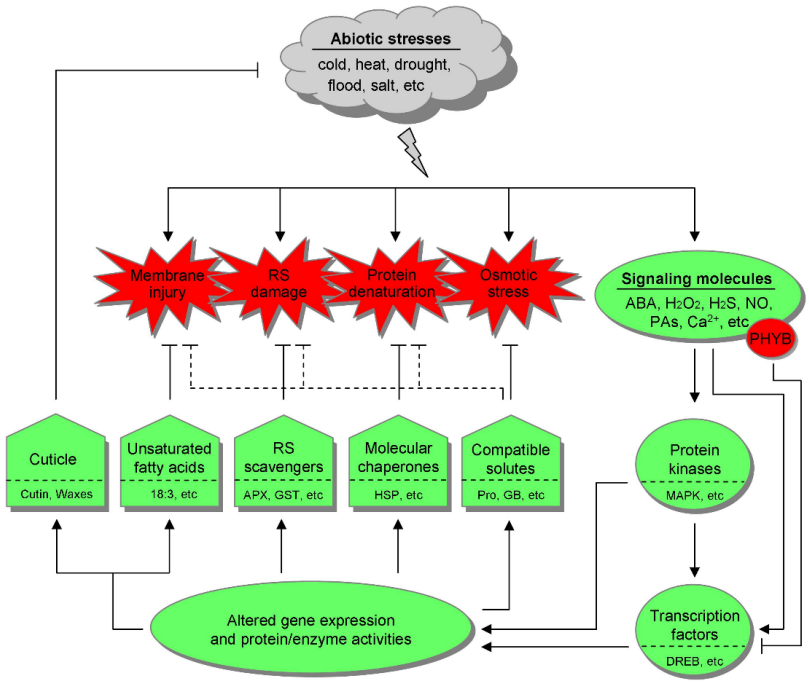


Figure 1.1 The general defense systems and the underlying regulatory network in botanic responses to abiotic stresses [reproduced from (He *et al.*, 2018)].

Abiotic stress signaling is a multi-faceted phenomenon. Plants elicit specific abiotic stress signals in response to a particular stress condition. However, there is a significant overlap between the abiotic signals. This is attributed due to different stresses happening in conjunction or in sequence. For example, drought stress is linked to heat stress as high temperatures also lead to loss of water. The first phase in response to stress is signal perception which leads to the modification in the quantity of several secondary signals. These signals cause a protein phosphorylation cascade that lead to the activation of specific transcription factors (TFs) or target genes. The signaling molecules also serve as checkpoints for signals to flow in a particular direction. There are several signaling pathways that are reported to be triggered in response to abiotic stress (Pearce, 2007).

1.2.1 ROS Signaling

Oxygen, while being an essential element for plant development and growth, also serves as a substrate for Reactive Oxygen Species (ROS). Aerobic metabolic activities like photosynthesis and respiration lead to formation of ROS like Hydrogen peroxide (H_2O_2), hydroxyl radical, superoxide radical, and singlet oxygen. Stress triggers increased production of ROS from organelles such as mitochondria, peroxisomes, and chloroplast, which are corrosive in nature and leads to apoptosis or cellular damage by their reaction to nucleic acids, proteins, and lipids. There are several ROS homeostasis mechanisms involved to keep its level in check. Antioxidants like Catalase (CAT), monodehydroascorbate reductase (MDHAR), superoxide dismutase (SOD), etc. are ROS foraging

enzymes to detoxify its effects. Several non-enzymatic antioxidants like tocopherols, carotenoids, GSH, etc. also take part in maintaining homeostasis. Other mechanisms to prevent ROS overproduction include photosynthetic apparatus rearrangement, leaf movement and leaf curling (Gill & Tuteja, 2010).

1.2.2 Calcium Signaling

Calcium plays an important role in a variety of abiotic stress responses. They are central signaling molecules which govern many functions. After its activation, calcium ions travel through specific calcium ion channels, cell membranes or organelles into the cytosol. They activate various calcium-dependent proteins like calmodulin, calcium dependent protein kinases (CDPKs) and calmodulin-dependent phosphatases. Increase in local calcium levels occur in specific organelles like the chloroplasts thus triggering organelle specific actions. Ca^{2+} plays an important role in both ABA-dependent and ABA-independent pathways for stress response. ABA or osmotic stress induces a spike in the levels of intracellular Ca^{2+} which likely affects calcium dependent protein kinases (CDPKs) that are involved in typical ABA responses like stomatal closure and gene expression control. Another important calcium dependent kinase family is the CBL-interacting protein kinase (CIPK)/SnRK3 family which plays a significant role in drought and osmotic stress signaling.

Together with ROS signaling, calcium waves are also central for systemic signaling (**Figure 1.2**). Pathogen infection and wounding, as well as abiotic stresses like drought, salt, cold and heat elicit systemic responses such that locally applied stress causes responses not only locally but in distal tissues as well. This results in a systemic acquired acclimation (SAA). Figure 1.2 illustrates the mechanism of systemic signaling wherein local exposure to stress results in the increase in levels of Ca^{2+}. The Ca^{2+} signals then activate CDPKs and CBLs-CIPKs which phosphorylates and activate Respiratory burst oxidase homolog protein D (RbohD). The activated RbohD generates H_2O_2 which diffuses through the cell wall to neighboring cells inducing Ca^{2+} signals through Receptor Like Kinases (RLKs) like Guard cell Hydrogen peroxide-Resistant 1 (GHR1). The H_2O_2 enters the cell wall as well through PIP water channels and activates Ca^{2+} signals intracellularly. This mutual activation between Ca^{2+} and H_2O_2 forms a self-propagating loop that can travel distal tissues at rapid speed causing SAA response. Stress triggered calcium and ROS signals can travel with speed exceeding 1000 μm per second enabling rapid transmission of signals from root to shoot (Choi et al., 2014).

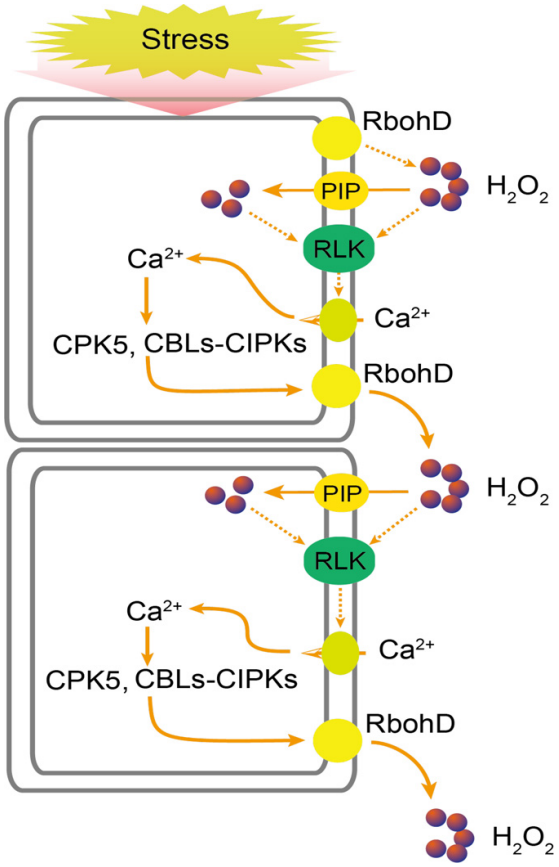


Figure 1.2 Systemic signaling model. Dashed lines indicate postulated regulation [reproduced from (Zhu, 2016)].

1.2.3 Salt Overly sensitive (SOS) signaling

Salinity results in severe osmotic pressure and scarcity of water causing ions like Na^+, K^+ and Mg^+ to accumulate in the root tissue of plants. The presence of high levels of ions causes ion toxicity and secondary stresses like oxidative stress. Salt tolerance is achieved by limiting the quantity of Na^+ ions. This is achieved by transporting these ions over long distances to leaves by the transpirational stream or pumping them out from the tissue. The gene SKC1/ HKT8 as well as the gene HKT1, both belonging to the High-affinity K+ Transporter (HKT) family are associated with long-distance trafficking of Na^+, hence maintaining a strong K^+/Na^+ balance in the plant (Golldack et al., 2002). Plants also use a calcium-dependent protein kinase pathway for salt stress tolerance known as the Salt-Overly-Sensitive (SOS) pathway (Zhu, 2002). In this pathway, the cytosolic calcium signal elicited by the salt stress is sensed by the EF-hand calcium-binding protein SOS3. SOS3 is expressed in the root, while an SOS3 paralog SCaBP8/CBL10 is expressed in the shoot (Quan et al., 2007). SOS3 in turn activates a serine/threonine protein kinase SOS2, which then phosphorylates and activates SOS1 (**Figure 1.3**). SOS1 is responsible for facilitating the outflow of Na^+ across the plasma membrane from the root epidermal cells. SOS1 is also expressed in the xylem parenchyma cells and takes part in loading Na^+ into the xylem for long distance transport to the leaves by the transpirational stream (Shi et al., 2002).

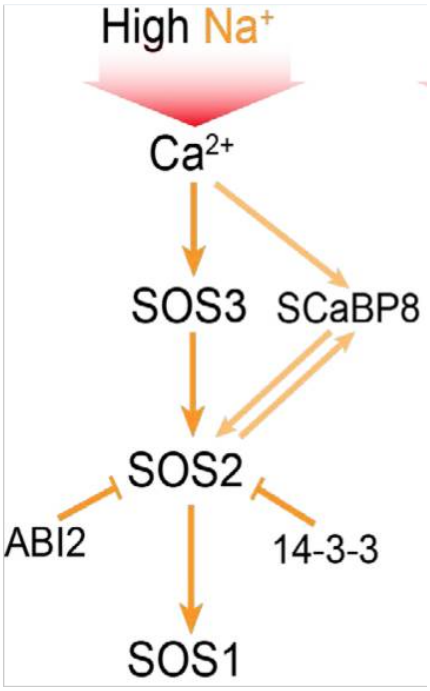


Figure 1.3 The Ca2+-CBL-CIPK module mediates signaling of ionic stresses. Arrows indicate activation, and bars indicate inhibition [reproduced from (Zhu, 2016)].

1.2.4 Phytohormone Signaling

Phytohormones are chemical messengers that help in coordination of cellular activities. They play important role in plants by regulating their growth, development, reproductive process, longevity and cell death. Phytohormones play a major role in triggering abiotic stress response in plants. They have the ability to either carry out their function in their synthesis site or move to their active location. Plant phytohormones broadly constitute of cytokinin (CK), indole acetic acid (IAA), salicylic acid (SA), gibberellins (GAs), ethylene, brassinosteroids (BRs), abscisic acid (ABA) and jasmonic acid (JA).

Cytokinins are adenine derivatives grouped into isoprenoid cytokinin and aromatic cytokinin. The presence of highly active growth substance, meta-topolin, belonging to aromatic CKs is suggestive of their important role as plant growth regulators as well as modulation of drought stress response. The endogenous levels of IAA regulated by auxin biosynthesis pathway has been reported to positively influence plant's drought stress tolerance. SA which has been majorly studied for its role in biotic stress has also potentially shown to regulate few abiotic stresses like drought stress by regulating stomatal aperture and transcriptional regulation of drought stress related genes. However, depending on the SA dosage, it implicates drought tolerance as well as sensitivity. The GAs belonging to the group of tetracyclic diterpenoid carboxylic acids, play active role in abiotic stress tolerance by regulating plant growth and development by affecting the cell division and cell elongation machinery. Ethylene is another important phytohormone central to plant growth and development. Ethylene plays a role in biotic and abiotic stresses by regulating flower senescence,

fruit ripening and abscission of leaves and petals. BRs are an important class of steroidal derived phytohormones that play important role in drought stress tolerance as it is involved in multiple physiological responses like pollen tube and fertility, ethylene synthesis, cell elongation, leaf bending and seed germination. ABA and JA are key stress response hormones. Stress responsive transcription factors like ABFs and MYCs are direct components involved in ABA and JA signaling.

ABA Signaling Pathway

ABA is the most significant phytohormone involved in abiotic stress response in plants. It mediates functions like root growth inhibition, leaf senescence and stomatal closure in response to stress. Abiotic stress promotes the biosynthesis of ABA (Seki et al., 2002). ABA is synthesized from β-carotene which requires the enzymatic activities of 9-cis-epoxycarotenoid dioxygenase (NCED), abscisic aldehyde oxidase (AAO), cytosolic short-chain dehydrogenase/reductase (SDR), and molybdenum cofactor sulfurase (MCSU). The transcription of these genes along with many stress-responsive transcription factors belonging to bZIP, MYC, NAC, AP2/ERF and MYB families, which are involved in ABA biosynthesis, are increased in response to abiotic stresses such as drought and osmosis (Zong et al., 2016).

The ABA signal transduction pathway involves three major components: the Pyrabactin Resistance 1 PYR/PYL/RCAR-type ABA receptor, 2C-type protein phosphatase (PP2C), and SNF1-related protein kinase 2 (SnRK2) (Fujita et al., 2013). These activate target transcription factors like the basic leucine zipper (bZIP) and ABA-responsive element binding factor (ABF) transcription factors which regulate ABA-responsive genes (ABREs) establishing stress-specific transcription. The ABFs are activated through direct phosphorylation from ABA-activated protein kinases comprising of the SnRK2 family of genes. However, in the absence of ABA, PP2Cs inhibit the kinase activity of SnRK2s thus blocking ABF activation (Umezawa et al., 2009). Hence, the activity of PP2Cs needs to be inhibited for the signal transduction to follow through. This is achieved with the PYR/PYL/RCAR protein which has a hydrophobic pocket that can recognize ABA as a ligand. The ABA binding alters the conformation of PYR/PYL/RCAR protein by closing the “gate and latch” structure over the ABA pocket, enabling it to interact and inhibit the phosphatase activity of PP2Cs. Hence, in case of abiotic stress, the synthesis of ABA is promoted leading to the formation of PYR/PYL-PP2C complex which inhibits PP2Cs activity allowing SnRK2s to phosphorylate and activate ABFs (**Figure 1.4**). The activated ABFs then bind to ABREs, activating the transcription of other stress-responsive transcription factors like NACs and AP2/ERF which are responsible for expression of stress responsive genes.

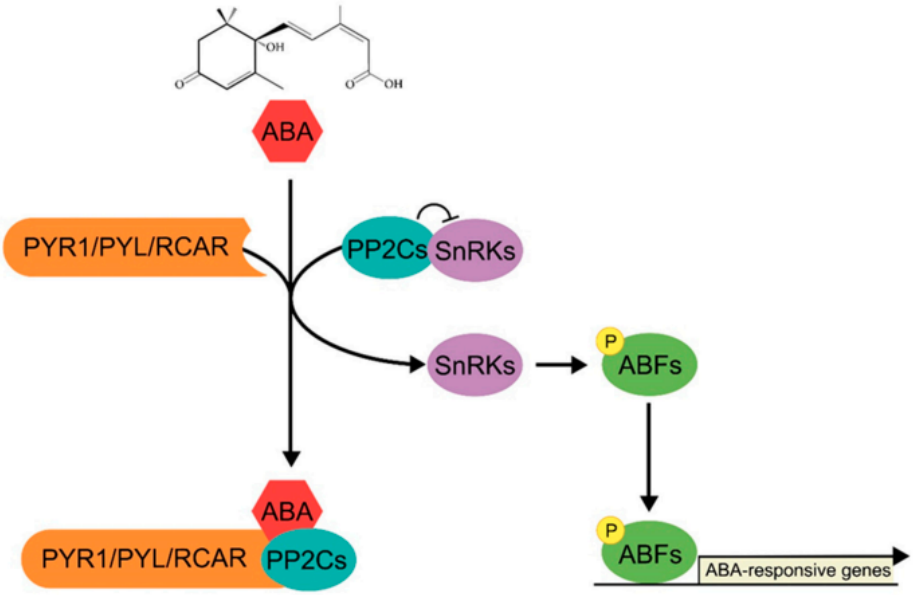


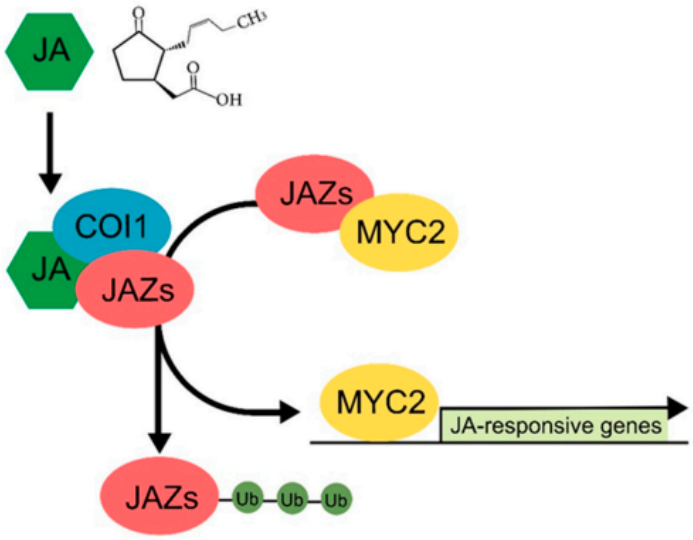
Figure 1.4 A schematic representation of ABA signaling pathway [reproduced from (Yoon *et al*., 2020)].

JA Signaling Pathway

JA is another phytohormone playing an important role in plant responses to environmental stresses (Wasternack & Hause, 2013). It is a cyclopentane fatty acid that was first isolated as a methyl ester from *Jasminum grandiflorum*. The biosynthesis of JA from linolenic acid takes place via the octadecanoid pathway involving various enzymes like lipoxygenase (LOX), allene oxide synthase (AOS), allene oxide cyclase (AOC), and 12-oxo-PDA-reductase (OPR). JA is further metabolized into its active form into JA-isoleucine conjugate (JA-Ile) or methyl jasmonate (MeJA) through the activity of jasmonate-amido synthetase 1 (JAR1) and jasmonate methyl transferase (JMT) respectively. Abiotic stress promotes the biosynthesis of JA by the regulation of JA biosynthesis genes. For example, in case of cold temperatures, JA signals are activated by inducing expression of JA biosynthesis genes like LOX, AOS, AOC in *Arabidopsis* and rice (Du et al., 2013; Hu et al., 2013). Exogenous treatment of JA has been shown to improve tolerance towards cold. Overexpression of JA biosynthesis gene AOC1 (TaAOC1) in the case of wheat showed enhanced tolerance to salt stress due to the role of JA (Qiu et al., 2014).

The JA signal transduction pathways involve two major components, the JA receptor Coronatine Insensitivity 1 (COI1) and the JA signaling repressors Jasmonate ZIM-domain proteins (JAZs) (Chini et al., 2007; Yan et al., 2009). The activated form of JA, mainly involving JA-Ile, activates JA signaling via their interaction with COI1. This interaction results in the proteolysis of JAZs. JAZs would suppress the activity of MYC2 transcription factor, which regulate the expression of JA-dependent stress-responsive genes. With the degradation of JAZs, MYC2 is liberated which activates

JA-responsive genes (**Figure 1.5**). MYC2 plays a pivotal role in JA signaling and is reported to be involved in tolerance to abiotic stress such as oxidative stress (Sasaki-Sekimoto et al., 2005). The MYC transcription factors are characterized by a basic helix-loop-helix (bHLH) domain that belongs to the bHLH family (Kazan & Manners, 2013). In rice, the gene OsbHLH148 has a similar function to MYC2. It directly interacts with JAZ proteins, thereby acting as a JA signaling component (Seo et al., 2011). Expression of OsbHLH148 dynamically increased with the various abiotic stresses like drought, salinity, cold and wounding, as well as with exogenous JA treatment.



Various bHLH transcription factors including MYC2 and OsbHLH148 are also involved in ABA-mediated stress tolerance. Their expressions are strongly upregulated with ABA treatment and ABA-related stresses. Results show the promotion of ABA biosynthesis by the activation of JMT involved in JA production (Kim et al., 2009). Interaction between the ABA receptor PYL6/RCAR9 and MYC2 has also been observed (Aleman et al., 2016). This suggests an interaction between JA and ABA to modulate stress response where the bHLH transcription factors play a central role in JA-ABA crosstalk.

1.3 Transcriptome Studies

The central dogma of molecular biology explains the flow of genetic information in a biological system. It involves a two-step process, the transcription of DNA to form RNA and the translation of a subset of RNAs called the messenger RNAs, mRNAs into proteins. The DNA is a heritable genetic information repository which is acted upon by RNA polymerase enzymes to form RNA which is a short-lasting information carrier. The transcriptome is defined as the complete set of all RNA transcripts in a cell, a set of cells or in a complete organism. Transcriptome analysis is hence the study of the transcriptome, or the study of the complete set of RNA transcripts that have been produced by an organism under certain circumstances or in its specific cells. Transcriptome profiling monitors the changes in behavior of a cell considering the variations in the complete set of RNA transcripts, making it useful in biomedical research in the areas of biomarker discovery, diagnosis of

diseases as well as risk assessments of new drugs or environmental impact. This method is particularly useful in characterizing gene functions and identifying pathways that respond to environmental stresses in plants. This is achieved by comparing pair of samples based on the condition or tissue that needs to be studied. For example, the transcriptomes are compared for samples obtained from the same organism when exposed to different external conditions (environmental conditions or healthy and diseased states) to study the effects of the condition. The transcriptome can also be characterized for organisms and tissues at various stages of development to understand the processes involved in embryonic development or cellular differentiation.

There are primarily two approaches used for transcriptome analysis – microarrays and RNA-Seq analysis. Microarrays were an early approach to study transcriptomes and mostly represented mRNAs, the genes that translated to proteins. Microarray array works on the principle that complementary nucleic acids will hybridize. It uses a glass slide where DNA molecules are fixed in an orderly manner at specific locations called probes or spots. These spots are printed on the glass using different technologies like robot spotting or photolithography. The DNA spot either contains a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA or anti-sense RNA sample called the target. These spots also referred to as targets can be hybridized by two samples (probes) labelled with different fluorescent dyes simultaneously. This probe-target hybridization is detected and quantified by fluorophore, silver or chemiluminescence labeled targets using confocal laser scanners to determine the abundance of nucleic acid in each of the sample. The separate scanned image capturing the intensity of signals at different spot locations for the two samples are then combined and pseudo colored by means of computer software to denote the relative expression levels. (**Figure 1.6**). Hence microarrays are able to analyze the expression of several genes in a single reaction in a quick and efficient manner. Microarrays however has the limitation of needing transcriptome to be available beforehand. Hence it is unable to find any novel transcripts, hampering the information of the regulatory gene network involved in stress response.

Earlier, several gene expression studies concerning stress-response and tolerance in plants have been carried out using microarrays. In rice, one of the earliest studies for monitoring global expression profiles in different stress condition like cold, drought and high salinity used cDNA microarray and RNA Gel blot analysis (Rabbani et al., 2003). *Rabbani* was able to identify 73 genes as stress inducible out of which 58 were novel and unreported genes in rice. A general flow for characterizing genes in plants involved the introduction of new traits in transgenic plants using gain-of-function mutagenesis, followed by transcriptional profiling using microarrays. In plants it has been used extensively to characterize downstream targets of stress responsive genes (Dubouzet et al., 2003; Li et al., 2019). The gene OsDREB1 was characterized to be potentially useful in response to drought, salt and cold stress. A NAC transcription factor, SNAC1 was identified and reported as a positive regulator of drought resistance in rice.

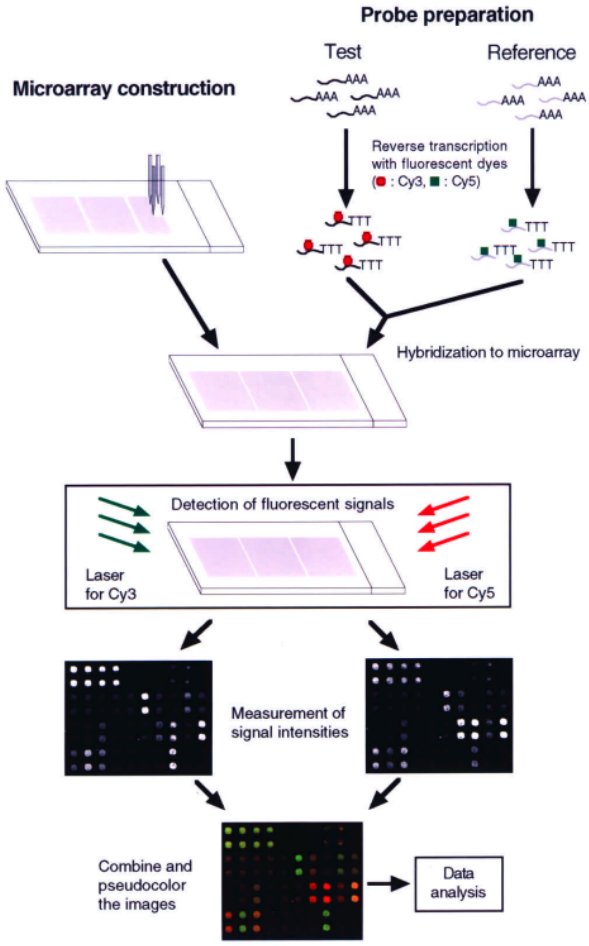


Figure 1.5 Principle of the cDNA microarray analysis system [reproduced from (Iida & Nishimura, 2002)].

With the availability of high-throughput techniques like the whole genome transcriptome analysis, small RNA sequencing analysis (RNA-Seq), proteomic analysis, epigenetic sequencing analysis and metabolomic analysis, an in depth understanding of complex regulatory gene network involved in stress response can be achieved. These high-throughput techniques use sequence-based approaches instead of hybridization-based approaches allowing them to determine, map and quantify transcript sequences from new genomes directly. RNA-Seq method in particular is superior due to its high coverage of genome, global expression of transcripts and information about allele specific expressions and alternate splicing. It is a throughput technique which involves isolation of the total RNA, conversion into a library of cDNA fragments by attaching adaptors to one of both ends, followed by sequencing of each molecule to obtain short reads (of length 20-300bp) from one end (single-end) or both ends (paired-end) (**Figure 1.7**). After the sequencing, the reads are aligned and mapped to the reference genome. The number of reads aligned to each gene is called ‘counts’ and it serves as a digital measure of gene expression levels in the investigated sample. The counts are then

normalized and corrected for different sources of bias like sequencing-depth of a sample and gene length bias. Differential expression are then evaluated based on statistical significant change in gene expression in the compared conditions. Common sequencing platform include Illumina IG, Roche 454 and Applied Biosystems.

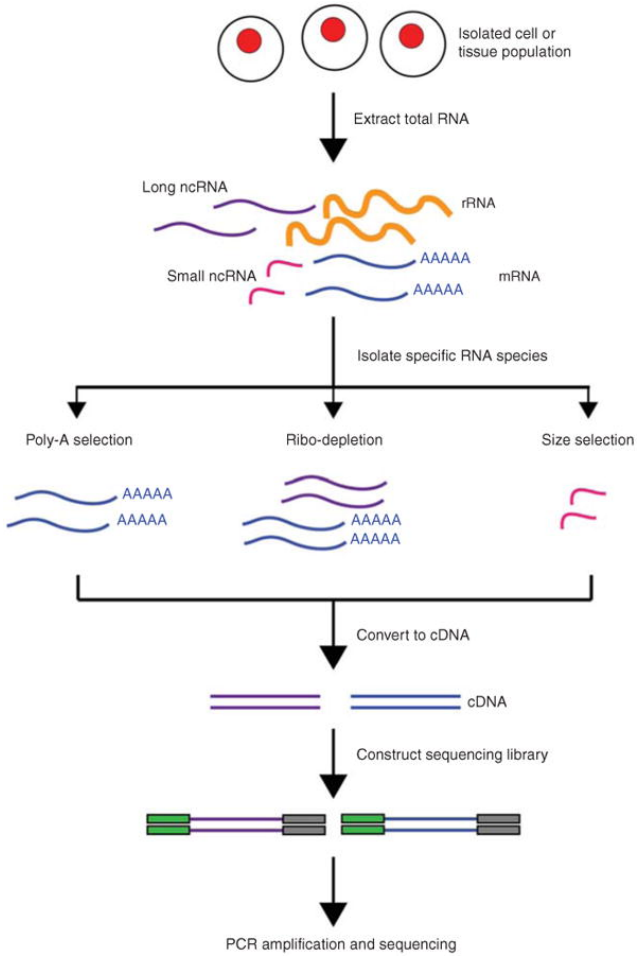


Figure 1.6 Schematic representation of RNA-Seq protocol [reproduced from (Kukurba & Montgomery, 2015)].

There have been several recent studies on rice using RNA-seq analysis concerning different aspects like genetic plasticity to varying environmental conditions (Kumar et al., 2022), characterization of genes involved in aerobic adaptations (Phule et al., 2019) and abiotic stresses (He et al., 2015), studying pathways involved in host-pathogen interactions (Liao et al., 2019), etc.

The rapid accumulation of large omics data has allowed the analysis of biological information and understanding of the underlying complex traits. The most popular method to interpret such large-scale data is the use of co-expression networks, which provides a visual and analytical approach for associating genes based on their expression values. A co-expression gene network is constructed by defining scores (like Pearson's correlation coefficient, or mutual information) between pair of genes based on the similarity of their expression profiles. A gene pair is considered co-expressed if its score

crosses a certain threshold. Modules of co-expressed genes can be identified and characterized for their biological function from the co-expression network. The prediction of gene function based on their co-expression analysis obeys the "guilt-by-association" principle (Wolfe et al., 2005). This is useful for hypothesizing the function of an unknown gene based on the co-expressed module it belongs to. While the construction of these networks is straightforward, the resulting network may become highly complex and may impede its biological interpretation. Several approaches are present to enhance the interpretability of these networks. Augmenting the network with data from external resources like gene ontology, pathway information and transcription factor annotations may further enhance gene prioritization and improve the elucidation of gene regulatory relationship. The workflow of co-expression network analysis is given in **Figure 1.8**. The correlation between expression values of gene pairs is evaluated. These pairwise correlation is represented as a network which can be clustered into modules. Analysis involving functional enrichment, identification of hub genes and regulators, differential co-expression analysis are carried out for these modules to understand its characteristic and change in behavior under different condition. Potential genes involved in the condition can be identified using the 'guilt-by-association' principle as the genes that are co-expressed with other genes involved in the condition (disease-associated module in the figure).

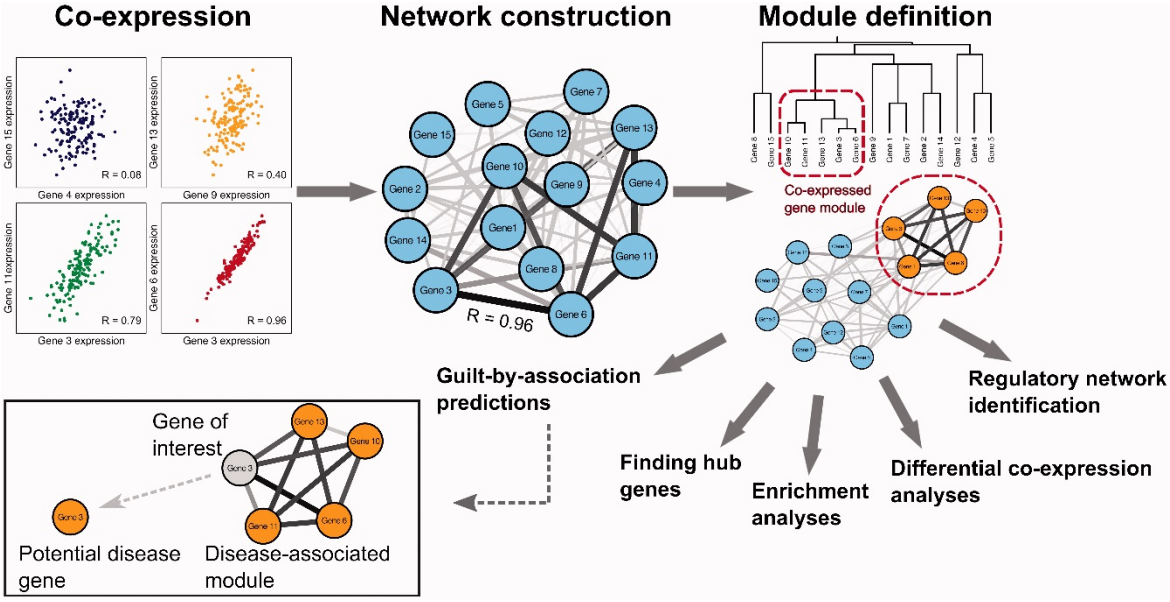


Figure 1.7 Schematic representation of the co-expression network Analysis workflow [reproduced from (van Dam et al., 2018)].

Several studies make the use of microarray or RNA-seq to evaluate gene expression data for samples exposed to a particular condition or a particular growth stage depending on the objective of study. Using the expression data, genes that are differentially expressed are evaluated. Co-expression networks are constructed from these differentially expressed genes (DEGs) and clustering is done to identify coordinated biological processes. One such study involving transcriptome analysis in salt tolerance has been conducted by Zhu in the paper titled "WGCNA Analysis of Salt-Responsive Core Transcriptome Identifies Novel Hub Genes in Rice" (Zhu et al., 2019). It has provided a meta-

analysis of 3 transcriptome datasets (SRP114666, SRP076274, and SRP083700) related to salinity and control conditions to explore the molecular mechanisms as well as key genes involved in salt stress response in rice. Using the aggregated dataset, it has identified a total of 28,432 expressed genes out of which 457 core differentially expressed genes were shortlisted as genes which responded to salt stress regardless of tissue, genotype or stress duration. These genes are analyzed based on functional enrichment from other sources like Gene Ontology (GO) and Kyoto Encyclopedia of Genes (KEGG), as well as using co-expression networks to identify regulatory mechanisms and hub genes involved in salt stress response.

The study uses RNA sequence data that is aggregated for control and salt-stress exposed rice cultivars from the National Center for Biotechnology Information (NCBI) Sequence Read Archive. RNA-seq transcriptome data provides the advantage of higher resolution and accuracy for low-abundant transcripts. The raw reads are preprocessed to remove low-quality and noisy data, and then STAR (Anders et al., 2013) is used to map these reads to the rice genome to obtain read counts. Comparing the read counts with the control sample, the differentially expressed genes (DEGs) are screened out using the edgeR software enforcing the condition of fold change ≥ 2, q value ≤ 0.05 with a significant false discovery rate-adjusted p value (FDR) < 0.05. A total of 15,596 unique DEGs are identified across the 3 datasets out of which 457 genes common among them are used for further analysis. GO enrichment analysis on these set of genes reveals the most significantly enriched biological process which is “peroxidase activity” followed by “response to stress”. Involvement in peroxidase activity suggests the importance of regulating antioxidant activity in salt stress response. KEGG pathway enrichment identified Phenylpropanoid biosynthesis and glutathione metabolism as the most significant pathways suggesting key roles of phenylpropanoid and glutathione in salt stress response in rice.

To ascertain key genes and processes in salt stress response, gene co-expression analysis is then done. Three differentially co-expressed modules are identified after conducting WGCNA on the set of 457 core DEGs. The modules are assigned color names, Blue, Grey and Turquoise. The Blue module contained 196 DEGs, the Grey module contained 32 DEGs and the Turquoise module contained 229 DEGs. All the three modules are found to be positively correlated with salt stress suggesting that the up regulation of the genes play a role in providing salt tolerance. Observing the gene expression values of the modules for the three datasets, it is found that the blue module genes are more responsive to higher salt concentration treatment (sea water ~600mM NaCl). The turquoise module genes are induced by 200-300 mM NaCl treatment than seawater treatment. The grey module is weakly induced in both the conditions when compared to blue and turquoise module. GO and KEGG enrichment analysis are performed for each module separately to understand the individual biological processes associated with them. The blue module is enriched in GO biological processes “response to oxidative stress” and “peroxidase activity”, suggesting genes involved in oxidative stress tolerance via reactive oxygen species scavenging. KEGG enrichment in the blue module indicated the involvement of “phenylpropanoid biosynthesis” and “plant hormone signal transduction” suggesting that salt tolerance is regulated by regulating phenylpropanoid related metabolites and plant hormones. Turquoise module is enriched in GO biological process “ADP

binding" suggesting the role of energy metabolism in salinity stress tolerance. KEGG analysis further revealed "phenylpropanoid biosynthesis" and "glutathione metabolism" as the most enriched pathways indicating the module regulating phenylpropanoid related metabolites as well as antioxidant glutathione for salinity stress response. GO analysis of the grey module identified "iron ion binding" as the most significant biological process, while KEGG analysis displayed various metabolic pathways that were uniformly enriched. Iron being a crucial micronutrient involved in chlorophyll biosynthesis and energy transfer, it is speculated that the genes in this module favor an optimum supply of Fe during salt stress.

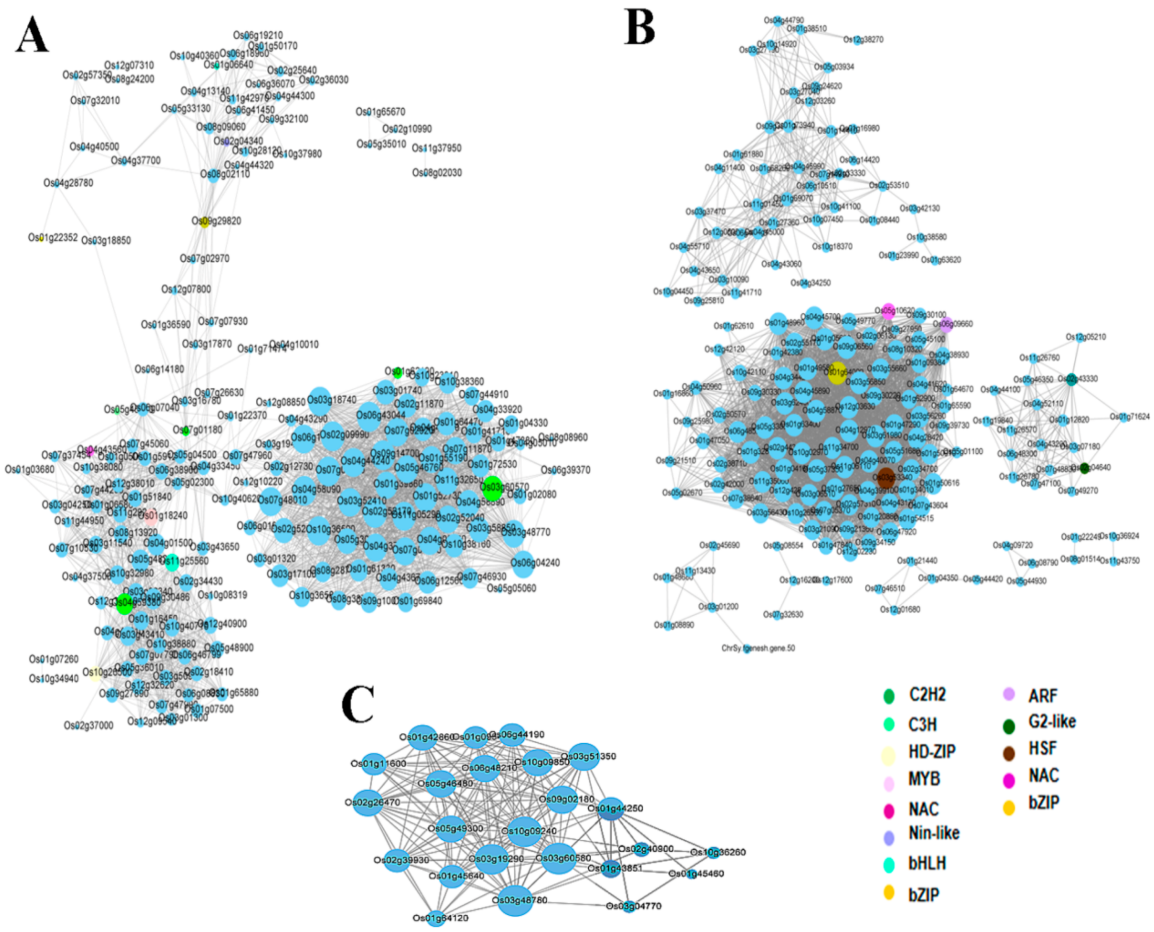


Figure 1.1.8 Representation of Co-expression networks for the three WGCNA modules A) Blue, B) Turquoise and C) Grey identified in salt stress condition in rice [reproduced from (Zhu et al., 2019)].

The network of co-expressed modules is plotted to identify hub genes which are a smaller subset of genes that have interactions with many other genes in the module. It is estimated that hub genes are more likely to be essential and representative of the module processes than genes having lesser co-expressed gene partners. The network is plotted with nodes denoting genes and edge denoting the presence of co-relation between the expression of gene nodes that it connects. The nodes are colored in sky blue except for transcription factors to highlight them. The size of the node circle is made

proportional to its number of interacting partners. The networks for the three modules have been depicted in **Figure 1.9**. For the blue module, we observe the module getting divided to three clusters having high interconnected genes. 15 hub genes are identified in all the three modules having genes encoding different proteins which include DNA binding protein, carboxyesterase, late embryosis abundant LEA4-5, calmodulin binding protein and salt responsive proteins. Two unknown genes (LOC_Os05g27340, LOC_Os01g72009) as well as two proteins with unknown domains (DUF630/632 and DUF581) are also observed. Furthermore, most of the known genes have not been reported for being involved in salt stress response in rice and are ideal candidates for such investigation.

Transcription factors (TFs) play an important role in stress response by regulating other stress responsive genes and hence their presence in the modules is investigated. The blue module contains TFs spanning various gene families including C2H2-type zinc finger (LOC_Os03g60570, LOC_Os01g62190, LOC_Os04g59380, and LOC_Os07g01180), basic helix-loop-helix bHLH (LOC_Os11g25560), basic leucine-zipper bZIP (LOC_Os09g29820), myeloblastosis MYB (LOC_Os01g18240), NAC (acronym from 3 families NAM, ATAF1/2 and CUC2) (LOC_Os04g43560), and plant regulator RWP-RK (LOC_Os02g04340) protein family (**Figure 1.9A**). The turquoise module contains TFs from families heat shock factor HSF (LOC_Os03g53340), auxin response factor ARF (LOC_Os06g09660), bZIP (LOC_Os01g64000), NAC (LOC_Os05g10620), homeobox (LOC_Os02g43330) and MYB (LOC_Os02g04640). The grey module does not contain any transcription factors. Several of these TF families have already been reported to play key role in salt stress response. OsMYB6 from the MYB transcription factor family, OsbZIP71 from the bZIP transcription factor family and ONAC022 from the NAC transcription factor family have been reported to confer salt and drought tolerance in rice (Hong et al., 2016; Liu et al., 2014; Tang et al., 2019). Rice plants overexpressing SNAC1 gene have also shown improved salt tolerance. The novel unreported transcription factors in the core set of DEGs could hence be considered for functional characterization using reverse genetic experiments, to gain insights into salinity stress response. Altogether, this study has laid a strong foundation into understanding processes involved in salt stress response and in exploring the functions of unknown proteins such as CHL27, PP2-13, DUF630/632, and DUF581 in rice.

1.4 Resources

Several web-based resources have been developed for rice for transcriptome analysis. With the high abundance in genomic data, more processing and reformatting of the data is required for further application or analysis. These web-based resources help in aggregating, organizing and presenting the high throughput data in a systematic manner which is relevant for research. The construction and utility of a few notable resources is briefly described below.

RiceNet v2 (Lee et al., 2015) is a genome-scale gene network prioritization web-server for rice (http://www.inetbio.org/ricenet/). The resource uses two complementary network prioritization

algorithms, network direct neighborhood and context-associated hubs (**Figure 1.10** shows visualization of context-associated hubs). In network direct neighborhood candidate genes are ranked by the sum of weight scores for all edges to the direct neighbors of guide genes which are already known to be involved in query phenotype. In context-associated hubs, each predefined subnetwork consists of a central hub and its neighbor genes. If the neighbor genes overlap significantly with the DEGs from a query phenotype context then the subnetwork hub genes is assigned as a context-associated hub. The gene network of rice is built over a published midsize network of 100 rice stress responsive proteins constructed through protein interaction mapping (Seo et al., 2011). It uses 41,203 non transposable element (TE)-related protein coding genes annotated from TIGR Rice Genome Annotation Release 5. Due to limitation of genome-scale datasets available for rice, it has made the use of gene orthology relationships to transfer datasets from other organisms. Thus, it contains evolutionarily conserved gene-gene linkages for rice by using datasets from other organisms like *Saccharomyces cerevisiae*, *C. elegans*, *Homo sapiens*, and *A. Thaliana*. RiceNet has been used to predict associations for diverse biological processes in rice. For example, the authors predicted and experimentally validated three previously unknown regulators of resistance mediated by a rice pattern recognition receptor XA21 which is responsible for innate immune response (Ronald & Beutler, 2010). Significant predictive power for identifying genes for other crops like maize was also observed.

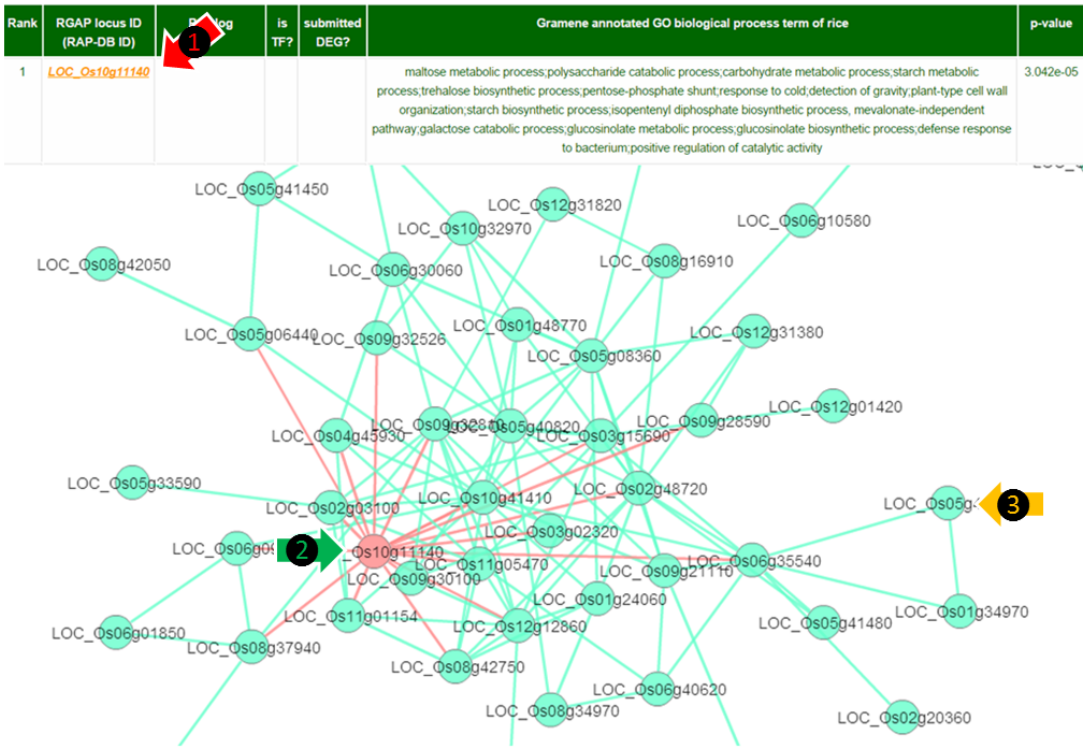


Figure 1.9 Candidate gene (hub) visualization in RiceNet v2. Network of query genes are shown in cyan (1) Clicking locus ID provides network visualization of connections between candidate hub genes and their DEGs partners, (2) Hub genes are denoted in red (3) Query genes are denoted in cyan [reproduced from (Lee et al., 2015)].

Oryza Express, formerly known as Rice Gene Expression Network (RGEN) (Hamada et al., 2011) is a web resource used for evaluating the statistical similarities of genome-wide gene expression patterns from aggregated large-scale gene expression data from various experimental conditions (http://plantomics.mind.meiji.ac.jp/OryzaExpress/). It is one of the first rice databases to provide information on both Gene Expression Networks and omics annotation. The microarray data used for the resource construction is obtained with Rice Gene Expression Microarray 4x44K (Agilent) and GeneChip(R) Rice Genome Array (Affymetrix) and are assembled from 1,893 samples in the NCBI GEO database. The similarities of gene expression patterns between genes are calculated using correspondence analysis (CA) which uses Pearson Correlation Coefficient (PCCs), Mutual Ranks (MRs) and partial correlation coefficient (PAC). The resource provides the option of conducting functional enrichment of the queried set of genes with Gene Ontology and KEGG Pathways. It also has support for visualizing the expression profile and co-expression network for the queried probes (**Figure 1.11**). The expression networks consist of edges comprising both similar and reciprocal expression patterns. Reciprocal expression pattern are inverse expression profiles that are effective in identifying repressor or downstream genes. While the database provides comprehensive gene expression networks integrated with various omics information from public database, it is limited in its specificity and dynamic range when compared to recent databases consisting of RNA-seq data.

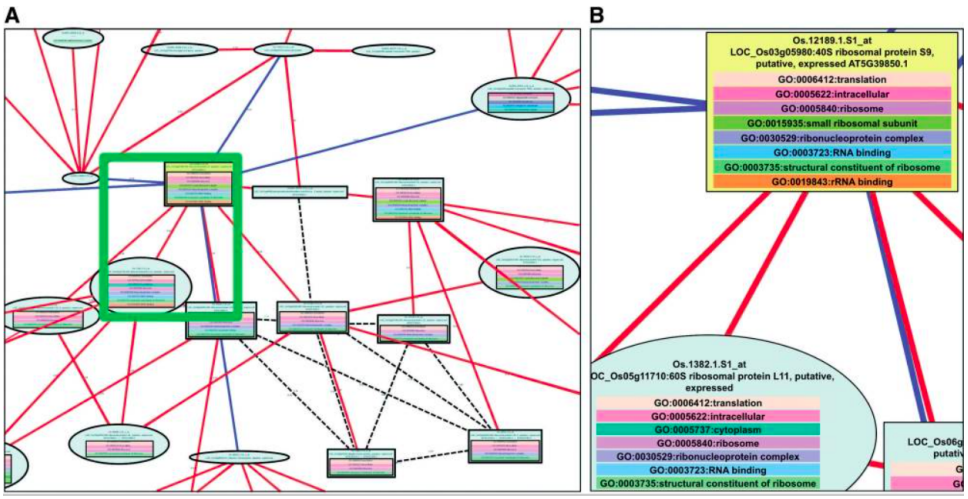


Figure 1.10 (A) Gene Expression Network image where red and blue edges indicate similar and reciprocal expression patterns, respectively. (B) Zoom-in of box in A showing the GO terms and metabolic pathways [reproduced from (Hamada et al., 2011)].

PlantArrayNet, formerly known as RiceArrayNet (RAN) (Lee et al., 2009) is the earliest database for rice which provides information on gene co-expression based on correlation coefficients (r values) (http://bioinfo.mju.ac.kr/arraynet/). It has the option for visualizing the degree of closeness between genes in a relzational tree and a relational network format (**Figure 1.12**). Figure 1.11 A represents the coexpression relationships of the ribosomal protein Os10g0124000, in the form of a clusterdiagram or network where genes with similar expression profiles form a cluster. A set of 36 genes are retrieved under the condition of $r \geq 0.6$ and depth = 1. Figure 1.11 B depicts the correlation

network under the same condition where the central protein Os10g0124000 is represented by an asterisk. The genes are denoted as circles and are placed such that the distance between them represents the closeness of their relationships. Expression data from 183 microarrays with 50 different treatments have been aggregated for this resource. Information suggesting biochemical pathways and cis-regulatory elements of clustered genes have been additionally provided through KEGG and PLACE databases, respectively. The RAN database has been shown to be a useful tool which offers insights into a particular gene by examining its co-expressed neighbors. As a case study, the authors have shown a correlation pattern being captured between the 16-member 17Ae ribosomal protein family within rice. A drought responsive element-binding transcription factor (Os01g0968800), a trehalose-6-phosphate synthase (Os02g0790500) and a small heat shock factor (Os06g0219500) were found to be correlated indicating their presence in regulating the same biological process.

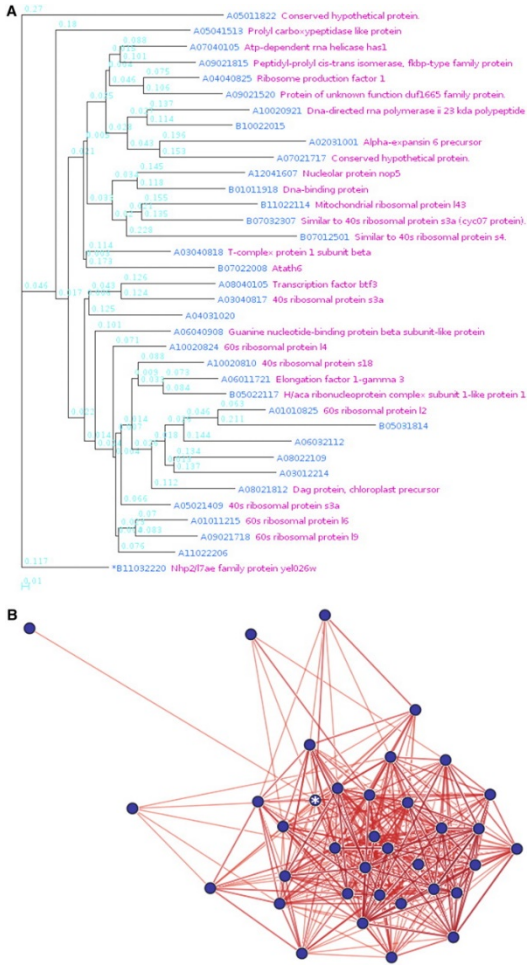


Figure 1.11 Graphical presentation of co-expressed genes in Oryza Express. (A) Relational tree of gene expression with a ribosomal protein, B11032220. (B) Relational network using B11032220 where the network consists of 36 genes under the parameters of r ≥ 0.6 and depth = 1 [reproduced from (Lee et al., 2009)].

Rice Interaction Viewer (RIV) (Ho et al., 2012) is a web-based interactome network visualization tool hosting predicted protein-protein interaction network in rice that can be queried for subnetworks (http://bar.utoronto.ca/rice_interactions/cgi-bin/rice_interactions_viewer.cgi). The network is predicted based on the universality of conserved protein function in different organisms, under the assumption that evolutionarily conserved orthologous proteins would retain their interactions with other conserved proteins. A total of 37112 interactions among 4567 rice protein were predicted using an ortholog prediction algorithm with ortholog gene matching with at least of the 11 reference organisms used for prediction. The tool also supports the identification of sub-cellular localization information in the network (**Figure 1.13**). Figure 1.12 is the visualization of a predicted interactome. The color of each protein is based on the predicted subcellular localization. The edges are coloured based on their correlation values and its thickness depends on the confidence value (CV).

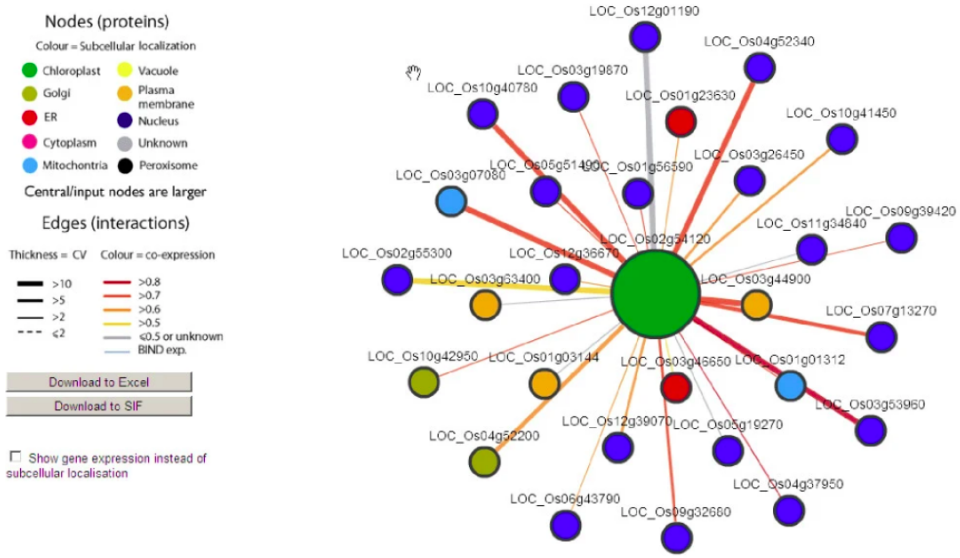


Figure 1.12 Visualization of rice interactome from Rice Interaction Viewer. Nodes are colored according to its subcellular localization, edges are colored according to its co-expression correlation. Edge thickness correlates to Confidence Value (CV) [reproduced from (Ho et al., 2012])

Rice Environment Co-expression Network (RECoN) (Krishnan et al., 2017) is a web resource for exploratory analysis of abiotic stress response in rice (https://plantstress-pereira.uark.edu/RECoN/). It consists of co-expression networks of 1744 abiotic stress-specific gene modules covering 28,421 rice genes obtained from 129 samples, which are involved in response to some environmental stress. This data is aggregated from 29 publicly available gene expression datasets of the Affymetrix rice GeneChip from NCBI GEO and ArrayExpress. These gene modules consist of genes that are tightly coexpressed across many environmental stresses and are likely to be functionally coherent. RECoN offers the functionality to identify modules that are substantially perturbed in the user provided differential expression profile, thereby suggesting functional and regulatory mechanisms based on enriched functions of the predefined modules. Hence it is useful in exploring new data from all abiotic

stresses and uncovering gene candidates involved in stress tolerance. One of the drought-tolerance clusters identified in this resource is Cluster0079 containing 71 genes which is depicted in **Figure 1.14A.** It includes a receptor-like cytoplasmic kinase OsRLCK253, a phosphatase OsPP108, a dehydrin OsLea3-1, an enzyme OsUGE-1 and TFs like OsDREB2A, OsNAC6, CMYB1 and ZFP182 which are known drought tolerant genes. Most genes in the cluster are up-regulated by drought in all its developmental stages and are suggestive of these genes to have a dual role as developmentally regulated and stress responsive. **Figure 1.14B** represents the Cluster0424 containing 20 genes enriched with reproductive drought and they contain an ABRE-like motif.

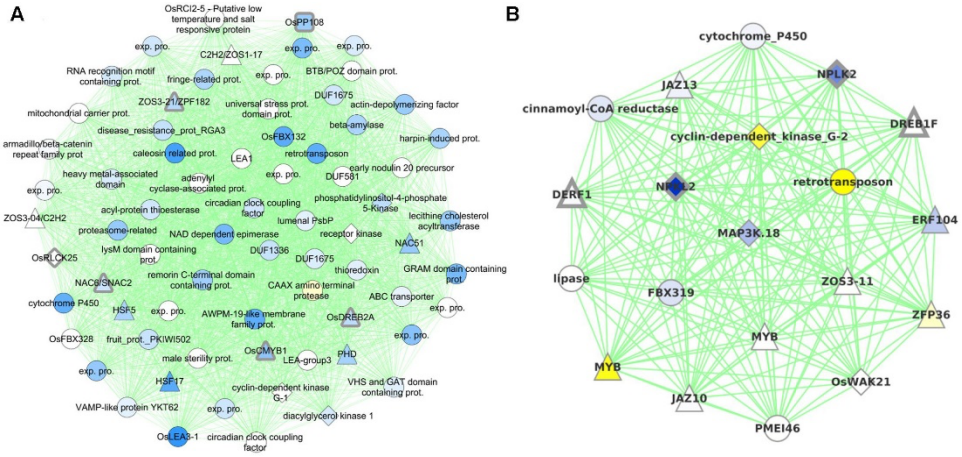


Figure 1.13 Graphical visualization of (A) 71 genes in Cluster0079 that contains six drought tolerance genes (with thick gray borders) and (B) 20 genes in Cluster0424 that contains four drought tolerance genes in RECoN [reproduced from (Krishnan et al., 2017)].

1.5 Organization of thesis

The thesis focuses on the analysis and understanding of abiotic stress tolerance in *Oryza sativa*. Chapter 2 describes the material and methods that have been used in the analysis. Genes which are differentially regulated under different abiotic stress conditions (drought, flood, cold and osmotic) and exogenous hormone (ABA and JA) treatment conditions are evaluated. These genes are possible players in regulating stress response. Since it is a large data, co-expression networks are constructed which allows for visual and analytical approaches into associating genes based on their expression values. To further reduce the dimensionality of the co-expression network, clustering of the genes is done by grouping genes having similar expression profile into the same module. Chapter 3 describes the development of NetREx, a web-based resource, for querying these condition-dependent co-expression networks, which provides different viewers to visualize the co-expression relation between the queried genes, as well as compare their expression values in a time-point based manner. A case study showcasing its usability in predicting novel genes related to the abscisic acid pathway involved in drought stress response is presented. Chapter 4 describes the differentially expressed genes and the metabolic processes they are involved in is compared across the different abiotic stresses. The co-expressed modules are further used to understand the response pathways involved

in osmotic stress response at the different time points. The thesis is summarized and concluded In chapter 5, the conclusion of the thesis and future directions is given.

Chapter 2

Materials and Methods

In this chapter we discuss the methodology used in the analysis of RNA-seq data of rice under various abiotic stress conditions (drought, flood, cold and osmotic) and exogenous hormone (ABA & JA) treatment. The datasets used in this study is given in section 2.1. In section 2.2 the preprocessing of the datasets is briefly described. Identifying the differentially expressed genes (DEGs) based on raw read counts in different stress conditions is described in section 2.3. To identify correlated genes, the co-expression network is constructed based on highest reciprocal rank (HRR) algorithm given in section 2.4. Section 2.5 describes the WGCNA method to identify co-expressed gene modules from the HRR network. The resources used in this work are briefly described in section 2.6.

2.1 Datasets

The raw mRNA-seq read data from the root and shoot tissues of *Oryza sativa L.* (cv. Nipponbare) seedlings under four abiotic conditions (drought, cold, osmotic, and flood) and two phytohormone treatment conditions [abscisic acid (ABA) and jasmonic acid (JA)] are obtained from TENOR dataset (Transcriptome ENcyclopedia Of Rice, http://tenor.dna.affrc.go.jp) (Kawahara et al., 2016) downloaded from DDBJ Sequence Read Archive (DRA000959). TENOR database is a large repertoire of genome-wide time-course transcriptomic studies of rice shoot and root tissues while carefully using standardized laboratory conditions across all experiments (same platform, same rice genotype, etc.) which allow us to compare/integrate data across the conditions. The data was collected for several time points ranging from 3hr up to about 1-3days. The reads of length 76bp were generated by single-read sequencing using the Illumina GAIIx platform under uniform library conditions. The details of the total number of samples, the number of biological replicates for each time point and for the six stress/treatment conditions are given in **Table 2.1**.

Table 2.1 Number of samples and biological replicates for the mRNA-Seq data from DDBJ-SRA (DRA000959) considered for multi-stress analysis are summarized.

Conditions	Treatment	Time-points (No. of biological replicates)		No. of Samples
		Root	Shoot	
Cold	4°C	0h (3), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	0h (3), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	Root:13 Shoot:13
Drought	Grown without medium	0h (2), 1h (2), 3h (2), 6h (2) 12h (2), 1d (2)	0h (2), 1h (2), 3h (3), 6h (2), 12h (2), 1d (2)	Root:12 Shoot:13
Osmotic	0.6 M Mannitol	0h (3), 1h (2), 3h (2), 6h (2), 12h (2)	0h (3), 1h (2), 3h (2), 6h (2), 12h (2)	Root:11 Shoot:11
Flood	Completely submerged in medium	0h (3), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2), 3d (2)	0h (3), 1h (2), 3h (2), 6h (2), 12h (2), 1d (3), 3d (2)	Root:15 Shoot:16
ABA	100 µM	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	Root:12 Shoot:12
JA	100 µM	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2)	Root:12 Shoot:12
No Treatment (NT)	-	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2), 3d (2), 4d (2), 5d (2), 10d (2)	0h (2), 1h (2), 3h (2), 6h (2), 12h (2), 1d (2), 3d (2), 4d (2), 5d (2), 10d (2)	Root:20 Shoot:20
Total				**Root:95** **Shoot:97**

2.2 Pre-processing

High-throughput data are prone to having noisy data, missing values and other inconsistencies. Preprocessing them resolves such issues and makes it cleaner and efficient to perform data analysis. Raw reads are filtered to remove low quality and uninformative reads. Using Cutadapt (version 1.15),

the adapter sequences and the low-quality bases (having Phred-score Q <15) present at the 5' and 3' ends of the reads are trimmed (Martin, 2011). Reads of length less than 20bp after trimming may not be reliable and are discarded. Using HISAT2 (version 2.1.0), a graph-based alignment program, the reads for each sample are aligned to the rice reference genome (Os-Nipponbare-Reference-IRGSP-1.0) (Kim et al., 2015). Gene annotations are taken from RAP-DB database (version 2017-04-14) (Sakai et al., 2013). The alignment details like the total number of base pair before and after filtering and the percentage of reads aligned to the genome are given in appendix (**Appendix Tables 1 to 7)**. The percentage of mapped reads gives insight towards the sequence accuracy and the presence of contamination in the samples (Conesa et al., 2016). As observed in **Table 2.2**, the average percentage of reads mapped ranged between ~92% (flood stress) to ~98.2% (ABA and JA treatment) in case of shoot tissue, and it ranged between ~75.1% (JA treatment) to ~91.9% (ABA treatment) in case of root tissue. To compute gene expression from raw read counts, 'featureCounts' tool from SubRead package was used (1.6.0) (Liao et al., 2014) and the gene annotations were obtained from RAP-DB database (version: 2017-04-14) (Sakai et al., 2013). The average percentage of reads aligned and assigned to the genes for each sample is given in **Table 2.2**.

Table 2.2 Average percentage of reads mapped to genes for different stress conditions after pre-processing.

Conditions	Root		Shoot	
	Avg. read align rate using HISAT2 (%)	**Avg. percentage of reads assigned to genes using featureCounts**	**Avg. read align rate using HISAT2 (%)**	**Avg. percentage of reads assigned to genes using featureCounts**
Cold	83.3	56.9	97.5	58.9
Drought	79.0	53.4	98.0	62.0
Osmotic	91.5	58.9	92.9	64.5
Flood	86.8	56.9	92.0	58.4
ABA	91.8	64.8	98.2	66.3
JA	75.1	48.9	98.2	62.8

2.3 Estimating Read Counts and Differential Gene Expression Analysis

Read counts are a measure of gene expression as it is the sum of reads associated with each of the exons that belong to the gene. A count matrix is a table constructed from the read counts for all samples, with genes in rows and the samples in columns. Final count matrix across various stress/hormone treatments is constructed by considering only genes that exhibited raw read counts ≥

5 in at least 50% of the total number of samples across all the conditions and time-points. Genes exhibiting 2-fold change and p-value ≤ 0.5 are obtained using the Bioconductor package DESeq2 (Love et al., 2014) and are referred to as differentially expressed genes (DEGs). Differentially expressed genes (DEGs) are identified in a condition-specific manner for every time-point by considering the expression values at 0h as control for each stress/treatment condition. Next, for each stress/treatment condition, only genes that are differentially expressed in at least 2 time-points are considered for analysis and summarized in **Table 2.3**. The condition-specific DEGs that have an overlap with the developmental time-points (NT) are filtered out as these mainly correspond to diurnal or developmental changes. In RNA-seq data, most strongly expressed genes show large variations across samples compared to those with lower expression profiles (heteroscedasticity). On the other hand, most common methods of multi-dimensional data analysis like clustering work best with homoscedastic data (variance is independent of mean). To achieve this approximate homoscedasticity, the combined count matrix with 13695 genes in root tissue and 13717 genes in shoot is normalized using variance stabilizing transformation (VST) in DeSeq2 to obtain a relatively flat trend of variance as a function of mean (Anders & Huber, 2010). This makes the data approximately homoscedastic and appropriate for network and clustering analysis. The DEGs are then considered for rank-based network construction.

Table 2.3 Number of genes that are differentially expressed in at least two time-points for various stress/treatment conditions in root tissue. Here, NT refers to DEGs obtained in the "no-treatment" condition.

Tissue	DEGs in at least 2 time-points							Total DEGs used for Network Construction
	Drought	Cold	Osmotic	Flood	ABA	JA	NT	
Root	3817	6327	8040	3651	9340	8835	1127	13695
Shoot	10576	6267	4540	6955	8898	6531	1116	13717

2.4 Network Construction

A network-based approach is used to capture the associations between genes that are up- or down-regulated under various stress conditions in *Oryza sativa*. There are two main approaches available when constructing a gene expression database. These approaches have been termed "targeted" (also known as "guided-gene" and "directional") and "non-targeted" (also known as "global" and "non-directional") approach in literature (Aoki et al., 2007). For the "targeted" approach, a network is constructed with genes that are correlated with a prior known set of genes (also known as bait genes) based on literature knowledge. The probable functions of these query genes can then be predicted based on the biological processes and pathways the bait genes are involved in. This approach has been used to identify candidate genes associated with the biosynthesis of cell wall (Rao et al., 2019; Sibout et al., 2017). For example, to screen transcription factors involved in lignin biosynthesis in

switchgrass, 14 known genes involved in this condition were considered as bait genes. Then genes exhibiting correlated expression patterns with these bait genes were searched across public microarray datasets to construct a "targeted" network related to the lignin biosynthesis.

The "non-targeted" approach does not use any prior known set of genes. The co-expression network is constructed based on pairwise co-expression values for the complete set of genes. Co-expressed modules can be identified from the network based on its topological structure. Modules having high cohesion and connectivity may be associated with a coordinated function and regulate the same biological process. This approach provides a global scenario of the co-expressed genes, however, it is computationally expensive. A few examples of this approach involve screening of genes involved in biotic and abiotic stress (Amrine et al., 2015; Ransbotyn et al., 2015), characterization of genes involved in seed maturation in Arabidopsis (Silva et al., 2016) and seed longevity in Medicago (Righetti et al., 2015).

In this work, a "non-targeted" or a global approach is used for constructing gene co-expression network by computing pairwise correlation values for the DEGs across all samples for each tissue. Calculating the similarity score for the correlation of gene pairs is central to co-expression network construction. There are several methods which have been employed and their comparison has been well described in numerous studies (Kryuchkova-Mostacci & Robinson-Rechavi, 2017; Saelens et al., 2018). The statistical method can be chosen based on the biological concern that is intended to be analyzed and the sample size, number of genes and expression value distribution in the dataset. The most popular statistical method is Pearson's correlation coefficient (PCC) which captures linear relationship between the expression level of two genes. A more robust method which can handle outliers is Spearman's correlation coefficient (SCC). It is a non-parametric measure of rank correlation (Usadel et al., 2009). While PCC is sensitive to outliers it is more powerful than SCC. Other statistical measures such as mutual information content and partial correlation have been used for quantifying correlation between genes (de la Fuente et al., 2004; Lim et al., 2007).

There can be false predictions due to different parameters and outliers in the data or the method used for analyzing the data. To minimize error in predictions, several experiments have been conducted to assess the construction and analysis of co-expression networks. Liesecke carried out comparative analysis of co-expression networks constructed using several distance measurements methods on Arabidopsis thaliana microarray and RNA-seq data (Liesecke et al., 2018). Pearson correlation coefficient with highest reciprocal rank (HRR) approach was shown to be reliable, resulting in highest correlation between the global co-expression network and enrichment of gene ontology annotation. Using HRR allows the use of ranked correlation coefficients instead of raw values. Ranking correlation coefficient implies that for every gene, the correlation calculated with the remaining genes are ranked from 1 to N (where N is the total number of genes). For defining a co-expressed edge instead of using a hard threshold as in the case of PCC, HRR allows the use of rank cut-off which has been shown to be more relevant for biological networks and less prone to loss of information.

For the construction of co-expression network, Pearson correlation coefficients (PCCs) are computed between every pair of differentially expressed genes across all the stress and treatment conditions, for root and shoot tissues separately. Positively correlated genes with p-value ≤ 0.05 are considered for the construction of Highest Reciprocal Rank (HRR)-based co-expression network proposed by *Mutwil et al* (Mutwil et al., 2010), both for root and shoot tissues. The HRR score between genes *A* and *B* is given by:

$$\mathrm{HRR}(A, B) = \max(r(A, B), r(B, A))$$

where $r(A, B)$ is correlation rank of gene *B* in gene *A*'s co-expression list and $r(B, A)$ is correlation rank of gene *A* in gene *B*'s co-expression list. For this study, the root and shoot networks are constructed with HRR values ≤ 100 (i.e., only top 100 neighbours for each gene are considered) and is termed as the global 'HRR-100' network. Every pair of connected genes in an HRR-100 network would imply that genes are within the top 100 correlated neighbours of each other. Corresponding stress-specific sub-networks are derived from the global HRR-100 network for each of the tissues using respective stress-specific DEGs in **Table 2.3**.

Number of DEGs across various stress conditions in the Pearson correlation network and HRR-100 network are given in **Tables 2.4 and 2.5** respectively. It may be noted that the major advantage of considering the HRR-100 network over the PCC network is a significant reduction in the number of edges.

Table 2.4 Number of Nodes (DEGs) and Edges across various conditions in Pearson Correlation Coefficient (PCC) network.

Tissue \ Stress		PCC network (with PCC > 0 and p-value < 0.05)						
		Total Network	Drought	Cold	Osmotic	Flood	ABA	JA
Root	Nodes	13695	2970	5005	6311	2347	7460	7565
	Edges	2,62,69,920	16,86,453	36,37,333	62,86,236	8,95,590	83,63,871	96,29,038
Shoot	Nodes	13717	8809	3688	5280	5791	7546	5523
	Edges	2,91,80,730	1,44,25,817	24,30,142	49,56,792	57,93,144	95,82,721	53,20,944

Table 2.5 Number of Nodes (DEGs) and Edges across various conditions in HRR-100 network.

Tissue \ Stress		HRR-100						
		Total Network	Drought	Cold	Osmotic	Flood	ABA	JA
Root	Nodes	13695	2970	5005	6311	2347	7460	7565

	Edges	21,38,990	1,30,207	3,52,715	3,53,129	63,630	4,63,175	4,66,528
Shoot	**Nodes**	13717	8809	3688	5280	5791	7546	5523
	Edges	22,72,758	6,57,747	1,78,198	3,03,953	3,35,654	5,00,651	3,21,672

2.5 Weighted Gene Co-expression Network Analysis (WGCNA)

Clustering helps in reducing the dimension of the networks by grouping the large set of differentially expressed genes into a few manageable co-expressed modules that possibly regulates the same biological function. There are several clustering algorithms available for this purpose. These algorithms can be broadly classified into hierarchical and non-hierarchical clustering algorithms. Hierarchical clustering uses an iterative approach to cluster genes by assigning them to a cluster in each step. Initial weights are assigned to network vertices using the calculated correlation coefficient. Clusters are then constructed using high weight vertices and progressively expanding them by considering their neighbors. Hence the number of final clusters may vary and depends on the chosen threshold. There are several hierarchical clustering methods available which includes Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder & Horvath, 2008), Markov Cluster Algorithm (MCL) (Enright et al., 2002), Improved Principal Component Analysis (IPCA) (Li et al., 2008) and Normalization Engine for Matching Organization (NeMo) (Rivera et al., 2010). Non-hierarchical approaches include K-mean clustering (Stuart et al., 2003) which assigns entities to pre-defined k clusters. The criteria for selecting a particular clustering algorithm should depend on the functional coherence of the predicted modules (Lysenko et al., 2011).

WGCNA is a systems biology method for describing the correlation patterns among genes. It is used for finding clusters or modules of highly correlated genes. These modules each have a central hub gene known as the module eigengene. The function of the hub gene reflects the properties of the module, as well as their relationship with other modules. The closeness of any gene with the eigengene of its module defines its module membership. WGCNA is used in our analysis to further reduce the dimension of the HRR-100 network by associating highly co-expressed set of genes into modules.

The analysis steps used in WGCNA involves network construction and identification of modules. For network construction, an adjacency matrix a_{ij} needs to be defined. An intermediate quantity called co-expression similarity s_{ij} is used to define $a_{ij.}$ s_{ij} is measured as the absolute value of correlation coefficient between the expression profiles of gene i and gene j ($s_{ij} = |cor(x_i, x_j)|$). Weighted networks allow the adjacency to take continuous values between 0 and 1. It is defined by raising the co-expression similarity to a power β:

$$a_{ij} = s^{\beta}_{ij}$$

with $\beta \geq 1$. By applying an approximate scale-free topology criterion the threshold parameters for the network is defined. Gene modules are identified from this network based on groups of nodes

with high topological overlap. Topological overlap of two nodes defines their relative connectedness. A similarity measure based on topological overlap is defined by the topological overlap matrix (TOM) $\Omega = [\omega_{ij}]$ which is defined as:

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}}$$

where $l_{ij} = \Sigma_u a_{iu} a_{uj}$, and $k_i = \Sigma_u a_{iu}$ is the node connectivity. In the case of hard thresholding, l_{ij} is same as the number of nodes to which both gene i and j are connected. This matrix is used in hierarchical based clustering to define gene modules. WGCNA uses soft thresholding approach by defining β, which avoids information loss which otherwise PCC with hard thresholding is prone to.

The patterns of co-expression often point at coordinated biological processes the genes may be involved in and representing genes as co-expressed clusters help in reducing the dimensionality of the data. For this purpose, two 'signed' gene co-expression networks are constructed with 13695 DEGs from root tissue and 13717 DEGs from the shoot tissue. The advantage of a 'signed' network over an 'unsigned' network is that it is able to capture the association of genes that are activated or repressed. Unsigned networks use absolute value of correlation $s^{ij}_{unsigned} = |\text{cor}(x_i , y_j)|$ which fail to distinguish between gene activation ($s^{ij}_{unsigned} = 1$) and gene repression ($s^{ij}_{unsigned} = 1$) leading to loss of biological information (Mason et al., 2009). The similarity measure of a 'signed' network is defined as follows:

$$s_{signed} = \frac{\left(1 + cor\left(x_i, y_j\right)\right)}{2}$$

where x_i and y_j represent the expression values of genes i and j, respectively. $s^{ij}_{signed} = 1$ denotes a positive correlation, while $s^{ij}_{signed} = 0$ denotes a negative correlation, and $s^{ij}_{signed} = 0.5$ denotes no correlation. Clustering based on this similarity measure is able to distinguish between positively and negatively correlated genes.

Using the *block-wiseModules* function in WGCNA R package (Langfelder & Horvath, 2008) hierarchical clustering of genes is carried out by Dynamic Tree Cut approach (Langfelder et al., 2008) with maximum block size = 14000, minimum module size = 50, "cut height" = 0.995 and "deep split" = 2. The parameters used in the construction of weighted gene co-expression network are given in **Table 2.6**. Here, for a weighted network, "**β"** is the soft thresholding power to which co-expression similarity is raised to calculate adjacency, thereby emphasizing high correlations at the expense of low correlations. And "**k**" refers to the connectivity or sum of the connection strengths with other genes in the network. The weighted topological overlap matrix in WGCNA allows us to compute various degree centrality measures which can be useful in screening important genes. Specifically, we computed the k_{IM}, the within-module degree or intramodular connectivity of a gene. Intramodular connectivity is a measure of how co-expressed a gene is with the genes of its module, and is a measure of module membership.

Table 2.6 Parameters used for the construction of signed, weighted gene co-expression network using WGCNA R package is summarized.

Tissue	No. of Samples	No. of Genes	β Cut-off	R^2 Scale-free fit	Mean k	Median k	Max k	No. of Modules
Root	69	13695	26	0.91	15.4	8.7	151	22
Shoot	72	13717	26	0.81	36.4	18.2	262	18

The co-expressed modules (22 for root and 18 for shoot) are tested for their biological relevance. For this, genes of the individual clusters are submitted for over-representation analysis in GO consortium using PANTHER classification system (v14.0) (Mi et al., 2019) for *Oryza sativa*. Results with Fischer's Exact Test and Bonferroni correction for multiple testing (p-value ≤ 0.05) are retrieved for each module. Significant GO terms indicate that genes of a given cluster are more often associated to certain biological functions than what would be expected in a random set of genes. Furthermore, we queried the top 100 highly connected genes (based on intra-modular connectivity) in an independent database, STRING DB (Szklarczyk et al., 2019) to check for protein-protein interactions (PPI). Since co-expression clusters often point to coordinated biological processes where there maybe physical interactions between proteins, enrichment of PPIs further validated the biological relevance of the clusters. The percentages of DEGs across modules and results from the functional databases (GO, STRING DB) are discussed in detail in Chapter 4. A web-based resource NetREx is constructed using the co-expression networks and the additional data that have been integrated for gene prioritization. The construction and usefulness of this resource is given in Chapter 3.

2.6 Resources Used

Various resources that have been used in the construction of NetREx and in the analysis of gene expression data under different stress conditions is described below.

STRING (Search Tool for the Retrieval of Interacting Genes) v11.5

STRING database integrates all the known and predicted interactions between proteins (Szklarczyk et al., 2019). These interactions include both physical interactions and functional associations that are collected and scored from various sources including (i) text mining from published literature, (ii) databases of interaction experiments and annotated pathways, (iii) predicted interactions from co-expression or conserved genomic context and (iv) transfer of interactions evidence from one organism to other based on orthologs mapping. It currently consists of data from over 14000 organisms covering more than 67 million proteins.

MapMan v4

MapMan is a framework which was initially developed with the aim to facilitate the visualization of omics data on plant pathways (Schwacke et al., 2019). Presently, it is used alongside gene ontology and KEGG to map genes to its corresponding metabolic and regulatory functions/processes. It uses a hierarchical tree structure called "bins" representing biological contexts. There are 27 top-level bins representing major biological processes and each child represent a focused subprocess or context within the parent bin. Proteins with complex and diverse functions may be mapped to multiple bins. The Mapman framework has evolved overtime to include bins representing regulatory processes, signaling pathways and biotic and abiotic stress response in addition to metabolic processes. Though it was initially developed for *Arabidopsis Thaliana*, it currently supports all land plant species.

Panther GO (Protein ANalysis THrough Evolutionary Relationships Gene Ontology) v14

Panther GO is a resource used for the functional and evolutionary classification of genes from organisms (Mi et al., 2019). It consists of a large, curated database of proteins and genes, their family information along with their functionally related subfamilies. The evolutionary classification has three levels based on their specificity: protein class, family, and subfamily. There exists over 15000 families and 80000 subfamilies. For functional classification, Gene ontology is used which is obtained from the Gene Ontology Consortium (available at http://geneontology.org). Currently, it hosts information of over 900 different genomes.

Ensembl Plants v2016

Ensembl Plants is an integrated platform for genome-scale information for various sequenced plant species (currently 33 in number) which include the genome sequences, gene models, functional annotations, and polymorphic loci (Bolser et al., 2016). It also includes information on variation and, genotype and phenotype data. Thus enabling comparative genomics and aiding in the analysis of transcriptional regulation and variation in sequences. It also supports orthology relationships across plant species that can facilitate in augmenting information and validating relationships.

funRiceGenes

funRiceGenes database is a comprehensive dataset of over 2800 functionally characterized rice genes (Yao et al., 2018) containing information of rice families, their regulatory connections, and their interaction network. The data is obtained by integrating information from available databases and reviewing publications on rice functional genomic studies.

RiceSRTFDB (Rice Stress-Responsive Transcription Factor Database)

RiceSRTFDB is a database of annotated rice transcription factors (Priya & Jain, 2013). Transcription factors are regulatory elements of a genome and are essential targets for engineering stress tolerance. The database consists of comprehensive expression information for the rice TFs during drought and salinity stress as well as through various stages of development. It also includes

curated information for cis-regulatory elements present in the promoter region of these TFs which is important for the study of binding proteins.

IC4R (Information Curation for Rice) v2.0

IC4R is a curated knowledgebase that integrates multiple omics rice data through community-driven modules (Consortium et al., 2016; Sang et al., 2020). It incorporates a variety of rice data which includes, expression profiles of genes at different developmental stages and tissues in rice obtained from RNA-seq data, genomic variations from re-sequencing data of rice varieties, plant homologs, post-translational modifications, rice-related literature, and community-contributed gene annotations. It is a useful resource for analyzing the expression profile of a gene across different tissues/conditions.

RAP-DB (Rice Annotation Project DataBase)

RAP-DB is a database providing the genome sequence assembly of the International Rice Genome Sequencing Project (IRGSP) as well as curated annotations of the sequence and other genomic information that can be useful in providing insights into rice biology (Sakai et al., 2013). It contains a variety of annotation data like the clone positions, structures and functions of 31429 genes which have been validated by RNA, cDNA genes detected by the massive parallel signature sequencing (MPSS) technology. Other validation methods include sequence similarity, transposable elements and flanking sequences of mutant lines. This tool has been proved useful for comprehensive understanding of the rice biology.

KEGG (Kyoto Encyclopaedia of Gene and Genomes)

KEGG is a collection of databases which deals with genomes, biological pathways, drugs, diseases and enzymes (Kanehisa et al., 2017). These databases are categorized into systems, genomic, chemical and health information. System information consists of pathway, modules and hierarchical information that helps in classifying genes. Its systems information is particularly useful for mapping and grouping genes into their related biological pathways.

Cytoscape

Cytoscape is an open-source software for network visualization of biological data (Franz et al., 2016). It allows the integration of various sources of data and offers visualization aids to ease the analysis. Various plugins are available that provides network and molecular profiling analyses, new layouts, connection with databases and searching in large networks. Its integration with the STRING-DB application eases the clustering and functional enrichment of subnetworks of the studied PPIs. Plugins helping in construction, visualization and clustering of gene coexpression networks are particularly useful in functional enrichment of genes and narrowing down their size based on functional filters for efficient analysis.

Gephi

Gephi is an open-source platform for network analysis and visualization (*Gephi - The Open Graph Viz Platform*, n.d.). It provides the user with various network layout options and integrates biological information into the network via visual aids like the node size, colour and shape. It provides broad access to network data and allows for spatializing, filtering, navigating, manipulating and clustering of nodes.

Chapter 3

NetREx – Network-based Rice Expression

3.1 Introduction

Studying Gene Co-expression Networks (GCNs) have been shown to be useful in inferring novel biological functions of genes and the activation and repression of biological processes under specific conditions. GCNs can be used for candidate gene prioritisation responsible for phenotypic differences, identification of regulatory genes and functional gene annotation. Owing to the dense nature of GCNs it is difficult to extract meaningful information from these networks and visualization approaches have been proposed to aid in their analysis. Rice co-expression networks constructed from the differentially expressed genes under different abiotic conditions would offer insights into regulatory response and stress resistant phenotypic differences. Querying and visualisation of these gene networks with respect to the temporal fold change, module membership and pathway information of genes would further enhance the analysis and understanding of stress responsive genes.

Various functional resources have been developed for the plant community exploring genomic relations based on co-expression. At the global scale, condition-independent co-expression networks have been constructed from large scale microarray datasets integrated from different databases with varying environmental conditions, platforms, tissues and developmental stages. For example, resources like MaizeNet, AraNet, RiceNet are examples of condition-independent co-expression networks specific to maize, *Arabidopsis thaliana* and rice respectively. For comparative genomics and evolutionary pathway analysis, resources which host pan-species co-expression network are required. Examples of such resources are ATTED-II (Obayashi et al., 2011), STRING database (Szklarczyk et al., 2019), PlaNet (Proost & Mutwil, 2017), PhytoNet (Ferrari et al., 2018), FamNet (Ruprecht et al., 2016), etc. Numerous rice specific resources have been developed such as Oryza Express and RiceFREND which are condition-independent co-expression networks constructed using various microarray data. ReCoN is an important rice specific resource that assembles 29 different gene expression databases with a unifying biological theme to construct a gene co-expression network with a focus on abiotic stresses. Although these resources allow us to derive associations from a large sample size, the merging of different experimental conditions and platforms might incorporate complexities in the network which may be difficult to account. On the other hand,

condition-dependent networks that are derived from datasets corresponding to specific conditions provide opportunities to explore context specific associations.

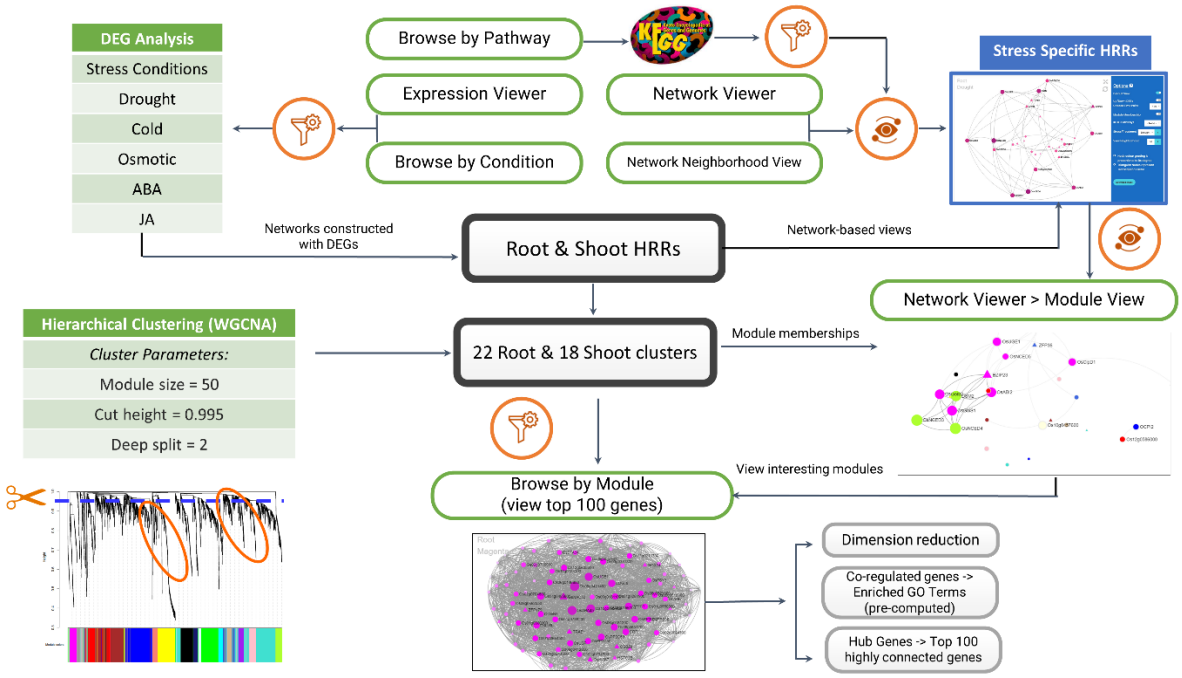


Figure 3.1 Schematic representation of NetREx.

With this objective, we developed Network-based Rice Expression Analysis Server (NetREx), a web-based network querying and visualization resource for *Oryza Sativa L. cultivar* (Sircar et al., 2022). **Figure 3.1** is a schematic representation of functionalities and construction of NetREx. The central black boxes represent the processed data on which the resource is constructed and the green boxes denote the functionalities accessible in the resource for analysis. In NetREx, data for four stress conditions (drought, cold, flood and osmotic stress) and two hormonal treatments (Abscisic Acid ABA and Jasmonic Acid JA) from root and shoot tissues of rice seedlings (10 days after germination) obtained from TENOR database (Transcriptome ENcyclopedia Of Rice, http://tenor.dna.affrc.go.jp) (Kawahara et al., 2016) have been considered for network construction. First, a global rank-based stress network across four stress conditions and two hormonal treatments, separately for root and shoot tissues is constructed. Stress-specific networks are then derived from these global networks using differentially expressed genes. These networks along with functional information are hosted in the database. Using NetREx one can view relationships between query genes and analyze them based on different supported visualizations like the network viewer, the network neighborhood viewer and the expression viewer. It also provides the option to browse through the complete database based on tissues, modules, stress conditions/ hormone treatment and KEGG pathways. Since the data obtained for gene expression values for different stress conditions have been obtained under uniform laboratory condition, this tool is particularly sensitive towards comparative analysis of gene behavior across stress conditions. The visualization of time-point wise variation of gene expression data across queried genes in a co-expression network further offers

insights in understanding the regulatory flow. The network neighborhood network helps us explore novel candidate genes that are absent in the initial query set.

3.2 Visualization Module Construction

HRR Networks based on gene-coexpression are evaluated from RNA-seq data obtained from rice when exposed to various stress conditions in a tissue-specific manner. The construction of this network has been described in detail in Section 2.4. This HRR Network forms the basis of this resource. To find the co-expression network of the queried genes, the resource extracts the subnetwork of this global HRR co-expression network. The nodes of the network represent the queried genes whereas an edge between any two nodes or genes, represent that their expression values are correlated. The HRR networks are processed, and additional omics data are stored along with the basic node and edge information so that they could be used and depicted as visual features in different viewers to enhance analysis. The processed information enables a faster query response time as it avoids database lookup for these features after the query is sent. This processing is achieved using the Gephi Software (*Gephi - The Open Graph Viz Platform*, n.d.) where the network information is assembled in a tissue-specific manner for differentially expressed genes in each stress and WGCNA module. The color and size of the nodes as well as the layout of the network are evaluated using Gephi.

The layout of a large network plays an important role as it defines how the network is perceived. A layout that incorporates both local and global structure, while being scalable to large graphs is essential. Hence the default layout is set to OpenOrd (Martin et al., 2011) which is a force-directed layout that uses average-link clustering to define the positions of the nodes in the network. The default size and colour of the nodes are set to be proportional to their degree in the global network. This helps in visual interpretation of high degree nodes that may be more important because of their higher connectivity with other genes.

The processed network is exported in JSON format and is used by the web application to query the database. The web application provides the option to retrieve co-expression network for the queried set of genes for chosen tissue and stress/hormone treatment condition. The nodes of the network are the queried genes and edges of the network are then filtered based on the condition that both the connecting nodes are present in the query list. For the specified tissue and stress/hormone treatment condition, if the query gene is differentially expressed in at least two time-points, then it is considered as ‘valid’ input for visualization, else it is filtered out as ‘invalid’ gene.

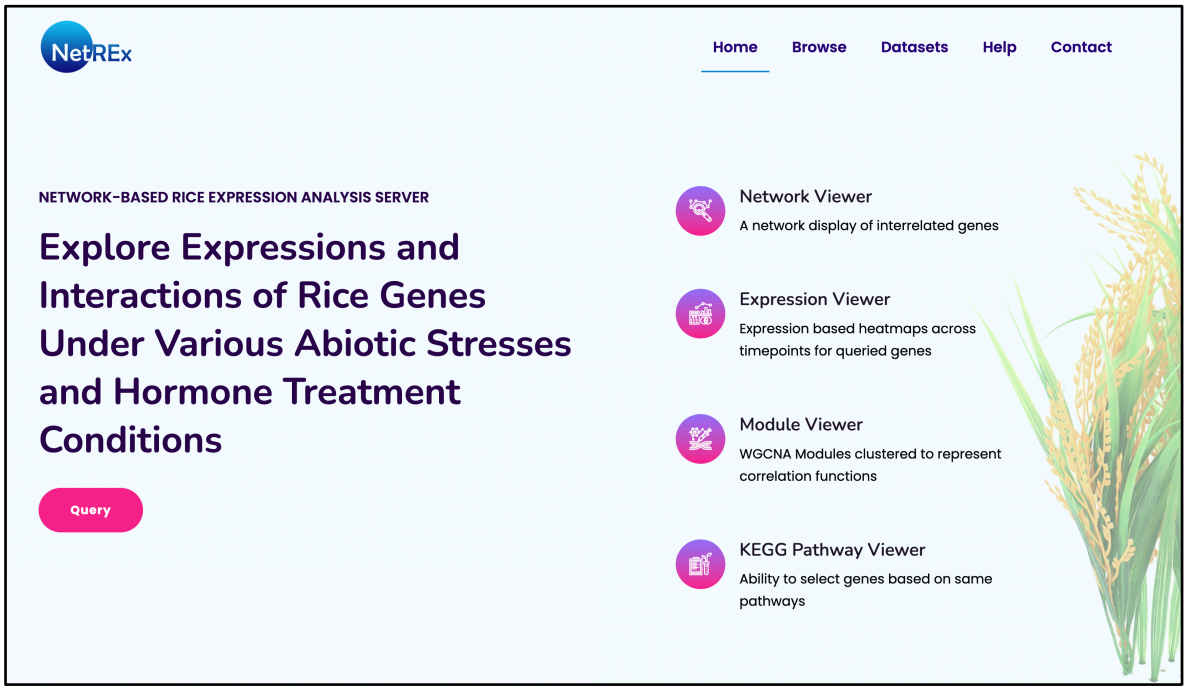


Figure 3.2 Screenshot of Home page in NetREx.

Querying NetREx: A user can submit up to a maximum of 300 genes and query their expression profiles across any of the four abiotic stresses, *viz.*, drought, cold, osmosis and flood, and two phytohormone treatments, abscisic acid and jasmonic acid, for root or shoot tissues. By clicking the query button on the homepage (**Figure 3.2**) the user can access the Query form (**Figure 3.3**) to perform the search. In the dropdown menu, the user may provide rice RAPDB (Rice Annotation Project Database) IDs. Alternatively, the user may also query using Ensembl Stable IDs for Arabidopsis, wheat (*Triticum aestivum*), maize (*Zea mays*), barley (*Hordeum vulgare*) or sorghum (*Sorghum bicolor*), and the genes will be mapped to the corresponding rice orthologs based on the Ensembl Plants database. This allows the user to investigate involvement of genes in other crops under abiotic stress and phytohormone treatments through the interaction networks of their orthologs in rice. Submitting the query directs it to the validation page (**Figure 3.4**), which displays all the valid query genes. The validation page provides option to access different viewers for the valid set of query genes.

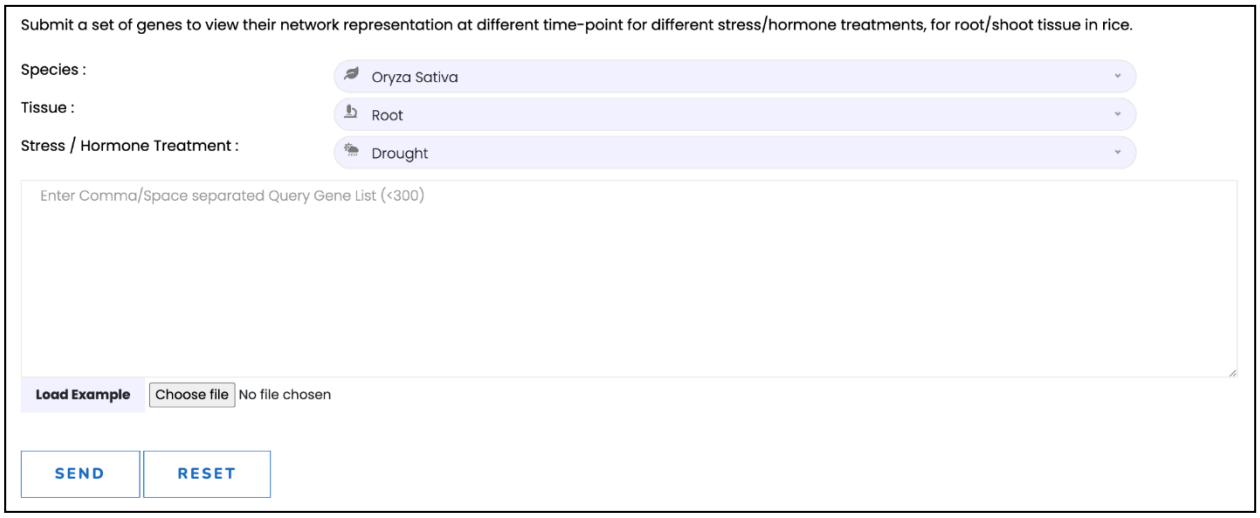


Figure 3.3 Screenshot of Query Form to query the genes in NetREx.

3.3 Visualization

Visualization is an essential aspect in the analysis of complex relationships between genes. It can help in narrowing the focus to important functional genes and their relationship with other genes in biological pathway regulation. NetREx provides multiple visualization options and features for such analyses. Three viewers are provided – Network Viewer, Network Neighborhood Viewer and Expression Viewer, to explore set of query genes in different tissues and abiotic stress/hormone treatment conditions. These viewers are enriched with features such as fold-change values, WGCNA module membership, KEGG pathway membership, as well as comparative study between different abiotic stress/treatment conditions. The network viewers explore the relationship across query genes based on co-expression while the expression viewer is heatmap display for comparative study of expression changes across genes at different time points. The viewers can be accessed via buttons provided in the validation page **(Figure 3.4)**.

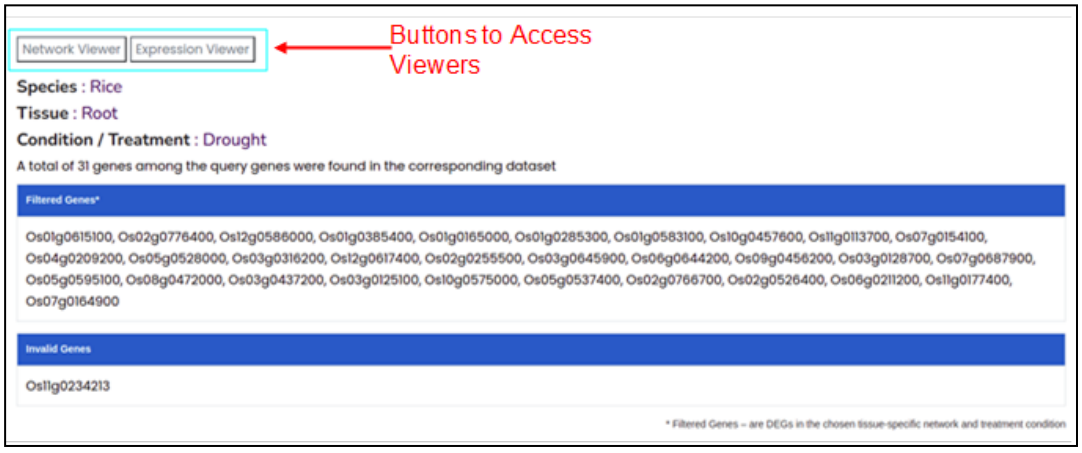


Figure 3.4 Screenshot of the Validation Page in NetREx.

3.3.1 Network Visualization

This module displays the interactions between the filtered set of queried genes, extracted from the HRR network of differentially expressed genes for the selected tissue and stress/treatment condition, shown in **Figure 3.5**. The nodes correspond to the genes and the edge represents a correlation between the two genes if they have a similar expression profile across time points. It can be accessed by clicking the Network Viewer button on the validation page. Network visualization have been provided to display co-expression networks of query genes, their neighbors and WGCNA modules as part of three viewers – Network Viewer, Network Neighborhood Viewer and Module Viewer respectively.

The network viewer is divided into 3 panels, the view panel displaying the network, the options panel which lets one interact with the network features and the table panel which consists of a table with detailed information about the genes constituting the network (**Figure 3.4**). A brief description of these features is given below.

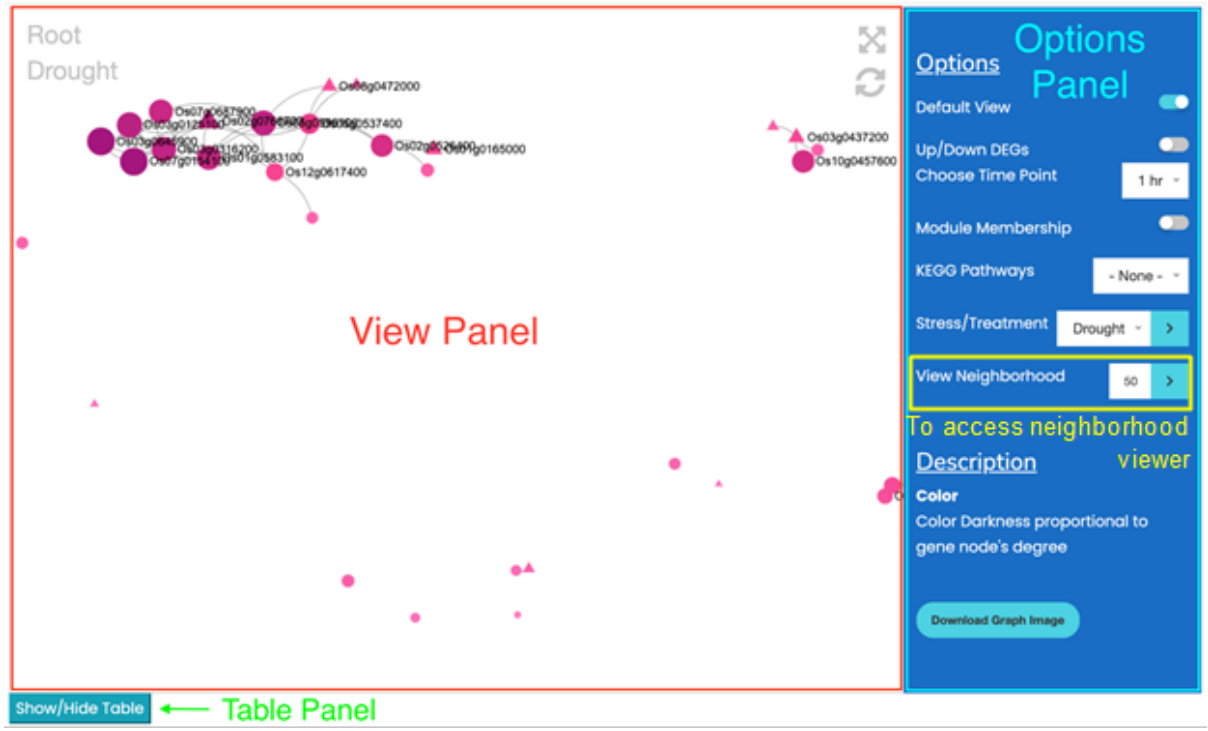


Figure 3.5 Screenshot of Network Viewer in NetREx.

3.3.1.1 View Panel

The interactions between queried genes is displayed in the View panel (**Figure 3.5**). It allows the user to interact with it, such as zooming, repositioning, and hovering over genes. The nodes in the network correspond to the filtered genes after validation and the interactions between them is extracted from HRR100 Network. In the Network Neighbourhood Viewer, apart from the filtered genes, their top 50 neighbours (max=100) are also shown. The view panel is enriched with visual and interactive features that aid in prioritizing and analyzing gene interactions.

Visual Features

The visual features constitute the layout of the nodes, the shape, color and size of the nodes and edges.

Layout: The layout of the network is defined by the ForceAtlas2 algorithm, a force-directed algorithm for network spatialization (Jacomy et al., 2014). The algorithm simulates a physical system wherein the nodes repel each other using force law equivalent to charged particles. The edges act like springs between nodes. These forces in turn directs the movement of the nodes such that they converge to a balanced state. These conflicting forces ensures that highly connected genes appear together while poorly connected nodes drift apart. This algorithm does not take into account any node attributes and is solely focused on turning the network structural proximities into visual proximities which aids in interpretation of the network. The ForceAtlas2 algorithm supports the lay outing of scale-free networks ranging from 10 to 10000 nodes in size and hence is suitable in this case. For NetREx, this algorithm is run on the queried subnetwork for specific time duration, which is set such that it is proportional to the number of nodes in the queried subnetwork. This ensures that for a large network, the algorithm runs for a larger time as compared to a small network, and the disconnected node clusters are well separated while not drifting too apart. It also avoids the issue of the clusters being too closely spaced that the edges are not decipherable. One can see the continuous lay outing algorithm run in action and the nodes taking their respective positions when loading the network viewers.

Size and Color: The size of the gene nodes is proportional to its overall degree in the complete HRR100 Network for the given tissue and stress condition. That is, nodes with higher degree are drawn larger in size to indicate their relative importance in the network. The default coloring scheme is also proportional to its degree in the complete HRR Network. A dark shade represents a high degree and vice versa. However various options are provided in the options panel to toggle the color of the gene nodes based on fold change across timepoints, involvement in biological pathways and their WGCNA module membership. In Network Neighborhood Viewer, the query genes are encircled by Lime green colour to distinguish them from its neighbours.

Shape: The gene nodes are represented by solid circle, while transcription factors are represented by solid triangles. Transcription factors regulate the expression of number of genes, and so representing them with distinct shape in the network gives an idea of the genes that they may be regulating. Since co-expression networks are in general dense, the edges are curved for better visualization (**Figure 3.6**).

Node Label: To avoid overcrowding and improve readability in the network, the labels of all the nodes are not shown all at once. The labels of only those nodes that cross a certain node size threshold across different zoom levels are displayed. However, hovering on a node displays its label.

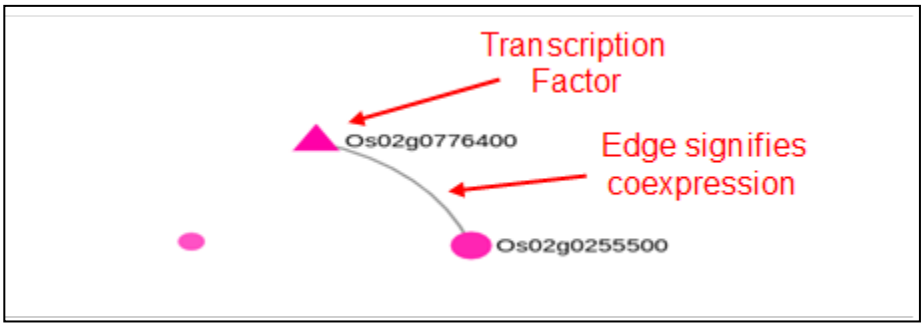


Figure 3.6 Visualization features of the network present in network viewers in NetREx.

Interactive Features

The interactive features provided in the viewer to aid the user in better visualization and analysis of the network is briefly described below.

Zoom and Pan: To focus on a subset of genes, the network can be zoomed in by hovering the cursor over the subset and scrolling the mouse wheel up and down to zoom in and out respectively. One can click and drag at the empty area in the network to move the complete network in the view panel.

Moving a Node: Upon clicking and dragging on a node, it moves relative to the network layout. This feature allows the user to rearrange the layout in order to highlight some regions of the network.

Hover: The first neighbours of a node get highlighted upon hovering over it. This can be used to explore the direct connections of a gene when the network is too dense. The node also displays the same feature if the corresponding row in the table is hovered on (**Figure 3.7**).

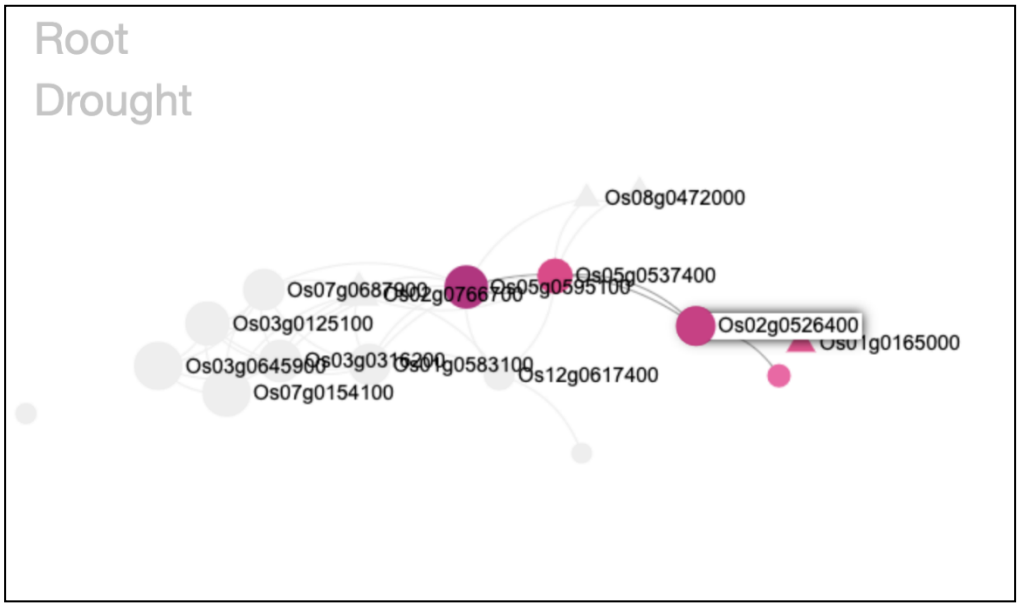


Figure 3.7 Screenshot depicting the hovering feature over a node in NetREx. The first neighbors of Os02g0526400 is highlighted.

3.3.1.2 Options Panel

To augment the visual features, options panel is provided to highlight or toggle the coloring of the nodes to incorporate additional information. It is located to the left of the View panel and its appearance can be toggled by clicking on the expand button present on the top right corner of the view panel (**Figure 3.5**). The different options provided are briefly described below.

Colouring Switches

The nodes in the graph can be coloured with three different schemes upon toggling the corresponding switches.

Default View: By default, the nodes in the graph are shaded based on their degree in the complete network. A darker shade corresponds to a higher degree and vice versa.

Up/Down DEGs: This feature allows the user to view the genes that are up- or down-regulated at different time-points. From the drop-down menu, the user can select the time-point and toggle the button to enable colouring according to up or down regulation of genes. Nodes coloured 'Red' indicates up-regulated genes while those coloured 'Blue' indicates down-regulated genes. Thus, by comparing across different time-points, the user can analyse which genes go up or down as a function of time. This functionality helps in identifying 'early' or 'late' responsive genes. This feature can also be used to compare differential expression of genes across tissues, or across different stress/treatment conditions.

Module Membership: This feature allows the user to view the genes based on their WGCNA module membership. If majority of query genes are part of the same co-expressed module, then it is highly likely that they represent the same biological process and based on their up or down regulation, we can know whether the associated process is activated or repressed.

KEGG Pathways

Using this feature the user can highlight genes that belong to a specific KEGG pathway. From the drop-down menu the user can select the KEGG pathway. On selecting this option, genes that belong to the selected pathway remain coloured while others are greyed out. This option can be used along with the colouring switches to understand the up/down regulation or module membership in a particular KEGG pathway. However, the colour switches should be toggled before the selection of the pathway.

Stress / Treatment

Using this feature the user can see the connectivity and expression status of the queried genes under any stress condition/hormone treatment. From the drop-down menu, the condition can be selected by clicking on the right arrow button. The network for the new stress/treatment condition is loaded in a new tab for comparison.

View Neighborhood

This feature is only provided in the Options Panel of the Network Viewer. In the text box enter the number of neighbours of the queried genes (default = 50, max=100) the user wishes to display. Clicking the right arrow button next to the text box loads the Neighbourhood Viewer (in the same tab).

Download

The user can capture the snapshot of the network in the View Panel as a high-quality publishable JPG image by clicking the Download Image button. The image of the displayed network in the view panel will be downloaded.

3.3.1.3 Table

Additional information, based on "Node" and "Edge" attributes, is provided in a tabular format by clicking on the "Show/Hide Table" button present below the View Panel in **Figure 3.5**. It provides information of the genes in the network view and their interactions, and are listed under the categories, Nodes and Edges.

Edges

Each row corresponds to an edge specifying the source and target gene as well as its PCC and HRR value.

Nodes

The node section consists of information about the gene nodes displayed in the network. This information has been further divided into three sections for better clarity.

Description: Consists of Gene Name, Transcription Factor (TF) Labelling, Module Name, General Description, MSU_ID, k_{total} and link to IC4R Expression. On clicking over a module name, the complete network of the corresponding module (Module Viewer) gets loaded in a new tab. On clicking the "Click!" button under IC4R Expression, the IC4R expression profile of that gene gets loaded in a new tab. Access to IC4R allows the users to compare expression profiles of the gene across a larger set of RNA-seq experiments (24 projects) across various growth stages, tissues, and conditions (The IC4R Project Consortium, 2016).

Function: For this attribute, functional annotations from GO, Mapman, and KEGG pathway is provided.

Fold Change: Fold-change ($\log_2$ Fold Change) and *p*-values of the genes across different time-points is provided in this attribute.

Hover

Upon hovering over a row in the Node Section, the corresponding node and its first neighbours get highlighted in the View Panel. To access this functionality, hover over a particular row in the table to select the corresponding gene and then move the cursor outside the table panel and scroll up to the view panel to see the selected node and its neighbour.

Search

A search bar is provided on the top right part of the table. A Keyword search of the entered text is done across all sections and only those rows which match the search are displayed.

Download Table

The complete table can be downloaded in different formats (*Copy* / *Excel* / *CSV* / *PDF*) by selecting the corresponding button at the end of the table.

3.3.2 Gene Expression Visualization

Early and late stress-responsive genes are known to have distinct roles in stress response. At the same time, genes that are ubiquitously differentially expressed across all the stages of a plant may have some essential roles (Mi et al., 2019). Earlier studies have shown tissue-specific roles of stress-responsive genes indicating that divergence in the expression patterns of differentially expressed genes is an important indicator of their functions. Particularly, for uncharacterized genes, stage and tissue-specific expression profiles can give important cues regarding their functions (Sircar & Parekh, 2015). The user can analyze such stress, tissue and time-point specific information through heat maps provided in the Expression Viewer module (**Figure 3.8**). For the user-provided gene set, differential expression of genes can be observed based on fold-change and *p*-values at various time-points for the chosen stress and tissue. Fold-change heatmap assist in the comparison of expression changes across multiple genes, while the *p*-value heatmap provides the significance to the expression values. The user is also provided with the option to sort the ordering of genes or time-points in the heatmap based on the expression value, thus allowing to observe general trends across them. Other features like zooming and panning are also made available.

Download

The snapshot of both the heatmaps can be downloaded as a high-quality publishable image in both SVG and PNG format.

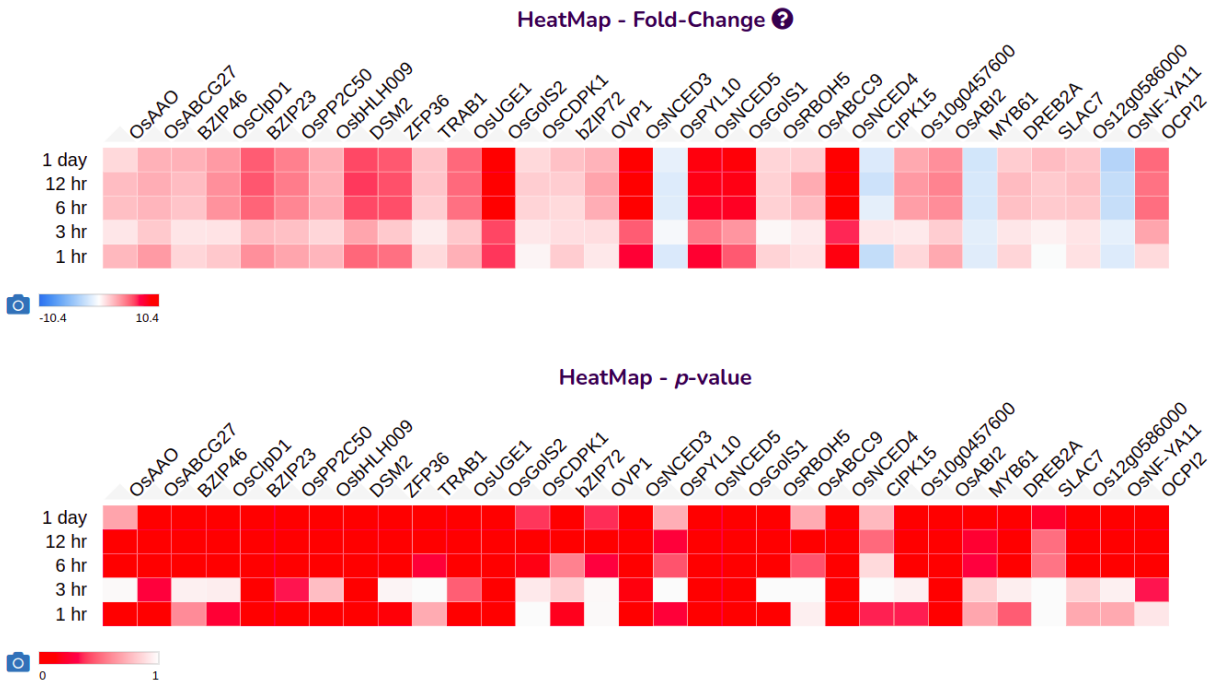


Figure 3.8 Screenshot of the Expression Viewer in NetREx.

3.4 Data Browse Options

Instead of querying for a particular set of genes, the user may browse through the complete dataset. Three browsing options are provided, namely, browsing based on stress/treatment condition, WGCNA module and KEGG pathway. This can be accessed from the navigation bar and provides an alternate way to select a gene set.

3.4.1 Condition-wise

To explore important stress-responsive genes, the user can browse NetREx using the browse option: "Condition-wise." On selecting the tissue and stress/hormone treatment, the user can fetch the list of DEGs (in tabular format) for the corresponding tissue and condition. Tables containing gene information and link to IC4R expression database, gene function (GO, MapMan and KEGG) and fold-change along with *p*-value across time-points are provided. The fold-change tables can be sorted by fold-change or *p*-value to identify most significant up- or down-regulated genes for the chosen condition at different time-points.

3.4.2 WGCNA Module-wise

Using WGCNA R package 22 co-expressed gene modules for root and 18 for shoot HRR networks were identified (**Table 2.6**). These gene modules can be accessed by clicking "Module-wise" under the "Browse" menu and choosing the tissue and module name from drop down menu. On this page, top 100 highly connected genes based on their within-module connectivity can viewed on a graded

colour scale, based on the "colour name" of the module. The top 100 genes of the module are also listed in tabular format along with their node and edge attributes, functional annotation (from GO, MapMan and KEGG) and GO enrichment. For a given module in a tissue-specific network, these genes represent the core components whose functions may be representative functions of the respective module. Further, GO enrichment terms for "biological processes" are provided with the fold-enrichment and FDR values to infer the overall function of the co-expressed functional cluster. The network of the top 100 genes can also be viewed by clicking the "Module Viewer" button present under this browse section (**Figure 3.9**). This network has the same functionality as present in the other network viewers (**Figure 3.10**).

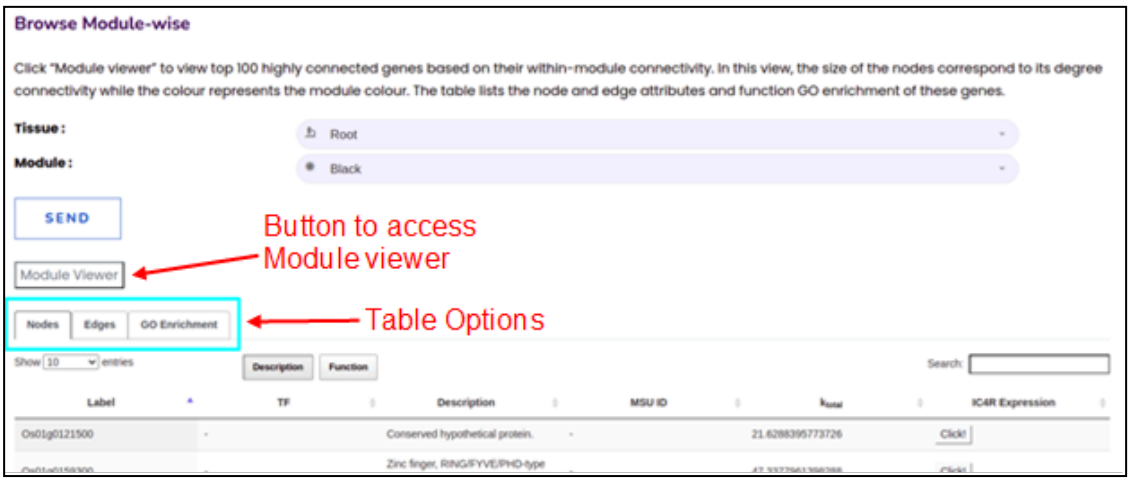


Figure 3.9 Screenshot of the Module-wise browse page in NetREx.

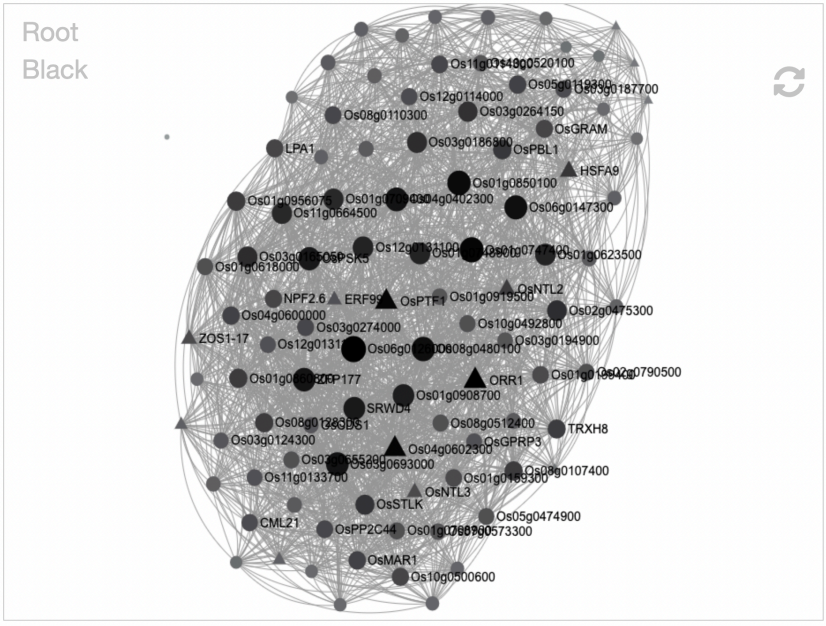


Figure 3.10 Screenshot of the Module-Viewer in NetREx where the WGCNA Black module for the root tissue is depicted.

3.4.3 KEGG Pathway-wise

This is one of the attractive features of NetREx by which hierarchical KEGG pathways can be explored. After selecting the appropriate tissue and condition, the user may select a certain pathway of interest. Genes of the selected pathway that are DEGs for at least two time-points (for chosen stress/treatment condition) can be filtered for further network analysis. A multi-select tree showing the hierarchically arranged KEGG pathways and the associated genes are displayed **(Figure 3.11)**. The checkboxes provided next to each pathway, sub pathways and genes may be ticked by the user for selection. Upon selection, the genes from the chosen pathways (max = 300) are displayed in the right-hand side panel. On submitting the selected genes, these are sent as query to the network and expression viewers and the user can analyze these genes as discussed above for any set of genes provided by the user. A search bar option is also provided for keyword search for KEGG pathways. Thus, this feature allows the user to view genes of the select pathway(s) and their representation in the chosen tissue and stress/treatment condition.

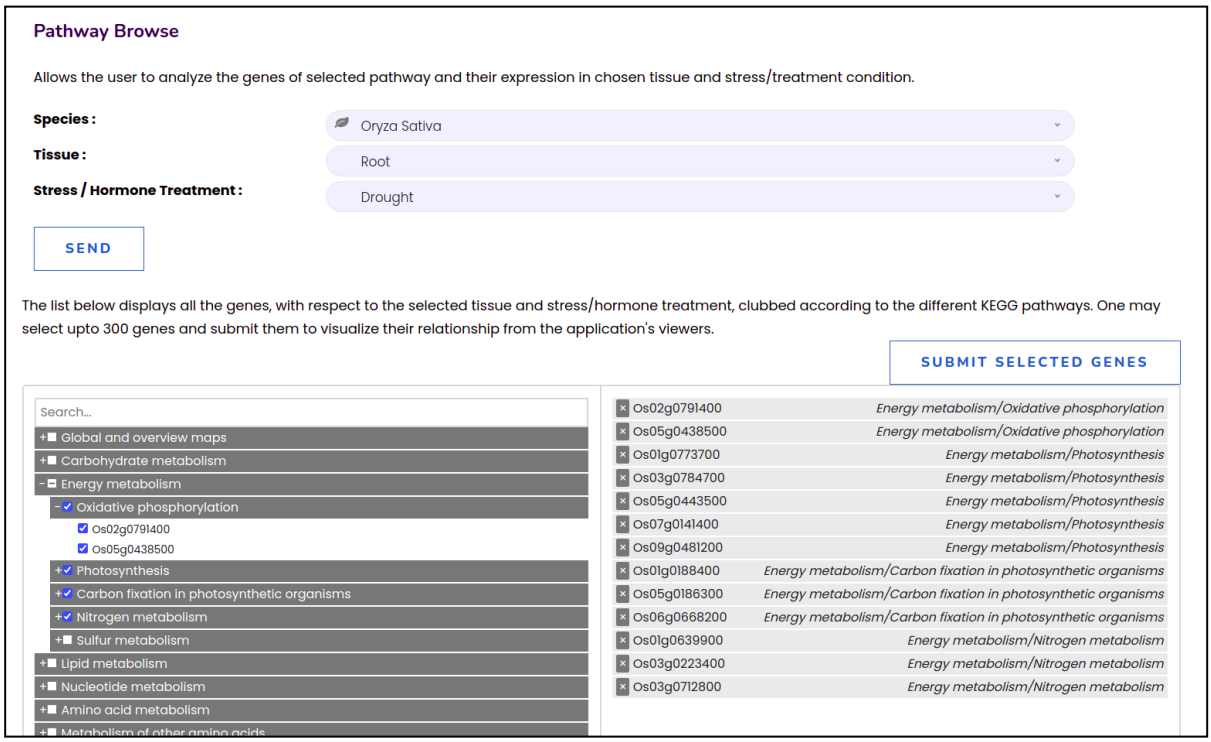


Figure 3.11 Screenshot depicting how to select genes in Pathway-wise browse option in NetREx.

3.5 Ortholog Network Analysis

Ortholog network leverages the use of existing experimental data in one species to identify similar parallel relationships in another species. Analyzing these networks can also aid in improving the accuracy of an inferred network. To investigate the involvement of genes in other crops under abiotic stress and phytohormone treatments, an option is provided to query the interaction networks of their orthologs in rice. The webserver allows querying of orthologous from Arabidopsis, wheat, maize,

barley, and sorghum. To access this feature, the appropriate species should be specified from the dropdown menu and a comma-separated list of Ensembl Stable Gene IDs to be queried should be provided in the same query form as used for *Oryza Sativa*. (**Figure 3.12**). For example, for Arabidopsis: AT2G43490, for Barley: HORVU3Hr1G015620, for Maize: Zm00001d040234, for Sorghum: SORBI_3001G319500, and for Wheat: TraesCS3B02G102400. The genes are mapped to the corresponding rice orthologs based on the Ensembl Plants database. Upon submission, the validation page opens displaying information of the orthologous mapping and the unmapped genes (**Figure 3.13**).

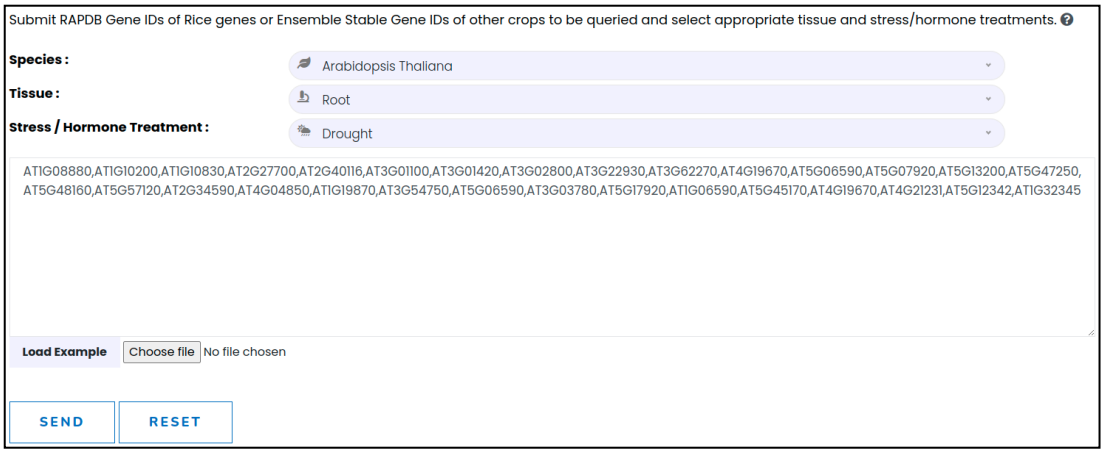


Figure 3.12 Example for ortholog query in NetREx.

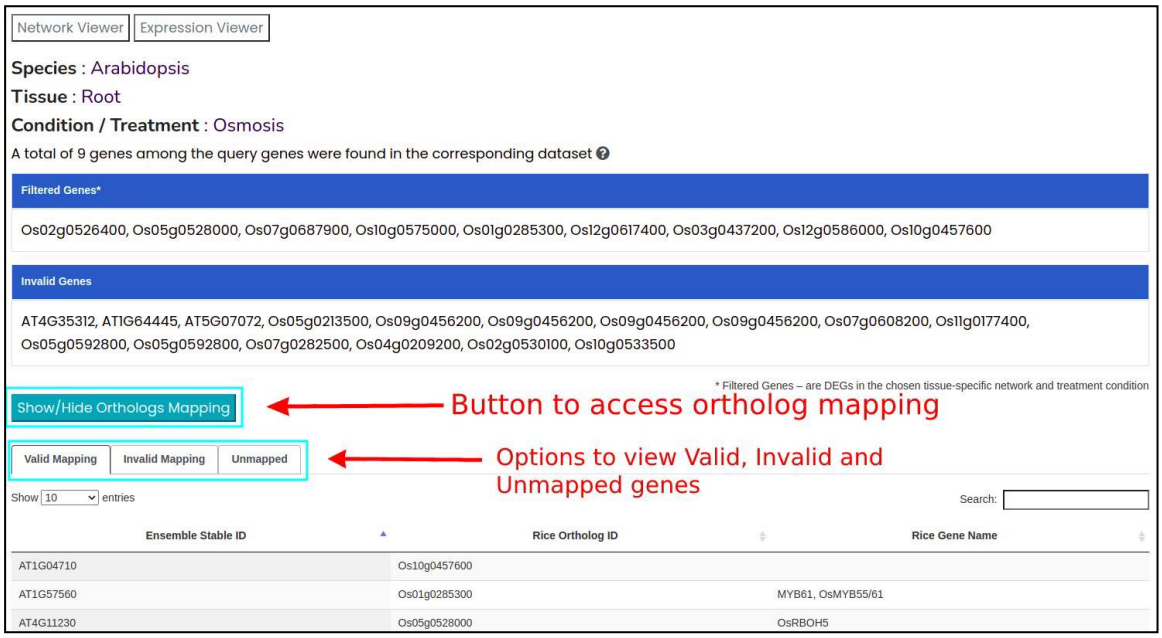


Figure 3.13 Validation Page in NetREx displaying the *Arabidopsis thaliana - Oryza Sativa* gene mapping.

3.6 Dataset

The complete raw datasets used for constructing NetREx can be downloaded by the user from the Navigation Bar on the Home page. It contains data of the Node Lists for Root and Shoot tissues for various stress conditions/hormone treatments along with the fold-change values for different time-points as shown in **Figure 3.14**.

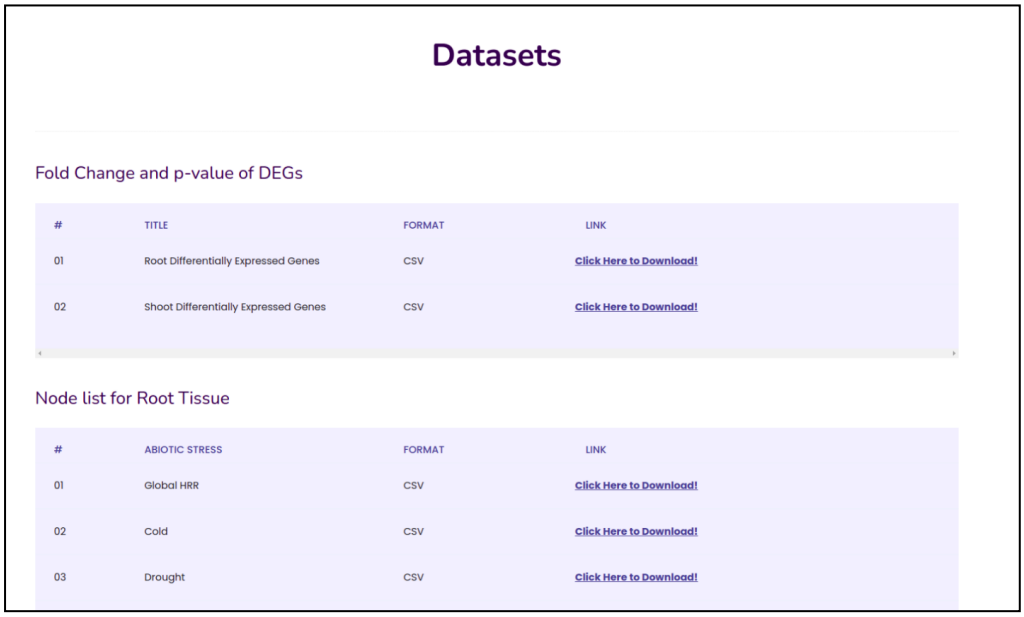


Figure 3.14 Screenshot of the dataset download page in NetREx.

3.7 Implementation

An overview of the functional components of the application is given in **Figure 3.15** and includes various query and browse options, network visualization, heatmap display and functional annotation. Analysis and visualization of large networks is a difficult task for web applications. NetREx has tackled this issue by precomputing various aspects of the network and storing them in the backend as JSON files so that minimal computation time is spent after loading. Network attributes like the initial position of the nodes and edges, their size, shape and color are computed in the Gephi Application (*Gephi - The Open Graph Viz Platform*, n.d.) and the gene attributes are stitched together and exported as one queryable file for each stress condition and tissue type. The backend computation now just involves selection of the appropriate file based on user's input, and then filtering the edges such that only the query genes are considered. For the various visualization features, Javascript plugins have been used. To stitch all of them together into one application, the backend is built on Express JS which is a Node.js web application framework. Express JS functions are used to communicate between the user and the JSON database thereby handling all logical computation. These computations include identifying the query, fetching and filtering of data from the complete graph and extracting relationship between queried genes for chosen tissue and stress/treatment

conditions. The frontend uses HTML, wherein the logical switches and backend calls are handled using the jQuery library.

For Network Visualization, Sigma JS is used which is a javascript library dedicated to graph drawing. It allows the incorporation of features like network lay outing, coloring switches and interactability while also being light weight and computationally efficient. For heatmap display of Expression Viewers, Clustergrammer JS is used which is a front end javascript library for clustergram visualization. It provides interactive visualization features such as sorting and zooming (Fernandez et al., 2017). The KEGG pathway tree in the pathway browse option is constructed using the Tree Multiselect jQuery plugin. Hence the complete application uses a unified language Javascript which makes it efficient and scalable.

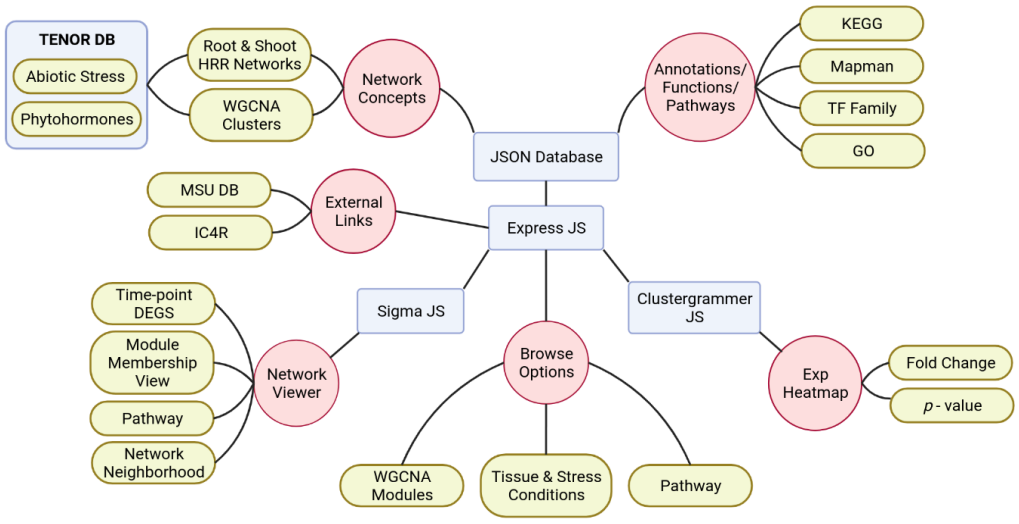


Figure 3.15 The implementation architecture of NetREx.

3.8 Querying NetREx - A Case Study

Below we demonstrate the usefulness of NetREx by quering a set of drought-responsive genes in rice. The selected genes belong to the ABA signalosome complex comprising of PYR/PYL receptors, PP2Cs, SnRK2 kinases and ABF/bZIP transcription factors, obtained from the KEGG database (pathway ID: dosa04075) (Kanehisa & Goto, 2000; Umezawa et al., 2009) for this case study. The ABA signal transduction pathway is one of the key mechanisms by which plants respond to environmental stresses like drought. Several studies indicate that the central signaling module comprises three protein classes: Pyracbactin Resistance/Pyracbactin resistance-like/Regulatory Component of ABA Receptor (PYR/PYL/RCARs) proposed to be the ABA receptors and the regulatory proteins, *viz*., Protein Phosphatase 2Cs (PP2Cs) which act as negative regulators together with SNF1-related protein kinases 2 (SnRKs) which are positive regulators (**Figure 3.16**). Increase in ABA levels during stress leads to the PYR/PYL/RCAR-PP2C complex formation causing inhibition of PP2C activity, thereby allowing activation of SnRKs which target the functional proteins

like membrane proteins, ion channels and transcription factors, and facilitate transcription of ABA-responsive genes. Network-based resources such as NetREx can facilitate querying the coordinated interactions of regulatory genes and their functional targets which further trigger downstream processes.

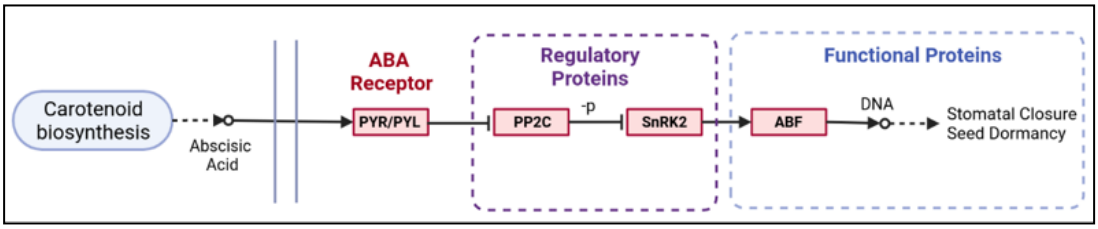


Figure 3.16 A schematic representation of the ABA signaling pathway from KEGG database.

A total of 41 rice genes from KEGG (pathway ID: dosa04075) are queried in NetREx in "root" and "shoot" tissues under "drought" stress. Of these 13 and 17 genes respectively mapped to the root and shoot networks and are listed in **Table 3.1**. The filtered gene sets (13 and 17) are a union of DEGs across all time points for the chosen condition (drought) and tissue (root and shoot). The "invalid genes" on the other hand either have very low expression values or are not differentially expressed in at least two timepoints and hence were not considered in the network construction in NetREx.

Table 3.1 Shoot- and root-specific subnetworks of ABA Signalosome Complex extracted using NetREx under drought stress.

RAP-DB Id	Gene Name	Component of the ABA Signalosome	Degree
Shoot Network			
Os09g0325700	OsSIPP2C1	PP2C phosphatase	7
Os01g0583100	OsPP2C06	PP2C phosphatase	6
Os03g0268600	OsPP2C30	PP2C phosphatase	6
Os05g0537400	OsPP2C50	PP2C phosphatase	6
Os01g0656200	OsPP2C08	PP2C phosphatase	5
Os01g0869900	SAPK4	SnRK2 protein kinase	5
Os02g0766700	OsbZIP23	bZIP TF	5
Os06g0211200	OsAREB1	bZIP TF	2
Os07g0622000	SAPK2	SnRK2 protein kinase	1
Os05g0213500	OsPYL5	ABA receptor	1

Os01g0867300	OsbZIP10	bZIP TF	1
Os02g0551100	SAPK6	SnRK2 protein kinase	1
Os01g0813100	HBF2	bZIP TF	0
Os10g0564500	SAPK3	SnRK2 protein kinase	0
Os03g0390200	SAPK1	SnRK2 protein kinase	0
Os05g0437700	OsbZIP40	bZIP TF	0
Os02g0255500	OsPYL3	ABA receptor	0
Root Network			
Os01g0656200	OsPP2C08	PP2C phosphatase	7
Os09g0325700	OsSIPP2C1	PP2C phosphatase	7
Os05g0537400	OsPP2C50	PP2C phosphatase	7
Os03g0268600	OsPP2C30	PP2C phosphatase	7
Os01g0583100	OsPP2C06	PP2C phosphatase	5
Os01g0867300	OsbZIP10	bZIP TF	5
Os02g0766700	OsbZIP23	bZIP TF	5
Os06g0211200	OsAREB1	bZIP TF	5
Os02g0255500	OsPYL3	ABA receptor	0
Os02g0551100	SAPK6	SnRK2 protein kinase	0
Os07g0622000	SAPK2	SnRK2 protein kinase	0
Os10g0573400	OsPYL1	ABA receptor	0
Os09g0456200	OsbZIP72	bZIP TF	0

In the **Expression Viewer**, fold-change of the filtered (valid) genes is displayed as heatmaps, shown in **Figure 3.17**. For the root tissue (**Figure 3.17 (A)),** it is observed that majority of the DEGs are strongly up-regulated as early as 1 hr time-point. However, at 3 hr, a decrease in fold-change is observed which is probably due to transcriptomic and metabolic reprogramming. For the shoot tissue (**Figure 3.17 (B))** it is observed that most genes are not induced at 1hr time-point but gradually the fold-change increases at later time-points. This indicates that response to drought stress is induced in root tissue earlier compared to shoot tissue.

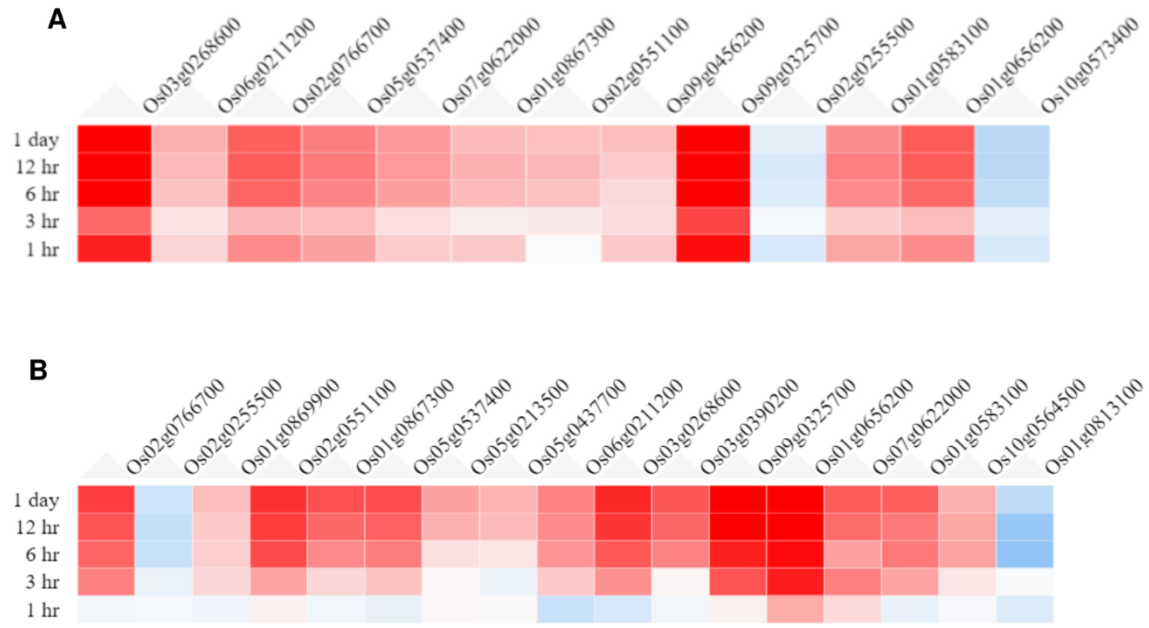


Figure 3.17 Heatmaps from Expression Viewer for (A) Root tissue and (B) Shoot tissue under Drought stress. Genes in red having high positive fold-change and those in blue having high negative fold-change with respect to control (0h, no stress).

Using "**Network Viewer**" module in NetREx, user can observe stress and tissue-specific view of the HRR-100 network. The expression status of the queried genes can be viewed in a time-specific manner by choosing the time-points 1h (default), 3h, 6h, 12h and 1day. In **Figure 3.18**, the "default view" indicates that among the 13 and 17 DEGs that mapped to the drought HRR-100 network in root (**Figure 3.18 (A)**) and shoot (**Figure 3.18 (B)**) respectively, 8 genes are seen to form the largest connected component in both these networks.

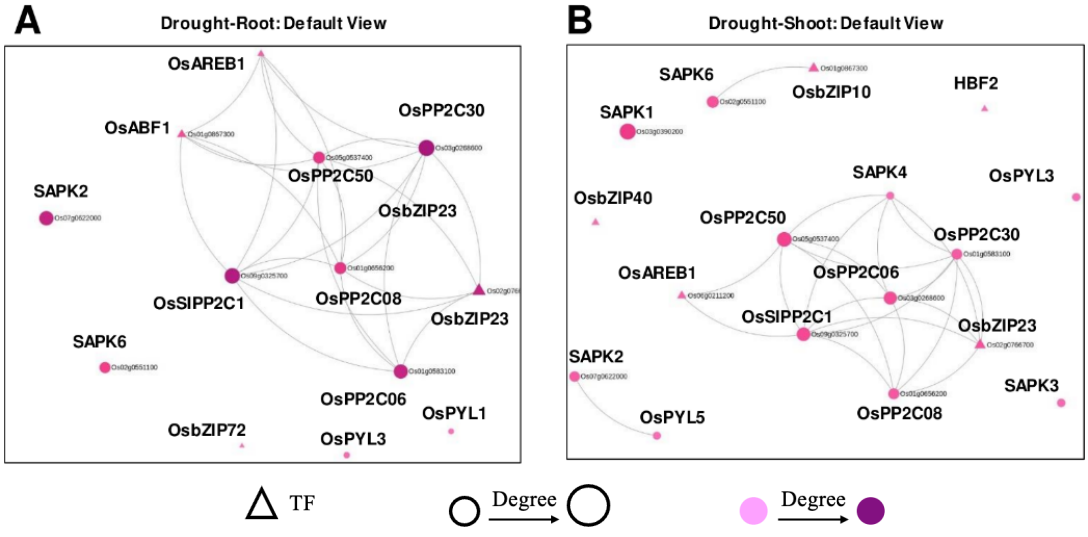


Figure 3.18 The network of drought-responsive DEGs in involved in ABA signaling pathway shown. (A) Root tissue (13 DEGs) and (B) Shoot tissue (17 DEGs).

In **Figure 3.19** is shown the connectivity information between genes, using the colour coding scheme: up- regulated (red) and down-regulated (blue) genes at 1h, 3h and 1day. For example, in the shoot network, the high-degree gene OsSIPP2C1 is observed to be up-regulated at 3h of stress, while it is up-regulated even early at 1h in root network. Interestingly, it was observed in previous studies that OsSIPP2C1 is negatively regulated by ABL1 which is involved in abiotic stress and panicle development in rice (Li et al., 2013). In the root network, the components are more tightly connected with the PP2Cs *viz.*, OsSIPP2C1, OsPP2C50, OsPP2C30 and OsPP2C08 with the highest degree genes interacting with the TFs OsAREB1, OsbZIP23 and OsABF1. Also, most of the genes are up-regulated at 1h time-point (**Figure 3.19**) and are probably early response genes in root tissue, except OsAREB1 which seems to be a late response gene activated only at 6h. For the shoot network, the response is delayed as discussed above and the SNRK2 protein kinase, SAPK4, is not induced until 12h. While majority of the valid genes are up-regulated in shoot and root, a few genes are observed to be down-regulated. These include OsPYL3 (Os02g0255500), down-regulated both in root and shoot tissues. Indeed, this gene has been shown to be down-regulated in drought-susceptible rice genotype under abiotic stress conditions, while over-expression of this gene in Arabidopsis led to improved tolerance in cold and drought stress conditions (Li et al., 2015). The shoot-specific TF HBF2 is also down-regulated in the shoot network (6h, not shown in the figure), while the root-specific OsPYL1 gene is down-regulated at 1h in the root network (**Figure 3.19A**). Among the 13 root and 17 shoot, 11 genes are common and differentially expressed in the two tissues. Interestingly, all the five 2C protein phosphatase (PP2C) proteins are common to both root and shoot, indicating their ubiquitous role of negative regulation of ABA (via SnRK2s and PYR/PYL/RCARs) in both the tissues (Fujii et al., 2009). However, in terms of network-concepts, connectivity between the common gene sets differ between the two tissues as observed in **Figure 3.18**. Among the six shoot-specific DEGs, the bZIP TF HBF2 has been shown to be highly expressed in shoot apical meristem (Fujii et al., 2009). On the other hand, the root-specific gene OsPYL1 was shown to interact with OsABIL2 that has a role in root development (Li et al., 2015).

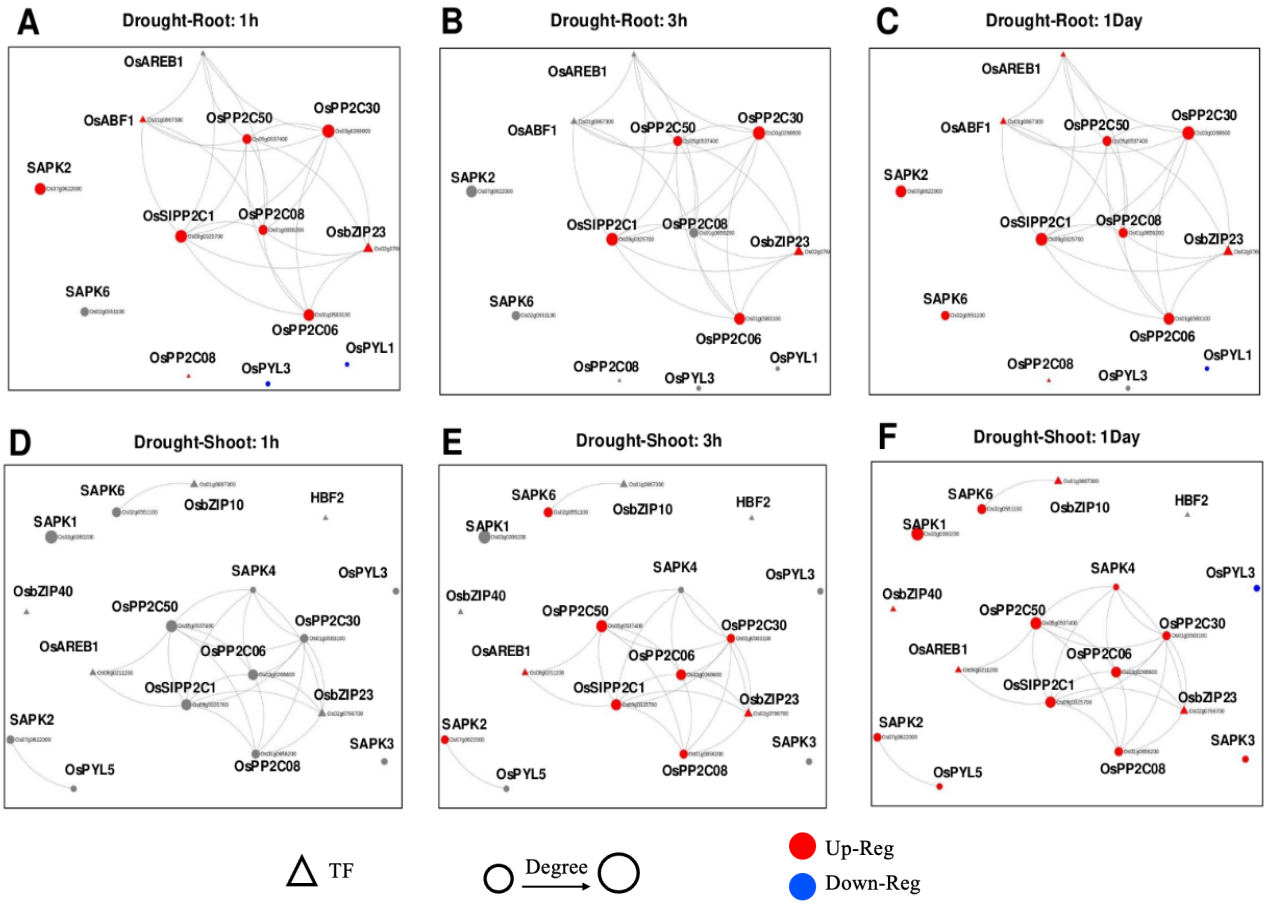


Figure 3.19 Time-point specific networks of drought-responsive DEGs in Root and Shoot tissues respectively at 1h (A, D), 3h (B, E) and 1Day (C, F). The figure provides a comparative view of the transcriptional changes for the different time-points along with tissue-specific gene connectivity.

Clusters of co-expressed genes often represent coordinated biological processes. Analysis of gene co-expression networks have been helpful in functional annotation of uncharacterized genes (Chandran et al., 2018; Sircar & Parekh, 2015), prioritization of candidate genes (Schaefer et al., 2018; Zeng et al., 2018), inferring biological processes, e.g., metabolic pathways (Kautsar et al., 2017; Wisecaver et al., 2017), stress response mechanisms (Nounjan et al., 2018; Tan et al., 2017), cell wall metabolism (Ferreira et al., 2016; Wang et al., 2012), etc. For example, out of the 13 genes in the root network, 8 genes belong to the root-specific Magenta module, shown in **Figure 3.20**. Incidentally, they also form the largest component in the network (**Figure 3.20 (A)**). The remaining five genes with zero degree belong to GreenYellow, Yellow (2 genes), Blue and the Brown module. To obtain further details on the root-specific Magenta module, we use the "Browse" option in NetREx for "Root" tissue. Some of the significant GO terms include "regulation of transcription, DNA-templated" (GO:0006355, FDR= 8.56e-04) and "abscisic acid-activated signaling pathway" (GO:0009738, FDR= 7.64e-03) indicating the relevance of Magenta module in drought stress. Similarly, for the shoot network, out of 17 genes, 8 belong to the shoot-specific Turquoise and 4 to the shoot-specific Salmon modules. The Salmon module harbours genes involved in

Dephosphorylation (GO:0016311, FDR= 1.98e-02), while the Turquoise module is involved in a number of stress-responsive processes including ER-associated misfolded protein catabolic process (GO:0071712, FDR= 8.83E-03), regulation of response to stress (GO:0080134, FDR=3.51e-05), etc. The Module Viewer allows further exploration of genes belonging to the respective modules. For example, the above analysis indicates that genes of root-specific Magenta module maybe biologically relevant for drought stress. On exploring the top 100 highly connected genes of the module, we see several known TF family genes like bZIPs and HSFs are important hubs of this module. Along with these, several "Conserved hypothetical proteins' lacking detailed functional annotations are also part of the hubs. The associations of these genes with known TFs can be further queried in NetRex along with their expression profiles across different conditions and different tissues (IC4R Expressions).

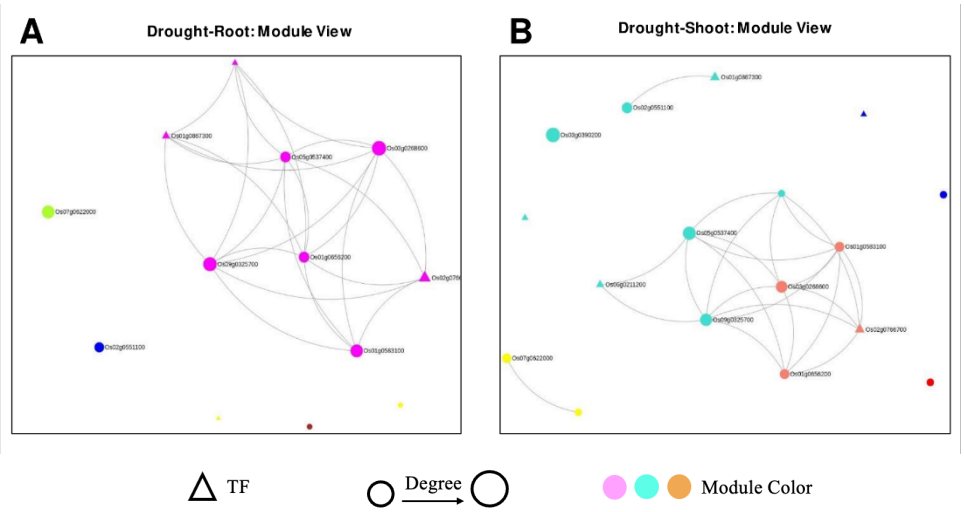


Figure 3.20 The module-wise colored network of drought-responsive DEGs in involved in ABA signaling pathway shown. (A) Root tissue (13 DEGs) and (B) Shoot tissue (17 DEGs).

In systems biology, network neighbourhood analysis is an important aspect as it facilitates a "guilt-by-association" strategy by which we can find interesting genes which are closely interacting/co-expressed with the initial "seed genes". In the **Neighborhood View** on the right panel, top 100 neighbours based on k_{Total}, the connectivity in the whole network, of the 13 root-specific seed genes are fetched. The "seed genes" are encircled in green to distinguish from other neighbourhood genes in **Figure 3.21 (A)**. To infer the overall function of the subnetwork (13 query genes and their respective 100 neighbours), we performed GO enrichment analysis. As expected, positive regulation of abscisic acid-activated signalling pathway (GO:0009789, FDR=9.50E-03) was the most enriched term. Two major clusters are clearly discernible (**Figure 3.21 (B)**) in this subnetwork.

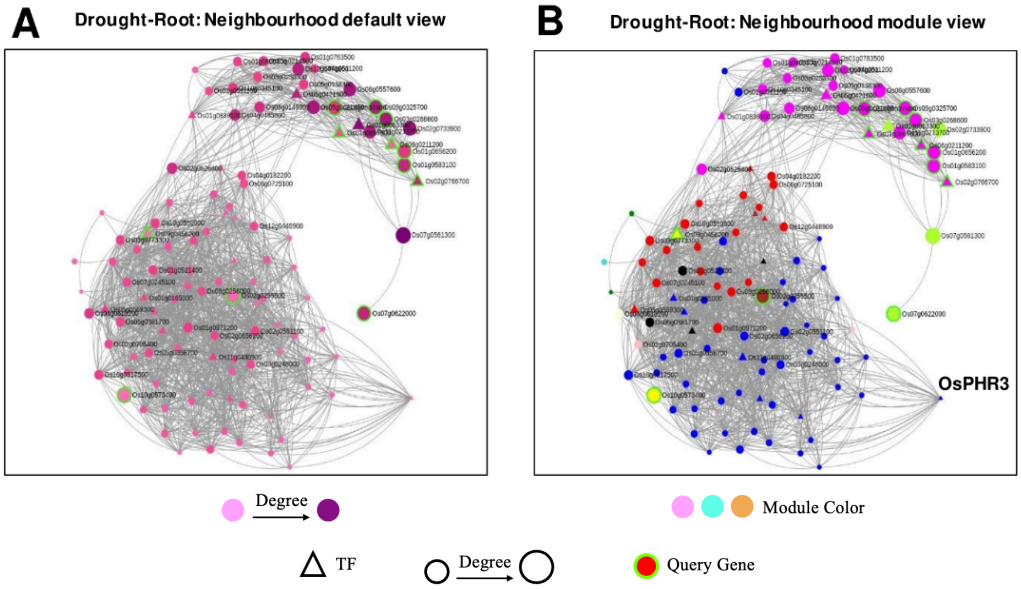


Figure 3.21 Neighborhood View of 13 drought-responsive DEGs in root tissue: (A) default view and (B) Module view.

The first set consists of neighbourhood genes that majorly belong to the Magenta module (16 genes) and these genes are in fact up-regulated as early as 1h of drought stress (**Figure 3.22 (A)**). As indicated above, the Magenta module is involved in the ABA signalling and stress responsive pathways. The other cluster consists of mostly genes belonging to Blue and Red modules. These sets of genes are late response genes that are over-expressed as a result of downstream cellular and metabolic adjustments after the signalling components have been induced in the early time-points (**Figure 3.16 and Figure 3.22** (**B**)). The Blue module includes stress-responsive genes that aid in autophagy of damaged proteins and cellular organelles (GO:0044805, FDR=6.74E-03) (Su et al., 2020). The Red module harbours genes involved in the methylerythritol 4-phosphate (MEP) pathway of isoprenoid biosynthetic process leading to the production of carotenoids and various other secondary metabolites (GO:0019288, FDR= 1.32E-03).

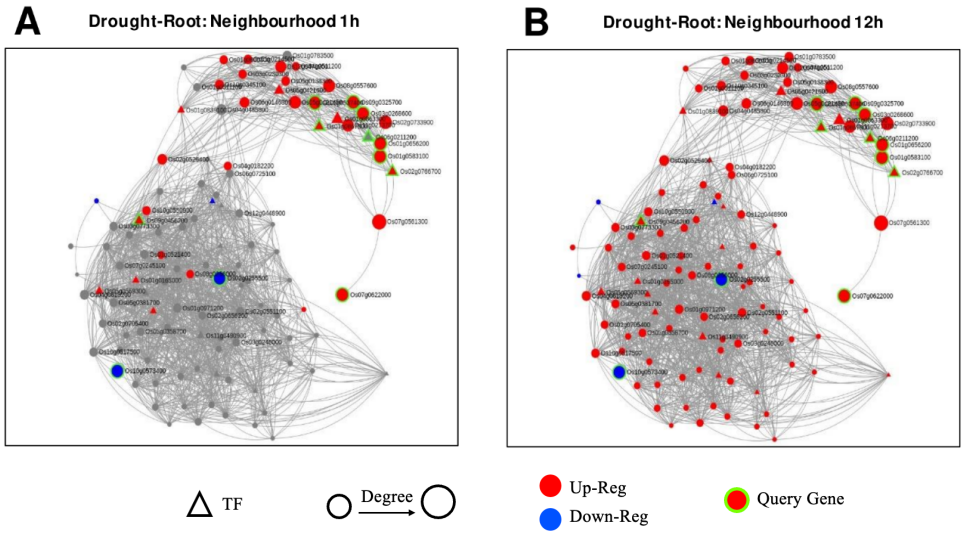


Figure 3.22 Time-point specific views for the extended neighbourhood of drought-responsive genes in root tissue: (A) early timepoint of 1h and (B) late time-point of 12h.

As discussed above, network neighbourhood helps us explore novel candidate genes that are absent in the initial query gene set. For example, the gene with the highest k_{Total} (weighted connectivity of a gene in the whole root network) is the transcription factor OsPHR3 (Os02g0139000) implicated in low Pi stress and in regulating Nitrogen homeostasis (Sun et al., 2018). To explore the transcriptomic dynamics of this gene in other stress conditions, we queried in NetREx again. Using the option to check the expression profile in 'other conditions' provided on the right panel in "Network Viewer", we observed that this gene is up-regulated in osmotic stress (3-6h), flood stress (1h), ABA (1h to 1 day) and JA (1 and 3h) while it was down-regulated in osmotic stress at 12h, flood stress (3-6h) and JA (6h to 1 day). Additionally, to explore the expression of this TF across rice growth stages and tissues, we used the IC4R link in the "Nodes Description" table. From **Figure 3.23** it may be noted that OsPHR3 exhibits higher expression in root and leaf tissues. On scanning the upstream 1kb of this gene using the PlantPan v2.0 database (Chow et al., 2016), we observed several bZIP binding motifs especially in the 500 kb upstream region. Further, concurrent with this, several binding sites for WRKY TFs were also detected in and around the same region, indicating that this TF may also be a target of biotic stress signalling cascades (**Appendix Table 8**). The network neighbourhood view of OsPHR3 was next explored. All its neighbours are observed to be up-regulated at 6h in root tissue under drought stress and all of them belong to the Blue module. Moreover, the top 10 neighbours are involved in functions like transferring phosphorus-containing groups, carbon-nitrogen ligase activity as well as drought and biotic stress (**Appendix Table 8).** Literature survey revealed that this TF has not been functionally characterized in multiple stress conditions such as drought and warrants further investigation.

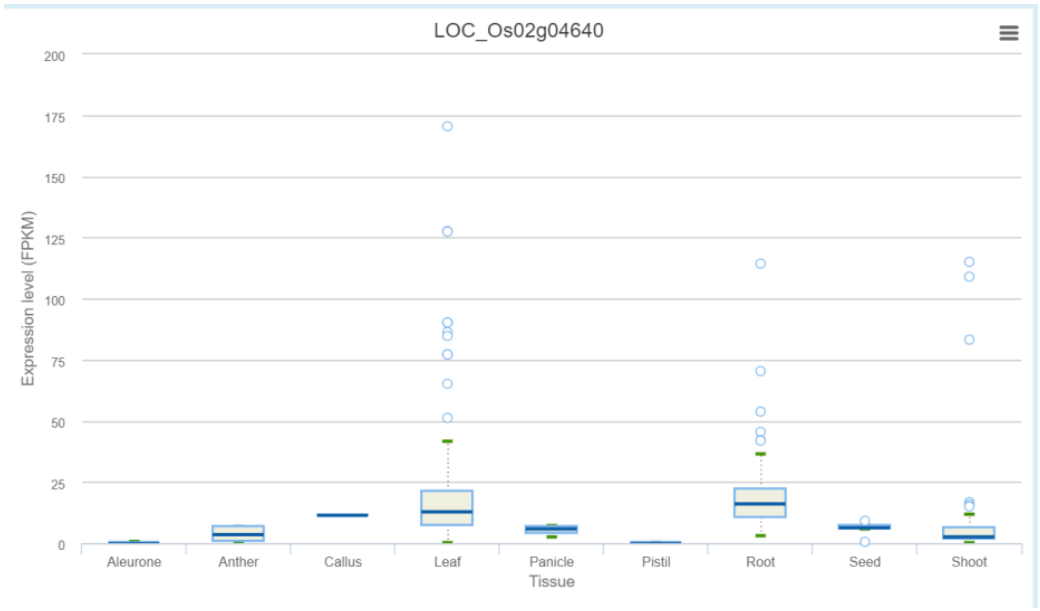


Figure 3.23 Expression profile of OsPHR3 from IC4R across different tissues and development stages.

A similar analysis was carried out with 17 shoot-specific DEGs and the Network View is shown in **Figure 3.24 (A)**. It may be noted from the module view in **Figure 3.24 (B)** that majority of the genes belong to the shoot-specific Turquoise Module while a separate cluster is formed by genes of the Blue Module. Some of the most significant functions of the subnetwork include protein serine/threonine phosphatase activity (GO:1905183, FDR=6.58E-03), ABA signalling (GO:0009789, FDR=8.95E-03), cellular response to nitrogen starvation (GO:0006995, FDR=1.28E-03), etc. Here, the gene with the highest k_{Total} is OsAtg8 (Os07g0512200), which is a well characterized gene involved in autophagy and protein degradation (Su et al., 2006). However, the role of this gene with respect to drought is yet to be explored. An important point to be noted is that we extracted extended root and shoot networks under drought stress using the same "seed" genes involved in ABA signalling. Needless to say, both the subnetworks had GO terms enriched for "abscisic acid-activated signaling pathway" (GO:0009738), "regulation of response to water deprivation" (GO:2000070), "cellular response to hormone stimulus" (GO:0032870) and so on. However, more specific tissue-specific GO terms like "photosynthesis and dark reaction" (GO:0019685, FDR= 3.15E-02), "gluconeogenesis" (GO:0006094, FDR= 3.60E-03), "cellular response to nitrogen levels" (GO:0043562, FDR= 4.21E-05), etc. for the shoot tissue and "cellular response to reactive oxygen species" (GO:0034614, FDR= 3.57E-02), "fatty acid oxidation" (GO:0019395, FDR= 5.41E-03), etc. for the root tissue were noted. Another major difference in the **Neighborhood View** of these genes in root and shoot tissues is the presence of down-regulated genes in the shoot network as compared to root and the gradual activation/repression of these genes in the shoot (**Figure 3.25 (A to D))** in contrast to root (**Figure 3.22 (A** and **B).** The down-regulated genes in shoot (21 genes) majorly belong to the **Blue** module (**Figure 3.24 (B**)) with at least 6 genes annotated to be involved in photosynthesis, indicating that this process is preferentially switched-off

in the green tissues under drought stress. Further exploration of the Blue module of the shoot HRR network using the "Browse Module Wise" page in NetREx revealed several interesting shoot-specific GO terms like "photosystem II repair" (GO:0010206, FDR= 9.62E-04), "photosystem II assembly" (GO:0010207, FDR= 4.41E-03), "regulation of photosynthesis, light reaction" (GO:0042548, FDR= 3.80e-03), etc. Adverse effects of abiotic stress conditions like drought on the photosynthetic machinery with harmful effects on the overall growth and yield of the crop are well documented (Sircar & Parekh, 2019; Zhou et al., 2007). This confidently explains the functional differences of the root and shoot subnetworks from the extended ABA signalosome analysis discussed above. Moreover, the expression profiles of the genes across time-points and their corresponding tissue-specific network connectivity enables one to confidently explore the temporal and functional space of the genes and arrive at relevant conclusions.

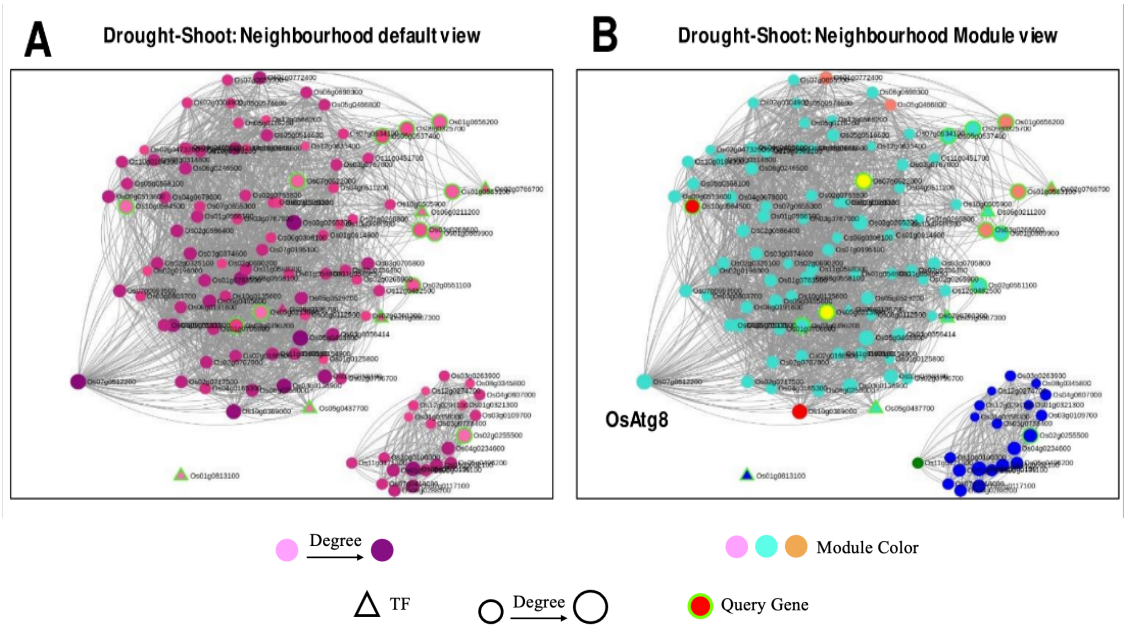


Figure 3.24 Neighborhood View of 17 shoot-specific genes with (A) default view and (B) Module views under Drought stress.

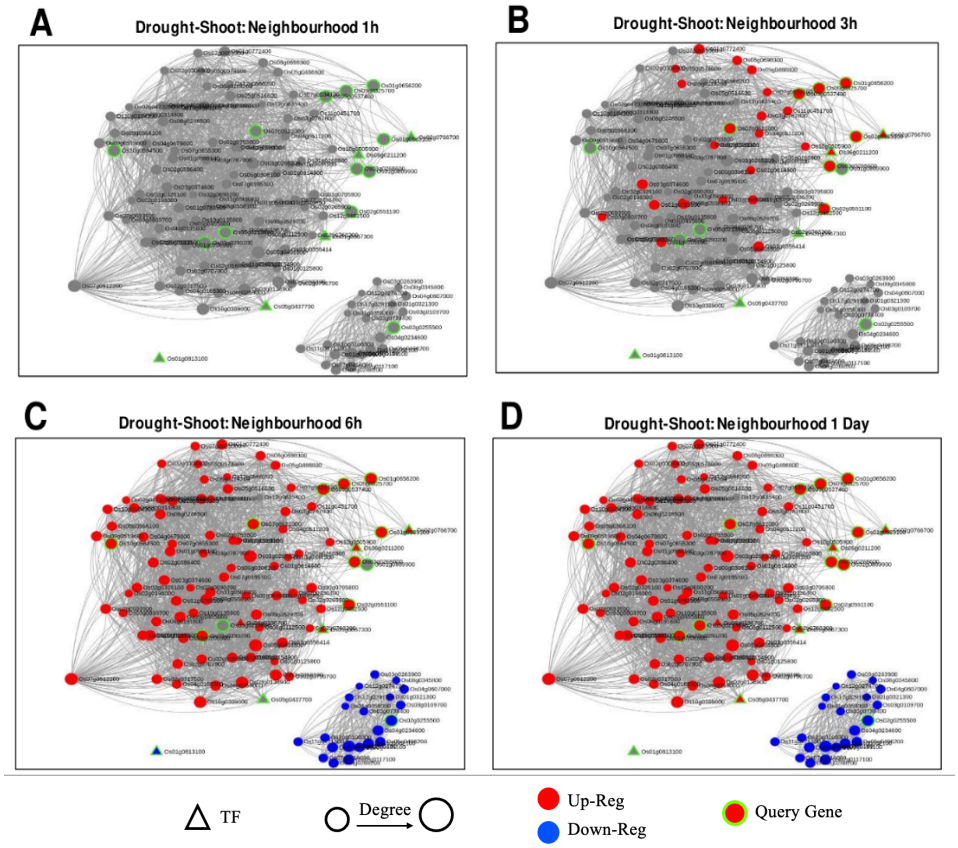


Figure 3.25 Time-point specific views for the extended neighbourhood for shoot-specific genes at early timepoints (1 and 3h) and late time-points (6h and 1day).

3.9 Conclusion

NetREx is a freely accessible web-server for biologists to conveniently explore the global rank-based stress networks in a tissue-specific manner. The resource has been constructed using high-quality RNAseq data from the TENOR database generated using homogeneous experimental protocols. In NetREx, substantial emphasis has been given to explore the networks through various perspectives such as exploring gene expression profiles (Expression Viewer heatmaps and Network Viewer in a time-point specific manner), network connectivity (Network Viewer and Neighbourhood Viewer), identification of novel stress-responsive candidates (Neighbourhood Viewer), functional analysis of genes (browsing NetREx by modules and pathways) and comparative analysis across stress conditions (supported in Network Viewer mode). The gene attributes displayed in the different modules have been extensively cross-linked to various other resources to provide additional information to the users. Our analysis indicates that the rank-based networks in NetREx are biologically relevant wherein the tissue and stress-specific information is effectively retained. Network-based subnetwork analysis and gene prioritization using NetREx will therefore be a significant resource to study complex phenotypes associated with stress-response in rice.

Chapter 4

Results and Discussions

In this chapter we discuss the results obtained from the processed RNA-seq data of rice under various abiotic stress conditions, followed by a thorough analysis into the processes involved in osmotic stress response in rice. The analysis on the timepoint wise count of differential expressed genes under various conditions is given in section 4.1. Section 4.2 discusses the distribution of DEGs based on their WGCNA module membership, thereby shedding some light on the coordinated biological processes that occur in the various stress conditions in a timepoint wise manner. Section 4.3 describes the distribution of genes of the differentially expressed WGCNA modules based on the metabolic pathways they take part in. Finally, Section 4.4 gives an in depth analysis of the processes and important genes involved in osmotic stress in rice.

4.1 Differential Expression Analysis in Root

A comparative analysis of Differentially Expressed Genes (DEGs) across different time-points and different conditions is conducted to get insights into the response time associated with different environmental stresses. The number of up and down-regulated genes observed at different time-points for various stress conditions and hormone treatments is shown in **Figure 4.1**. It is observed that drought exhibits significant number of DEGs from 6hr time-point, and the number of up-regulated genes are much higher than down-regulated genes. Osmotic stress exhibits a similar behavior but with significantly higher number of DEGs than in the case of drought case and from 1hr onwards indicating an early response to stress. Flood stress exhibits negligibly small number of DEGs that are mainly down-regulated. Cold stress and ABA and JA treatment exhibit a similar trend of DEGs from 1hr onwards with higher number down- regulated compared to up-regulated genes. While osmotic stress and ABA treatment show a uniform number of up-regulated DEGs from 3h onwards, flood stress and JA treatment have a slight drop in DEGs towards the final time-points with a peak at around 12h for cold stress and 6h – 12h for JA treatment.

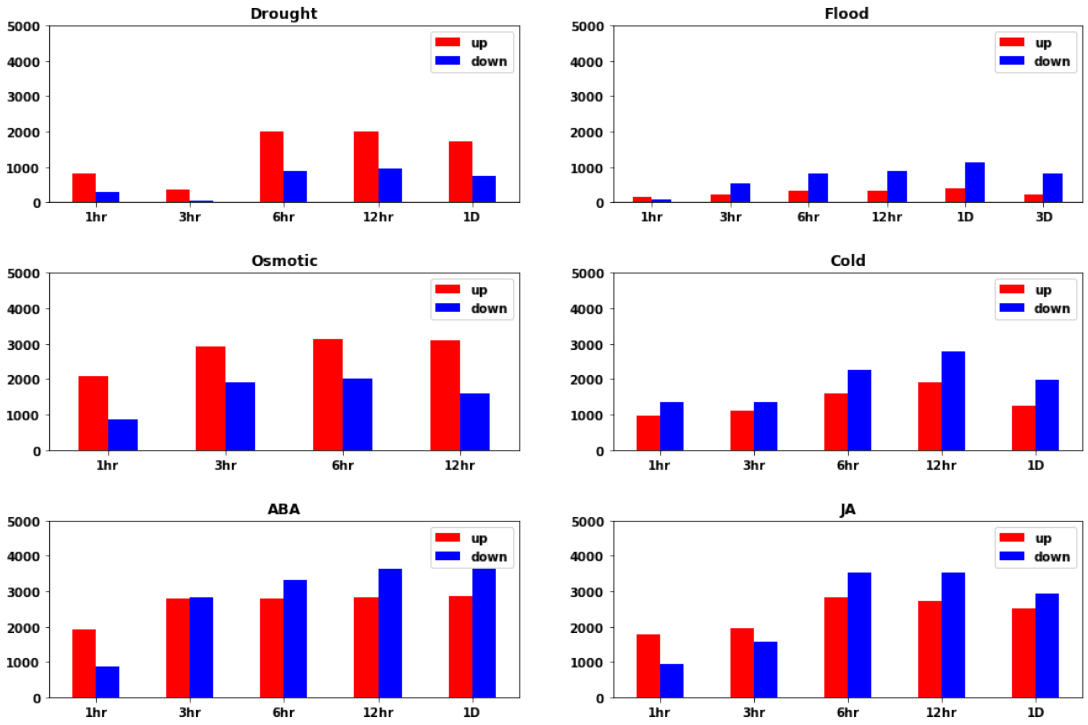


Figure 4.1 Number of up and down-regulated genes as a function of time is depicted for various stress/treatment conditions in root tissue. WGCNA Module Analysis

4.2 WGCNA Module Analysis

To understand the processes that are activated or repressed in different stress conditions we analyze the distribution of DEGs across co-expressed modules. These can point towards important processes/pathways that are affected in response to stress. To obtain co-expressed modules, we use Weighted Gene Co-expression Network Analysis (WGCNA) algorithm, the process of which is explained in **Section 2.5**. It results in a total of 22 co-expressed modules for the HRR100 Network of the root tissue. The distribution of DEGs across co-expressed modules under different conditions is tabulated in Table 4.1. The 'signed' network gives insight into how different coordinated biological processes represented by modules are expressed in different conditions. The differentially expressed genes in different time points is evaluated for each condition module-wise. The time point where the maximum percentage for DEGs are observed for a module is taken into consideration.

Cells with a considerable number of DEGs (>30%) are coloured. The red colour denotes the presence of more up-regulated genes while the blue colour denotes the presence of more down-regulated genes in that module. Cells are coloured grey if the difference in the number of up-regulated and down-regulated genes is not more than 20. Functionally Enriched GO terms from the PANTHER classification system are evaluated for each of the modules by submitting the corresponding set of genes in the GO Consortium website geneontology.org. Co-expressed modules are representative of

coordinated biological function; hence their functional enrichment tells us the potential function, pathways or biological process that the module is involved in.

Table 4.1 Co-expressed modules with %age DEGs across time-points in root tissue (h: hour, d: day).

Module	Drought	Flood	Osmotic	Cold	ABA	JA	GO
Black (521)	64.3% 12hr (335,0)	9.2% 1hr (48,0)	91.2% 3hr (475,0)	44.9% 1day (214,20)	51.6% 1day (260,9)	40.1% 6hr (165,44)	regulation of signal transduction, regulation of signaling, regulation of cell communication
Blue (1653)	18.6% 1day (244,64)	6.7% 3day (27,84)	50.3% 6hr (806,25)	37.7% 12hr (80,544)	64.9% 1day (1068,4)	54.6% 6hr (777,126)	late nucleophagy, autophagy of nucleus, autophagy of mitochondrion
Brown (1478)	28.3% 12hr (6,413)	18.9% 1day (178,102)	49.7% 3hr (3,731)	48.1% 12hr (32,679)	53.3% 1day (40,748)	31.1% 12hr (244,215)	arabinose metabolic process, L-arabinose metabolic process, cellular response to light stimulus
Cyan (341)	20.5% 1day (5,65)	5.0% 6hr (13,4)	15.5% 6hr (37,16)	89.4% 12hr (0,305)	17.6% 1day (8,52)	21.4% 1day (31,42)	chromatin remodeling, chromatin remodeling, chromatin organization
Green (807)	10.9% 12hr (7,81)	25.9% 3day (156,53)	46.7% 12hr (166,211)	37.4% 1day (282,20)	60.5% 12hr (4,484)	41.1% 6hr (52,280)	photosystem II assembly, photosystem II repair, NAD(P)H dehydrogenase complex assembly
Green Yellow (419)	65.6% 12hr (273,2)	15.3% 3day (10,54)	78.0% 1hr (327,0)	33.2% 12hr (102,37)	48.4% 1day (56,147)	40.1% 1day (69,99)	carotene metabolic process, hydrocarbon metabolic process, terpene metabolic process
Grey (963)	15.9% 1day (59,94)	14.0% 3day (32,103)	34.2% 12hr (235,94)	40.1% 12hr (110,276)	39.8% 1day (148,235)	29.2% 1day (128,153)	nucleosome positioning, negative regulation of DNA recombination, negative regulation of DNA metabolic process
Grey60 (263)	17.1% 1day (35,10)	35.4% 1day (75,18)	52.5% 6hr (137,1)	39.2% 1day (62,41)	31.2% 1day (64,18)	47.1% 12hr (93,31)	No significant terms
Light Cyan (274)	5.5% 12hr (9,6)	20.8% 1day (41,16)	37.2% 12hr (95,7)	20.1% 1day (38,17)	71.5% 1day (196,0)	68.2% 12hr (152,35)	autophagy of peroxisome, process utilizing autophagic mechanism, autophagy

Light Green (229)	5.2% 1day (1,11)	37.6% 1day (1,85)	27.1% 12hr (30,32)	55.0% 12hr (126,0)	16.6% 1day (19,19)	28.4% 1day (22,43)	mitochondrial electron transport, ubiquinol to cytochrome c, mitochondrial electron transport, cytochrome c to oxygen, mitochondrial ATP synthesis coupled electron transport
Light Yellow (216)	86.1% 12hr (186,0)	12.0% 1day (3,23)	75.0% 1hr (162,0)	37.5% 12hr (16,65)	33.8% 1day (52,21)	45.8% 6hr (75,24)	chorismate metabolic process, aromatic amino acid family biosynthetic process, dicarboxylic acid metabolic process
Magenta (510)	71.6% 12hr (363,2)	12.2% 3day (16,46)	75.9% 3hr (387,0)	36.9% 1day (85,103)	82.4% 3hr (420,0)	42.5% 12hr (116,101)	positive regulation of abscisic acid-activated signaling pathway, tryptophan biosynthetic process, indolalkylamine biosynthetic process
Midnight Blue (306)	51.6% 12hr (158,0)	7.2% 3day (9,13)	50.7% 1hr (154,1)	92.2% 12hr (282,0)	28.4% 1day (60,27)	21.6% 1day (34,32)	No significant terms
Pink (509)	27.7% 12hr (129,12)	13.9% 1day (10,61)	35.2% 12hr (126,53)	35.6% 12hr (159,22)	86.8% 12hr (442,0)	45.6% 12hr (154,78)	acetyl-CoA biosynthetic process from acetate, acetate metabolic process, citrate metabolic process
Purple (484)	14.3% 12hr (1,68)	9.7% 3day (29,18)	25.6% 3hr (1,123)	19.0% 1day (40,52)	11.8% 1day (38,19)	70.2% 6hr (76,264)	small nucleolar ribonucleoprotein complex assembly, box C/D snoRNP assembly, positive regulation of transcription by RNA polymerase I
Red (626)	34.0% 12hr (201,12)	10.9% 3day (23,45)	43.5% 12hr (206,66)	36.1% 12hr (35,191)	38.2% 1day (125,114)	76.2% 6hr (399,78)	purine nucleoside catabolic process, isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway, nucleobase-containing small molecule catabolic process
Royal Blue (206)	57.3% 6hr (118,0)	25.7% 12hr (4,49)	55.3% 1hr (113,1)	39.8% 1day (52,30)	47.1% 1day (10,87)	49.0% 6hr (70,31)	CDP biosynthetic process, xylan acetylation, CDP metabolic process

Salmon (343)	47.5% 12hr (0,163)	13.7% 1day (41,6)	28.9% 12hr (88,11)	84.8% 1day (0,291)	13.4% 1day (37,9)	15.7% 1day (24,30)	biological_process, organic substance metabolic process
Tan (360)	29.2% 1day (58,47)	12.8% 3day (12,34)	35.8% 3hr (52,77)	43.1% 12hr (19,136)	84.7% 6hr (305,0)	35.3% 12hr (47,80)	sulfur compound metabolic process, biological_process
Turquoise (2158)	12.6% 1day (57,215)	26.1% 1day (10,553)	39.3% 12hr (230,618)	26.9% 1day (234,347)	86.0% 1day (3,1853)	71.7% 12hr (275,1272)	syncytium formation, mitotic chromosome condensation, meiotic chromosome condensation
Yellow (1028)	9.4% 6hr (15,82)	20.5% 1day (9,202)	19.2% 12hr (91,106)	27.3% 12hr (256,25)	27.0% 1day (50,228)	72.9% 6hr (71,678)	positive regulation of translational fidelity, snRNA pseudouridine synthesis, ribosomal large subunit assembly

The number of up and down-regulated modules in each stress condition can be summarized as in **Table 4.2**. It is observed that the number of up-regulated modules always exceeds the number of down-regulated modules. For the flood stress condition, only one module is observed to be up-regulated (Grey60 - 35.4%) and one down-regulated (Light Green - 37.6%) with a very low percentage of DEGs. For drought and osmotic stresses, large number of up-regulated modules are observed in agreement with **Figure 4.1**. For cold stress and ABA, JA phytohormone treatment, the number of up and down-regulated modules are almost equal.

Table 4.2 Number of differentially expressed modules across different stress conditions in the root tissue.

Conditions	**Number of differentially expressed modules**	
	Up-regulated	**Down-regulated**
Drought	7	1
Flood	1	1
Osmosis	12	4
Cold	8	8
ABA	8	6
JA	9	6

The understand the time-profile of co-expressed modules, in **Table 4.3** the modules that enriched with up or down regulated genes at different time points under different stress conditions is summarized. This would help us identify early/late-stage processes that are activated or repressed in response to different abiotic stresses. Under drought and cold stresses, differentially expressed modules are observed only from 12hr time-point. On the other hand, Osmotic stress shows an early response, with the presence of a number of differentially expressed modules as early as 1hr time point. In response to ABA treatment, single differentially expressed modules are observed from 3hr time-point, while at the end of 1D, 5 modules are up-regulated and 5 down-regulated. While in response to JA treatment, 8 and 6 modules are differentially regulated 6hr and 12hr timepoints respectively.

Black and Magenta are modules central to regulation of signaling and contains the highest percentage of transcription factors (> 10%). Magenta module is highly enriched in the function positive regulation of abscisic acid-activated signaling pathway. Both these modules have high number of up regulated genes in case of drought (12hr) and osmotic (3hr) stress conditions. In the case for ABA hormone treatment, the Magenta module is observed to have high number of up-regulated genes at 3hr time point, while the Black module has its highest number of up-regulated genes at 1 day time point. For Cold and JA, while Black module shows distinctively high number of up-regulated genes, Magenta module has almost equal number of up and down-regulated genes. In all the cases except for cold, these modules have high number of up-regulated genes at the early timepoints (1hr-12hr) indicating their role in initiation of stress response.

Light Yellow, Royal Blue and Red modules, which are functionally enriched in metabolic processes like chorismate metabolism, CDP metabolism and purine nucleoside catabolism respectively, are observed to be highly up-regulated at almost the same timepoints in case of drought, osmotic and JA hormone treatment stress conditions. It has been reported that exogenous treatment of JA increases the antioxidative capacity of plants under drought stress by enhancing the activity of antioxidant enzymes (Nafie et al., 2011). Chorismate metabolism has also been reported to accumulate antioxidants by promoting salicylic acid response pathway (Jan et al., 2021). Hence, we observe these modules to take part in oxidative response for these conditions. Green Yellow module which is functionally enriched in carotene metabolism is up-regulated in the case of drought and osmotic stress condition. Carotene metabolism is the central step towards the synthesis of ABA for an ABA mediated response towards drought stress. Pink module enriched in acetate and citrate metabolic processes is up-regulated at the 12-hour time point for Cold, Osmotic, ABA and JA stress conditions. Acetate plays an important role in acetyl-CoA biosynthesis which participates in fatty acid metabolism as part of stress response (Lin & Oliver, 2008). Down-regulation of the Tan and Brown module in osmotic and cold conditions at the same time points, is indicative of the down regulation of their enriched metabolic functions, sulphur compound metabolism and arabinose metabolism. The demand for the uptake of sulphur is reduced in extreme condition and hence its metabolism is observed to be down-regulated in case of osmotic and cold conditions. However, in the case of ABA hormone treatment Tan module is highly up regulated at 6-hour time point signifying the importance of sulphur metabolism in its response. ABA has been reported to increase the levels

of mRNA encoding cytosolic O-acetylserine(thiol)lyase which is a key enzyme in Sulphur assimilation and metabolism (Kopriva, 2006).

Autophagy plays an important role in stress response for osmotic, ABA and JA hormone treatment, as is evident from the high number of up-regulated genes in Blue and Light Cyan modules, which are functionally enriched in processes like nucleophagy, autophagy of nucleus, mitochondrion and peroxisome. External stress leads to the formation of oxidative waste like reactive oxygen species (ROS) which needs to be removed or recycled from the system. This is achieved through autophagy which recycling of damaged molecules of cell organelles.

Table 4.3 Up/Down-regulated modules across different stress conditions time point wise in the root tissue.

		Drought	Flood	Osmotic	Cold	ABA	JA
1hr	Up	-	-	Green Yellow (78.0%), Light Yellow (75.0%), Royal Blue (55.3%), Midnight Blue (50.7%)	-	-	-
	Down	-	-	-	-	-	-
3hr	Up	-	-	Black (91.2%), Magenta (75.9%)	-	Magenta (82.4%)	-
	Down	-	-	Brown (49.7%), Tan (35.8%)	-	-	-
6hr	Up	Royal Blue (57.3%)	-	Grey60 (52.5%), Blue (50.3%)	-	Tan (84.7%)	Red (76.2%), Blue (54.6%), Royal Blue (49.0%), Light Yellow (45.8%), Black (40.1%)
	Down	-	-	-	-	-	Yellow (72.9%), Purple (70.2%), Green (41.1%)

12hr	**Up**	Light Yellow (86.1%), Magenta (71.6%), Green Yellow (65.6%), Black (64.3%), Midnight Blue (51.6%), Red (34.0%)	-	Red (43.5%), Light Cyan (37.2%), Pink (35.2%), Grey (34.2%)	Midnight Blue (92.2%), Light Green (55.0%), Pink (35.6%), Green Yellow (33.2%)	Pink (86.8%)	Light Cyan (68.2%), Grey60 (47.1%), Pink (45.6%), Brown (31.1%)
	Down	Salmon (47.5%)	-	Green (46.7%), Turquoise (39.3%)	Cyan (89.4%), Brown (48.1%), Tan (43.1%), Grey (40.1%), Blue (37.7%), Light Yellow (37.5%), Red (36.1%)	Green (60.5%)	Turquoise (71.7%), Tan (35.3%)
1day	**Up**	-	Grey60 (35.4%)	-	Black (44.9%), Royal Blue (39.8%), Grey60 (39.2%), Green (37.4%)	Light Cyan (71.5%), Blue (64.9%), Black (51.6%), Light Yellow (33.8%), Grey60 (31.2%)	-
	Down	-	Light Green (37.6%)	-	Salmon (84.8%)	Turquoise (86.0%), Brown (53.3%), Green Yellow (48.4%), Royal Blue (47.1%), Grey (39.8%)	Green Yellow (40.1%)

4.3 Metabolic Processes

Mapman analysis is performed to provide insights into various metabolic processes involved in stress response. Differentially expressed genes corresponding to top up and down regulated modules for each stress condition (as in **Table 4.2**) are mapped to Mapman bins. Mapman ontology consists of a set of 34 bins structured as a tree that may represent a metabolic pathway (e.g., glycolysis), a

cellular function (e.g. protein synthesis), biological response (e.g. stress) or even large protein families (e.g. cytochrome P450). The plots in **Figures 4.2 - 4.6** represent the percentage of differentially expressed genes that are mapped to a bin after normalization by the total number of reference genes of the respective bin. Some of the important processes affected are discussed below.

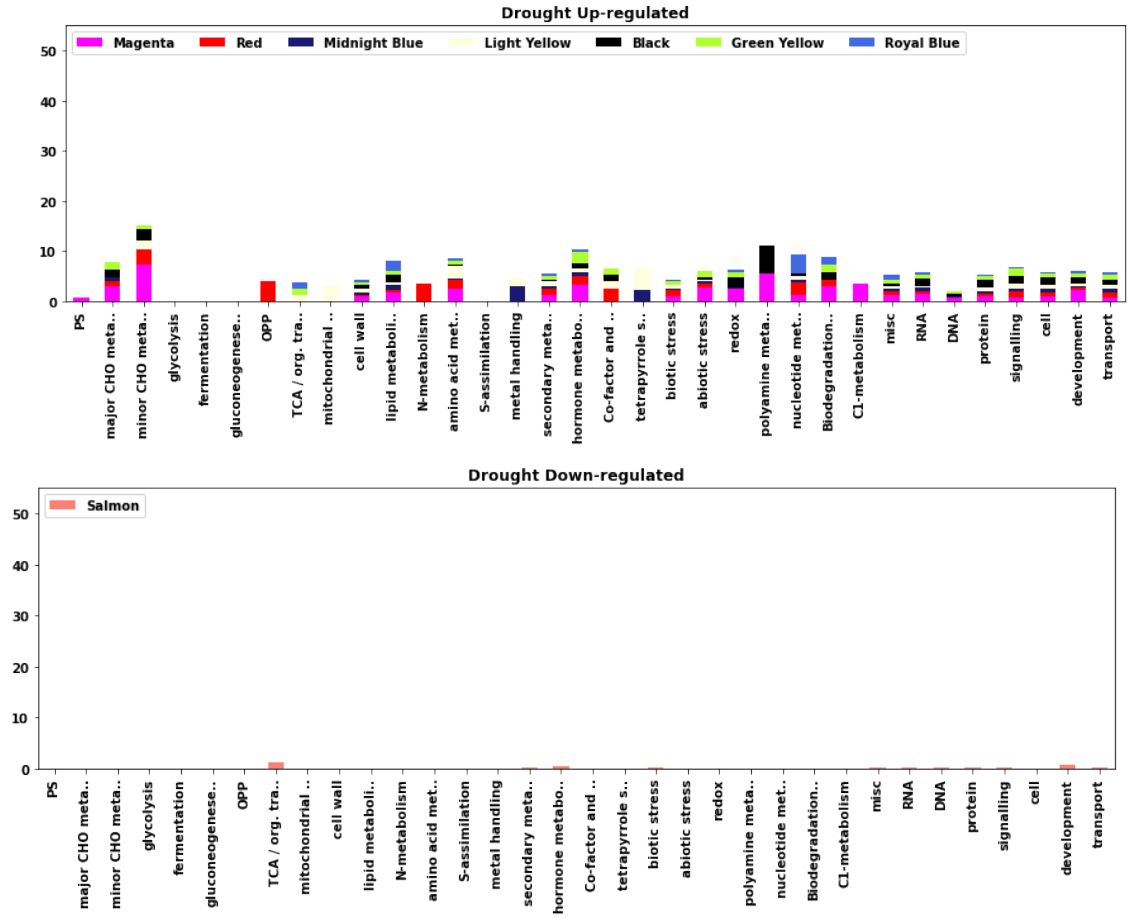


Figure 4.2 Mapman bin distribution for top modules involved in drought stress which are (a) up-regulated and (b) down-regulated in the root tissue. The y axis represents percentage of DEGs that have been mapped to a corresponding bin after normalization.

In case of drought, the up-regulated module distribution is almost evenly distributed across mapman bins (**Figure 4.2 a**). We observe that only bins corresponding to metabolisms like minor CHO, hormone, polyamine and nucleotide metabolism have more than 20% genes mapped to it. All the bins have a high number of differentially expressed genes present from the magenta module. None of the metabolic processes are observed to be downregulated except from Tricarboxylic Acid Cycle (TCA) in the Salmon module. The TCA cycle is responsible for release of stored energy through oxidation of acetyl-CoA (**Figure 4.2 b**).

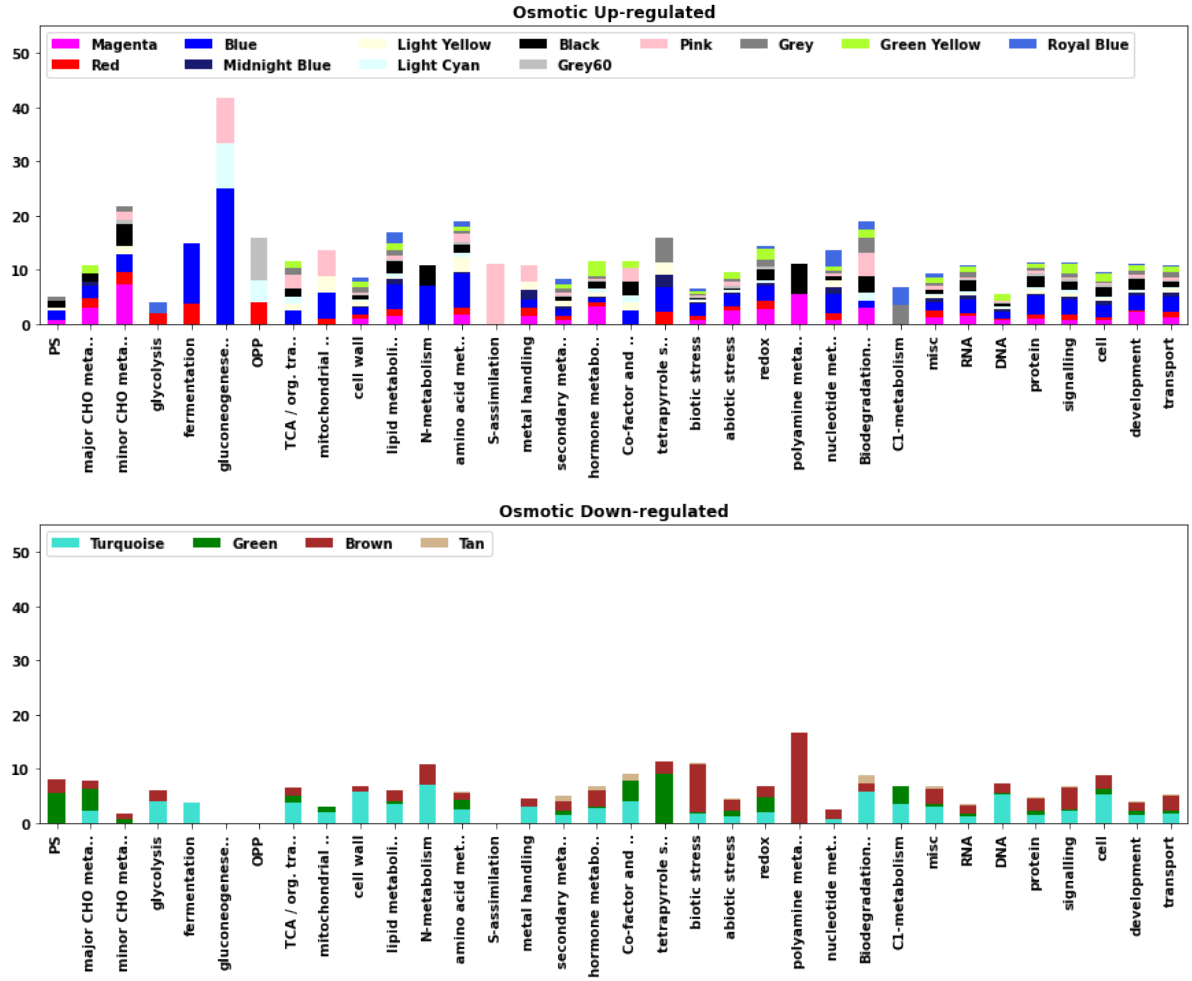


Figure 4.3 Mapman bin distribution for top modules involved in osmotic stress which are (a) up-regulated and (b) down-regulated in the root tissue.

A high fraction of up regulated genes (>40%) are observed to map to gluconeogenesis bin in the case of osmotic stress (**Figure 4.3 a**). Other bins have nearly uniform distribution, except for metabolism bins corresponding to minor CHO, lipid and amino acid metabolism which have more than 20% genes mapped to it. All the bins mapped from up-regulated genes have a high number of differentially expressed genes present from the Blue, Pink, Light Cyan and Magenta module. In the case of mapping of down-regulated genes, we observe the highest percentage mapped to the polyamine metabolism bin (~20%). The mapped down-regulated genes are mostly from the Turquoise module, however we also find genes from Brown Module to be in high percentage for the polyamine metabolism and biotic stress bins. (**Figure 4.3 b**).

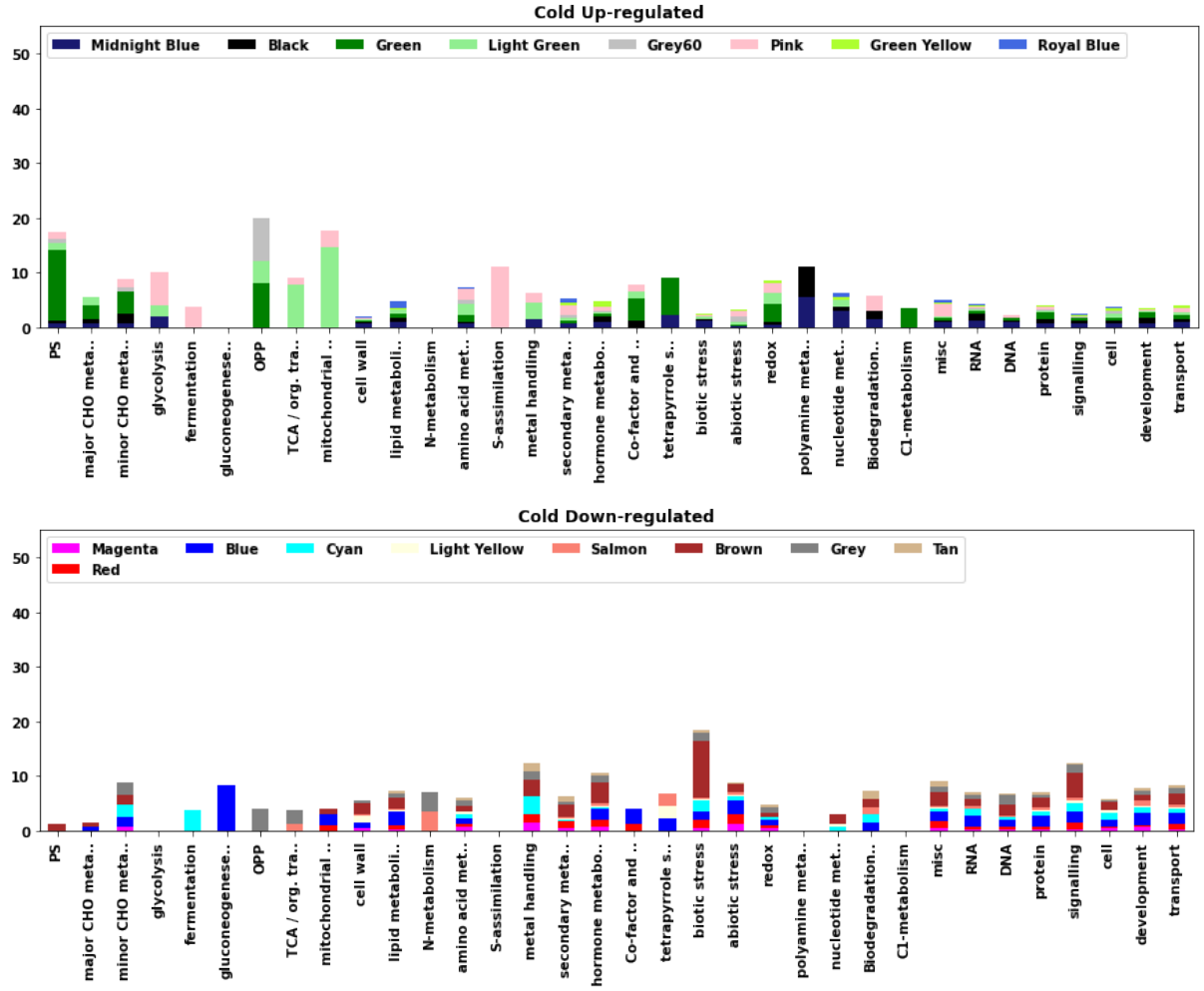


Figure 4.4 Mapman bin distribution for top modules involved in cold stress which are (a) up-regulated and (b) down-regulated in the root tissue.

Distribution of up-regulated genes under cold stress across MapMan bins is shown in **Figure 4.4a**. We observe high percentage (~20%) of DEGs mapped to the bins Photosynthesis (PS), Oxidative Pentose Phosphate (OPP) cycle and Mitochondrial electron transport. While the genes belonging to the PS bin are predominantly from the Green module, we observe the presence of genes from Light Green module in the Mitochondrial electron transport bin. Genes involved in processes involved in biotic stress response, signaling and metal handling are observed to be down-regulated under cold stress in root (**Figure 4.4b**). The highest percentage of down-regulated genes are mapped to biotic stress bin, the genes of which predominantly belong to the Brown module.

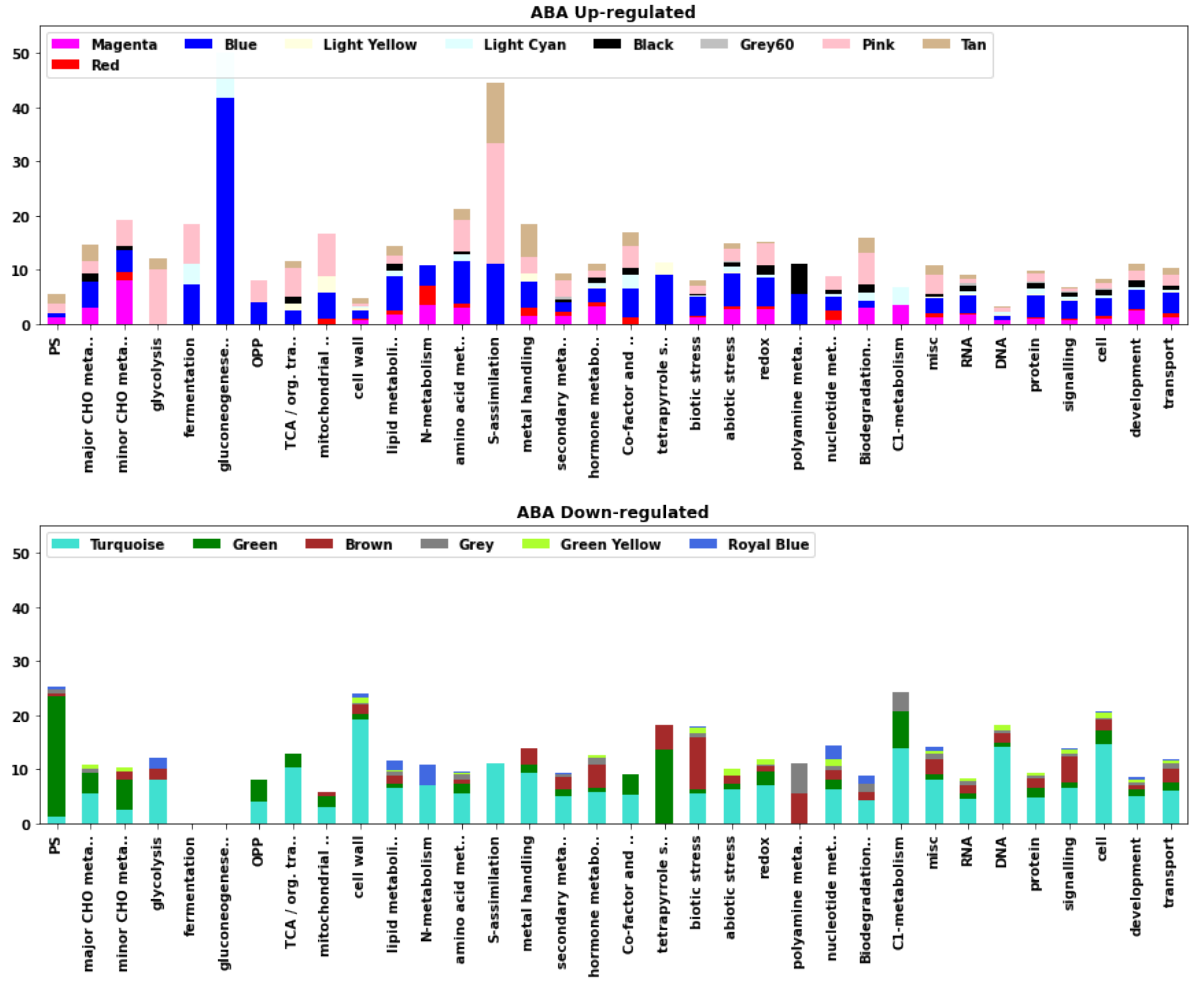


Figure 4.5 Mapman bin distribution for top modules involved in ABA hormone treatment stress which are (a) up-regulated and (b) down-regulated in the root tissue.

Both the hormone treatment conditions have a similar MapMan distribution for up and down regulated module genes. Gluconeogenesis and S-assimilation have a high percentage presence (~50% in ABA and ~35% in JA) for up-regulated genes (**Figure 4.5a & 4.6a**). Both these bins contain genes predominantly from the Blue and Pink module. Polyamine metabolism also has a high percentage (>20%), but only in the case for JA hormone treatment condition. For down-regulated genes, C1 metabolism bin has a high percentage presence in both the hormone treatment (**Figure 4.5b & 4.6b**). Photosynthesis and cell wall bins also have a similar percentage in the case for ABA down-regulated genes. In the case for JA, S-assimilation and nucleotide metabolism bins have a high percentage of mapped down-regulated genes. The mapped down regulated genes are mostly from the Turquoise module.

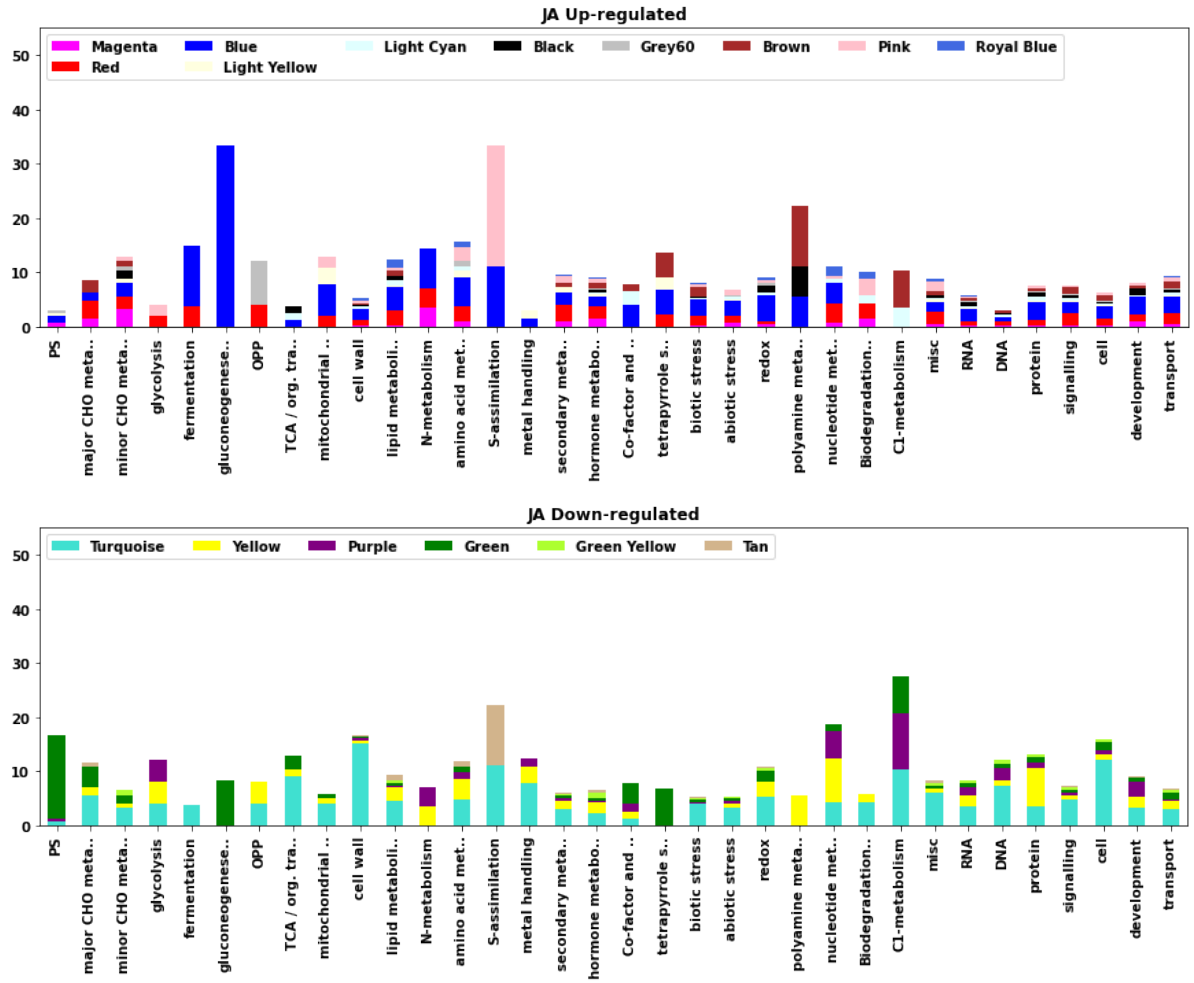


Figure 4.6 Mapman bin distribution for top modules involved in JA hormone treatment stress which are (a) up-regulated and (b) down-regulated in the root tissue.

In all the stress conditions, except for JA hormone treatment, the abiotic stress bin has a higher percentage than the biotic stress bin for up-regulated genes, and a reverse trend is observed for down-regulated genes. Gluconeogenesis bin has an exceptionally high percentage in the case of Osmotic stress, JA and ABA hormone treatment conditions for up-regulation. Gluconeogenesis is a process involving the synthesis of glucose from non-carbohydrate carbon substrates like lactate, amino acids and glycerol. In all the three stress conditions (osmotic, ABA and JA), an external agent is applied resulting in an initial osmotic shock at the root. Gluconeogenesis can be associated to be involved in mechanism in response to this shock. It has been observed that salt stress induces the accumulation of sucrose, glucose and fructose, through the gluconeogenesis pathway in tobacco plant (Zhang et al., 2011). Photosynthesis is predominantly down-regulated in Osmotic stress, JA and ABA hormone treatment condition as observed from the higher percentage in its bin for down-regulated gene, while an opposite trend is observed in Cold stress. This photosynthesis process is essential to plant to generate oxygen and carbohydrates like sugar and starch as a form of chemical energy to fuel the

organism's activities. In case of Osmotic, JA and ABA treatment condition, this mechanism is shut down to counter the accumulation of reactive oxygen species (ROS) and avoid the loss of water through stomata. The required glucose levels are maintained through gluconeogenesis. However in case of Cold stress, the immediate effect is not perceived at the root tissue and it increases its photosynthetic processes to prepare for the extreme conditions.

4.4 Osmotic Stress Analysis

A plant experiences osmotic stress when there is a significant change in water potential in its environment, leading to the disruption of normal cellular activities that may even result in plant death. Several other stresses like drought, high salinity and cold, results in the imbalance of water potentials in a cell and impose osmotic stress. In drought stress, the shortage of water in the soil results in the root cell to experience an osmotic shock due to the increased internal pressure caused by the difference in solute concentration. Similarly salt stress also results in an imbalance of ionic concentration across cell membranes disrupting ionic and osmotic homeostasis. Drought and salinity are the major stresses in the realm of agriculture which are responsible for limiting plant growth and productivity. Hence, an understanding in the mechanisms of osmotic stress can provide insights into development of stress tolerant species.

For the root tissue, we study the up and down regulated WGCNA modules for osmotic stress in a time-point wise manner. **Table 4.4** is an extract from table 4.5 for the osmotic stress condition. For the drought stress condition, we observe that the WGCNA modules are differentially expressed at almost a single timepoint (Royal Blue up-regulated at 6 hour; Light Yellow, Magenta, Green Yellow, Black, Midnight Blue, Red up-regulated at 12 hour), making it difficult to make a timepoint based analysis of processes. For the osmotic stress we observe a similar set of WGCNA modules, however they are differentially expressed in a broad range of timepoints (1hour till 12 hour). The GO enriched functions of these modules present in **Table 4.1** can be referenced alongside their time-point based differential regulation (**Table 4.4**) to get a potential sequence of biological processes that are triggered as a response to osmotic stress.

Table 4.4 Up/Down-regulated modules across osmotic stress conditions time point wise in the root tissue.

1hr		3hr		6hr		12hr	
Up	**Down**	**Up**	**Down**	**Up**	**Down**	**Up**	**Down**
Green Yellow (78.0%), Light Yellow (75.0%), Royal Blue (55.3%),	-	Black (91.2%), Magenta (75.9%)	Brown (49.7%), Tan (35.8%)	Grey60 (52.5%), Blue (50.3%)	-	Red (43.5%), Light Cyan (37.2%), Pink (35.2%), Grey (34.2%)	Green (46.7%), Turquoise (39.3%)

Midnight Blue (50.7%)							

Since very little is known about defense mechanism pathways activated/repressed prior to the abscisic acid signaling, analysis of the early response modules in response to osmotic stress may offer insights towards some early regulatory mechanisms. For this, 756 genes up-regulated at 1-hour time point from the four modules (GY, LY, RB and MB) are evaluated. These genes may possibly be involved in osmotic stress tolerance or take part in the relay of stress signals to the corresponding stress tolerant genes. An integrative approach combining co-expression networks with protein-protein interaction networks and metabolic pathway information can enhance the analysis and help in linking genotype to phenotype. Genes which have a high degree in the protein-protein interaction (PPI) network may point towards its significance in regulating the stress. Hence a PPI network of these 756 genes is constructed using the STRING database (Szklarczyk et al., 2019). The PPI network consisting of 740 nodes and 883 edges with a confidence score ≥ 0.4 and p-value < 1e-16 is obtained. To identify biological processes/pathways associated with these genes, we identify gene clusters (> 5 nodes) using the Markov Cluster algorithm (MCL) with an inflation value of 3. MCL algorithm is an unsupervised graph (network) clustering algorithm whose objective is to keep highly connected nodes in one cluster which weakly connected nodes in different cluster. It is based on simulation of stochastic flows in graph wherein it alternates between an expansion step and an inflation step until an equilibrium is reached. A total of 14 clusters are obtained whose size, module membership and functional enrichment details are given in **Table 4.5**. The functional relevance of a few of these clusters is briefly discussed below.

Table 4.5 Biological processes associated with the early responsive up-regulated gene clusters in response to Osmotic stress.

Cluster	Associated Biological Process	Total Genes	GY	LY	RB	MB
U1	Phosphate metabolic process, Intracellular protein transport, Response to abiotic stimulus, Chromatin organization	62	29	14	6	13
U2	Jasmonic Acid Signaling Pathway	25	7	3	10	5
U3	Glycerophospholipid metabolism	18	9	2	6	1
U4	Aromatic amino acid family biosynthetic process	15	0	9	4	2
U5	Alpha-amino acid metabolic process	14	7	3	3	1
U6	Phenylpropanoid biosynthesis	14	3	6	3	2
U7	Amino sugar and nucleotide sugar metabolism	10	2	6	2	0
U8	Polysaccharide biosynthetic process	10	4	0	4	2

U9	-	9	5	0	2	2
U10	Valine, leucine and isoleucine degradation	7	0	5	1	1
U11	Hormone activity and di-glucose binding with ER	6	4	1	1	0
U12	Terpenoid metabolic process	6	5	0	0	1
U13	-	6	2	0	1	1
U14	Endocytosis	6	3	1	2	0

Cluster U1: Response to stimulus, MAPK Cascade, Protein transport, Chromatin organization

Any environmental stimulus transmitted as signals into the cellular machinery is identified by receptors, and secondary signals are triggered so that an appropriate response may be generated. Mitogen Activated Protein Kinase (MAPK) cascades are signaling pathways that are central to regulation of a variety of cellular stimulated processes like proliferation, differentiation, apoptosis and stress response. The signals are transmitted via a cascade effect from the cellular wall to the nucleus where they modulate different processes like the activity of transcription factors, chromatin remodeling and protein transport.

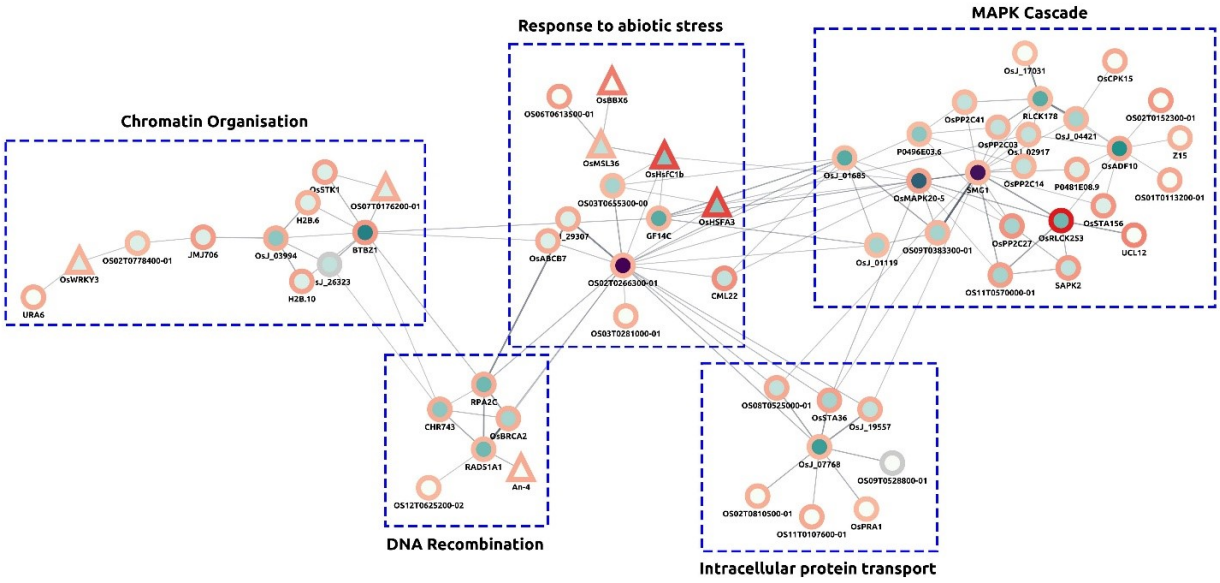


Figure 4.7 PPI network of cluster U1. The nodes are colored proportional to their degree (darker shade represent high degree). Border color represents fold-change (darker shade represents higher fold-change value). Triangular nodes indicate transcription factors.

Cluster U1, consisting of 62 genes with 119 interactions **(Figure 4.7)**, clearly captures MAPK signaling and secondary signal response interactions. Four subgroups based on their enriched functions is depicted in the figure. We observe the largest subcluster of genes is enriched in

“Phosphate containing compound metabolic process”. It majorly comprises kinases and phosphatases belonging to the families MAPK, Stress Activated Protein Kinase SAPK, Receptor-like Kinase RLK and Protein Phosphate 2C PP2Cs. The role of SAPK and PP2Cs in ABA mediated stress response is well known. However, the PYL-PP2C-SnRK2 core formed in the ABA signaling pathway is observed to regulate the MAPK cascade comprising of the MAP3Ks MAP3K17/18, the MAP2K MKK3 and the MAPKs MPK1/2/7/14 (de Zelicourt et al., 2016). While PP2Cs are suppressed in the ABA mediated pathway, few PP2Cs (like BpPP2C1 in Betula platyphylla and *ZmPP2C-A in maize*) have been observed to be up-regulated in response to salt stress, as they promote autophagy (Memisoglu et al., 2019), thereby improving the antioxidant defense system (He et al., 2019; Xing et al., 2021). RLKs are plasma membrane-localized signaling molecules that perceive endogenous and exogenous signals via sequential phosphorylation and function as upstream of MAPKs in regulating plant development (Z. Wang & Gou, 2020). The subcluster consists of 4 kinases that are part of the MAPK signaling pathway in plants, OsMAPK20-5/MPK7, SMG1/MKK4, SAPK2 and OS09T0383300-01 wherein they transmit the environmental signal into the nucleus of the cell. We observe a direct connection between the MAP kinase OsMAPK20-5 with two heat shock transcription factors HSF11 and HSFA3, which are a part of the subcluster enriched with “Response to abiotic stress”. These heat shock protein factors are transcriptional regulators that bind to heat shock protein elements (HSEs). Environmental stress disrupts protein folding and results in an increase in reactive oxygen species (ROS) or induces oxidative stress in the cell (Hu et al., 2009). The heat shock proteins act as molecular chaperones which prevent protein misfolding and help in promoting cell survival in adverse conditions. Incidentally these two transcription factors report one of the highest fold change values in this cluster.

MAPK cascade involves transportation of macromolecules inside the nucleus via nuclear pore complexes (NPCs). The function of such intracellular proteins has been captured in one of the subclusters. The DNA is packed into chromatin wherein the double stranded DNA is coiled around a core of histone proteins. These are dense structures which make it inaccessible to other proteins like transcription factors. Hence a “decompaction” of the chromatin is required for transcriptional activity. Consequently, we observe a subcluster of chromatin organization and DNA recombination which considers chromatin remodeling in order to make target genes accessible. This process includes histone acetylation, phosphorylation, DNA conformational change and poly ADP ribosylation.

A high number of signaling genes (13) are observed in this cluster. These involve 2 calcium and 7 G protein signaling genes. Together they aid in transmitting signal from outside stimuli to the interior of a cell. The cluster is rich in transcription factors (TFs) with a total of 7 transcription factors present in this cluster. The KEGG enrichment mapping assigns 5 genes to be directly involved in the plant-pathogen interaction pathway (CML22, RLCK178, OsCPK15, MKK4 and Os01g0899000). Calcium signaling triggered by abiotic stress induces the proteins CML22 and OsCPK15 which initiates the plant-pathogen interaction pathway. These genes in turn induce defense-related genes like NHO1 and PR1 (PR1-72, PR1-73, PR1-101) which have been observed to be up-regulated in

osmotic stress at a later timepoint. Various transcription factors in other plants have been reported to enhance pathogenesis related genes like PR1 in response to abiotic stress (Liu et al., 2022; W. X. Liu et al., 2013). Abiotic stresses have been shown to have a pronounced effect on plant-pathogen interactions (Zarattini et al., 2021). Plant pathogen interactions have been shown to be influenced by abiotic stress via factors like plant metabolism, cell viability, signaling (Kissoudis et al., 2014) and transcriptomic regulations. There is multi-faceted crosstalk between the signaling pathways involved in abiotic and biotic stress responses. Several pathogenesis-related proteins have been shown to be differentially expressed following drought and osmotic stress (Haider et al., 2017; Hatmi et al., 2014). For example, 11 out of 35 WRKY genes involved in *Fusarium udum* infection in pigeonpea were also induced by salt stress (Kumar et al., 2019).

The gene with the highest degree (18) in the cluster, OS02T0266300-01, is associated with two HSFs, 5 genes in abiotic stress response, 4 genes of the MAPK cascade, 4 genes involved in intracellular protein transport and 3 in DNA recombination. This is suggestive of its role in being a connecting factor in regulating these processes and playing an important role in osmotic stress response. The second highest degree node is SMG1 belonging to the MAPK Cascade subcluster and is connected to 17 genes. It has been reported to influence the grain size in rice and is involved in arsenic and wounding stress response (Duan et al., 2014; Rao et al., 2011; Yoo et al., 2014). The gene with the highest fold change is OsRLCK253 with a fold change value of ~10.2. It is a receptor-like cytoplasmic kinase that has been reported to interact via A20 zinc-finger and improve water-deficit and salt tolerance in transgenic Arabidopsis plants (Giri et al., 2011).

Thus, we observe this cluster captures the mechanisms that are activated at the very early time-point. It involves the perception of the stress signals by the RLKs and the transfer of signals from the cell wall to the nucleus via the MAPK cascade. This is followed by the response to stress aided by the activation of transcription factors like the heat shock proteins, which requires the remodeling of chromatin to make it more accessible. Signaling genes as well as genes involved in plant-pathogen interactions are also seen to be activated, which aid in countering the initial hyperosmotic shock.

Cluster U2: Jasmonic Acid Signaling Pathway

Jasmonic acid (JA) plays an integral role in abiotic stress response by regulating stress response pathways like the JA signal transduction pathway, ABA signal transduction pathway and the oxylipin pathways (Genva et al., 2018). The genes involved in JA biosynthesis are hence dramatically up-regulated when the plant is exposed to environmental stress. JA signaling components primarily involve two gene families, the JA receptor Coranatine Insensitive 1 (COI1) and the JA signaling repressors Jasmonate ZIM-Domain proteins (JAZs). JA is metabolized into its active form JA isoleucine (JA-Ile) which interacts with the receptor COI1. This initiates proteosomal degradation of JAZs. MYC2 transcription factor which is otherwise suppressed with its interaction with JAZ proteins, is now released and leads to the activation of JA-responsive genes (Chini et al., 2007).

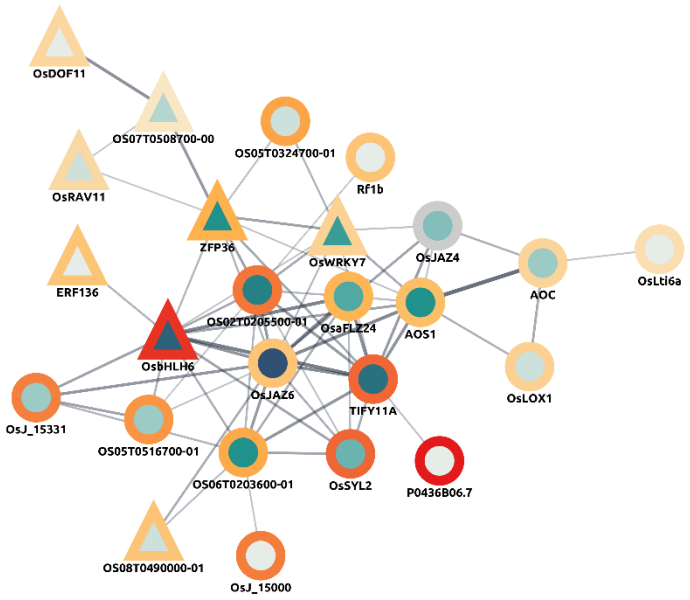


Figure 4.8 PPI network of cluster U2. The nodes are colored proportional to their degree (darker shade represent high degree). Border color represents fold change (darker shade represents higher fold change value). Triangular nodes indicate transcription factors.

Cluster U2, consisting of 25 genes with 64 interactions (**Figure 4.8**), is enriched with genes involved in the jasmonic acid signaling pathway and stress response. A STRING cluster of 15 genes which is annotated as jasmonic acid signaling pathway and stress response is observed at the core of this cluster. KEGG mapping also annotates 3 JAZ genes (JAZ4, JAZ6 and JAZ9) to the plant hormone signal transduction pathway which undergoes ubiquitin mediated proteolysis in the jasmonic acid signaling pathway. We also observe the presence of genes containing zinc finger domains like ZFP36 which fall in the JAZ family. Incidentally, their fold change is observed to decrease with time which is justified as they act as JA signaling repressors which are degraded with the action of JA-Ile during stress response. A jasmonic acid biosynthesis gene, CYP74A1 is also observed, which results in the increase in the concentration of JA.

This cluster is rich in transcription factors with a total of 10 genes out of 25 annotated as TFs. These TFs are mainly dominated by two families, the basic Helix-Loop-Helix family (OsbHLH6 and Os08g0490000) and the zinc finger family (OsDOF11, ZFP36, Os07g0508900 and JAZ9). Being rich in TFs, this cluster plays an integral role in regulation of stress response. ZFP36, belonging to the C2H2 zinc finger transcription factor family, has been recently reported to be a key player in ABA induced stress response for oxidative stress tolerance in rice (Zhang et al., 2014). It has direct connections with 4 other transcription factors OsRAV11, OsWRKY7, OsbHLH6 and OS07T0508700-00 suggesting its role in being a master regulator. An important transcription factor having both a high degree (12) as well as a high fold change (~9.7) is OsbHLH6. Also known as RERJ1, it is a jasmonic acid responsive gene that has been reported to be involved in wounding and drought stress response in rice (Kiribuchi et al., 2004, 2005). The high fold change and high degree of both ZFP36 and bHLH6 indicates their important role in early osmotic stress response.

Cluster U3: Glycerophospholipid metabolism

Lipid signaling plays a vital role in abiotic stress response in plants. Being one of the major components of plasma membrane it serves as the interface between the cell and its environment. The membrane lipids also act as substrates for various signaling lipids like phosphatidic acid, sphingolipids, phosphoinositide, triacylglycerol, and others. The synthesis and metabolism of these signaling lipids are influenced by abiotic stress signals and take part in plant adaptation process. In osmotic stress signaling pathway, abscisic acid (ABA) signaling plays a significant role. The SNF1-related protein kinase 2 (SnRK2) family of proteins are vital intermediates in the ABA signal transduction pathway as they mediate the phosphorylation of ABA binding factors (ABFs) which in turn promotes the expression of ABA responsive genes (ABREs). The osmotic stress induced activity of SnRK2 family (SnRK2.4 and SnRK2.10) is localized to the cellular membranes through their interaction with lipid like Phosphatidic Acid (PA) (McLoughlin et al., 2012).

It was shown by Yu and Li that in Arabidopsis thaliana subjected to drought stress, with the increase in concentration of plastid lipids and double bond index, the fluidity of membranes improved thereby increasing water stress tolerance (B. Yu & Li, 2014). Increased rates of storage lipids like Triacylglycerol were an essential adaptive response to high temperature stress in plants (Mueller et al., 2015). Phosphatidylserine and Triacylglycerol have also been shown to be involved in salt stress response in sweet potato leaves (Yu et al., 2019).

Cluster U3 with 18 genes and 50 interactions (**Figure 4.9**) is enriched with genes involved in glycerophospholipid metabolism. Annotation obtained from KEGG mapping reported 7 genes involved in glycerolipid metabolism and 6 genes involved in glycerophospholipid metabolism. Observing the pathways of the mapped genes, they are involved in biosynthesis of lipids like triacylglycerol, diacylglycerol phosphate, phosphatidyl glycerol, cardiolipin and inositol phosphate. These lipids take part in stress responsive signaling cascade. For example, triacylglycerol serves to sequester toxic lipid intermediates that get formed due to lipid remodeling induced due to stress. Triacylglycerol is enclosed by lipid droplets that serve as binding sites and substrate for synthesis of other stress responsive compounds.

The high degree genes of this cluster are OsJ_02930 (12) and OS05T0502200-01 (11). OsJ_02930 has been reported in STRING to encode a phosphatidic acid phosphatase beta-like protein which plays major role in lipid homeostasis. OS05T0502200-01 has been reported in STRING to encode a putative 1-acylglycerol-3-phosphate acyltransferase protein that serves as intermediate enzymes involved in the biosynthesis pathways of glycerophospholipids and triacylglycerol. We also observe two genes from the MGD family (OsMGD2 and MGD1) and 2 genes from the PLD family (OsPLDα5 and OsPLDα4) are part of the core of this cluster. Previous studies have demonstrated that OsMGD2 plays an important role in grain quality while also helping in plant growth and development. Increased accumulation of OsMGD was observed by ethephon, gibberellin, drought, and salt treatment (Qi et al., 2004). Phospholipase D (PLD) in plants hydrolyzes phospholipids to generate phosphatidic acid (PA) which is an important secondary messenger

mediating the generation of ROS and activating MAPK cascade (Yamaguchi et al., 2009). Hence, PLD genes play a significant role in abiotic and biotic stress response.

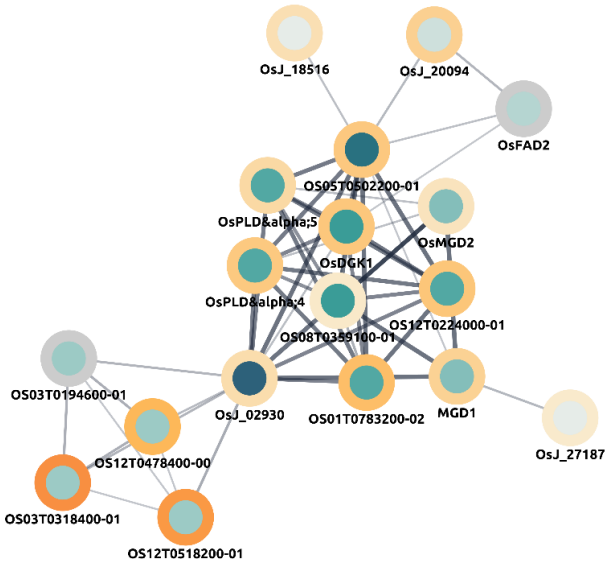


Figure 4.9 PPI network of cluster U3. The nodes are colored proportional to their degree (darker shade represent high degree). The border color represents fold change (darker shade represents higher fold change value).

Early responsive down regulated modules

From **Table 4.4**, we observe that the early down-regulated modules observed under Osmotic stress are Brown and Tan modules with about 49.7% and 35.8% of their genes down-regulated at the 3hr time-point. These modules are enriched in metabolic processes like arabinose metabolism and sulphur compound metabolism. At 12hr time-point, Green and Turquoise modules are observed to be down-regulated with 46.7% and 39.3% of their genes down-regulated. Their enriched functions include syncytium formation, chromosome condensation and photosystem II assembly and repair.

We analyzed the genes of the early responsive down-regulated modules Brown and Tan to understand the biological processes that are repressed in response to osmotic stress. There are 731 genes down-regulated in Brown module and 77 genes down-regulated in Tan module. The expression profile of the down-regulated genes in these modules are plotted in **Figure 4.10**.

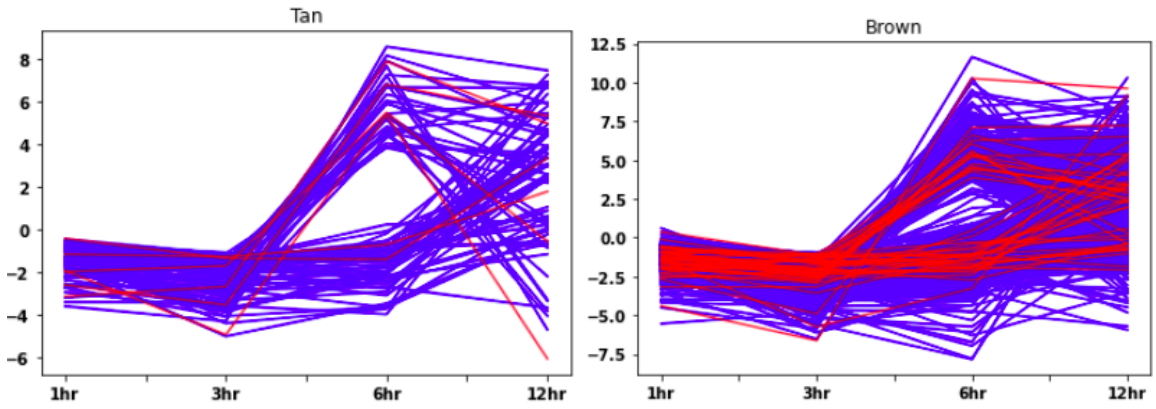


Figure 4.10 Expression Profiles of the down regulated genes at the 3-hour timepoint in the Brown and Tan modules for osmotic stress condition (Red indicates TF, Blue indicates other genes).

The Brown module has a higher number of transcription factors compared to the Tan module. In both the modules, two patterns of expression profiles are observed. There are a set of genes which exhibit a sudden increase in their fold-change values at 6hr time-point followed by a dip at 12hr time-point (we refer to these as Down Regulated Genes from 0-3hr, or, DRG0-3). Another set of genes observed exhibit a gradual increase in their fold-change values. These are down-regulated from 0-6hr and exhibit an increase in their fold-change at 12hr time-point (we refer to these as Down Regulated Genes from 0-6hr DRG0-6). These genes can be viewed as genes which are late responsive as they get upregulated with a noticeably high fold change (2.5 – 12 $\log_2$FC) at late timepoints (6 - 12 hour) in spite of being downregulated at the earlier timepoints.

We investigate the function of these two sets of genes as their respective genes are likely to have common transcriptional regulation. The first set, DRG0-3, consists of 268 genes and the other set, DRG0-6, consists of 540 genes. Protein-protein interaction networks are constructed for these two sets of genes using STRING DB application. This network establishes a relationship between co-expressed set of genes based on experimental data, computational predicted methods and public texts and provides for a more reliable analysis.

Analysis of DRG0-3 PPI

The set of 268 genes corresponding to DRG0-3 were submitted to STRING DB to infer their PPI network. A total of 266 genes mapped to their corresponding proteins in STRING DB and resulted in a PPI network of 266 nodes and 40 edges. The edge confidence is set to a score ≥ 0.4 and the network has a PPI enrichment p-value of 4.13e-05. The network is imported in Cytoscape for further analysis. The MCL algorithm (with inflation value = 4) is used to identify closely connected group of genes in this PPI network. Five clusters with size greater than 2 are obtained. The associated biological processes for these clusters and the module membership of their genes have been listed in **Table 4.6**.

Table 4.6 Biological processes of the early responsive down regulated gene clusters pertaining to DRG0-3 in Osmotic stress.

Cluster	Associated Biological Process	Total Genes	Brown	Tan
D1	Leucine-rich repeat	10	8	2
D2	Oxidation reduction process, alpha-amylase inhibitors, lipid transfer	7	1	6
D3	Lipid and amino acid metabolism	5	4	1
D4	Oxidation reduction process, Monooxygenase activity	4	3	1
D5	Protein phosphorylation	3	2	1

Cluster D1: Actin Organization and Panicle Development/ Leucine-rich repeat (LRR)

Actin filaments and microtubules are an integral part of the plant cytoskeleton and play crucial role in cell growth and developmental processes like cell division, motility and intracellular organization. Panicles form the plant inflorescence and are important for initiating the reproductive phase of rice development. Exposure to abiotic stresses results in limited resources for the plant hampering growth and development by stopping processes like actin organization and panicle development (Wang et al., 2011; Wei et al., 2017). Cluster D1 having 10 genes and 10 interactions (**Figure 4.11**) has a central hub gene OsSCAR2 directly correlated with 7 of the 9 genes in this cluster. OsSCAR2 encodes a SCAR/WAVE domain and is associated with TUT1 which shows a pleiotropic phenotype characterized by short roots, reduced plant height, and development of pollen grains and anthers. TUT1 is a functional SCAR protein that has been reported to show an important role in panicle development (Bai et al., 2015). At the earlier timepoints, it is down-regulated indicating that defense measure like hampering plant growth is not turned on. Only after the 6-hour timepoint these measures become active to counter the abiotic stress.

The cluster D1 is also enriched with Leucine-rich repeat genes. Leucine rich repeat mainly constitutes receptor like kinases (RLKs) which are potential cell-wall sensors. They play the role of key regulators which perceive and process external stimuli to cellular signaling molecules. The extracellular signals are perceived through the LRR domain and are transmitted via the serine/threonine domains. Several LRR-RLKs have been identified to be involved in abiotic stress responses in plants. For example, the Srlk gene in *Medicago truncatula* was shown to improve salt tolerance in plants by accumulating fewer Na^+ ions (de Lorenzo et al., 2009). The OsSIK1 gene in rice, which has an LRR domain, activated the antioxidative system and improved salt and drought tolerance (Ouyang et al., 2010). However, there also exists certain LRRs like the SIF1 and SIF2 that negatively regulate plant salt resistance (Yuan et al., 2018). The LRR-RLKs function by improving the activities of ROS scavenger, thereby reducing the levels of malondialhedyde (MDA) and ROS. The early down regulaton and the late upregulation of the genes in this cluster is suggestive of the importance of the presence of ROS for stress signaling, as they take part in systemic signaling

pathway and transfer signals across the plant. However, with the increase in the ROS levels, the LRR proteins become up-regulated and aid in oxidative homeostasis.

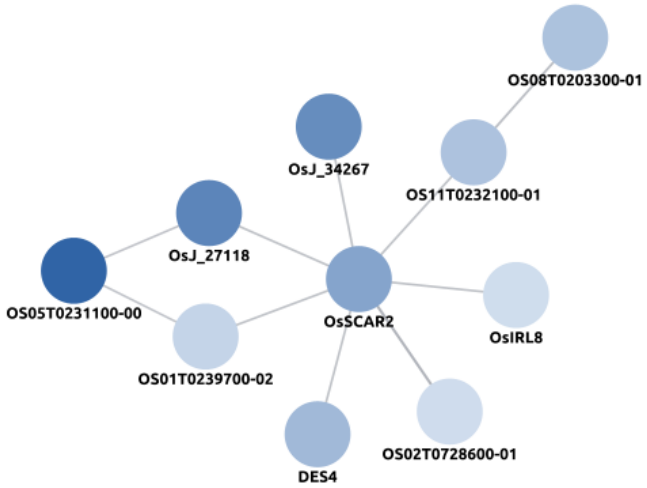


Figure 4.11 . PPI Network of Cluster D1. The nodes are colored proportional to their fold change (the darker the shade, the lower is the fold change).

Other Clusters

The other clusters obtained had a size lower than 8 and did not have a highly distinctive enriched function. GO and KEGG annotations of a few genes in the clusters indicated functions like oxidation-reduction process, alpha-amylase inhibitors, lipid metabolism and protein phosphorylation. Due to accumulation of reactive oxygen species as a result of hyperosmotic shock, to maintain oxidative homeostasis, oxidation reduction processes are necessary but should act at late timepoints so as to not disrupt the oxidative signaling. Alpha-amylase inhibitors have been reported to be differentially expressed in response to biotic and abiotic stresses in *Amaranthus hypochondriacus* (Sánchez-Hernández et al., 2004). They act as defense systems against pathogens and are observed to be up-regulated with exogenous treatment of MeJA or ABA, suggesting their involvement in ABA and JA signaling pathways. The role of lipid metabolism in abiotic stress has been discussed earlier in the 1-hour timepoint up regulated cluster (U3), however here we observe the small lipid metabolism cluster (D3) being down-regulated at the 3-hour timepoint followed by up-regulation at the 6-hour timepoint. The central gene annotated for lipid metabolism in this cluster is OsLCB2a1. OsLCB2a1 has been reported to encode serine palmitoyltransferase (SPT), a key enzyme responsible for sphingolipids biosynthesis (Begum et al., 2016). The disruption of sphingolipid metabolism affects plant growth and responses to abiotic stresses. Hence, the down-regulation of OSLCB2a1 affects the biosynthesis of sphingolipid hindering plant growth, thereby enhancing stress tolerance.

Analysis of DRG0-6 PPI

The DRG0-6 comprises set of 540 genes that are down regulated from 0-6hr time-points, after which these are up-regulated at 12hr time-point. The PPI network of this set of 540 genes resulted in a PPI network of 532 nodes and 263 edges using STRING database. The edge confidence is set to 0.4 and the network has a PPI enrichment p-value of 1.0e-16. The network was imported in Cytoscape for further analysis. Using the MCL algorithm (with an inflation value = 4), the network is grouped into closely connected group of nodes. Five clusters with a size greater than 5 are obtained and are listed in **Table 4.7** along with associated biological processes and module membership of their genes.

Table 4.7 Biological processes of the early responsive down regulated gene clusters pertaining to DRG0-6 in Osmotic stress.

Cluster	Associated Biological Process	Total Genes	Brown	Tan
D6	Protein Phosphorylation	59	54	5
D7	Signaling G-proteins, Cell vesicle transport	7	6	1
D8	N-Glycan Biosynthesis	7	7	0
D9	Gene Silencing by RNA	6	5	1
D10	Auxin Biosynthetic process, polyamine metabolism	6	6	0

Cluster D6: Gibberellin (GA) Signaling Response

Gibberellins are plant hormones playing crucial role in plant growth and development like flowering, leaf senescence and stem elongation. When exposed to stress, plants take measures to save resources like reducing its growth rate. Hence, we also observe a relationship between GA levels and amount of abiotic stress experienced. Studies have shown that plant exposure to abiotic stresses like salt, cold and osmotic stress results in restriction of plant growth by the repression of GA signaling (Achard et al., 2006; Vettakkorumakankav et al., 1999). Cluster D6 having a total of 59 genes with 72 interactions (**Figure 4.12**) has a central hub gene OsPKG which has been reported to play a role in mediating the gibberellin (GA) response in rice (Shen et al., 2019). OsPKG has the highest degree and is connected to all other genes in the cluster. PKG has many potential targets including GAMYB, a transcription factor involved in GA signaling, and has broad effects mainly in the salt stress response. PKGs are a family of genes that are structurally unique having an additional type 2C protein phosphatase domain. OsPKG possesses both protein kinase and phosphatase activities. cGMP activates its kinase activity which inhibits its phosphatase activity. Its downregulation in the early timepoints (1 – 6 hour) during osmotic stress represent the repression of GA signaling which hinders the plant growth thereby enhancing stress tolerance.

The cluster D6 is enriched in genes involved in “Protein Phosphorylation”. Protein phosphorylation is a crucial step to synthesize nascent stress responsive proteins by subjecting a protein to an array of posttranslational modifications (PTMs). These activated proteins aid in rapid stress response. Protein phosphorylation is a reversible form of modification that has a significant

role in signaling cascades. We observe a high presence of calcium-dependent protein kinases (CDPKs), receptor-like kinases (RLKs) and cGMP-protein kinases (PKG). These kinases are activated based on the fluctuation of cytosolic calcium levels and reactive oxygen species and affect various signaling pathways. RLKs are integral proteins that sense cell wall signals and relay them to the appropriate functional proteins. They also take part in systemic signaling for the transmission of local signals to distal tissues. We also observe the presence of molecules involved in SOS signaling pathway which includes calcineurin B-like gene CBL7 which is an EF-hand calcium-binding protein, and SOS1 which is a Na^+/H^+ antiporter at the plasma membrane. Together they help in coping with ionic stress. The cluster also consists of two transcription factors ZOS8-14 and OslDD11 which belong to the C2H2 zinc finger protein family. Since the central hub gene OsPKG in linked with all the other genes in the cluster, the other genes may be suggested to take part in GA signaling pathway. The genes in this cluster are late responsive genes, as they become up-regulated post 12hr time-point. Many of these genes play a role in plant growth and development and are down-regulated for resource allocation upon stress treatment.

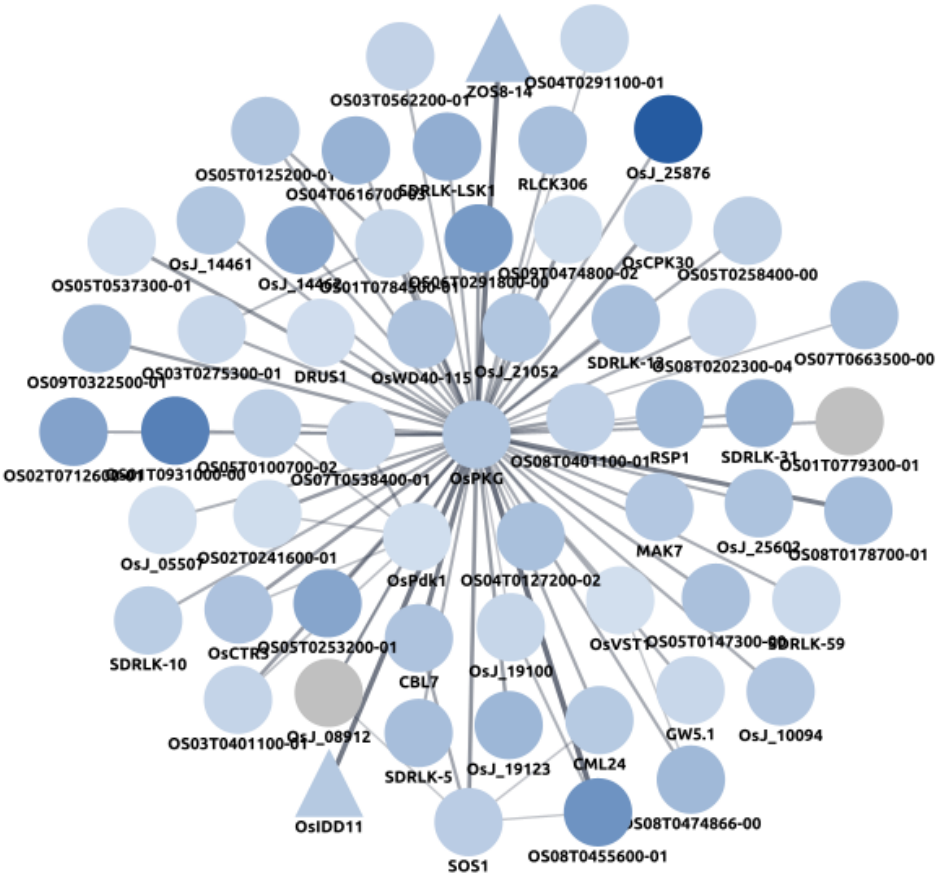


Figure 4.12 PPI Network of Cluster D6. The nodes are colored proportional to their fold change (the darker the shade, the lower is the fold change). Triangular nodes indicate transcription factors.

Cluster D7: Signaling G-protein and Cell Vesicle transport

This cluster consisting of 7 genes (**Figure 4.14**) is enriched with signaling G-proteins and genes with functions like cell vesicle transport. Heterotrimeric G proteins are shown to play a significant role in drought and salt stress (Y. Wang & Botella, 2022). These G proteins are composed of three subunits α, β, and γ and they serve as universal signaling molecules mediating the response to various internal and external signals. Some groups have identified involvement of Gα in salt stress however both positive and negative roles are associated with this subunit. The other subunit Gβ combined with Gγ regulates the osmotic and ionic stresses by increasing the levels of ROS scavengers and osmoprotectants (**Figure 4.13**).

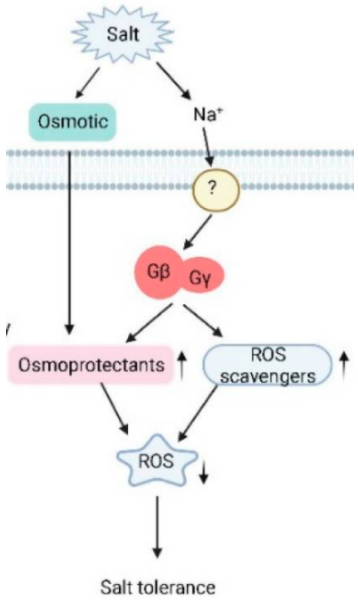


Figure 4.13 Role of heterotrimeric G proteins in regulating salt stress [reproduced from (Y. Wang & Botella, 2022)].

While 3 of the highest degree genes in this cluster are signaling G-proteins (OsRac3, Os08T0537600 and Os11T0303400), the other 4 genes, all of which are connected to G proteins are related to transport. Stress response involves complex cellular work requiring cooperation among distinctive organelles within a cell and even neighboring cells. This is achieved by exchanging content between them using small membranous containers called vesicles. Vesicles are loaded with content from a donor organelle which then moves to the target site to release them. The ROS scavenging activity is required only at a later timepoint as an optimum concentration of reactive oxygen is required for signaling. Hence, they are down-regulated until the 6-hour timepoint.

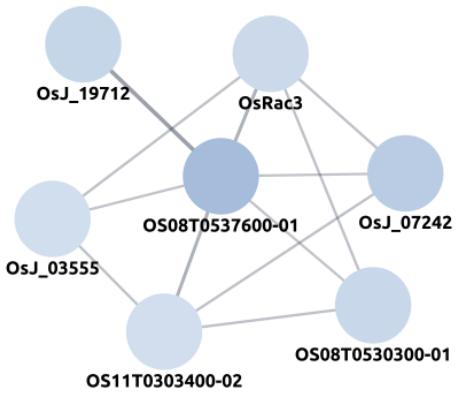


Figure 4.14 PPI network of Cluster D7. The nodes are colored proportional to their fold change (the darker the shade, the lower is the fold change).

Other clusters

The other clusters were enriched in functions like N-Glycan biosynthesis, gene silencing and auxin biosynthetic process. N-glycans have been shown to play a role in salt stress response in *Arabidopsis thaliana*. A membrane anchored endoglucanase KORRIGAN1/RADIAL SWELLING2 (KOR1/RSW2) with functions in cellulose biosynthesis is involved in salt stress response in *Arabidopsis thaliana*. N-glycans are required for their activation thus linking their role to salt/osmotic stress (Schaewen et al., 2008). The phytohormone auxin regulates the developmental plasticity of the plant root in salinity and drought stress (Korver et al., 2018; Naser & Shani, 2016). Auxin-mediated growth inhibition thus helps plants adapt to the changing environment. The presence of genes enriched in auxin biosynthetic process are down regulated till the 6-hour time point and is only activated at the 12-hour time point, pointing the role of auxin in late response to the osmotic stress.

4.5 Conclusion

With this analysis, we can suggest a causal flow of mechanisms and signaling pathways that are activated just after the plant experiences osmotic stress within the 1-hour to 3-hour range. At the 1-hour timepoint, the Green Yellow, Light Yellow, Midnight Blue and Royal Blue modules are activated which regulate functions like the perception and transmission of cell signaling via the MAPK cascade. Protein transport and chromatin remodeling takes place to allow for transcriptional activity. Important glycerophospholipids are synthesized which take part in lipid signaling as well as enhance the production of secondary messengers like phosphatidic acid. JA signaling is also activated earlier than the ABA signal transduction pathway. At the 3-hour timepoint, the Brown and Tan modules are down regulated. In these modules we observe two sets of genes according to their expression profile. The first set DRG0-3 is down-regulated till the 3-hour timepoint followed by up-regulation at the 6-hour timepoint. Associated functions which are suppressed until 3-hour timepoint involve the function of LRR-RLKs in mediating ROS scavenging activity as well as clusters involved in oxidation reduction process. This indicates the need of reactive oxygen at the earlier timepoints followed by their control mechanisms through the function of LRRs. The second set DRG0-6 is

down-regulated until the 6-hour timepoint thereafter being up-regulated. The clusters captured in the second set were found to be enriched in functions like protein phosphorylation and G-protein signaling. These genes are involved in plant growth and development and were hence down-regulated so that resources may be allocated for stress response. Other clusters enriched in functions such as auxin biosynthesis, gene silencing and N-glycan biosynthesis indicated that these mechanisms were only activated post 12-hour timepoint. Thus, we obtain an overall picture of the early activated and suppressed functions for coping with osmotic stress in rice.

Chapter 5

Conclusions

With the advances in high-throughput data, RNA-seq or transcriptome analysis has become a highly potential approach to gain insights into the complex machinery involved in plant regulatory response during stress conditions. It has helped us to understand the molecular mechanisms involved in the physiological and biochemical changes pertaining to stress response and has helped in establishing hierarchical relationships between signaling components and downstream effector genes to cope with the stress. The integration of information from other 'omics' data further increases the impact in identifying candidate genes. Rice being an important staple crop feeding over half of the world's population, requires such analysis for development of multi-stress tolerant hybrids.

In this thesis, we have used publicly available RNA-seq dataset to understand the mechanisms involved in different abiotic stress response in rice. The data is obtained from the TENOR database which ensures uniform library conditions in collecting timepoint-wise data for different abiotic stresses in a tissue specific manner, thereby allowing a comparative analysis across stresses, tissues and timepoints. The second chapter describes the pre-processing steps of this data to obtain read counts, alignment of the reads in the genome, evaluating differentially expressed genes in different condition and tissue, and the construction of a global co-expression network for analysis.

The third chapter discusses the construction of NetREx which is a web-server to query the global rank-based stress co-expression network in a tissue-specific manner. The coexpressed network is further enhanced by integrating various genomic data like gene ontology, pathway information, transcription factor annotations and aggregated omics data from different tissues. Various tools and functionalities have been developed for exploring the networks such as exploring gene expression profiles, network connectivity, identification of novel stress-responsive candidates, functional analysis of genes and comparative analysis across stress conditions. We demonstrate that NetREx can be used to identify novel candidate genes and tissue-specific interactions under stress conditions

and can aid in the analysis and understanding of complex phenotypes linked to stress response in rice.

The fourth chapter discusses the functional analysis of coexpressed modules across the different stress and hormone treatment conditions in the root tissue. The metabolic processes involved in different stress conditions based on the coexpressed modules are evaluated and compared. This is followed by an in-depth analysis of the early responsive modules active in the osmotic stress condition for the root tissue. We decipher the flow of signaling events and metabolic processes that follow osmotic shock within the 3-hour timepoint. Events like MAPK cascade, chromatin reorganisation, JA signaling pathway and lipid metabolism are activated within one hour. ROS scavenging activity and oxidation reduction processes are prevented until the 6-hour timepoint. This is followed by the activity of kinases and leucine rich repeats in the activation of required proteins and signal transmission.

The various analysis and studies have paved the way for understanding stress response in rice. With the abundance of omics data, there is a tremendous need for resources that can perform data summarization, integration of cross-platform resources and visualization which can aid in a thorough analysis. While an endeavour has been made in this direction by the creation of NetREx for the analysis of abiotic stress in rice, there is always a room for improvement. Integration of the gene expression data with ChIP-seq data specific to stress condition can offer insights into the regulatory relationship between genes and can be of great scientific value. Furthermore, with the increasing size of biological networks, machine learning techniques can aid at extracting useful information.

Appendix

Appendix Table 1 (a): Results from Shoot samples, after performing pre-processing and alignment for Drought Stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	907432856	748230345	82.5	97.8	62.3
	2	698212228	660235232	94.6	98.3	66.1
1 Hour	**1**	429886400	364153310	84.7	97.9	64.2
	2	488040612	462559660	94.8	98.3	68.5
3 Hour	**1**	443888868	419224548	94.4	98.4	67.7
	2	904912848	856110119	94.6	98.5	66.6
	3	829786620	788641587	95	98.6	67.0
6 Hour	**1**	792933080	657271011	82.9	97.7	62.6
	2	527915304	500109092	94.7	98.3	68.1
12 Hour	**1**	562146008	475743643	84.6	98.0	59.3
	2	849001548	811314106	95.6	98.2	54
	3	1394994896	1328231947	95.2	98.5	54.1
Day 1	**1**	43390216	371896482	85.7	98.1	60.1
	2	1183385056	1130694952	95.5	98.4	50.2

Appendix Table 1 (b): Results from Root samples, after performing pre-processing and alignment for Drought Stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate Using HISAT 2 (%)	Percentage reads assigned using featureCounts
Control	**1**	579818744	472441799	81.5	61.6	40.4
	2	653996948	622465056	95.2	97.9	67.4
1 Hour	**1**	660945324	529637126	80.1	62.8	43.3

	2	821572844	782628646	95.3	61.9	43.6
3 Hour	**1**	672104328	639437593	95.1	96.2	67.1
	2	992408456	945272961	95.3	82.1	57.6
6 Hour	**1**	521790464	417097626	79.9	66.2	44.8
	2	433799260	410262586	94.6	95.4	66.5
12 Hour	**1**	812957636	650448954	80.0	70.9	47.0
	2	507000180	480936808	94.9	91.4	58.7
Day 1	**1**	294636876	280972422	95.4	96.0	58.9
	2	760328092	720016435	94.7	66.1	45.7

Appendix Table 2 (a): Results from Shoot samples, after performing pre-processing and alignment for Cold stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	881646740	837345090	95.0	98.3	64.3
	2	607299280	409957913	67.5	96.3	61.4
	3	924780768	743574467	80.4	97.8	65.1
1 Hour	**1**	345985212	248857129	71.9	96.3	61.5
	2	510530000	486500479	95.3	98.3	59.0
3 Hours	**1**	333445592	229563105	68.8	95.9	59.9
	2-1	498210780	473058264	95.0	98.4	57.2
	2-2	720808548	689046581	95.6	98.5	
6 Hours	**1**	437651472	293448032	67.1	95.7	58.2
	2-1	380045904	359614311	94.6	98.4	63.3
	2-2	509582280	471474376	92.5	98.5	
12 Hours	**1**	272828068	192763474	70.7	96.3	58.3
	2-1	553511344	524999664	94.8	98.5	60.8
	2-2	976304132	918486440	94.1	98.6	
1 Day	**1**	276600632	195902523	70.8	95.6	48.0
	2-1	277888148	265037913	95.4	98.3	49.1
	2-2	553242456	533541649	96.4	98.4	

Appendix Table 2 (b): Results from Root samples, after performing pre-processing and alignment for Cold stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	1	1249143296	1196454327	95.8	84.5	59.5
	2	276101160	199192472	72.1	80.1	54.8
	3	714701644	580553617	81.2	73.7	50.1
1 Day	1	357673480	341518903	95.5	98.0	63.0
	2	496436560	475882762	95.9	98.2	62.9
1 Hour	1	378665060	280841448	74.2	84.5	57.2
	2	1117230096	1032506789	92.4	85.0	59.6
12 Hours	1	358160260	265437514	74.1	79.5	53.4
	2-1	498152640	478243958	96.0	68.7	46.8
	2-2	910753752	866966264	95.2	68.6	
3 Hours	1	350773896	263572983	75.1	89.5	60.1
	2	501562532	479464015	95.6	86.2	59.9
6 Hours	1	423490772	314674463	74.3	82.3	54.8
	2	609968172	582853590	95.6	85.7	58.6

Appendix Table 3 (a): Results from Shoot samples, after performing pre-processing and alignment for Osmotic stress

Time Points	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	1	881646740	837345090	95.0	97.8	64.3
	2	607299280	409957913	67.5	98.5	61.4
	3	924780768	743574467	80.4	98.3	65.1
1 Hour	1	507308512	420254646	82.8	92.2	63.9
	2	1189371880	1133232364	95.3	98.0	65.4
3 Hours	1	663097340	544580638	82.1	98.3	65.6
	2	1003019576	955463930	95.3	82.1	66.1
6 Hours	1	670896460	545132502	81.3	80.1	66.2
	2	1475798248	1412506265	95.7	88.5	64.3
12 Hours	1	856743364	701064272	81.8	91.8	65.8
	2	706020696	677760447	96.0	96.1	61.7

Appendix Table 3 (b): Results from Root samples, after performing pre-processing and alignment for Osmotic stress

Time Points	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	1	1249143296	1196454327	95.8	73.7	59.5
	2	276101160	199192472	72.1	97.7	54.8
	3	714701644	580553617	81.2	84.5	50.1
1 Hour	1	801065004	645567855	80.6	98.0	61.9
	2	672343652	636437233	94.7	89.3	62.3
3 Hours	1	768674640	638012019	83.0	97.9	61.5
	2	315982996	300973297	95.2	89.3	61.3
6 Hours	1	640814140	530178952	82.7	89.6	59.0
	2	1231838400	1178655304	95.7	89.3	60.6
12 Hours	1	722609292	597354743	82.7	98.4	60.4
	2	838759712	805733850	96.1	98.4	56.6

Appendix Table 4 (a): Results from Shoot samples, after performing pre-processing and alignment for Flood stress

Time points	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate Using HISAT2 (%)	Percentage reads assigned using featureCounts
Control	1	881646740	837345090	95	96.7	64.3
	2	607299280	409957913	67.5	73.9	61.4
	3	924780768	743574467	80.4	97.2	65.1
1 Hour	1	378832716	284913476	75.2	98.3	60.6
	2	981018564	934575631	95.3	89.9	65
3 Hours	1	399905008	290472408	72.6	71.3	62.9
	2-1	428738952	404668112	94.4	98.5	5.3
	2-2	746166936	717084707	96.1	96.3	
6 Hours	1	464446032	323790694	69.7	98.2	62
	2	534791784	507435056	94.9	98.4	65.4
12 Hours	1	421422356	304440347	72.2	98.5	63.2
	2	622551036	589909044	94.8	96.1	65.7
1 Day	1	599438980	568786021	94.9	98.4	67.5
	2	1400898728	1335515577	95.3	84.5	66.5
	3	1559714864	1477659865	94.7	80.1	69.1

3 Days	**1**	449239192	427564265	95.2	88.4	45
	2	849432620	795310796	93.6	97.8	45.3

Appendix Table 4 (b): Results from Root samples, after performing pre-processing and alignment for Flood stress

Time points	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate Using HISAT2 (%)	Percentage reads assigned using featureCounts
Control	**1**	1249143296	1196454327	95.8	98.5	59.5
	2	276101160	199192472	72.1	98.5	54.8
	3	714701644	580553617	81.2	89.3	50.1
1 Hour	**1**	521348144	374075481	71.8	73.7	61.9
	2	583115244	554188903	95	89.6	55.2
3 Hours	**1**	381853716	280223572	73.4	98.6	62.6
	2	499614348	476575955	95.4	70.3	63.6
6 Hours	**1**	477663192	350059938	73.3	80.3	62.2
	2	539177288	513709573	95.3	69.4	63.9
12 Hours	**1**	563410420	405280074	71.9	97.2	61
	2	556642620	525369170	94.4	96.1	52.1
1 Day	**1**	293880904	208107732	70.8	90.6	48.7
	2	1091446944	1039665987	95.3	96.5	50.1
3 Days	**1**	449626488	306217413	68.1	91.8	49.1
	2	293268496	280137456	95.5	80.6	57.1
	2	556556664	534978329	96.1	80.9	

Appendix Table 5 (a): Results from Shoot samples, after performing pre-processing and alignment for ABA treatment

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	753215252	596184260	79.2	97.8	65.2
	2	654536396	626980227	95.8	98.5	64.4
1 Hour	**1**	467703620	383556982	82.0	98.0	66.6
	2	1228120100	1176548952	95.8	98.4	65.2
	1	546360808	440643637	80.7	97.9	66.2

3 Hours	**2**	1167969976	1118014644	95.7	98.4	64.8
6 Hours	**1**	634916844	507891903	80.0	97.8	68.4
	2	553,242,456	533,541,649	96.4	98.4	65.6
12 Hours	**1**	696719664	565731763	81.2	98.0	68.4
	2	702398384	650248606	92.6	98.5	66.2
1 Day	**1**	518302140	425912740	82.2	97.9	68.1
	2	1488841216	1388650985	93.3	98.5	66.8

Appendix Table 5 (b): Results from Root samples, after performing pre-processing and alignment for ABA treatment

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	526740192	418298786	79.4	97.6	66.5
	2	1013412424	961743522	94.9	98.0	64.9
1 Hour	**1**	410940208	332103360	80.8	97.6	68.3
	2	2023243880	1913180023	94.6	97.8	67.5
3 Hours	**1**	683697672	540174516	79.0	91.4	65.7
	2	2395719044	2279602092	95.2	89.6	63.6
6 Hours	**1**	720925436	559086146	77.6	92.9	67.5
	2	781650880	746498551	95.5	85.2	61.5
12 Hours	**1**	840271656	666711614	79.3	84.2	61.5
	2	1909279752	1820182384	95.3	93.5	66.8
1 Day	**1**	537581744	431861037	80.3	88.5	62.9
	2	1274690316	1215312374	95.3	84.8	60.3

Appendix Table 6 (a): Results from Shoot samples, after performing pre-processing and alignment for Jasmonic Acid stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	775009924	623535884	80.5	97.9	64.1
	2	591917412	561662378	94.9	98.4	65.7
1 Hour	**1**	472502412	394086934	83.4	98.1	63.1
	2-1	206822524	196980365	95.2	98.5	65.5
	2-2	615613452	589283627	95.7	98.5	
	1	617565816	509392336	82.5	98.0	63.7

3 Hours	**2**	710704272	676235442	95.2	98.5	49.5
6 Hours	**1**	684308788	559104471	81.7	97.9	63.4
	2	564344764	536078111	95.0	98.4	66.5
12 Hours	**1**	617565816	509392336	82.5	98	64.9
	2	598872552	570990066	95.3	98.4	65.6
1 Day	**1**	416125460	350651386	84.3	98.0	64.5
	2-1	368522936	352140211	95.6	98.5	57.4
	2-2	715759412	674233142	94.2	98.5	

Appendix Table 6 (b): Results from Root samples, after performing pre-processing and alignment for Jasmonic Acid stress

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	640495092	495521484	77.4	80.73	54.7
	2	767064124	578479558	75.4	79.42	54.5
1 Hour	**1**	657721100	497652715	75.7	85.5	51.9
	2	564178856	442018072	78.3	85.21	53.9
3 Hour	**1**	712076908	543145747	76.3	80.73	49.7
	2	796485624	602178232	75.6	81.87	53.6
6 Hour	**1**	547832396	408481798	74.6	78.81	46.7
	2	656757876	495806661	75.5	85.7	58.4
12 Hour	**1**	658917416	499218092	75.8	73.76	48.8
	2	651628788	494919522	76.0	62.53	41.2
1 Day	**1**	797808708	603254588	75.6	55.19	37.9
	2-1	410933368	320443062	78.0	66.49	41.1
	2-2	1314611824	1236515010	94.1	66.21	

Appendix Table 7 (a): Results from Shoot samples, after performing pre-processing and alignment for Developmental Time-points (No Treatment)

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	612382388	439576373	71.8	97.8	64.5
	2	667381764	468784510	70.2	97.9	65.3
1 Hour	**1**	586393124	417592326	71.2	97.8	65.1
	2	545417800	387111565	71.0	97.9	63.9

3 Hours	**1**	562898408	404914931	71.9	97.8	65.4
	2	683050380	474308234	69.4	97.4	65.9
6 Hours	**1-1**	471812104	337387754	71.5	97.8	66.6
	1-2	852563744	821155586	96.3	98.8	
	2	563744288	394827485	70.0	97.8	64.5
12 Hours	**1**	743380776	529798020	71.3	97.8	65
	2	758414868	536401596	70.7	97.8	63.2
1 Day	**1-1**	661066240	471464376	71.3	97.8	65.4
	1-2	697630448	482239862	69.1	97.4	
	2	783969944	545432861	69.6	97.5	64.4
3 Days	**1-1**	529668472	379653448	71.7	97.8	65.6
	1-2	989260612	942400264	95.3	98.8	
	2	589092644	412502587	70.0	97.7	64.8
4 Days	**1**	547062668	389567508	71.2	97.8	65.5
	2	788784012	551376946	69.9	97.8	64.1
5 Days	**1**	799590300	537377148	67.2	97.4	59.9
	2	497319224	352040957	70.8	97.5	65.1
10 Days	**1**	505663796	355650859	70.3	97.5	64.5
	2	716291336	493806845	68.9	97.3	66.1

Appendix Table 7 (b): Results from Root samples, after performing pre-processing and alignment for Developmental Time-points (No Treatment)

Time Point	Replicates	Raw Data (bp)	Filtered Data (bp)	Percentage Filtered (bp)	Read Align Rate using HISAT2 (%)	Percentage reads assigned using FeatureCounts
Control	**1**	7518132	405345006	70.9	90.2	63.0
	2	617656636	419525465	67.9	83.5	58.6
1 Hour	**1**	570453264	398425992	69.8	90.3	63.5
	2-1	603476100	408556653	67.7	84.1	59.7
	2-1	1655964304	1533531564	92.6	85.7	
3 Hour	**1**	610521908	419308518	68.7	84.0	56.2
	2	630330016	430205696	68.3	83.8	57.3
6 H2ur	**1**	634707464	471869517	74.3	89.1	62.5
	2	773149520	550770174	71.2	81.0	56.8
12 Hour	**1**	611680224	451950522	73.9	81.0	56.2
	2	643527948	463123951	72.0	91.1	63.1
1 Day	**1-1**	553411784	403435538	72.9	81.7	57.3
	1-2	588325880	412189345	70.1	81.4	
	2	645960024	472308295	73.1	79.7	56.1
3 Days	**1**	650700752	487138147	74.9	70.0	50.1

	2	674622132	487532682	72.3	77.0	54.5
4 Days	**1**	655813120	478550089	73.0	78.0	55.2
	2	650388772	468334563	72.0	73.2	52.1
5 Days	**1**	586760052	429279737	73.2	67.6	47.8
	2	713029796	523884566	73.5	72.1	50.4
10 Days	**1**	684959196	497710256	72.7	76.8	54.2
	2	863996196	609010264	70.5	79.8	56.2

Appendix Table 8: Top 10 neighbours of OsPHR3 (gene with highest degree in the neighbourhood of root ABA signalling network)

Source	Target	PCC	HRR	Function
Os02g0139000	Os11g0536800	0.919	1	Molecular Function: carbon-nitrogen ligase activity, with glutamine as amido-N-donor (GO:0016884)
Os02g0139000	Os11g0551800	0.916	1	Similar to Yippee-like protein 1 (DGL-1) (Mdgl-1).
Os02g0139000	Os07g0637300	0.939	2	Transferase activity, transferring phosphorus-containing groups
Os02g0139000	Os04g0398000	0.932	3	ERF transcription factor, A member of APETALA2/Ethylene-Responsive Element Binding Protein (AP2/EREBP) family, Regulation of drought stress response, Innate immunity, Pathogenesis-related transcriptional factor and ERF domain containing protein.
Os02g0139000	Os02g0265900	0.922	3	Reticulon family protein., Hypothetical conserved gene., Similar to Reticulon.
Os02g0139000	Os04g0321800	0.916	3	OSIGBa0097I24.2 protein., Protein phosphatase 2C domain containing protein.
Os02g0139000	Os02g0139000	0.84	3	Protein.degradation.ubiquitin.E3.RING
Os02g0139000	Os03g0231600	0.926	5	Cysteine and methionine metabolism, Valine, leucine and isoleucine degradation, Valine, leucine and isoleucine biosynthesis, Pantothenate and CoA biosynthesis, Metabolic pathways, Biosynthesis of secondary metabolites, 2-Oxocarboxylic acid metabolism, Biosynthesis of amino acids
Os02g0139000	Os12g0583700	0.918	6	RNA.regulation of transcription.C2H2 zinc finger family'
Os02g0139000	Os02g0817700	0.916	7	Fatty acid degradation, Valine, leucine and isoleucine degradation, alpha-Linolenic acid metabolism, Biosynthesis of unsaturated fatty acids, Metabolic pathways, Biosynthesis of secondary metabolites, Fatty acid metabolism, Peroxisome

Related Publications

Sircar, S., Musaddi, M., & Parekh, N. (2022). NetREx: Network-based Rice Expression Analysis Server for abiotic stress conditions. *Database*, *2022*. https://doi.org/10.1093/DATABASE/BAAC060

Bibliography

Achard, P., Cheng, H., de Grauwe, L., Decat, J., Schoutteten, H., Moritz, T., van der Straeten, D., Peng, J., & Harberd, N. P. (2006). Integration of plant responses to environmentally activated phytohormonal signals. *Science*, *311*(5757), 91–94. https://doi.org/10.1126/SCIENCE.1118642/SUPPL_FILE/ACHARD.SOM.PDF

Aleman, F., Yazaki, J., Lee, M., Takahashi, Y., Kim, A. Y., Li, Z., Kinoshita, T., Ecker, J. R., & Schroeder, J. I. (2016). An ABA-increased interaction of the PYL6 ABA receptor with MYC2 Transcription Factor: A putative link of ABA and JA signaling. *Scientific Reports*, *6*. https://doi.org/10.1038/SREP28941

Amrine, K. C. H., Blanco-Ulate, B., & Cantu, D. (2015). Discovery of Core Biotic Stress Responsive Genes in Arabidopsis by Weighted Gene Co-Expression Network Analysis. *PLOS ONE*, *10*(3), e0118731. https://doi.org/10.1371/JOURNAL.PONE.0118731

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols 2013 8:9*, *8*(9), 1765–1786. https://doi.org/10.1038/nprot.2013.099

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*(10), 1–12. https://doi.org/10.1186/GB-2010-11-10-R106/COMMENTS

Aoki, K., Ogata, Y., & Shibata, D. (2007). Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant & Cell Physiology*, *48*(3), 381–390. https://doi.org/10.1093/PCP/PCM013

Bai, J., Zhu, X., Wang, Q., Zhang, J., Chen, H., Dong, G., Zhu, L., Zheng, H., Xie, Q., Nian, J., Chen, F., Fu, Y., Qian, Q., & Zuo, J. (2015). Rice TUTOU1 Encodes a Suppressor of cAMP Receptor-Like Protein That Is Important for Actin Organization and Panicle Development. *Plant Physiology*, *169*(2), 1179. https://doi.org/10.1104/PP.15.00229

Banerjee, A., & Roychoudhury, A. (2017). Abscisic-acid-dependent basic leucine zipper (bZIP) transcription factors in plant abiotic stress. *Protoplasma*, *254*(1), 3–16. https://doi.org/10.1007/S00709-015-0920-4/FIGURES/1

Bibliography

Begum, M. A., Shi, X. X., Tan, Y., Zhou, W. W., Hannun, Y., Obeid, L., Mao, C., & Zhu, Z. R. (2016). Molecular characterization of rice OsLCB2a1 gene and functional analysis of its role in insect resistance. *Frontiers in Plant Science*, *7*(DECEMBER2016), 1789. https://doi.org/10.3389/FPLS.2016.01789/FULL

Bilgin, D. D., Zavala, J. A., Zhu, J., Clough, S. J., Ort, D. R., & Delucia, E. H. (2010). Biotic stress globally downregulates photosynthesis genes. *Plant, Cell & Environment*, *33*(10), 1597–1613. https://doi.org/10.1111/J.1365-3040.2010.02167.X

Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomics data. In *Methods in Molecular Biology* (Vol. 1374, pp. 115–140). Humana Press Inc. https://doi.org/10.1007/978-1-4939-3167-5_6

Botanical classification of Rice | agropedia. (n.d.). Retrieved June 12, 2022, from http://agropedia.iitk.ac.in/content/botanical-classification-rice

Cantrell, R. P., & Reeves, T. G. (2002). The rice genome. The cereal of the world's poor takes center stage. *Science (New York, N.Y.)*, *296*(5565), 53. https://doi.org/10.1126/SCIENCE.1070721

Chandran, A. K. N., Bhatnagar, N., Yoo, Y.-H., Moon, S., Park, S.-A., Hong, W.-J., Kim, B.-G., An, G., & Jung, K.-H. (2018). Meta-expression analysis of unannotated genes in rice and approaches for network construction to suggest the probable roles. *Plant Molecular Biology*, *96*(1–2), 17–34. https://doi.org/10.1007/s11103-017-0675-8

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F. M., Ponce, M. R., Micol, J. L., & Solano, R. (2007a). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature 2007 448:7154*, *448*(7154), 666–671. https://doi.org/10.1038/nature06006

Chini, A., Fonseca, S., Fernández, G., Adie, B., Chico, J. M., Lorenzo, O., García-Casado, G., López-Vidriero, I., Lozano, F. M., Ponce, M. R., Micol, J. L., & Solano, R. (2007b). The JAZ family of repressors is the missing link in jasmonate signalling. *Nature 2007 448:7154*, *448*(7154), 666–671. https://doi.org/10.1038/nature06006

Choi, W. G., Toyota, M., Kim, S. H., Hilleary, R., & Gilroy, S. (2014). Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(17), 6497–6502. https://doi.org/10.1073/PNAS.1319955111/-/DCSUPPLEMENTAL/SM08.MOV

Chow, C.-N., Zheng, H.-Q., Wu, N.-Y., Chien, C.-H., Huang, H.-D., Lee, T.-Y., Chiang-Hsieh, Y.-F., Hou, P.-F., Yang, T.-Y., & Chang, W.-C. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Research*, *44*. https://doi.org/10.1093/nar/gkv1035

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology 2016 17:1*, *17*(1), 1–19. https://doi.org/10.1186/S13059-016-0881-8

Consortium, I. P., Hao, L., Zhang, H., Zhang, Z., Hu, S., & Xue, Y. (2016). Information Commons for Rice (IC4R). *Nucleic Acids Research*, *44*(D1), D1172–D1180. https://doi.org/10.1093/nar/gkv1141

de la Fuente, A., Bing, N., Hoeschele, I., & Mendes, P. (2004). Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics (Oxford, England)*, *20*(18), 3565–3574. https://doi.org/10.1093/BIOINFORMATICS/BTH445

de Lorenzo, L., Merchan, F., Laporte, P., Thompson, R., Clarke, J., Sousa, C., & Crespi, M. (2009). A Novel Plant Leucine-Rich Repeat Receptor Kinase Regulates the Response of Medicago truncatula Roots to Salt Stress. *The Plant Cell*, *21*(2), 668. https://doi.org/10.1105/TPC.108.059576

de Torres Zabala, M., Bennett, M. H., Truman, W. H., & Grant, M. R. (2009). Antagonism between salicylic and abscisic acid reflects early host–pathogen conflict and moulds plant defence responses. *The Plant Journal*, *59*(3), 375–386. https://doi.org/10.1111/J.1365-313X.2009.03875.X

de Zelicourt, A., Colcombet, J., & Hirt, H. (2016). The Role of MAPK Modules and ABA during Abiotic Stress Signaling. *Trends in Plant Science*, *21*(8), 677–685. https://doi.org/10.1016/J.TPLANTS.2016.04.004

Du, H., Liu, H., & Xiong, L. (2013). Endogenous auxin and jasmonic acid levels are differentially modulated by abiotic stresses in rice. *Frontiers in Plant Science*, *4*(OCT), 397. https://doi.org/10.3389/FPLS.2013.00397/ABSTRACT

Duan, P., Rao, Y., Zeng, D., Yang, Y., Xu, R., Zhang, B., Dong, G., Qian, Q., & Li, Y. (2014). SMALL GRAIN 1, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. *The Plant Journal : For Cell and Molecular Biology*, *77*(4), 547–557. https://doi.org/10.1111/TPJ.12405

Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., Seki, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2003). OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *The Plant Journal : For Cell and Molecular Biology*, *33*(4), 751–763. https://doi.org/10.1046/J.1365-313X.2003.01661.X

Enright, A. J., van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575–1584. https://doi.org/10.1093/NAR/30.7.1575

Fernandez, N. F., Gundersen, G. W., Rahman, A., Grimes, M. L., Rikova, K., Hornbeck, P., & Maayan, A. (2017). Clustergrammer, a web-based heatmap visualization and analysis tool for high-dimensional biological data. *Scientific Data 2017 4:1*, *4*(1), 1–12. https://doi.org/10.1038/sdata.2017.151

Ferrari, C., Proost, S., Ruprecht, C., & Mutwil, M. (2018). PhytoNet: comparative co-expression network analyses across phytoplankton and land plants. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky298

Ferreira, S. S., Hotta, C. T., de Carli Poelking, V. G., Leite, D. C. C., Buckeridge, M. S., Loureiro, M. E., Barbosa, M. H. P., Carneiro, M. S., & Souza, G. M. (2016). Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant Molecular Biology*, *91*(1–2), 15–35. https://doi.org/10.1007/s11103-016-0434-2

Fich, E. A., Segerson, N. A., & Rose, J. K. C. (2016). The Plant Polyester Cutin: Biosynthesis, Structure, and Biological Roles. *Http://Dx.Doi.Org/10.1146/Annurev-Arplant-043015-111929*, *67*, 207–233. https://doi.org/10.1146/ANNUREV-ARPLANT-043015-111929

Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., & Bader, G. D. (2016). Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics (Oxford, England)*, *32*(2), 309–311. https://doi.org/10.1093/bioinformatics/btv557

Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S.-Y., Cutler, S. R., Sheen, J., Rodriguez, P. L., & Zhu, J.-K. (2009). In vitro reconstitution of an abscisic acid signalling pathway. *Nature 2009 462:7273*, *462*(7273), 660–664. https://doi.org/10.1038/nature08599

Fujita, Y., Yoshida, T., & Yamaguchi-Shinozaki, K. (2013). Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. *Physiologia Plantarum*, *147*(1), 15–27. https://doi.org/10.1111/J.1399-3054.2012.01635.X

Gao, J. P., Chao, D. Y., & Lin, H. X. (2007). Understanding Abiotic Stress Tolerance Mechanisms: Recent Studies on Stress Response in Rice. *Journal of Integrative Plant Biology*, *49*(6), 742–750. https://doi.org/10.1111/J.1744-7909.2007.00495.X

Genva, M., Obounou Akong, F., Andersson, M. X., Deleu, M., Lins, L., & Fauconnier, M. L. (2018). New insights into the biosynthesis of esterified oxylipins and their involvement in plant defense and developmental mechanisms. *Phytochemistry Reviews 2018 18:1*, *18*(1), 343–358. https://doi.org/10.1007/S11101-018-9595-8

Gephi - The Open Graph Viz Platform. (n.d.). https://gephi.org/

Gill, S. S., & Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiology and Biochemistry*, *48*(12), 909–930. https://doi.org/10.1016/J.PLAPHY.2010.08.016

Giri, J., Vij, S., Dansana, P. K., & Tyagi, A. K. (2011). Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. *New Phytologist*, *191*(3), 721–732. https://doi.org/10.1111/J.1469-8137.2011.03740.X

Golldack, D., Su, H., Quigley, F., Kamasani, U. R., Muñoz-Garay, C., Balderas, E., Popova, O. v., Bennett, J., Bohnert, H. J., & Pantoja, O. (2002). Characterization of a HKT-type transporter in rice as a general alkali cation transporter. *The Plant Journal*, *31*(4), 529–542. https://doi.org/10.1046/J.1365-313X.2002.01374.X

Graph clustering by flow simulation. (n.d.). Retrieved June 24, 2022, from https://dspace.library.uu.nl/handle/1874/848

Hahn, A., Kilian, J., Mohrholz, A., Ladwig, F., Peschke, F., Dautel, R., Harter, K., Berendzen, K. W., & Wanke, D. (2013). Plant Core Environmental Stress Response Genes Are Systemically Coordinated during Abiotic Stresses. *International Journal of Molecular Sciences 2013, Vol. 14, Pages 7617-7641*, *14*(4), 7617–7641. https://doi.org/10.3390/IJMS14047617

Haider, M. S., Zhang, C., Kurjogi, M. M., Pervaiz, T., Zheng, T., Zhang, C., Lide, C., Shangguan, L., & Fang, J. (2017). Insights into grapevine defense response against drought as revealed by biochemical, physiological and RNA-Seq analysis. *Scientific Reports 2017 7:1*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-13464-3

Hamada, K., Hongo, K., Suwabe, K., Shimizu, A., Nagayama, T., Abe, R., Kikuchi, S., Yamamoto, N., Fujii, T., Yokoyama, K., Tsuchida, H., Sano, K., Mochizuki, T., Oki, N., Horiuchi, Y., Fujita, M., Watanabe, M., Matsuoka, M., Kurata, N., & Yano, K. (2011). OryzaExpress: an integrated database of gene expression networks and omics annotations in rice. *Plant & Cell Physiology*, *52*(2), 220–229. https://doi.org/10.1093/PCP/PCQ195

Hatmi, S., Trotel-Aziz, P., Villaume, S., Couderchet, M., Clément, C., & Aziz, A. (2014). Osmotic stress-induced polyamine oxidation mediates defence responses and reduces stress-enhanced grapevine susceptibility to Botrytis cinerea. *Journal of Experimental Botany*, *65*(1), 75–88. https://doi.org/10.1093/JXB/ERT351

He, F., Liu, Q., Zheng, L., Cui, Y., Shen, Z., & Zheng, L. (2015). Rna-seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Frontiers in Plant Science*, *6*(DEC), 1136. https://doi.org/10.3389/FPLS.2015.01136/BIBTEX

He, M., He, C. Q., & Ding, N. Z. (2018). Abiotic stresses: General defenses of land plants and chances for engineering multistress tolerance. *Frontiers in Plant Science*, *871*, 1771. https://doi.org/10.3389/FPLS.2018.01771/BIBTEX

He, Z., Wu, J., Sun, X., & Dai, M. (2019). The Maize Clade A PP2C Phosphatases Play Critical Roles in Multiple Abiotic Stress Responses. *International Journal of Molecular Sciences*, *20*(14). https://doi.org/10.3390/IJMS20143573

Ho, C. L., Wu, Y., Shen, H. bin, Provart, N. J., & Geisler, M. (2012). A predicted protein interactome for rice. *Rice*, *5*(1), 1–14. https://doi.org/10.1186/1939-8433-5-15/FIGURES/6

Hong, Y., Zhang, H., Huang, L., Li, D., & Song, F. (2016). Overexpression of a stress-responsive NAC transcription factor gene ONAC022 improves drought and salt tolerance in rice. *Frontiers in Plant Science*, *7*(JAN2016), 4. https://doi.org/10.3389/FPLS.2016.00004/BIBTEX

Hu, Y., Jiang, L., Wang, F., & Yu, D. (2013). Jasmonate Regulates the INDUCER OF CBF EXPRESSION–C-REPEAT BINDING FACTOR/DRE BINDING FACTOR1 Cascade and Freezing Tolerance in Arabidopsis. *The Plant Cell*, *25*(8), 2907–2924. https://doi.org/10.1105/TPC.113.112631

Iida, K., & Nishimura, I. (2002). Gene expression profiling by DNA microarray technology. *Critical Reviews in Oral Biology and Medicine : An Official Publication of the American Association of Oral Biologists*, *13*(1), 35–50. https://doi.org/10.1177/154411130201300105

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS One*, *9*(6). https://doi.org/10.1371/JOURNAL.PONE.0098679

Jan, R., Khan, M. A., Asaf, S., Lubna, Lee, I. J., & Kim, K. M. (2021). Over-Expression of Chorismate Mutase Enhances the Accumulation of Salicylic Acid, Lignin, and Antioxidants in Response to the White-Backed Planthopper in Rice Plants. *Antioxidants*, *10*(11). https://doi.org/10.3390/ANTIOX10111680

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. https://doi.org/10.1093/nar/gkw1092

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102409&tool=pmcentrez&rendertype=abstract

Kautsar, S. A., Suarez Duran, H. G., Blin, K., Osbourn, A., & Medema, M. H. (2017). plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Research*, *45*(W1), W55–W63. https://doi.org/10.1093/nar/gkx305

Kawahara, Y., Oono, Y., Wakimoto, H., Ogata, J., Kanamori, H., Sasaki, H., Mori, S., Matsumoto, T., & Itoh, T. (2016). TENOR: Database for Comprehensive mRNA-Seq Experiments in Rice. *Plant and Cell Physiology*, *57*(1), e7–e7. https://doi.org/10.1093/PCP/PCV179

Kazan, K., & Manners, J. M. (2013). MYC2: The Master in Action. *Molecular Plant*, *6*(3), 686–703. https://doi.org/10.1093/MP/SSS128

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/NMETH.3317

Kim, E. H., Kim, Y. S., Park, S. H., Koo, Y. J., Choi, Y. do, Chung, Y. Y., Lee, I. J., & Kim, J. K. (2009). Methyl jasmonate reduces grain yield by mediating stress signals to alter spikelet development in rice. *Plant Physiology*, *149*(4), 1751–1760. https://doi.org/10.1104/PP.108.134684

Kim, T. H., Böhmer, M., Hu, H., Nishimura, N., & Schroeder, J. I. (2010). Guard Cell Signal Transduction Network: Advances in Understanding Abscisic Acid, CO2, and Ca2+ Signaling. *Annual Review of Plant Biology*, *61*, 561. https://doi.org/10.1146/ANNUREV-ARPLANT-042809-112226

Kiribuchi, K., Jikumaru, Y., Kaku, H., Minami, E., Hasegawa, M., Kodama, O., Seto, H., Okada, K., Nojiri, H., & Yamane, H. (2005). Involvement of the Basic Helix-Loop-Helix Transcription Factor RERJ1 in Wounding and Drought Stress Responses in Rice Plants. *Bioscience, Biotechnology, and Biochemistry*, *69*(5), 1042–1044. https://doi.org/10.1271/BBB.69.1042

Kiribuchi, K., Sugimori, M., Takeda, M., Otani, T., Okada, K., Onodera, H., Ugaki, M., Tanaka, Y., Tomiyama-Akimoto, C., Yamaguchi, T., Minami, E., Shibuya, N., Omori, T., Nishiyama, M., Nojiri, H., & Yamane, H. (2004). RERJ1, a jasmonic acid-responsive gene from rice, encodes a basic helix–loop–helix protein. *Biochemical and Biophysical Research Communications*, *325*(3), 857–863. https://doi.org/10.1016/j.bbrc.2004.10.126

Kissoudis, C., van de Wiel, C., Visser, R. G. F., & van der Linden, G. (2014). Enhancing crop resilience to combined abiotic and biotic stress through the dissection of physiological and molecular crosstalk. *Frontiers in Plant Science*, *5*(MAY), 207. https://doi.org/10.3389/FPLS.2014.00207/BIBTEX

Kopriva, S. (2006). Regulation of sulfate assimilation in Arabidopsis and beyond. *Annals of Botany*, *97*(4), 479–495. https://doi.org/10.1093/AOB/MCL006

Korver, R. A., Koevoets, I. T., & Testerink, C. (2018). Out of Shape During Stress: A Key Role for Auxin. *Trends in Plant Science*, *23*(9), 783. https://doi.org/10.1016/J.TPLANTS.2018.05.011

Krishnan, A., Gupta, C., Ambavaram, M. M. R., & Pereira, A. (2017). RECoN: Rice environment co-expression network for systems level analysis of abiotic-stress response. *Frontiers in Plant Science*, *8*, 1640. https://doi.org/10.3389/FPLS.2017.01640/BIBTEX

Kryuchkova-Mostacci, N., & Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, *18*(2), 205–214. https://doi.org/10.1093/BIB/BBW008

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, *2015*(11), 951. https://doi.org/10.1101/PDB.TOP084970

Kumar, G., Bajpai, R., Sarkar, A., Mishra, R. K., Kumar Gupta, V., Singh, H. B., & Sarma, B. K. (2019). Identification, characterization and expression profiles of Fusarium udum stress-responsive WRKY transcription factors in Cajanus cajan under the influence of NaCl stress and Pseudomonas fluorescens OKC. *Scientific Reports 2019 9:1*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-50696-x

Kumar, S., Seem, K., Kumar, S., & Mohapatra, T. (2022). RNA-seq analysis reveals the genes/pathways responsible for genetic plasticity of rice to varying environmental conditions on direct-

sowing and transplanting. *Scientific Reports 2022 12:1*, *12*(1), 1–22. https://doi.org/10.1038/s41598-022-06009-w

Kuromori, T., Seo, M., & Shinozaki, K. (2018). ABA Transport and Plant Water Stress Responses. *Trends in Plant Science*, *23*(6), 513–522. https://doi.org/10.1016/J.TPLANTS.2018.04.001

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 1–13. https://doi.org/10.1186/1471-2105-9-559/FIGURES/4

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, *24*(5), 719–720. https://doi.org/10.1093/BIOINFORMATICS/BTM563

Lata, C., Yadav, A., & Prasad, M. (2011). Role of Plant Transcription Factors in Abiotic Stress Tolerance. *Abiotic Stress Response in Plants - Physiological, Biochemical and Genetic Perspectives*. https://doi.org/10.5772/23172

Lee, T. H., Kim, Y. K., Pham, T. T. M., Song, S. I., Kim, J. K., Kang, K. Y., An, G., Jung, K. H., Galbraith, D. W., Kim, M., Yoon, U. H., & Nahm, B. H. (2009). RiceArrayNet: A Database for Correlating Gene Expression from Transcriptome Profiling, and Its Application to the Analysis of Coexpressed Genes in Rice. *Plant Physiology*, *151*(1), 16–33. https://doi.org/10.1104/PP.109.139030

Lee, T., Oh, T., Yang, S., Shin, J., Hwang, S., Kim, C. Y., Kim, H., Shim, H., Shim, J. E., Ronald, P. C., & Lee, I. (2015). RiceNet v2: an improved network prioritization server for rice genes. *Nucleic Acids Research*, *43*(W1), W122–W127. https://doi.org/10.1093/NAR/GKV253

Li, C., Shen, H., Wang, T., & Wang, X. (2015). ABA Regulates Subcellular Redistribution of OsABI-LIKE2, a Negative Regulator in ABA Signaling, to Control Root Architecture and Drought Resistance in Oryza sativa. *Plant and Cell Physiology*, *56*(12), 2396–2408. https://doi.org/10.1093/PCP/PCV154

Li, M., Chen, J. E., Wang, J. X., Hu, B., & Chen, G. (2008). Modifying the DPClus algorithm for identifying protein complexes based on new topological structures. *BMC Bioinformatics*, *9*(1), 1–16. https://doi.org/10.1186/1471-2105-9-398/TABLES/5

Li, X., Chang, Y., Ma, S., Shen, J., Hu, H., & Xiong, L. (2019). Genome-Wide Identification of SNAC1-Targeted Genes Involved in Drought Response in Rice. *Frontiers in Plant Science*, *10*, 982. https://doi.org/10.3389/FPLS.2019.00982/BIBTEX

Li, Y. S., Sun, H., Wang, Z. F., Duan, M., Huang, S. D., Yang, J., Huang, J., & Zhang, H. S. (2013). A novel nuclear protein phosphatase 2C negatively regulated by ABL1 is involved in abiotic stress and panicle development in rice. *Molecular Biotechnology*, *54*(2), 703–710. https://doi.org/10.1007/s12033-012-9614-8

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Liao, Z. X., Ni, Z., Wei, X. L., Chen, L., Li, J. Y., Yu, Y. H., Jiang, W., Jiang, B. le, He, Y. Q., & Huang, S. (2019). Dual RNA-seq of Xanthomonas oryzae pv. oryzicola infecting rice reveals novel insights into bacterial-plant interaction. *PloS One*, *14*(4). https://doi.org/10.1371/JOURNAL.PONE.0215039

Liesecke, F., Daudu, D., de Bernonville, R. D., Besseau, S., Clastre, M., Courdavault, V., de Craene, J.-O., Crèche, J., Giglioli-Guivarc'h, N., Glévarec, G., Pichon, O., & de Bernonville, T. D. (2018). Ranking genome-wide correlation measurements improves microarray and RNA-seq based global and targeted co-expression networks. *Scientific Reports*, *8*(1), 10885. https://doi.org/10.1038/s41598-018-29077-3

Lin, M., & Oliver, D. J. (2008). The Role of Acetyl-Coenzyme A Synthetase in Arabidopsis. *Plant Physiology*, *147*(4), 1822–1829. https://doi.org/10.1104/PP.108.121269

Lim, W. K., Wang, K., Lefebvre, C., & Califano, A. (2007). Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics*, *23*(13), i282–i288. https://doi.org/10.1093/BIOINFORMATICS/BTM201

Liu, C., Mao, B., Ou, S., Wang, W., Liu, L., Wu, Y., Chu, C., & Wang, X. (2014). OsbZIP71, a bZIP transcription factor, confers salinity and drought tolerance in rice. *Plant Molecular Biology*, *84*(1–2), 19–36. https://doi.org/10.1007/S11103-013-0115-3/METRICS

Liu, T., Chen, T., Kan, J., Yao, Y., Guo, D., Yang, Y., Ling, X., Wang, J., & Zhang, B. (2022). The GhMYB36 transcription factor confers resistance to biotic and abiotic stress by enhancing PR1 gene expression in plants. *Plant Biotechnology Journal*, *20*(4), 722–735. https://doi.org/10.1111/PBI.13751

Liu, W. X., Zhang, F. C., Zhang, W. Z., Song, L. F., Wu, W. H., & Chen, Y. F. (2013). Arabidopsis Di19 Functions as a Transcription Factor and Modulates PR1, PR2, and PR5 Expression in Response to Drought Stress. *Molecular Plant*, *6*(5), 1487–1502. https://doi.org/10.1093/MP/SST031

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Lysenko, A., Defoin-Platel, M., Hassani-Pak, K., Taubert, J., Hodgman, C., Rawlings, C. J., & Saqi, M. (2011). Assessing the functional coherence of modules found in multiple-evidence networks from Arabidopsis. *BMC Bioinformatics*, *12*(1), 1–14. https://doi.org/10.1186/1471-2105-12-203/FIGURES/6

Maiti, R. K., & Satya, P. (2014). Research advances in major cereal crops for adaptation to abiotic stresses. *Https://Doi.Org/10.4161/21645698.2014.947861*, *5*(4), 259–279. https://doi.org/10.4161/21645698.2014.947861

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

Martin, S., Brown, W. M., Klavans, R., & Boyack, K. W. (2011). OpenOrd: an open-source toolbox for large graph layout. *Https://Doi.Org/10.1117/12.871402*, *7868*, 45–55. https://doi.org/10.1117/12.871402

Mason, M. J., Fan, G., Plath, K., Zhou, Q., & Horvath, S. (2009). Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics*, *10*(1), 1–25. https://doi.org/10.1186/1471-2164-10-327/TABLES/5

McLoughlin, F., Galvan-Ampudia, C. S., Julkowska, M. M., Caarls, L., van der Does, D., Laurière, C., Munnik, T., Haring, M. A., & Testerink, C. (2012). The Snf1-related protein kinases SnRK2.4 and SnRK2.10 are involved in maintenance of root system architecture during salt stress. *The Plant Journal*, *72*(3), 436–449. https://doi.org/10.1111/J.1365-313X.2012.05089.X

Melotto, M., Underwood, W., Koczan, J., Nomura, K., & He, S. Y. (2006). Plant Stomata Function in Innate Immunity against Bacterial Invasion. *Cell*, *126*(5), 969–980. https://doi.org/10.1016/J.CELL.2006.06.054

Memisoglu, G., Eapen, V. v., Yang, Y., Klionsky, D. J., & Haber, J. E. (2019). PP2C phosphatases promote autophagy by dephosphorylation of the Atg1 complex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(5), 1613–1620. https://doi.org/10.1073/PNAS.1817078116

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019a). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419–D426. https://doi.org/10.1093/NAR/GKY1038

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019b). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research*, *47*(D1), D419–D426. https://doi.org/10.1093/nar/gky1038

Mittler, R. (2006). Abiotic stress, the field environment and stress combination. *Trends in Plant Science*, *11*(1), 15–19. https://doi.org/10.1016/J.TPLANTS.2005.11.002

Mueller, S. P., Krause, D. M., Mueller, M. J., & Fekete, A. (2015). Accumulation of extra-chloroplastic triacylglycerols in Arabidopsis seedlings during heat acclimation. *Journal of Experimental Botany*, *66*(15), 4517–4526. https://doi.org/10.1093/JXB/ERV226

Mutwil, M., Usadel, B., Schütte, M., Loraine, A., Ebenhöh, O., & Persson, S. (2010). Assembly of an Interactive Correlation Network for the Arabidopsis Genome Using a Novel Heuristic Clustering Algorithm. *Plant Physiology*, *152*(1), 29. https://doi.org/10.1104/PP.109.145318

Nafie, E., Hathout, T., & al Mokadem, A. S. (2011). Jasmonic acid elicits oxidative defense and detoxification systems in Cucumis melo L. cells. *Brazilian Journal of Plant Physiology*, *23*(2), 161–174. https://doi.org/10.1590/S1677-04202011000200008

Naser, V., & Shani, E. (2016). Auxin response under osmotic stress. *Plant Molecular Biology 2016 91:6*, *91*(6), 661–672. https://doi.org/10.1007/S11103-016-0476-5

Nguyen, D., Rieu, I., Mariani, C., & van Dam, N. M. (2016). How plants handle multiple stresses: hormonal interactions underlying responses to abiotic stress and insect herbivory. *Plant Molecular Biology 2016 91:6*, *91*(6), 727–740. https://doi.org/10.1007/S11103-016-0481-8

Nounjan, N., Chansongkrow, P., Charoensawan, V., Siangliw, J. L., Toojinda, T., Chadchawan, S., & Theerakulpisut, P. (2018). High Performance of Photosynthesis and Osmotic Adjustment Are Associated With Salt Tolerance Ability in Rice Carrying Drought Tolerance QTL: Physiological and Co-expression Network Analysis. *Frontiers in Plant Science*, *9*, 1135. https://doi.org/10.3389/fpls.2018.01135

Obayashi, T., Nishida, K., Kasahara, K., & Kinoshita, K. (2011). ATTED-II updates: condition-specific gene co-expression to extend co-expression analyses and applications to a broad range of flowering plants. *Plant & Cell Physiology*, *52*(2), 213–219. https://doi.org/10.1093/pcp/pcq203

Ouyang, S. Q., Liu, Y. F., Liu, P., Lei, G., He, S. J., Ma, B., Zhang, W. K., Zhang, J. S., & Chen, S. Y. (2010). Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (Oryza sativa) plants. *The Plant Journal : For Cell and Molecular Biology*, *62*(2), 316–329. https://doi.org/10.1111/J.1365-313X.2010.04146.X

Pearce, R. S. (2007). Plant Abiotic Stress. Edited by M. A. Jenks and P. M. Hasegawa. Oxford: Blackwell Publishing (2005), pp. 270, £99.95. ISBN 1-4051-2238-2. *Experimental Agriculture*, *43*(2), 257–257. https://doi.org/10.1017/S0014479706224906

Phule, A. S., Barbadikar, K. M., Maganti, S. M., Seguttuvel, P., Subrahmanyam, D., Babu, M. B. B. P., & Kumar, P. A. (2019). RNA-seq reveals the involvement of key genes for aerobic adaptation in rice. *Scientific Reports 2019 9:1*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-41703-2

Priya, P., & Jain, M. (2013). RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. *Database : The Journal of Biological Databases and Curation*, *2013*, bat027. https://doi.org/10.1093/database/bat027

Proost, S., & Mutwil, M. (2017). PlaNet: Comparative Co-Expression Network Analyses for Plants. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1533, pp. 213–227). https://doi.org/10.1007/978-1-4939-6658-5_12

Qi, Y., Yamauchi, Y., Ling, J., Kawano, N., Li, D., & Tanaka, K. (2004). Cloning of a putative monogalactosyldiacylglycerol synthase gene from rice (Oryza sativa L.) plants and its expression in response to submergence and other stresses. *Planta*, *219*(3), 450–458. https://doi.org/10.1007/S00425-004-1245-2

Quan, R., Lin, H., Mendoza, I., Zhang, Y., Cao, W., Yang, Y., Shang, M., Chen, S., Pardo, J. M., & Guo, Y. (2007). SCABP8/CBL10, a Putative Calcium Sensor, Interacts with the Protein Kinase SOS2

to Protect Arabidopsis Shoots from Salt Stress. *The Plant Cell*, *19*(4), 1415. https://doi.org/10.1105/TPC.106.042291

Rabbani, M. A., Maruyama, K., Abe, H., Khan, M. A., Katsura, K., Ito, Y., Yoshiwara, K., Seki, M., Shinozaki, K., & Yamaguchi-Shinozaki, K. (2003). Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses. *Plant Physiology*, *133*(4), 1755. https://doi.org/10.1104/PP.103.025742

Ransbotyn, V., Yeger-Lotem, E., Basha, O., Acuna, T., Verduyn, C., Gordon, M., Chalifa-Caspi, V., Hannah, M. A., & Barak, S. (2015). A combination of gene expression ranking and co-expression network analysis increases discovery rate in large-scale mutant screens for novel Arabidopsis thaliana abiotic stress genes. *Plant Biotechnology Journal*, *13*(4), 501–513. https://doi.org/10.1111/PBI.12274

Rao, K. P., Vani, G., Kumar, K., Wankhede, D. P., Misra, M., Gupta, M., & Sinha, A. K. (2011). Arsenic stress activates MAP kinase in rice roots and leaves. *Archives of Biochemistry and Biophysics*, *506*(1), 73–82. https://doi.org/10.1016/J.ABB.2010.11.006

Rao, X., Chen, X., Shen, H., Ma, Q., Li, G., Tang, Y., Pena, M., York, W., Frazier, T. P., Lenaghan, S., Xiao, X., Chen, F., & Dixon, R. A. (2019). Gene regulatory networks for lignin biosynthesis in switchgrass (Panicum virgatum). *Plant Biotechnology Journal*, *17*(3), 580–593. https://doi.org/10.1111/PBI.13000

Righetti, K., Vu, J. L., Pelletier, S., Vu, B. L., Glaab, E., Lalanne, D., Pasha, A., Patel, R. v., Provart, N. J., Verdier, J., Leprince, O., & Buitink, J. (2015). Inference of Longevity-Related Genes from a Robust Co-expression Network of Seed Maturation Identifies Regulators Linking Seed Storability to Biotic Defense-Related Pathways. *The Plant Cell*, *27*(10), 2692–2708. https://doi.org/10.1105/TPC.15.00632

Rivera, C. G., Vakil, R., & Bader, J. S. (2010). NeMo: Network module identification in cytoscape. *BMC Bioinformatics*, *11*(SUPPLL.1), 1–9. https://doi.org/10.1186/1471-2105-11-S1-S61/FIGURES/3

Ronald, P. C., & Beutler, B. (2010). Plant and animal sensors of conserved microbial signatures. *Science (New York, N.Y.)*, *330*(6007), 1061–1064. https://doi.org/10.1126/SCIENCE.1189468

Rosell, C. M., & Marco, C. (2008). Rice. *Gluten-Free Cereal Products and Beverages*, 81–III. https://doi.org/10.1016/B978-012373739-7.50006-X

Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A. R., Nikoloski, Z., Persson, S., & Mutwil, M. (2016). FamNet: A framework to identify multiplied modules driving pathway diversification in plants. *Plant Physiology*, *170*(3), pp.01281.2015. https://doi.org/10.1104/pp.15.01281

Saelens, W., Cannoodt, R., & Saeys, Y. (2018). A comprehensive evaluation of module detection methods for gene expression data. *Nature Communications*, *9*(1). https://doi.org/10.1038/S41467-018-03424-4

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., Wakimoto, H., Yang, C. C., Iwamoto, M., Abe, T., Yamada, Y., Muto, A., Inokuchi, H., Ikemura, T., Matsumoto, T., Sasaki, T., & Itoh, T. (2013). Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant & Cell Physiology*, *54*(2). https://doi.org/10.1093/PCP/PCS183

Sánchez-Hernández, C., Martínez-Gallardo, N., Guerrero-Rangel, A., Valdés-Rodríguez, S., & Délano-Frier, J. (2004). Trypsin and α-amylase inhibitors are differentially induced in leaves of amaranth (Amaranthus hypochondriacus) in response to biotic and abiotic stress. *Physiologia Plantarum*, *122*(2), 254–264. https://doi.org/10.1111/J.0031-9317.2004.00398.X

Sang, J., Zou, D., Wang, Z., Wang, F., Zhang, Y., Xia, L., Li, Z., Ma, L., Li, M., Xu, B., Liu, X., Wu, S., Liu, L., Niu, G., Li, M., Luo, Y., Hu, S., Hao, L., & Zhang, Z. (2020). IC4R-2.0: Rice Genome Reannotation Using Massive RNA-seq Data. *Genomics, Proteomics and Bioinformatics*, *18*(2), 161–172. https://doi.org/10.1016/j.gpb.2018.12.011

Sasaki-Sekimoto, Y., Taki, N., Obayashi, T., Aono, M., Matsumoto, F., Sakurai, N., Suzuki, H., Hirai, M. Y., Noji, M., Saito, K., Masuda, T., Takamiya, K. I., Shibata, D., & Ohta, H. (2005). Coordinated activation of metabolic pathways for antioxidants and defence compounds by jasmonates and their roles in stress tolerance in Arabidopsis. *The Plant Journal : For Cell and Molecular Biology*, *44*(4), 653–668. https://doi.org/10.1111/J.1365-313X.2005.02560.X

Schaefer, R., Michno, J.-M., Jeffers, J., Hoekenga, O. A., Dilkes, B. P., Baxter, I. R., & Myers, C. (2018). Integrating co-expression networks with GWAS to prioritize causal genes in maize. *The Plant Cell*, tpc.00299.2018. https://doi.org/10.1105/tpc.18.00299

Schaewen, A. von, Frank, J., & Koiwa, H. (2008). Role of complex N-glycans in plant stress tolerance. *Plant Signaling & Behavior*, *3*(10), 871. https://doi.org/10.4161/PSB.3.10.6227

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., Gruden, K., Stitt, M., Bolger, M. E., & Usadel, B. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant*, *12*(6), 879–892. https://doi.org/10.1016/j.molp.2019.01.003

Seck, P. A., Diagne, A., Mohanty, S., & Wopereis, M. C. S. (2012). Crops that feed the world 7: Rice. *Food Security 2012 4:1*, *4*(1), 7–24. https://doi.org/10.1007/S12571-012-0168-1

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y., & Shinozaki, K. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *The Plant Journal : For Cell and Molecular Biology*, *31*(3), 279–292. https://doi.org/10.1046/J.1365-313X.2002.01359.X

Seo, J. S., Joo, J., Kim, M. J., Kim, Y. K., Nahm, B. H., Song, S. I., Cheong, J. J., Lee, J. S., Kim, J. K., & Choi, Y. do. (2011). OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. *The Plant Journal*, *65*(6), 907–921. https://doi.org/10.1111/J.1365-313X.2010.04477.X

Seo, Y. S., Chern, M., Bartley, L. E., Han, M., Jung, K. H., Lee, I., Walia, H., Richter, T., Xu, X., Cao, P., Bai, W., Ramanan, R., Amonpant, F., Arul, L., Canlas, P. E., Ruan, R., Park, C. J., Chen, X., Hwang, S., … Ronald, P. C. (2011). Towards establishment of a rice stress response interactome. *PLoS Genetics*, *7*(4). https://doi.org/10.1371/JOURNAL.PGEN.1002020

Shen, Q., Zhan, X., Yang, P., Li, J., Chen, J., Tang, B., Wang, X., & Hong, Y. (2019). Dual Activities of Plant cGMP-Dependent Protein Kinase and Its Roles in Gibberellin Signaling and Salt Stress. *The Plant Cell*, *31*(12), 3073–3091. https://doi.org/10.1105/TPC.19.00510

Shepherd, T., & Griffiths, D. W. (2006). The effects of stress on plant cuticular waxes. *New Phytologist*, *171*(3), 469–499. https://doi.org/10.1111/J.1469-8137.2006.01826.X

Shi, H., Quintero, F. J., Pardo, J. M., & Zhu, J. K. (2002). The Putative Plasma Membrane Na+/H+ Antiporter SOS1 Controls Long-Distance Na+ Transport in Plants. *The Plant Cell*, *14*(2), 465. https://doi.org/10.1105/TPC.010371

Sibout, R., Proost, S., Hansen, B. O., Vaid, N., Giorgi, F. M., Ho-Yue-Kuang, S., Legée, F., Cézart, L., Bouchabké-Coussa, O., Soulhat, C., Provart, N., Pasha, A., le Bris, P., Roujol, D., Hofte, H., Jamet, E., Lapierre, C., Persson, S., & Mutwil, M. (2017). Expression atlas and comparative co-expression network analyses reveal important genes involved in the formation of lignified cell wall in Brachypodium distachyon. *The New Phytologist*, *215*(3), 1009–1025. https://doi.org/10.1111/NPH.14635

Silva, A. T., Ribone, P. A., Chan, R. L., Ligterink, W., & Hilhorst, H. W. M. (2016). A Predictive Co-expression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in Arabidopsis thaliana. *Plant Physiology*, *170*(4), 2218–2231. https://doi.org/10.1104/PP.15.01704

Sircar, S., Musaddi, M., & Parekh, N. (2022). NetREx: Network-based Rice Expression Analysis Server for abiotic stress conditions. *Database*, *2022*. https://doi.org/10.1093/DATABASE/BAAC060

Sircar, S., & Parekh, N. (2015). Functional characterization of drought-responsive modules and genes in Oryza sativa: a network-based approach. *Frontiers in Genetics*, *6*, 256. https://doi.org/10.3389/fgene.2015.00256

Sircar, S., & Parekh, N. (2019). Meta-analysis of drought-tolerant genotypes in Oryza sativa: A network-based approach. *PLOS ONE*, *14*(5), e0216068. https://doi.org/10.1371/JOURNAL.PONE.0216068

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-co-expression network for global discovery of conserved genetic modules. *Science (New York, N.Y.)*, *302*(5643), 249–255. https://doi.org/10.1126/SCIENCE.1087447

Su, W., Bao, Y., Yu, X., Xia, X., Liu, C., & Yin, W. (2020). Autophagy and its regulators in response to stress in plants. In *International Journal of Molecular Sciences* (Vol. 21, Issue 23, pp. 1–12). MDPI AG. https://doi.org/10.3390/ijms21238889

Su, W., Ma, H., Liu, C., Wu, J., & Yang, J. (2006). Identification and characterization of two rice autophagy associated genes, OsAtg8 and OsAtg4. *Molecular Biology Reports*, *33*(4), 273–278. https://doi.org/10.1007/S11033-006-9011-0

Sun, Y., Luo, W., Jain, A., Liu, L., Ai, H., Liu, X., Feng, B., Zhang, L., Zhang, Z., Guohua, X., & Sun, S. (2018). OsPHR3 affects the traits governing nitrogen homeostasis in rice. *BMC Plant Biology*, *18*(1), 1–15. https://doi.org/10.1186/S12870-018-1462-7/FIGURES/8

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/NAR/GKY1131

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & von Mering, C. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/NAR/GKY1131

Tan, M., Cheng, D., Yang, Y., Zhang, G., Qin, M., Chen, J., Chen, Y., & Jiang, M. (2017). Co-expression network analysis of the transcriptomes of rice roots exposed to various cadmium stresses reveals universal cadmium-responsive genes. *BMC Plant Biology*, *17*(1), 194. https://doi.org/10.1186/s12870-017-1143-y

Tang, Y., Bao, X., Zhi, Y., Wu, Q., Guo, Y., Yin, X., Zeng, L., Li, J., Zhang, J., He, W., Liu, W., Wang, Q., Jia, C., Li, Z., & Liu, K. (2019). Overexpression of a myb family gene, Osmyb6, increases drought and salinity stress tolerance in transgenic rice. *Frontiers in Plant Science*, *10*, 168. https://doi.org/10.3389/FPLS.2019.00168/BIBTEX

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., Ishihama, Y., Hirayama, T., & Shinozaki, K. (2009a). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17588–17593. https://doi.org/10.1073/pnas.0907095106

Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F., Yamaguchi-Shinozaki, K., Ishihama, Y., Hirayama, T., & Shinozaki, K. (2009b). Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17588–17593. https://doi.org/10.1073/PNAS.0907095106/SUPPL_FILE/0907095106SI.PDF

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F. M., Bassel, G. W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., & Provart, N. J. (2009). Co-expression tools for plant biology:

opportunities for hypothesis generation and caveats. *Plant, Cell & Environment*, *32*(12), 1633–1651. https://doi.org/10.1111/J.1365-3040.2009.02040.X

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., & de Magalhães, J. P. (2018). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics*, *19*(4), 575–592. https://doi.org/10.1093/BIB/BBW139

Vettakkorumakankav, N. N., Falk, D., Saxena, P., & Fletcher, R. A. (1999). A Crucial Role for Gibberellins in Stress Protection of Plants. *Plant and Cell Physiology*, *40*(5), 542–548. https://doi.org/10.1093/OXFORDJOURNALS.PCP.A029575

Wang, C., Zhang, L. J., & Huang, R. D. (2011). Cytoskeleton and plant salt stress tolerance. *Plant Signaling & Behavior*, *6*(1), 29. https://doi.org/10.4161/PSB.6.1.14202

Wang, S., Yin, Y., Ma, Q., Tang, X., Hao, D., & Xu, Y. (2012). Genome-scale identification of cell-wall related genes in Arabidopsis based on co-expression network analysis. *BMC Plant Biology*, *12*(1), 1–12. https://doi.org/10.1186/1471-2229-12-138/FIGURES/4

Wang, Y., & Botella, J. R. (2022). Heterotrimeric G Protein Signaling in Abiotic Stress. *Plants*, *11*(7), 876. https://doi.org/10.3390/PLANTS11070876

Wang, Z., & Gou, X. (2020). Receptor-Like Protein Kinases Function Upstream of MAPKs in Regulating Plant Development. *International Journal of Molecular Sciences*, *21*(20), 1–27. https://doi.org/10.3390/IJMS21207638

Wasternack, C., & Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany*, *111*(6), 1021–1058. https://doi.org/10.1093/AOB/MCT067

Wei, H., Chen, C., Ma, X., Zhang, Y., Han, J., Mei, H., & Yu, S. (2017). Comparative analysis of expression profiles of panicle development among tolerant and sensitive rice in response to drought stress. *Frontiers in Plant Science*, *8*, 437. https://doi.org/10.3389/FPLS.2017.00437/BIBTEX

Wisecaver, J. H., Borowsky, A. T., Tzin, V., Jander, G., Kliebenstein, D. J., & Rokas, A. (2017). A Global Co-expression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *The Plant Cell*, *29*(5), 944–959. https://doi.org/10.1105/tpc.17.00009

Wojtaszek, P. (1997). Oxidative burst: an early plant response to pathogen infection. *Biochemical Journal*, *322*(3), 681–692. https://doi.org/10.1042/BJ3220681

Wolfe, C. J., Kohane, I. S., & Butte, A. J. (2005). Systematic survey reveals general applicability of "guilt-by-association" within gene co-expression networks. *BMC Bioinformatics*, *6*, 227. https://doi.org/10.1186/1471-2105-6-227

Xing, B., Gu, C., Zhang, T., Zhang, Q., Yu, Q., Jiang, J., & Liu, G. (2021). Functional Study of BpPP2C1 Revealed Its Role in Salt Stress in Betula platyphylla. *Frontiers in Plant Science*, *11*, 2215. https://doi.org/10.3389/FPLS.2020.617635/BIBTEX

Yamaguchi, T., Kuroda, M., Yamakawa, H., Ashizawa, T., Hirayae, K., Kurimoto, L., Shinya, T., & Shibuya, N. (2009). Suppression of a Phospholipase D Gene, OsPLDβ1, Activates Defense Responses and Increases Disease Resistance in Rice. *Plant Physiology*, *150*(1), 308–319. https://doi.org/10.1104/PP.108.131979

Yan, J., Zhang, C., Gu, M., Bai, Z., Zhang, W., Qi, T., Cheng, Z., Peng, W., Luo, H., Nan, F., Wang, Z., & Xie, D. (2009). The Arabidopsis CORONATINE INSENSITIVE1 Protein Is a Jasmonate Receptor. *The Plant Cell*, *21*(8), 2220–2236. https://doi.org/10.1105/TPC.109.065730

Yao, W., Li, G., Yu, Y., & Ouyang, Y. (2018). funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *GigaScience*, *7*(1), 1–9. https://doi.org/10.1093/GIGASCIENCE/GIX119

Yoo, S. J., Kim, S. H., Kim, M. J., Ryu, C. M., Kim, Y. C., Cho, B. H., & Yang, K. Y. (2014). Involvement of the OsMKK4-OsMPK1 Cascade and its Downstream Transcription Factor OsWRKY53 in the Wounding Response in Rice. *The Plant Pathology Journal*, *30*(2), 168–177. https://doi.org/10.5423/PPJ.OA.10.2013.0106

Yoon, Y., Seo, D. H., Shin, H., Kim, H. J., Kim, C. M., & Jang, G. (2020). The Role of Stress-Responsive Transcription Factors in Modulating Abiotic Stress Tolerance in Plants. *Agronomy 2020, Vol. 10, Page 788*, *10*(6), 788. https://doi.org/10.3390/AGRONOMY10060788

Yu, B., & Li, W. (2014). Comparative profiling of membrane lipids during water stress in Thellungiella salsuginea and its relative Arabidopsis thaliana. *Phytochemistry*, *108*, 77–86. https://doi.org/10.1016/J.PHYTOCHEM.2014.09.012

Yu, Y., Kou, M., Gao, Z., Liu, Y., Xuan, Y., Liu, Y., Tang, Z., Cao, Q., Li, Z., & Sun, J. (2019). Involvement of Phosphatidylserine and Triacylglycerol in the Response of Sweet Potato Leaves to Salt Stress. *Frontiers in Plant Science*, *10*, 1086. https://doi.org/10.3389/FPLS.2019.01086/BIBTEX

Yuan, F., Leng, B., & Wang, B. (2016). Progress in Studying Salt Secretion from the Salt Glands in Recretohalophytes: How Do Plants Secrete Salt? *Frontiers in Plant Science*, *7*, 977. https://doi.org/10.3389/FPLS.2016.00977/BIBTEX

Yuan, N., Yuan, S., Li, Z., Zhou, M., Wu, P., Hu, Q., Mendu, V., Wang, L., & Luo, H. (2018). STRESS INDUCED FACTOR 2, a Leucine-Rich Repeat Kinase Regulates Basal Plant Pathogen Defense. *Plant Physiology*, *176*(4), 3062–3080. https://doi.org/10.1104/PP.17.01266

Zarattini, M., Farjad, M., Launay, A., Cannella, D., SouliÃ, M. C., Bernacchia, G., & Fagard, M. (2021). Every cloud has a silver lining: how abiotic stresses affect gene expression in plant-pathogen interactions. *Journal of Experimental Botany*, *72*(4), 1020–1033. https://doi.org/10.1093/JXB/ERAA531

Zeng, L., Liu, X., Zhou, Z., Li, D., Zhao, X., Zhu, L., Luo, Y., & Hu, S. (2018). Identification of a G2-like transcription factor, OsPHL3, functions as a negative regulator of flowering in rice by co-

expression and reverse genetic analysis. *BMC Plant Biology*, *18*(1), 157. https://doi.org/10.1186/s12870-018-1382-6

Zhang, H., Liu, Y., Wen, F., Yao, D., Wang, L., Guo, J., Ni, L., Zhang, A., Tan, M., & Jiang, M. (2014). A novel rice C2H2-type zinc finger protein, ZFP36, is a key player involved in abscisic acid-induced antioxidant defence and oxidative stress tolerance in rice. *Journal of Experimental Botany*, *65*(20), 5795–5809. https://doi.org/10.1093/JXB/ERU313

Zhang, J., Zhang, Y., Du, Y., Chen, S., & Tang, H. (2011). Dynamic metabonomic responses of tobacco (Nicotiana tabacum) plants to salt stress. *Journal of Proteome Research*, *10*(4), 1904–1914. https://doi.org/10.1021/PR101140N/SUPPL_FILE/PR101140N_SI_002.PDF

Zhou, Y., Lam, H. M., & Zhang, J. (2007). Inhibition of photosynthesis and energy dissipation induced by water and high light stresses in rice. *Journal of Experimental Botany*, *58*(5), 1207–1217. https://doi.org/10.1093/jxb/erl291

Zhu, J. K. (2002). Salt and Drought Stress Signal Transduction in Plants. *Annual Review of Plant Biology*, *53*, 247. https://doi.org/10.1146/ANNUREV.ARPLANT.53.091401.143329

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell*, *167*(2), 313. https://doi.org/10.1016/J.CELL.2016.08.029

Zhu, M., Xie, H., Wei, X., Dossa, K., Yu, Y., Hui, S., Tang, G., Zeng, X., Yu, Y., Hu, P., & Wang, J. (2019). WGCNA Analysis of Salt-Responsive Core Transcriptome Identifies Novel Hub Genes in Rice. *Genes 2019, Vol. 10, Page 719*, *10*(9), 719. https://doi.org/10.3390/GENES10090719

Zong, W., Tang, N., Yang, J., Peng, L., Ma, S., Xu, Y., Li, G., & Xiong, L. (2016). Feedback Regulation of ABA Signaling and Biosynthesis by a bZIP Transcription Factor Targets Drought-Resistance-Related Genes. *Plant Physiology*, *171*(4), 2810–2825. https://doi.org/10.1104/PP.16.00469